Abstract

Cremer and McLean (1988) show that, when buyers' valuations are correlated, a seller with full knowledge of the distribution can extract the full surplus. We study whether this phenomenon persists when the seller has only partial knowledge of the distribution. We assume that the seller knows that the distribution is one of finitely many and has access to samples (independent draws) from the true distribution. We show that using samples directly in the mechanism is more effective than indirectly using samples via statistical inferences. Our main result is a tight bound on the number of samples needed for full surplus extraction, which is the difference between the number of distributions and the dimension of the linear space they span, plus one.
1 Introduction

Cremer and McLean (1988) show that if the values of the buyers in an auction are correlated, then generically the seller can design an auction that extracts the full surplus. The auction given by Cremer and McLean requires full knowledge of the distribution of buyers’ valuations, but in reality the seller rarely has access to such detailed information. This paper studies the possibility of full surplus extraction when the seller has only partial knowledge of the distribution of buyers’ valuations.

Our model of partial knowledge has two components. First, the seller knows that the true distribution of buyers’ valuations belongs to a given finite set of possible distributions. Second, the seller has access to external information regarding the identity of the true distribution. Such information can be obtained by market research or by observing historical bid data. We mainly focus on the case where the information consists of independent draws from the true distribution of values, which we refer to as samples.

We study the design of mechanisms that map the realized samples and the buyers’ bids to allocations and payments. Crucially, we assume that while the choice of the mechanism may affect the buyers’ bids, it does not affect the realized samples. This assumption models environments where the participants in the current and the past auctions have the same observable characteristics (therefore the distributions of values in the current and past auctions are identical), but are otherwise different individuals.

Our goal is to identify the number of samples that guarantees the existence of a mechanism that extracts the full surplus for all possible distributions. We interpret the number of samples that guarantee full surplus extraction as a proxy to the amount of information required to extract surplus without full knowledge of the distribution.

Our main result, Theorem 2, identifies the number of samples that are necessary and sufficient for the existence of a mechanisms that extracts the full surplus. The number of samples is equal to the number of distributions in the set, minus the dimension of the linear space spanned by them, plus one. This number is at least equal to one and at most equal to the number of distributions minus one. For example, if the distributions are linearly independent, the required number of samples is one.

The mechanism that extracts the full surplus with the minimum number of samples uses the samples directly (rather than via statistical inference). The mechanism is a second price auction plus side payments that depend on the sample realizations. Buyers bid before the

---

1 It is also necessary that full surplus extraction is possible for every distribution in the set.
samples are revealed, and therefore they perceive these payments as random. To extract the full surplus, the side payments for each buyer must be constructed to match the buyer’s utility from the second price auction in expectation. Randomized side payments are commonly used in extensions of Crémer-McLean (e.g. McAfee and Reny [1992], Liu [2017]), and are used in Lopomo et al. [2017] in a mechanism design setting with Knightian uncertainty.

To prove sufficiency of the number of samples identified in Theorem 2, the main step is to show that the number guarantees the existence of appropriate side payments. Such side payments exist if for each buyer, the conditional probability matrix that specifies the probability of a profile of other buyers’ values and samples conditioned on the buyer’s type (its value as well as the index of the distribution) has full rank. We crucially use the independence assumption on samples to decompose the conditional probability matrix as an outer product of conditional probability matrices of value profiles and samples. We then identify, for a given set of vectors (representing conditional probabilities for each value of a buyer), the number of times that each vector should be multiplied by itself (in the outer product sense) such that the resulting set of vectors are linearly independent.

A natural alternative approach to use the samples is via statistical inferences. That is, a mechanism may use the samples to infer the true distribution with high enough certainty, and then choose an auction (i.e., a mapping from bids to outcomes). For example, one may use maximum likelihood estimation or a statistical learning method. We show that inference based approaches are ineffective in their use of samples. In particular, by Theorem 2, the number of samples required to guarantee the existence of a full surplus extracting mechanism is at most the number of possible distributions in the set. On the other hand, inferring the true distribution with a given certainty over all sets of two distributions requires an unbounded number of samples.

We consider two classes of inferences-based mechanisms, show that they must infer the true distribution with high probability, and therefore are ineffective in their use of samples. First, we consider using the samples to select an auction that extracts the full surplus for the identified distribution. A problem is that such an auction may violate individual rationality for the true distribution, if the identified distribution is not the true distribution. Therefore individual rationality will be violated unless the true distribution is identified with certainty. Second, and to avoid violating the individual rationality constraint, we consider selecting an auction that respects individual rationality for all distributions in the set. The problem is that no single auction may extract the full surplus on all possible distributions. Even if we only require the mechanism to extract full surplus approximately, it must infer the true
distribution with high probability and requires a large number of samples.

Our results show that it is more effective to use samples directly in a mechanism than it is to use statistical inferences that are derived from samples. Given the number of samples identified in our main theorem as sufficient for full surplus extraction, it is not possible to infer the true distribution with high certainty. In particular, we construct a set of distributions for which, using the mechanism with random side payments, one sample suffices to extract the full surplus. However, as the number of distributions in the set grows or they get closer to each other, the probability of inferring the true distribution with one sample goes to zero.

Our work can be seen as a reexamination of features of commonly used auction models in order to identify a feature that is responsible for the unsettling prediction of full surplus extraction. Other papers have shown that full surplus extraction is not generically possible if the assumptions of risk neutrality and unlimited liability are relaxed (Robert, 1991). The particular feature we study is the assumption that the seller has full knowledge of the distribution of types. We show that this assumption is mostly innocent: full surplus extraction persists even with sample access to the true distribution.

1.1 Related Works

Other papers have studied the design of mechanisms without full knowledge of the distribution of types, and without access to any external information. In Bergemann et al. (2016), the seller knows the distribution of values but not the agent’s beliefs. Since no external information is available, the seller can not extract the full surplus for all possible distributions. As a result, Bergemann et al. (2016) focus on a worst case objective. In comparison, given the external information, we guarantee full surplus extraction for all possible distributions. A common approach in designing mechanisms with no external information is to use the agents’ reported types for inference. For example, in Goldberg et al. (2006); Balcan et al. (2008), the price to be offered to an agent is inferred from the bids of other agents (for a survey of the follow up literature see Nisan et al., 2007). In Segal (2003); Baliga and Vohra (2003), an agent’s demand, and consequently its virtual valuation, is inferred from bids of other agents.

The literature on prior-independent mechanism design often assumes access to samples (e.g. Fu et al. 2013; Cole and Roughgarden 2014; Morgenstern and Roughgarden 2016). These works assume that the buyers’ values are independent and focus on obtaining approximately optimal mechanisms. The most relevant to our work is Dhangwatnotai et al. (2010)’s single-sampling auction, which showed that with one sample from each bidder’s valuation.
distribution, the VCG auction with the samples as reserve prices gives a 4-approximation to the optimal revenue, when the distributions are regular. As an extension, Roughgarden and Talgam-Cohen (2013) gave a single-sampling mechanism for the more general interdependent value settings under various assumptions, although the benchmark is the optimal revenue under ex post individual rationality.

Our work is also related the literature on model misspecification and model uncertainty (Madarász and Prat, 2017; Bergemann and Schlag, 2011). A key difference is that in the model misspecification literature, it is assumed that the designer knows a distribution that is close to the true distribution, whereas in our setting, the true distribution belongs to a set of possible distributions. That is, in our setting the seller’s model is “partially specified”, rather than misspecified.

A related literature to our work studies genericity of priors that admit full surplus extracting mechanisms in universal type spaces (Heifetz and Neeman, 2006; Barelli, 2009; Chen and Xiong, 2013). These papers seek implementation in Bayes-Nash equilibria, whereas our solution concept is dominant strategy equilibria, and hence we do not explicitly model high order beliefs. More importantly, in those paper the distribution is known to the mechanism. That is, the question is whether a mechanism exists that can extract surplus for a given distribution. In our setting, on the other hand, we ask whether a single mechanism exists that can extract surplus for all distributions in a set (equipped with the additional power of having access to samples).

A large body of literature is dedicated to robust mechanism design (Bergemann and Morris, 2005; Chung and Ely, 2007), that is mechanisms that are less sensitive to details such as the agents’ higher order beliefs. Since we study implementation in dominant strategies, the buyers’ higher order beliefs need not be known. However, since our individual rationality condition is defined in expectation, a buyer’s entry decision does indeed depend on its beliefs, and therefore for a buyer to deduce that other buyers will participate in the mechanism, it must be common knowledge among the agents that the distribution belongs to the set of candidate distribution. Nevertheless, our common knowledge assumption is weaker than requiring all agents to agree on a single distribution. Interim individual rationality is an important feature in Cremer and McLean (1988) and its extensions, and one interpretation of our work is that it is interim individual rationality, and not full knowledge of distribution, that is crucial to existence of full surplus extracting mechanisms.
2 The Setting

A single indivisible item is to be assigned to at most one of $n$ potential buyers. Each buyer $i$ privately knows its value $v_i$. The value $v_i$ belongs to a finite set $V_i \subset \mathbb{R}^+$. Let $v = (v_1, \ldots, v_n)$ be a profile of values, and $V = \prod_i V_i$ the set of possible value profiles. For a buyer $i$, $v_{-i} \in V_{-i} = \prod_{i' \neq i} V_{i'}$ is a profile of values of buyers other than $i$.

A deterministic allocation is $x \in X = \{(x_1, \ldots, x_n) \in \{0, 1\}^n, \sum_i x_i \leq 1\}$, where $x_i$ is the indicator for buyer $i$’s allocation. A randomized allocation is $x \in \Delta(X)$, where $x_i$ is the probability that $i$ gets the item. Buyers have linear utilities. That is, the utility of buyer $i$ with value $v_i \in V_i$ for receiving the item with probability $x_i$ and paying $p_i$ is $v_i x_i - p_i$.

Although our results extend to more general settings, we restrict attention to this setting to focus on the key features of the model.

Let $S$ be a finite set of signals with elements $s \in S$. Let $\mathcal{F} = \{F^1, \ldots, F^m\}$ be a finite set of distinct joint distributions over value profiles $v \in V$ and signals $s \in S$. That is, $F^j \in \Delta(V \times S)$ for all $j$ where $F^j(v, s)$ is the probability of $(v, s)$ according to the $j$’th distribution. For each $j$, let $D^j$ be the marginal probability distribution of $F^j$ on value profiles, i.e., $D^j(v)$ is the probability of value profile $v$.

We study the design of mechanisms that map the buyers’ bids and the realized signal to allocation and payments. The signal is the external information available to the mechanism via market research. A buyer bids in the mechanism knowing only its own value, and before the signal is revealed. We assume that it is commonly known to all buyers that the true distribution belongs to $\mathcal{F}$. Our goal is to design a mechanism that extracts the full surplus in expectation over any distribution in $\mathcal{F}$, to be formalized throughout this section.

We mainly focus on a special case where the signal $s$ consists of independent draws from the distribution of values, defined below.

**Definition 1.** A set $\mathcal{F}$ is a $k$-sample set if $S = V^k$ and for each $(s^1, \ldots, s^k) \in S$ and $j$, $F^j(v, s) = D^j(v) \times D^j(s^1) \times \ldots \times D^j(s^k)$.

Given a $k$-sample set, we refer to each independent draw $s^i$ as a sample, and abusing notation, represent $\mathcal{F}$ as $\{D^1, \ldots, D^m\}$. Note that given the independence assumption, for

---

2In particular, it is possible to extend the setting to allow for interdependence of utilities, in the sense that a buyer’s willingness to pay depends on other buyers’ signals. It is also possible to extend a setting to a multi-alternative setting with multidimensional types. In such a setting, each buyer $i$ has a finite set of preference types $\Theta_i$. Each buyer $i$ has a valuation function $v_i : A \times \Theta_i \rightarrow \mathbb{R}$. The utility of buyer $i$ with preference type $\theta_i \in \Theta_i$ for alternative $a \in A$ and payment $p \in \mathbb{R}$ is $v_i(a, \theta_i) - p$.

3Since the buyers need not know the true distribution, the standard approach of extracting such information via scoring rules does not apply.
each distribution the signal is uninformative of the value profile.

We study mechanisms with dominant strategy equilibria. We invoke the revelation principle and focus on direct mechanisms. A (direct) mechanism consists of a pair of functions \((x, p)\). The function \(x\) is the allocation function mapping actions and signals, possibly at random, to allocations \(x : V_1 \times \ldots \times V_n \times S \rightarrow \Delta(X)\). The function \(p\) is the payment function mapping actions and signal to payments \(p : V_1 \times \ldots \times V_n \times S \rightarrow \mathbb{R}^n\).

A mechanism is dominant strategy incentive compatible (DSIC) if the following holds for all \(i, v_i, v'_i, v_{-i}, \) and \(s\):

\[
v_i x_i(v_i, v_{-i}, s) - p(v_i, v_{-i}, s) \geq v'_i x_i(v'_i, v_{-i}, s) - p(v'_i, v_{-i}, s).
\]

A mechanism is interim individually rational (IIR) for \(\mathcal{F}\) if, for any buyer \(i\), value \(v_i\), and distribution \(F^j \in \mathcal{F}\),

\[
E_{(v, s) \sim F^j} [v_i x_i(v, s) - p_i(v, s) | v_i] \geq 0.
\]

That is, knowing \(v_i\) but not \(v_{-i}\) or \(s\), buyer \(i\) expects non-negative utility from participation, regardless of which distribution is the true distribution.

**Definition 2.** A mechanism \((x, p)\) extracts full surplus on \(\mathcal{F}\) if

1. The mechanism is DSIC.
2. The mechanism is IIR for \(\mathcal{F}\).
3. In expectation for each distribution in \(\mathcal{F}\), the revenue of the mechanism equals the highest value. That is, for all \(j\),

\[
E_{(v, s) \sim F^j} \left[ \sum_i p_i(v, s) \right] = E_{(v, s) \sim F^j} \left[ \max_i v_i \right].
\]

We call a mechanism satisfying properties (1) and (2) above a \(\mathcal{F}\)-feasible mechanism. We say that full surplus extraction is possible for \(\mathcal{F}\) if there exists a mechanism \((x, p)\) that extracts the full surplus on \(\mathcal{F}\).

---

4By risk neutrality there is no loss in focusing on deterministic payment rules.

5Recall that the allocation \(x_i\) is allowed to be randomized. A more restrictive definition would be to require the condition to hold for every internal random choice of the mechanism. The mechanism we construct in our main theorem indeed satisfies the stronger condition.

6If a mechanism extracts full surplus on \(\mathcal{F}\), then it maximizes revenue in expectation over any possible distribution over \(\mathcal{F}\). As a result, our model need not include the seller’s prior belief over \(\mathcal{F}\).
A special case of our model is when the true distribution is known.

**Definition 3.** The true distribution is known if the set of distributions is a singleton \( \mathcal{F} = \{F^1\} \), and the signal \( s \) and the value profile \( v \) are independent in \( F^1 \).

[Cremer and McLean] (1988) study the case where the true distribution is known. Note that in this case, the signal \( s \) does not reveal any extra information about the distribution or the value profile and thus there is no gain in conditioning a mechanism on \( s \). In particular, there is no loss in assuming that \((x, p)(v, s) = (x, p)(v, s')\) for all \( v, s, \) and \( s' \). We will refer to such a mechanism as an *auction*, and use \( \mathcal{A} \) to denote the set of auctions.

[Cremer and McLean] (1988) show that, if the true distribution is known and under a correlation condition on the value distribution, there exists a DSIC and interim IR auction that extracts the full surplus. The auction is a second price auction with side payments where the side payment of agent \( i \) depends only on \( v - i \). With enough correlation, there is sufficient information in \( v - i \) about \( v_i \). As a result, the side payment can be constructed such that in expectation, it equals the utility that the agent receives in a second price auction.

To state the Crémé-McLean result, we start with some notation for distributions.

**Definition 4.** For \( j \in \{1, \ldots, m\} \), \( i \in \{1, \ldots, n\} \), and \( v_i \in V_i \),

1. let \( \vec{D}^j = \left( D^j(v) \right)_{v \in V} \) be the distribution \( D^j \) over \( v \), represented as a vector of size \( |V| \).
2. let \( \vec{D}^j_{v_i} = \left( D^j(v - i|v_i) \right)_{v - i \in V - i} \) be the distribution \( D^j \) over \( v - i \) conditioned on \( v_i \), represented as a vector of size \( |V - i| \).

We now state the condition used by Crémé-McLean.

**Definition 5.** A valuation distribution \( D^j \) satisfies the Crémé-McLean condition if, for each bidder \( i \), the \(|V_i|\) vectors in \( \{\vec{D}^j_{v_i}\}_{v_i \in V_i} \) are linearly independent.

We now restate the Crémé-McLean theorem in our setting.

**Theorem 0 ([Cremer and McLean] 1988).** Assume that the true distribution is known \([Definition 3]\). There exists an auction that extracts full surplus for \( \mathcal{F} = \{F^1\} \) if and only if the marginal distribution on value profiles \( D^1 \) satisfies the Crémé-McLean condition\(^8\).
Let us call a set of distributions $\mathcal{F}$ a Crémé-McLean set if for each $F^j \in \mathcal{F}$, the marginal distribution on values $D^j$ satisfies the Crémé-McLean condition. Directly from Definition 4, one can see that if a mechanism extracts the full surplus on $\mathcal{F}$, then it must also extract surplus on a singleton set $\{F^j\}$ for all $j$. Therefore, by Theorem 0, $\mathcal{F}$ must be a Crémé-McLean set. We later give conditions on the signal so that full surplus extraction extends to a set of distributions $\mathcal{F}$.

3 Why Statistical Inference is Ineffective

A natural approach to the problem would be to use the samples for inference. We show via an example that this approach ineffective, in the sense that it requires a large amount of information to perform well. To quantify the amount of information needed, throughout this section we focus on $k$-sample sets of distributions (Definition 1).

Example 1. Fix a positive real number $\epsilon$, $0 < \epsilon < 1$. Consider two buyers whose values are generated by the following process: two random variables $\nu_1, \nu_2 \in \{1, \ldots, H\}$ are independently and identically drawn such that $\Pr[\nu_i \geq h] = \frac{1}{h}$, for all $h \in \{1, \ldots, H\}$. With probability $1 - \epsilon$, the two values are defined as $v_1 = \nu_1$ and $v_2 = \nu_2$; with probability $\epsilon$, the higher of the two random variables is assigned to buyer 1, and the lower to buyer 2, $v_1 = \max(\nu_1, \nu_2)$, $v_2 = \min(\nu_1, \nu_2)$. Call the resulting correlated distribution $D^1$. Define another distribution $D^2$ by exactly the same procedure but eventually switching the values of the two buyers. As a result, buyer 1 is more likely to have a higher value than buyer 2 in $D^1$, and the opposite holds in $D^2$. The $k$-sample set of distributions is $\mathcal{F} = \{D^1, D^2\}$.

In Section 4, we show that given a single sample, there exists a mechanism that extracts the full surplus. On the other hand, in this section we show that inferring the true distribution requires a number of samples that grows to infinity as $\epsilon$ goes to zero. As a result, any inference based approach is ineffective in its use of samples.

The lemma below bounds the number of samples that are required to infer the true distribution to within a small probability of error. Formally, a function $h : S \to \Delta(\{1, \ldots, m\})$ infers the true distribution with error at most $\delta$ if $\Pr_{s \sim (\times D^j)^k}[h(s) \neq j] \leq \delta$ for $j \in \{1, \ldots, m\}$ (the expectation is also over the randomization of $h$). The lemma below identifies the number of samples required to infer the true distribution with error at most $\delta$.

Lemma 1. Consider the $k$-sample set of distributions in Example 1. If there exists a function $h : S \to \Delta(\{1, 2\})$ that infers the true distribution with error at most $\delta$, then the number of
samples \( k \) must be at least

\[
(1 - 2\delta) \log((1/\delta) - 1) \frac{1 - \epsilon}{\epsilon^2}.
\]

Note that as either the error \( \delta \) or the difference between the distributions \( \epsilon \) go to zero, the number of samples goes to infinity.

We additionally show that each distribution \( D^1 \) and \( D^2 \) satisfies the Crémé-McLean condition. Therefore, if the true distribution is known, there exists an auction that extracts full surplus.

**Lemma 2.** Both distributions \( D^1 \) and \( D^2 \) in \textbf{Example 1} satisfy the Crémé-McLean condition. Therefore, for each \( D^j \) there exists an auction that extracts full surplus.

### 3.1 Using Statistical Inference to Select a Full Surplus Extracting Auction

Motivated by the fact that each distribution admits a full surplus extracting auction (Lemma 2), it is natural to consider an approach that uses the samples to identify the true distribution, and then selects an auction that extracts full surplus for the identified distribution. We call such a mechanism an *inferring Crémé-McLean mechanism*. The lemma below shows that any inferring Crémé-McLean mechanism violates IIR. In addition, even if the IIR requirement is relaxed, any such mechanism requires many samples to be approximately feasible (to be defined below).

The main idea is that an auction that extracts full surplus on \( D^1 \) satisfies the IIR condition for \( D^1 \) with equality. On the other hand, any auction that extracts full surplus on \( D^2 \) violates the IIR condition for \( D^1 \) significantly. Therefore, for an inferring Crémé-McLean mechanism to not violate the IIR condition by much, it must infer the true distribution with small error. But from **Lemma 1** we know that the number of samples required to limit the misidentification probability is large.

Recall that \( \mathcal{A} \) is the set of auctions (i.e., signal invariant mechanisms). A mechanism can be specified by an inference rule \( L : S \rightarrow \mathcal{A} \) that selects an auction \( L(s) \) given the signal \( s \). We say that a mechanism *infers the true distribution with error at most \( \delta \) if there exist mutually disjoint sets \( A_1, \ldots, A_m \subseteq \mathcal{A} \) such that \( \Pr_{s \sim (x, D^j)^k}[L(s) \notin A_j] \leq \delta \) for \( j \in \{1, \ldots, m\} \).
A mechanism \((x, p)\) is \(\sigma\)-IIR for distribution \(F^j\) if for all \(i, v_i\)
\[
\mathbb{E}_{(v, s) \sim F^j} [v_i x_i(v, s) - p_i(v, s)|v_i] \geq -\sigma.
\]

**Proposition 1.** Consider the \(k\)-sample set of distributions \(\mathcal{F}\) in Example 1, and assume that the highest possible value for each buyer is \(H = 2\). Consider any inferring Crémers-McLean mechanism that is \(\sigma\)-IIR for each distribution in \(\mathcal{F}\). Then the mechanism must infer the true distribution with error at most \(\delta' = 4\sigma/(0.5 + \epsilon)\) and the number of samples \(k\) is at least
\[
(1 - 2\delta') \log((1/\delta') - 1) \frac{1 - \epsilon}{\epsilon^2}.
\]

### 3.2 Using Statistical Inference to Select a Feasible Auction

To avoid violating the IIR condition, one approach could be to use the samples to infer the true distribution, and then to select an \(\mathcal{F}\)-feasible auction (an auction that respects IIR for all distributions in \(\mathcal{F}\)). We show that to get a revenue that is close to full surplus on a \(k\)-sample \(\mathcal{F}\), any such mechanism must infer the true distribution with small error and therefore needs a large number of samples.

Formally, we call a mechanism \((x, p)\) a signal-feasible mechanism if for all \(s\), the auction \(x(\cdot, s), p(\cdot, s)\) is \(\mathcal{F}\)-feasible. That is, \((x, p)\) is DSIC, and for all \(v_i, s\), and \(D^j \in \mathcal{F}\),
\[
\mathbb{E}_{v \sim D^j} [v_i x_i(v, s) - p_i(v, s)|v_i, s] \geq 0.
\]

Note that effectively the only restriction imposed on this class is that individual rationality must be satisfied for each realization \(s\), whereas to satisfy IIR (Inequality \(1\)), a mechanism needs to satisfy the condition only in expectation over \(s\).

For Example 1, let the full surplus \(FS(H)\) be the expected maximum value in either distribution (the full surplus is identical for the two distributions).

**Theorem 1.** Consider the \(k\)-sample set of distributions in Example 1. Consider a signal-feasible mechanism such that
\[
\mathbb{E}_{v \sim D^j, s \sim (x D^j)^k} \left[ \sum_i p_i(v, s) \right] \geq (1 - \delta) FS(H), \forall j \in \{1, 2\}.
\]
Then the mechanism must infer the true distribution with error at most \( \delta' \),

\[
\delta' = \frac{\delta \text{FS}(H)}{\text{FS}(H) - (4 + 2\epsilon)}
\]

and the number of samples \( k \) is at least

\[
(1 - 2\delta') \log((1/\delta') - 1) \frac{1 - \epsilon}{\epsilon^2}.
\]

The full surplus \( \text{FS}(H) \) is at least the expected value of each buyer, which is roughly \( \log(H) \). In the limit as \( H \) gets large, \( \delta' \to \delta \) and therefore the bound of Theorem 1 simply becomes

\[
k \geq (1 - 2\delta) \log((1/\delta) - 1) \frac{1 - \epsilon}{\epsilon^2}.
\]

For a fixed \( \delta \), the rate of growth is quadratic in \( 1/\epsilon \), and for a fixed \( \epsilon \), the rate of growth is logarithmic in \( 1/\delta \).

### 4 Informational Requirement of Surplus Extraction

In this section we state and prove the main theorem of the paper. The theorem quantifies the amount of information required for full surplus extraction. The mechanism derived that extracts the full surplus uses samples directly rather than via statistical inference.

Let us define the terminology. Recall that a set of distributions \( \mathcal{F} \) is a \( k \)-sample set if a signal consists of \( k \) independent draws from the distribution of value profiles (Definition 1). Recall also that for each distribution \( D^j \) over value profiles, \( \vec{D}^j = (D^j(v))_{v \in V} \) is its representation as a vector (Definition 4). The dimension of a vector space is the cardinality of its basis. The dimension of a \( k \)-sample set of distributions \( \mathcal{F} \) is the dimension of the linear space spanned by \( \{\vec{D}^1, \ldots, \vec{D}^m\} \). Note that the dimension of \( \mathcal{F} \) is between 2 and \( m \)\(^9\). Recall that a set of distributions \( \mathcal{F} \) a Crémer-McLean set if for each \( F^j \in \mathcal{F} \), the marginal distribution on values \( D^j \) satisfies the Crémer-McLean condition.

**Theorem 2.** Consider any \( m \) and \( d \) such that \( 2 \leq d \leq m \). Full surplus extraction is possible for all Crémer-McLean \( k \)-sample sets of distributions \( \mathcal{F} \) of size \( m \) and span \( d \) if and only if

\(^9\)The dimension can not be 1. Otherwise, distributions must be a scaled versions of each other. This is not possible for probability distributions.
Recall from Section 2 that to get full surplus extraction on \( F \), it is necessary that each distribution \( D^j \) satisfies the Crémer-McLean condition. The theorem shows that additionally, having access to \( k \geq m - d + 1 \) samples is necessary and sufficient for full surplus extraction. Since \( 2 \leq d \leq m \), the sufficient number of samples is at least 1 and at most \( m - 1 \). The dimension of the space spanned by any two distinct distributions is 2. As a result, only 1 sample is sufficient to extract full surplus for any two distributions. Contrast this with Theorem 1 where the number of samples grows to infinity even for sets of two distributions.

We prove the necessary and sufficient directions of Theorem 2 separately in Section 4.3 (Proposition 3) and Section 4.4 (Proposition 4). We start by defining a class of mechanisms in Section 4.1 that extract full surplus give the conditions of Theorem 2.

### 4.1 Signal-dependent Payments

We first define a class of mechanisms that extend the Crémer-McLean construction to our setting, without requiring \( F \) to be \( k \)-sample. Similar to the Crémer-McLean construction and its extensions (McAfee and Reny, 1992; Lopomo et al., 2017), a mechanism in this class consists of two components. First, a second-price auction is run. The allocation of the second price auction is efficient, but buyers have positive expected utility from participation in it. Second, to extract the remaining surplus from the buyers, each buyer makes an additional side payment to the mechanism. In order to ensure that these side payments do not violate incentive compatibility, each buyer’s side payment depends only on the reports of other buyers and the realized signal (i.e., the payment does not depend on the buyer’s own report). The class is defined formally below. The side payments are unrestricted in the definition below, and will be constructed later to extract the full surplus.

**Definition 6.** A second-price auction with signal-dependent payments works as follows:

1. The allocation is efficient. That is, \( x_i(v,s) > 0 \) only if \( v_i = \max_j v_j \), and also \( \sum_i x_i(v,s) = 1 \) (ties among maximum bids are broken arbitrarily).

2. The payment consists of two parts. First, a second price payment \( p_i^{SPA}(v) \) that is the second highest value if \( i \) gets the item and zero otherwise. Second, a side payment \( q_i(v_{-i},s) \) that for each buyer \( i \) depends only on the values of other buyers. Each buyer’s payment is the sum of the two parts, \( p_i(v,s) = p_i^{SPA}(v) + q_i(v_{-i},s) \).
For each buyer $i$ and value profile $v$, let $u^{\text{SPA}}_i(v)$ be the ex post utility of buyer $i$ in the second price auction. For each buyer $i$ with value $v_i$ and distribution $j$, let

$$u_{i,j}^{\text{SPA}}(v_i) = \mathbb{E}_{v \sim D^j} [u^{\text{SPA}}_i(v)]$$

be the *interim* expected utility of buyer $i$ in the second price auction. Now assume that there exists side payments $q_i$ such that

$$\mathbb{E}_{(v,s) \sim F^j} [q_i(v,s) | v_i] = u_{i,j}^{\text{SPA}}(v_i), \forall i, v_i, j.$$  \hspace{1cm} (3)

Then, the interim utility of buyer $i$ with value $v_i$ is zero in the mechanism, for any distribution $j$. Since the allocation of the mechanism is efficient, this implies full surplus extraction for any distribution. Therefore, full surplus extraction is possible if side payments that satisfy Equation 3 exist.

The lemma below specifies conditions for existence of such side payment functions. It provides conditions on $F$ under which a solution exists for *any* right hand side $u^{\text{SPA}}$, without assuming any structure on $u^{\text{SPA}}$. In particular, it shows that for any $i$, the set of possible interim expected side payments $\{(\mathbb{E}_{(v,s) \sim F^j} [q_i(v,s) | v_i])_{j,v_i}\}_{q_i}$ is equal to $\mathbb{R}^{m \times |V_i|}$. The proof is standard. Similar to Definition 4, we first define $\vec{F}_{v_i}^j$ as the vector representation of the probability of $(v_{-i},s)$ under distribution $j$ conditioned on $v_i$, below.

**Definition 7.** For $j \in \{1, \ldots, m\}, i \in \{1, \ldots, n\}$, and $v_i \in V_i$,

1. let $\vec{F}_{v_i}^j = (F^j(v_{-i},s | v_i))_{v_{-i} \in V_{-i}, s \in S}$ be the distribution $F^j$ over $(v_{-i},s)$ conditioned on $v_i$, represented as a vector of size $|V_{-i}| \times S$.

Compare $\vec{F}_{v_i}^j$ with $\vec{D}_{v_i}^j$ in Definition 4. Whereas $\vec{F}_{v_i}^j$ is a distribution of $(v_{-i},s)$, $\vec{D}_{v_i}^j$ is simply a distribution of $v_{-i}$.

**Lemma 3.** Consider a set of distributions $F = \{F^1, \ldots, F^m\}$. There exists a Crémer-McLean mechanism with samples that extracts the full surplus on $F$, if for each bidder $i$, the set of $|V_i| \times m$ vectors $\{\vec{F}_{v_i}^j\}_{v_i \in V_i, j \in \{1, \ldots, m\}}$ are linearly independent.

**Example 2.** Let us revisit Example 1. For simplicity set $H = 2$ and $\epsilon = 0.5$. For buyer 1, the interim utilities in the second price auction are $u_{1,j}^{\text{SPA}}(1) = 0$ for all $j$, and $u_{1,j}^{\text{SPA}}(2) =$

\footnote{The utility is well defined regardless of how ties are broken, since in case of a tie, a buyer with maximum value has zero utility regardless of how the tie is broken.}

13
\( \Pr[v_2 = 1|v_1 = 2] \), which is equal to 3/5 if \( j = 1 \), and 1/3 if \( j = 2 \). We look for a side payment function \( q_1(v_2, s) \) to match the interim utilities, given a single sample \( s \). We in fact show that a side payment function exists that is zero if the buyer’s values match in the sample \( s \). That is, \( q_1(v_2, s) = 0 \) if \( s = (v_1, v_2) \), \( v_1 = v_2 \). So let us focus only on samples where the values do not match, and abusing notation, let \( s \in \{1, 2\} \) denote the buyer with the higher value. The probabilities are \( \Pr_{D^1}[s = 1] = 3/8, \Pr_{D^1}[s = 2] = 1/8, \Pr_{D^2}[s = 1] = 1/8, \Pr_{D^2}[s = 2] = 3/8 \). We need \( q_1 \) to solve the following system

\[
\begin{pmatrix}
(v_1, j) \\
(1, 1) (1, 2) (2, 1) (2, 2)
\end{pmatrix}
\begin{pmatrix}
q_1(1, 1) \\
q_1(1, 2) \\
q_1(2, 1) \\
q_1(2, 2)
\end{pmatrix}
= \begin{pmatrix}
0 \\
\frac{3}{5} \\
0 \\
\frac{1}{3}
\end{pmatrix}
\]

Note that the probabilities in each row do not sum up to 1 since we removed the samples where the values are equal. The solution to the system is \( q_1(1, 1) = -6, q_1(1, 2) = -6, q_1(2, 1) = 16, q_1(2, 2) = 0 \).

### 4.2 Inference is not Necessary for Surplus Extraction

Given Theorem 2, we argue that it is possible to extract full surplus without inferring the true distribution with a non trivial error. Consider the following extension of Example 1 with \( n \) buyers.

**Example 3.** The \( k \)-sample set of distributions is \( \mathcal{F} = \{D^1, \ldots, D^n\} \), where \( D^i \) is defined as follows. Draw \( n \) random variables \( \nu_1, \ldots, \nu_n \in \{1, 2\} \) independently and identically drawn such that \( \Pr[\nu_i \geq h] = 1/h \), for all \( h \in \{1, 2\} \). With probability \( 1-\epsilon \), the values are assigned directly \( \nu_i = \nu_{i'} \) for all \( i' \); with probability \( \epsilon \), buyer \( i \) is has the maximum value \( \nu_i = \max_{\nu_{i'}} \nu_{i'} \), and the remaining \( n-1 \) random variables are uniformly assigned to the remaining \( n-1 \) buyers.

The following proposition shows that the span of \( \mathcal{F} \) in Example 3 is \( n \), and therefore by Theorem 2 full surplus extraction is possible with a single sample. On the other hand, with a single sample it is not possible to infer the true distribution with high probability.
Proposition 2. Full surplus extraction is possible with 1 sample for the instance given in Example 3. With 1 sample, there exists no mechanism that learns the true distribution with error at most \((1 - \epsilon)(1 - 1/n)\).

4.3 The Upper Bound on the Number of Samples Needed

We now state and prove the sufficiency part of Theorem 2.

Proposition 3. Consider a Crémér-McLean \(k\)-sample set of distributions \(\mathcal{F}\) of size \(m\) and span \(d\). If \(k \geq m - d + 1\), then there exists a second-price auction with signal-dependent payments that extracts the full surplus on \(\mathcal{F}\).

Proposition 3 is a corollary of Lemma 3 and the lemma below.

Lemma 4. Consider a Crémér-McLean \(k\)-sample set of distributions \(\mathcal{F}\) of size \(m\) and span \(d\). If \(k \geq m - d + 1\), then for each bidder \(i\), the set of \(|V_i| \times m\) vectors \(\{\vec{F}_{V_i}^j\}_{v_i \in V_i, j \in \{1, \ldots, m\}}\) are linearly independent.

The rest of this subsection proves Lemma 4. Notation will be greatly simplified by using outer products on vectors. The outer product of two vectors \(A = (a_i)_{i \in I} \in \mathbb{R}^{|I|}\) of size \(|I|\) and \(B = (b_j)_{j \in J} \in \mathbb{R}^{|J|}\) of size \(|J|\), denoted \(C = A \otimes B\), is a vector \(C = (a_1 B, \ldots, a_{|I|} B)\) of size \(|I| \times |J|\). Outer products are bilinear and associative, but in general are not commutative.\(^{11}\)

We use the following standard property of outer products.

Lemma 5. Consider a set of linearly independent vectors \(A = \{A_1, \ldots, A_m\}\) and, for each \(j = 1, \ldots, m\), a set \(B_j\) of linearly independent vectors. The set of vectors in the set \(\{B \otimes A_j\}_{j \in \{1, \ldots, m\}, B \in B_j}\) (of size \(\sum_j |B_j|\)) are linearly independent.

We also establish the following property on independence of outer product of vectors. Let \((\otimes A)^k\) denote the outer product of \(k\) copies of \(A\).

Lemma 6. Consider a set of \(m\) vectors \(\{A^1, \ldots, A^m\}\). Let \(d\) be the dimension of the linear space spanned by \(\{A^1, \ldots, A^m\}\). The set of vectors in \(\{(\otimes A^1)^k, \ldots, (\otimes A^m)^k\}\) are linearly independent if \(k \geq m - d + 1\).

\(^{11}\)Two \(A \otimes B\) and \(B \otimes A\) are identical only up to permutations, for example \((1, 2) \otimes (3, 4) = (1(3, 4), 2(3, 4)) = (3, 4, 6, 8)\) and \((3, 4) \otimes (1, 2) = (3(1, 2), 4(1, 2)) = (3, 6, 4, 8)\).
Example 4. To illustrate Lemma 6, consider the following set of three vectors \( \{ A^1, A^2, A^3 \} \). The dimension spanned by these vectors is 2. Therefore, by Lemma 6, the set of vectors \( \{ (\otimes A^1)^2, (\otimes A^2)^2, (\otimes A^3)^2 \} \) are linearly independent.

\[
\begin{align*}
A^1 &= (1, 0), & (\otimes A^1)^2 &= (1, 0, 0, 0) \\
A^2 &= (0, 1), & (\otimes A^2)^2 &= (0, 0, 0, 1) \\
A^3 &= (0.3, 0.7), & (\otimes A^3)^2 &= (0.09, 0.21, 0.21, 0.49)
\end{align*}
\]

We next use Lemma 5 and Lemma 6 to prove Lemma 4.

Proof of Lemma 4. Fix a buyer \( i \). Recall from Definition 7 the definition of \( \vec{F} \), for all \( j \), and \( v_i \),

\[
\vec{F}_{v_i}^j = \left( F^j(v_{i-}, s|v_i) \right)_{v_{i-} \in V_{i-}, s \in S}.
\]

Since \( \mathcal{F} \) is \( k \)-sample, we have,

\[
\vec{F}_{v_i}^j = \left( D^j(v_{i-}|v_i) \times D^j(s^1) \times \ldots \times D^j(s^k) \right)_{v_{i-}, s}.
\]

Using the outer product notation, this simplifies to (see Definition 4 and Definition 7)

\[
\vec{F}_{v_i}^j = \vec{D}_{v_i}^j \otimes (\otimes \vec{D}^j)^k. \tag{4}
\]

Recall that by assumption, \( k \geq m - d + 1 \), where \( d \) is the dimension of the linear space spanned by \( \{ \vec{D}^1, \ldots, \vec{D}^m \} \). Therefore, by Lemma 6, the \( m \) vectors in \( \{ (\otimes \vec{D}^1)^k, \ldots, (\otimes \vec{D}^m)^k \} \) are linearly independent. Also, by the assumption that for each \( j \) the distribution \( D^j \) satisfies the Crémér-McLean condition, the \( |V_i| \) vectors in \( \{ \vec{D}_{v_i}^j \}_{v_i \in V_i} \) are linearly independent. We can then apply Lemma 5 to conclude that the vectors in \( \{ \vec{F}_{v_i}^j \}_{v_i \in V_i, j \in \{1, \ldots, m\}} \) are linearly independent. In particular, define the set \( \mathcal{A} = \{ (\otimes \vec{D}^1)^k, \ldots, (\otimes \vec{D}^m)^k \} \) of linearly independent vectors as argued above. Also, for each \( j \in \{1, \ldots, m\} \), define the set \( \mathcal{B}_j = \{ \vec{D}_{v_i}^j \}_{v_i \in V_i} \). Given the Crémér-McLean condition, each \( \mathcal{B}_j \) consists of linearly independent vectors. Now plug into Lemma 5 to conclude that the vectors in the set \( \{ B \otimes A^j \}_{j \in \{1, \ldots, m\}, B \in \mathcal{B}_j} = \{ \vec{D}_{v_i}^j \otimes (\otimes \vec{D}^j)^k \}_{j \in \{1, \ldots, m\}, v_i \in V_i} = \{ \vec{F}_{v_i}^j \}_{j \in \{1, \ldots, m\}, v_i \in V_i} \) are also linearly independent.
4.4 The Lower Bound on the Number of Samples Needed

We now show that the number of samples in Theorem 2 is necessary.

**Proposition 4.** Consider any $m$ and $d$ such that $2 \leq d \leq m$. If $k \leq m - d$ then there exists a Crémér-McLean $k$-sample set of distributions $\mathcal{F}$ of size $m$ and span $d$ such that full surplus extraction is not possible for $\mathcal{F}$.

Let us first point out a difficulty. Recall that Proposition 3 was established through Lemma 3 which ensured that for each buyer $i$, the set of possible interim expected side payments is equal to $\mathbb{R}^{m \times |V_i|}$. Thus, for any interim utilities $u_i^{SPA}$ of the second price auction, side payments $q_i$ exist that extract full surplus. To prove the converse of the theorem, it is not sufficient to show that the set of possible interim expected side payments is a strict subset of $\mathbb{R}^{m \times |V_i|}$. The reason is that the set of interim utilities $u_i^{SPA}$ is structured. For instance, consider distributions $D_1, D_2, \text{ and } D_3$ such that for a buyer $i$,
\[
D_3(v_i | v_i) = D_1(v_i | v_i)/2 + D_2(v_i | v_i)/2.
\]
Then it must be that $u_i^{SPA}_{3,1} = u_i^{SPA}_{1,1}/2 + u_i^{SPA}_{2,1}/2$. As a result, even though the set of interim expected side payments is not equal to $\mathbb{R}^{m \times |V_i|}$, a side payment may exists for each utility function satisfying $u_i^{SPA}_{3,1} = u_i^{SPA}_{1,1}/2 + u_i^{SPA}_{2,1}/2$.

We first prove the case where $d = 2$, and later discuss the generalization which is a simple extension. The proof is based on the following instance.

**Example 5.** Buyer 1 has two possible values, $v_1 \in \{2, 3\}$. There is only two profile of values that other buyers can possibly have, $v_{-1}^1$ and $v_{-1}^2$.[12] We only assume that the maximum value in $v_{-1}^1$ and $v_{-1}^2$ is 1, that is $\max_{j \neq i} v_j^1 = \max_{j \neq i} v_j^2 = 1$, and otherwise leave them unconstrained. Construct basis distributions $B_1, B_2$ as follows.

\[
B_1 = \frac{1}{3} \begin{pmatrix} v_{-i}^1 & v_{-i}^2 \\ 0 & 2/3 \end{pmatrix}, \quad B_2 = \frac{2}{3} \begin{pmatrix} 0 & 2/3 \\ 1/3 & 0 \end{pmatrix}.
\]

Consider $\alpha_1$ to $\alpha_m$, where $0 \leq \alpha_j \leq 1$, $\alpha_j \neq 1/2$. Construct each distribution $D_j$ in the set $\mathcal{F}$ as a convex combination of $D_1$ and $D_2$ with weight $\alpha_j$, that is, $D_j = \alpha_j B_1 + (1 - \alpha_j) B_2$.

Assume for contradiction that a full surplus extracting mechanism exists for Proposition 4. Then buyer 1 must be allocated regardless of the profile of values, since buyer 1 has

---

[12]Strictly speaking, $V_{-i}$ is any product set $v_{-i}^1$ and $v_{-i}^2$. We only consider $v_{-i}^1$ and $v_{-i}^2$ since they are the only profiles that may have positive probability.
the highest value. As a result, buyer 1’s utility from allocation, ignoring payments, is equal to its value. Therefore, to extract surplus the expected payment of each value must be equal to the value,

$$E_{v \sim D_j, s \sim (x \times D_j)^k} [p_1(v, s) | v_1] = v_1, \quad \forall j, v_1.$$ 

Since buyer 1 gets the item regardless of its value, incentive compatibility requires that the payment of buyer 1 does not depend on its report. We thus write the payment function of buyer 1 as \(p_1(v - 1, s)\). The above equality becomes

$$E_{v \sim D_j, s \sim (x \times D_j)^k} [p_1(v - 1, s) | v_1] = v_1, \quad \forall j, v_1.$$ 

Now consider any profile \(\beta = (\beta_{j,v_1})_{j,v_1}\) such that \(\sum_{j,v_1} \beta_{j,v_1} = 0\). Note that this implies that \(\sum_j \beta_{j,2} = -\sum_j \beta_{j,3}\). Assume further that \(\sum_j \beta_{j,2} \neq 0\). We must have

$$\sum_{j,v_1} \beta_{j,v_1} E_{v \sim D_j, s \sim (x \times D_j)^k} [p_1(v - 1, s) | v_1] = \sum_j \beta_{j,2} \cdot 2 + \sum_j \beta_{j,3} \cdot 3$$

$$= (2 - 3)(\sum_j \beta_{j,2}) \neq 0.$$ 

Summarizing the argument so far, we have shown that if a full surplus extracting mechanism exists, then there must exist a function \(p_1 : V_{-1} \times S \to \mathbb{R}\) such that for any profile \(\beta = (\beta_{j,v_1})_{j,v_1}\) satisfying i) \(\sum_{j,v_1} \beta_{j,v_1} = 0\) and ii) \(\sum_j \beta_{j,2} \neq 0\), we have

$$\sum_{j,v_1} \beta_{j,v_1} E_{v \sim D_j, s \sim (x \times D_j)^k} [p_1(v - 1, s) | v_1] \neq 0.$$ 

The next lemma shows that no such function \(p_1\) exists.

**Lemma 7.** Consider the set of distributions \(\mathcal{F} = \{D^1, \ldots, D^n\}\) defined in Example 5 and assume that the number of samples is \(k \leq m - 2\). There exists a profile \(\beta = (\beta_{j,v_1})_{j,v_1}\) such that i) \(\sum_{j,v_1} \beta_{j,v_1} = 0\) and ii) \(\sum_j \beta_{j,2} \neq 0\) such that any payment function \(p_1 : V_{-1} \times S \to \mathbb{R}\) satisfies

$$\sum_{j,v_1} \beta_{j,v_1} E_{v \sim D_j, s \sim (x \times D_j)^k} [p_1(v - 1, s) | v_1] = 0.$$
5 Discussion

The surplus-extracting auction of Crémer and McLean is often seen as a critique on the model of auction design. The (arguably) counterintuitive phenomenon of surplus extraction is often attributed to the unrealistic combination of several assumptions in the model: first, that the agents are risk neutral and only consider their expected linear utilities; second, that the auctioneer has exact knowledge on the underlying distribution of the agents' values; and third, that the distribution is commonly known to all buyers. The second and third assumptions are seen as a violation of the desired Wilson’s principle. Our result suggests that these two assumptions may not be the main driver of full surplus extraction: the assumption that the seller knows the distribution can be weakened, as long as sampling from the underlying distribution is available, and the number of samples does not have to be large; similarly, it is sufficient to assume it is commonly known to the buyers that the true distribution belongs to the set, without requiring the buyers to agree on the true distribution.

It would be interesting to investigate full surplus extraction on infinite sets of distributions. Even though our approach involves inverting matrices whose entries are probabilities of atom events, there may be hope to extend the approach to infinite-support distributions, since there have been such extensions to Crémer and McLean’s auction (McAfee and Reny 1992; Rahman 2014). This seems a prerequisite for possibly extending the approach further to infinite families of distributions.

References


A Missing Proofs

A.1 Proofs from Section 3

A.1.1 Proof of Lemma 1

The proof uses Kullback-Leibler (KL) divergence which is a measure of relative entropy of two distributions, defined as follows. For two probability distributions \( P \) and \( Q \) over elements of \( \Omega \), their KL divergence \( D_{KL}(P, Q) \) is

\[
D_{KL}(P, Q) = \sum_{i \in \Omega} \left( \log \frac{P(i)}{Q(i)} \right) P(i). \tag{5}
\]

The proof is based on bounding the KL divergence of the distributions of \( k \) independent samples from \( D^1 \) and \( D^2 \). Since distributions \( D^1 \) and \( D^2 \) are close to each other, we can provide an upper bound on the KL divergence of \((\times D^1)^k\) and \((\times D^2)^k\) directly from the definition. On the other hand, for the outcome of a function \( h \) to be significantly different under \((\times D^1)^k\) and \((\times D^2)^k\), the two distributions must be different enough. These two claims are formalized and proved in the following two lemmas, respectively.

We first state two standard properties of KL divergence, non-negativity of KL divergence (Theorem 2.6.3 in Cover and Thomas (2012)) and the chain rule (Theorem 2.5.3 in Cover and Thomas (2012)).

Claim 3. For any two distributions \( P \) and \( Q \), and random variables \( y_1, y_2 \), the following holds.

- (Information Inequality) \( D_{KL}(P(y_1), Q(y_1)) \geq 0 \), and \( D_{KL}(P(y_1), Q(y_1)) = 0 \) if \( P = Q \).
- (Chain Rule)

\[
D_{KL}(P(y_1, y_2), Q(y_1, y_2)) = D_{KL}(P(y_1), Q(y_1)) + E_{y_1} \left[ D_{KL}(P(y_2|y_1), Q(y_2|y_1)) \right].
\]

In particular, if \( (y_1, y_2) \) are independent under \( P \) and \( Q \),

\[
D_{KL}(P(y_1, y_2), Q(y_1, y_2)) = D_{KL}(P(y_1), Q(y_1)) + D_{KL}(P(y_2), Q(y_2)).
\]

We now provide an upper bound on the KL divergence of \((\times D^1)^k\) and \((\times D^2)^k\).
Lemma 8. The KL divergence of \((\times D^1)^k\) and \((\times D^2)^k\) is at most \(2k\epsilon^2/(1-\epsilon)\).

Proof. We first provide a bound on the KL divergence of \(D^1\) and \(D^2\). A sample \((v_1, v_2)\) from \(D^1\) or \(D^2\) is uniquely described by the unordered pair of values \(\{v_1, v_2\}\) and the sign of \(v_1 - v_2\), \(\text{sign}(v_1 - v_2)\). The distribution of \(\{v_1, v_2\}\) is the same regardless of whether \((v_1, v_2)\) is sampled from \(D^1\) or \(D^2\). The value of \(\text{sign}(x - y)\) equals +1 with probability \((1 + \epsilon)/2\) in the case of \(D^1\) and with probability \((1 - \epsilon)/2\) in the case of \(D^2\). By the chain rule for KL divergence (Claim 3),

\[
D_{KL}(D^1(v_1, v_2) || D^2(v_1, v_2)) =
D_{KL}(D^1(\{v_1, v_2\}) || D^2(\{v_1, v_2\})) + D_{KL}(D^1(\text{sign}(v_1 - v_2)) || D^2(\text{sign}(v_1 - v_2))).
\]

The first term on the right side is zero. The second term can be bounded using the definition of KL divergence as follows (Lemma 6.4 in Karp and Kleinberg (2007)).

\[
D_{KL}(D^1(\text{sign}(v_1 - v_2)) || D^2(\text{sign}(v_1 - v_2)))
= \left(\log \frac{1 + \epsilon}{1 - \epsilon}\right) \frac{1 + \epsilon}{2} + \left(\log \frac{1 - \epsilon}{1 + \epsilon}\right) \frac{1 - \epsilon}{2}
= \log 1 \times \frac{1}{2} + \log \left(\left(\frac{1 + \epsilon}{1 - \epsilon}\right)^2\right) \frac{\epsilon}{2}
= \log \left(1 + \frac{2\epsilon}{1 - \epsilon}\right) \times \epsilon \leq \left(\frac{2\epsilon}{1 - \epsilon}\right) \epsilon,
\]
where the inequality followed since \(\log(1 + z) < z\) for all \(z > 0\). Therefore,

\[
D_{KL}(D^1(v_1, v_2) || D^2(v_1, v_2)) \leq \frac{2\epsilon^2}{1 - \epsilon}.
\]

The bound for \((\times D^1)^k\) and \((\times D^2)^k\) follows from applying the chain rule to the above inequality \(\square\).

We next argue that if a function \(h\) exists such that its outcomes under \((\times D^1)^k\) and \((\times D^2)^k\) are significantly different, then the KL divergence of the two distributions must be high enough.

Lemma 9. Consider any two distributions \(P\) and \(Q\) over a random variable \(x \in \Omega\). If there
exists a function \( h : \Omega \rightarrow \Delta(\{1, 2\}) \) such that

\[
\Pr_{x \sim P}\left[h(x) = 1\right] \geq 1 - \delta, \quad \text{and} \quad \Pr_{x \sim Q}\left[h(x) = 1\right] \leq \delta,
\]

then,

\[
D_{KL}(P(x), Q(x)) \geq (1 - 2\delta) \log((1/\delta) - 1).
\]

**Proof.** Let us first make the randomization of \( h \) explicit. In particular, assume that \( h(x) = \bar{h}(x, a) \), where \( \bar{h} \) is a deterministic function, and \( a \) is the random seed of function \( h \) drawn independently from \( x \) from a distribution \( A \). Now define a random variable \( z = \bar{h}(x, a) \). Abusing notation, write \( P(x, a, z) \) as the joint distribution of \((x, a, z)\) when \( x \) is drawn from \( P \), \( a \) from \( A \), and \( z \) is defined using \( \bar{h} \). Similarly define \( Q(x, a, z) \). By the chain rule of KL divergence (Claim 3),

\[
D_{KL}(P(x, a, z), Q(x, a, z)) = D_{KL}(P(z), Q(z)) + D_{KL}(P(x, a|z), Q(x, a|z)) \geq D_{KL}(P(z), Q(z))
\]

where the above inequality followed form non-negativity of KL divergence,

\[
\geq (1 - \delta) \log((1 - \delta)/\delta) + \delta \log(\delta/(1 - \delta)) = (1 - 2\delta) \log((1/\delta) - 1).
\]

Note that \( z \) is deterministically defined from \((x, a)\) using \( \bar{h} \). Therefore,

\[
D_{KL}(P(x, a, z), Q(x, a, z)) = D_{KL}(P(x, a), Q(x, a)) = D_{KL}(P(x), Q(x)) + D_{KL}(P(a), Q(a))
\]

where the last equation followed from chain rule

\[
= D_{KL}(P(x), Q(x)) + D_{KL}(A(a), A(a)) = D_{KL}(P(x), Q(x)).
\]
We conclude,

\[ D_{KL}(P(x), Q(x)) \geq (1 - 2\delta) \log((1/\delta) - 1). \]

\[ \square \]

**Proof of Lemma 1**  The proof follows from combining Lemma 8 and Lemma 9. By Lemma 9

\[ D_{KL}(P, Q) \geq (1 - 2\delta) \log((1/\delta) - 1), \]

and by Lemma 8

\[ D_{KL}(P, Q) \leq 2k\epsilon^2/(1 - \epsilon). \]

Combining the two inequalities, we get

\[ k \geq (1 - 2\delta) \log((1/\delta) - 1) \frac{1 - \epsilon}{\epsilon^2}. \]

\[ \square \]

**A.1.2 Proof of Lemma 2**

**Proof of Lemma 2**  We first verify the Crémer-McLean condition for buyer 1 and \( D^1 \). Assume that for some profile \((\beta_{v_1})_{v_1 \in v_1}\),

\[ 0 = \sum_{v_1} \beta(v_1)D^1(v_2|v_1), \forall v_2 \]  

(6)

We need to show that \( \beta = 0 \). The conditional probability \( D^1(v_2|v_1) \) is,

\[ D^1(v_2|v_1) = \frac{D^1(v)}{D^1(v_1)} \]
The densities of the two distributions in Example 1 are as follows.

\[
D_1(v) = \begin{cases} 
(1 + \epsilon)D^0(v) & \text{if } v_1 > v_2, \\
D^0(v) & \text{if } v_1 = v_2, \\
(1 - \epsilon)D^0(v) & \text{if } v_1 < v_2.
\end{cases}
\]

\[ D_2(v) = \begin{cases} 
(1 - \epsilon)D^0(v) & \text{if } v_1 > v_2, \\
D^0(v) & \text{if } v_1 = v_2, \\
(1 + \epsilon)D^0(v) & \text{if } v_1 < v_2. 
\end{cases} \tag{7}
\]

Plugging in \(D_1(v)\) from definition in Equation 7, we have

\[
D_1(v_2 | v_1) = \begin{cases} 
\frac{(1+\epsilon)D^0(v_1,v_2)}{D^1(v_1)} & \text{if } v_1 > v_2, \\
\frac{D^0(v_1,v_2)}{D^1(v_1)} & \text{if } v_1 = v_2, \\
\frac{(1-\epsilon)D^0(v_1,v_2)}{D^1(v_1)} & \text{if } v_1 < v_2.
\end{cases}
\]

Given the above formula, and since \(D^0\) is a product distribution \(D^0(v_1,v_2) = D^0(v_1)D^0(v_2)\), Equation 6 becomes,

\[
0 = \left( \sum_{v_1 < v_2} \beta(v_1) \frac{(1 + \epsilon)D^0(v_1)D^0(v_2)}{D^1(v_1)} \right) + \beta(v_2) \frac{D^0(v_2)D^0(v_2)}{D^1(v_2)} \\
+ \left( \sum_{v_1 > v_2} \beta(v_1) \frac{(1 - \epsilon)D^0(v_1)D^0(v_2)}{D^1(v_1)} \right)
= D^0(v_2) \times
\left[ \left( \sum_{v_1 < v_2} \beta(v_1) \frac{(1 + \epsilon)D^0(v_1)}{D^1(v_1)} \right) + \beta(v_2) \frac{D^0(v_2)}{D^1(v_2)} + \left( \sum_{v_1 > v_2} \beta(v_1) \frac{(1 - \epsilon)D^0(v_1)}{D^1(v_1)} \right) \right].
\]

Define \(\gamma(v_1) = \frac{\beta(v_1)D^0(v_1)}{D^1(v_1)}\). Since \(D^0(v_2) > 0\), the above equation is equivalent to

\[
0 = \sum_{v_1 < v_2} (1 + \epsilon)\gamma(v_1) + \gamma(v_2) + \sum_{v_1 > v_2} (1 - \epsilon)\gamma(v_1).
\tag{8}
\]

We show that we must have \(\gamma(v_1) = 0\) for all \(v_1\), which in turn implies that \(\beta(v_1) = 0\), completing the proof. Subtract the above equation for \(v_2 + 1\) from \(v_2\),

\[
\epsilon \gamma(v_2) + \epsilon \gamma(v_2 + 1) = 0.
\]
As a result, $\gamma(1) = -\gamma(2) = \gamma(3) = \ldots$. If $|V_1|$ is even, Equation 8 for $v_2 = 1$ becomes

$$0 = \gamma(1) + (1 - \epsilon)\gamma(2) + \sum_{v_1 \geq 3, v_1 \text{ odd}} (1 - \epsilon)(\gamma(v_1) + \gamma(v_1 + 1)),$$

and if $|V_1|$ is odd, Equation 8 for $v_2 = 1$ becomes

$$0 = \gamma(1) + \sum_{v_1 \geq 2, v_1 \text{ even}} (1 - \epsilon)(\gamma(v_1) + \gamma(v_1 + 1)) = \gamma(1).$$

Each case would imply that $\gamma(1) = 0$, in turn implying that $\gamma(v_1) = 0$ for all $v_1$. This completes the proof for buyer 1. The proof for buyer 2 and $D^2$ is analogous.

A.1.3 Proof of Proposition 1

We first use following standard characterization of DSIC mechanisms. Since we consider finite sets of values, the payment function is not pinned down uniquely by the allocation function (even after fixing the utility of the lowest type). Instead, any monotone extension of the allocation function to real numbers can be used to define a payment function.

Recall that $V_i$ is the finite set of values of buyer $i$. Let $\underline{v}_i$ be the minimum element $\bar{v}_i$ be the maximum element in $V_i$.

Lemma 10. A mechanism $(x, p)$ is DSIC if and only if,

1. $x_i$ is monotone nondecreasing in $v_i$ for all $i$, $v_{-i}$, and $s$.

2. $p_i(v, s) = p_i^A (v_{-i}, s) + v_i x_i (v, s) - \int_{z=\underline{v}_i}^{\bar{v}_i} y_i(z, v_{-i}, s)dz$, for some function $y_i(\cdot, v_{-i}, s) : \mathbb{R} \to [0, 1]$, where $y_i(z, v_{-i}, s)$ is monotone non-decreasing in $z$ and coincides with $x_i$ on $V_i$, that is $y_i(v_i, v_{-i}, s) = x_i(v_i, v_{-i}, s)$ for all $v_i \in V_i$.

Proof. Assume that the mechanism $(x, p)$ is ex-post incentive compatible. Fix a buyer $i$, types of other buyers $v_{-i}$, and signal $s$. Define allocation function $y_i$ and payment function $p_i$ over type space $[\underline{v}_i, \bar{v}_i]$ as follows, $(y_i, \hat{p}_i)(z) = \max_{(x, p) \in \{(x(v_i), p(v_i)) | v_i \in V_i\}} z x - p$. Clearly $(y_i, p_i)$ satisfies the ex-post IC conditions. Therefore, by standard characterization, $y_i$ must be monotone nondecreasing and

$$\hat{p}_i(z, v_{-i}, s) = p_i^A (v_{-i}, s) + v_i x_i (v, s) - \int_{z=\underline{v}_i}^{\bar{v}_i} y_i(z, v_{-i}, s)dz.$$
The “only if” part of the lemma follows since \((x_i, p_i)\) must agree with \((y_i, \hat{p}_i)\) on \(V_i\).

We now prove the “if” part of the lemma. Assume that \(y_i\) exists satisfying the assumptions. Mechanism \((y_i, p_i)\) is incentive compatible for type space \([v_i, \bar{v}_i]\). Since \((x_i, p_i)\) agrees with \((y_i, \hat{p}_i)\) on \(V_i\), then the mechanism \((x_i, p_i)\) is ex-post incentive compatible over type space \(V_i\).

We now prove Proposition 1.

**Proof of Proposition 1.** By Lemma 10, the payments in any DSIC auction is
\[
p_i(v) = p_i^A(v - i) + p_i^B(v); \quad p_i^B(v) = v_i x_i(v) - \int_{z=v_i}^{v_i} y_i(z, v_i) dz.
\]
(An auction is by definition state invariant and therefore \(s\) is dropped.) Therefore, any DSIC auction can be split into an auction with the same allocation and with payment \(p_i^B(v)\) which is DSIC and ex post IR, plus a side payment \(p_i^A(v - i)\) for each buyer \(i\) that does not depend on the report of buyer \(i\). Now consider any full surplus extraction auction \((x, p)\) for \(D^1\). Let \(u_i(v)\) be the ex post utility of buyer \(i\) when the profile of values is \(v\). We must have
\[
E_{v \sim D^1} [p_i^A(v) | v_i = 1] = 0, \quad E_{v \sim D^1} [p_i^A(v) | v_i = 2] = E_{v \sim D^1} [u_i(v) | v_i = 2]. \tag{10}
\]
We now provide a lower bound on \(\sum_i E_{v \sim D^1} [u_i(v) | v_i = 2]\). Note that \(u_i(v) = \int_{z=v_i}^{v_i} y_i(z, v_i) dz\) so we provide an lower bound on \(\int_{z=v_i}^{v_i} y_i(z, v_i) dz\). Since \(y_i\) coincides with \(x_i\) on \(V_i\), we have \(y_i(1, v_i - i) = x_i(1, v_i - i)\). Since \(y_i\) is monotone non-decreasing, we have \(y_i(z, v_i - i) \geq x_i(1, v_i - i)\) for all \(z \in [1, 2]\). Therefore, \(\int_{z=v_i}^{v_i} y_i(z, v_i - i) dz \geq x_i(1, v_i - i)\). We conclude that in any auction with payment \(p_i^B\), the ex post utility of value 2 is \(u_i(2, v_i - i) \geq x_i(1, v_i - i)\). Therefore,
\[
\sum_i E_{v \sim D^1} [u_i(v) | v_i = 2] \geq \sum_i E_{v \sim D^1} [x_i(1, v_i - i)] = \frac{0.5 + \delta}{2}.
\]
Combining the above inequality with Equation 10, we conclude that in any auction that extracts full surplus for \(D^1\), there must exist a buyer \(i\) such that
\[
E_{v \sim D^1} [p_i^A(v) | v_i = 2] \geq \frac{0.5 + \delta}{4}.
\]
Let \((\hat{x}, \hat{p})\) be any auction that extracts full surplus for \(D^2\). Since the IIR condition of
auction \( (\hat{x}, \hat{p}) \) for distribution \( D^2 \) is tight,

\[
E_{v \sim D^2} \left[ \hat{p}_i^A(v) | v_i = 1 \right] = 0. \tag{11}
\]

Now consider any inferring Crémer-McLean mechanism. Recall that a inferring Crémer-McLean mechanism selects either a full surplus extraction auction for either distribution \( D^1 \) or \( D^2 \). Let \( h(s) \in \{1, 2\} \) identify the selected mechanism in the inferring Crémer-McLean mechanism. We have

\[
E_{(v,s) \sim F^2} [v_i x_i(v, s) - p_i(v, s) | v_i = 1] = - Pr_{s \sim (\times D^2)^k} [h(s) = 1] (E_{v \sim D^2} [\hat{p}_i^A(v) | v_i = 1])
\]

\[- Pr_{s \sim (\times D^2)^k} [h(s) = 2] (E_{v \sim D^2} [\hat{p}_i^A(v) | v_i = 1])
\]

\[= - Pr_{s \sim (\times D^2)^k} [h(s) = 2] (E_{v \sim D^1} [p_i^A(v) | v_i = 1]),
\]

where the last equality followed from Equation 11. By definition of the two distributions,

\[
E_{v \sim D^1} [p_i^A(v) | v_i = 1] = E_{v \sim D^2} [p_i^A(v) | v_i = 2] \geq \frac{0.5 + \delta}{4}.
\]

Therefore we have

\[
E_{(v,s) \sim F^2} [v_i x_i(v, s) - p_i(v, s) | v_i = 1] \leq - Pr_{s \sim (\times D^2)^k} [h(s) = 1] \frac{0.5 + \epsilon}{4}.
\]

Now combine the above inequality with the assumption that the inferring Crémer-McLean mechanism is \( \delta \)-IIR for distribution \( D^2 \),

\[
Pr_{s \sim (\times D^2)^k} [h(s) = 1] \frac{0.5 + \epsilon}{4} \leq \delta,
\]

and therefore,

\[
Pr_{s \sim (\times D^2)^k} [h(s) = 2] = 1 - Pr_{s \sim (\times D^2)^k} [h(s) = 1] \geq 1 - \frac{4\delta}{0.5 + \epsilon}.
\]

An identical argument implies that \( Pr_{s \sim (\times D^1)^k} [h(s) = 1] \geq 1 - \frac{4\delta}{0.5 + \epsilon} \). Now we can invoke Lemma 1 to finish the proof.
Proof of Theorem 1

We first provide an outline of the proof. To prove Theorem 1, we show that no \( F \)-feasible auction can achieve a large revenue for both distributions. Therefore, to perform adequately, any signal-feasible mechanism must limit the probability of misidentifying the distribution. We then invoke Lemma 1 to bound the number of samples.

Consider any auction \((x, p)\). By Lemma 10, the auction’s payment is the sum of two parts \( p_A \) and \( p_B \),

\[
p_i(v) = p_A^i(v) + p_B^i(v); \quad p_B^i(v) = v_i x_i(v) - \int_{z=v_i}^{v_i} y_i(z, v_i) dz,
\]

where \( y_i(\cdot, v_{-i}) \) is an extension of \( x_i \) with support over \([v_i, \bar{v}_i]\). The part \( p_A \) is the constant term that behaves as a lottery when \( v_{-i} \) is drawn at random. The part \( p_B \) is to maintain incentive compatibility for fixed \( v_{-i} \), and \( \int_{z=v_i}^{v_i} y_i(z, v_{-i}) dz \) is the information rent of type \( v_i \).

We bound the expectation of each part \( p_A \) and \( p_B \) in the following two lemmas.

The first lemma shows that the expectation of \( p_A \) can not be high simultaneously for both distributions \( D^1 \) and \( D^2 \). In particular, it shows that for each player, the sum of the expectation of \( p_A \) over \( D^1 \) and \( D^2 \) is at most 2. Note that the lemma does not directly imply a bound for either distribution in isolation. That is, \( \mathbb{E}_{v \sim D^1}[p_A^1(v)] \) may be a large positive number, but only if \( \mathbb{E}_{v \sim D^2}[p_A^1(v)] \) is a large negative number. Indeed, in the Crémér-McLean construction with a known distribution, \( p_A \) is constructed to match the expected utility of a buyer in the second price auction.

To describe our approach, let us first define \( D^0 \) to be the distribution resulting from drawing \( v_1 \) and \( v_2 \) independently such that \( \Pr[v_i \geq h] = 1/h \), for all \( h \in \{1, \ldots, H\} \).

The proof outline is as follows. The distribution \( D^0 \), which is a product distribution, is the average of \( D^1 \) and \( D^2 \). Therefore to prove the lemma we bound the expected payment in \( D^0 \). By IIR, the expected payment of the lowest value in \( D^j \) can not be too large, and since \( D^0 \) is close to \( D^j \), the expected payment of the lowest value is small in \( D^0 \) as well. But since \( D^0 \) is a product distribution, the expected payment of all values are equal and are small.

Lemma 11. In Example 1, any \( F \)-feasible auction \((x, p)\) satisfies

\[
\mathbb{E}_{v \sim D^1}[p_A^i(v)] + \mathbb{E}_{v \sim D^2}[p_A^i(v)] \leq 2, \forall i.
\]

Proof. Note that when \( v_i = 1 \), we have \( p_B^i(v) = 0 \) and therefore \( p_i(v) = p_A^i(v_{-i}) \). By interim IR condition of \( D^j \) for \( v_i = 1 \), the expected payment can not be higher than the probability
of allocation (multiplied by 1, since the value is 1), and therefore
\[ E_{v_i \sim D^j} \left[ p^A_i(v_{-i}) | v_i = 1 \right] = E_{v_i \sim D^j} \left[ p_i(v) | v_i = 1 \right] \leq E_{v_i \sim D^j} \left[ x_i(v) | v_i = 1 \right] \leq 1. \]

Multiplying both sides of the above inequality by \(D^j(v_i = 1)\), we have
\[ \sum_{v_{-i}} p^A_i(v_{-i}) D^j(1, v_{-i}) \leq D^j(v_i = 1). \]

Summing over \( j \),
\[ \sum_{v_{-i}} p^A_i(v_{-i}) (D^1(1, v_{-i}) + D^2(1, v_{-i})) \leq D^1(v_i = 1) + D^2(v_i = 1). \quad (13) \]

By definition \[ D^1(v) + D^2(v) = 2D^0(v) \] for all \( v \), and therefore the left hand side is
\[ \sum_{v_{-i}} p^A_i(v_{-i}) (D^1(1, v_{-i}) + D^2(1, v_{-i})) = \sum_{v_{-i}} p^A_i(v_{-i}) 2D^0(1, v_{-i}) = 2D^0(v_i = 1) E_{v_{-i} \sim D^0} \left[ p^A_i(v_{-i}) \right], \quad (14) \]
where the last equality followed since \( D^0 \) is a product distribution. Combining \[ 13 \] with \[ 14 \] we have
\[ 2D^0(v_i = 1) E_{v_{-i} \sim D^0} \left[ p^A_i(v_{-i}) \right] \leq D^1(v_i = 1) + D^2(v_i = 1). \]

Since \( 2D^0(v_i) = D^1(v_i) + D^2(v_i) \) for all \( v_i \), the above is equivalent to
\[ E_{v_{-i} \sim D^0} \left[ p^A_i(v_{-i}) \right] \leq 1. \quad (15) \]
Given $15$ and since $D^1 + D^2 = 2D^0$, we can finish the proof by writing

$$E_{v \sim D^1} \left[ p^A_i(v) \right] + E_{v \sim D^2} \left[ p^A_i(v) \right] = 2 \times E_{v \sim D^0} \left[ p^A_i(v) \right]$$

$$= 2 \times E_{v \sim D^0} \left[ p^A_i(v_{-i}) \right]$$

$$= 2 \times E_{v_{-i} \sim D^0} \left[ p^A_i(v_{-i}) \right] \leq 2.$$ 

\[ \square \]

We next bound $p^B$. This bound follows from noting that each of the two distributions $D^1$ and $D^2$ are close to the distribution $D^0$, and that the expected revenue under distribution $D^0$ can be bounded using Myerson’s characterization of optimal auctions for product distributions [Myerson, 1981]. The first lemma below follows from Myerson’s characterization. The revenue curve associated with $D^0$ is a constant at $1$, since the revenue of posting a price $p$ is $p \times \Pr[v_i \geq p] = p \times 1/p = 1$. Therefore the virtual value (marginal revenue) of any type is zero, and the expected virtual surplus is zero. Revenue is the expected virtual surplus plus the payment of the lowest type, which is $1$ for each player.

**Lemma 12.** In [Example 1], the expected revenue of any DSIC and ex post IR auction, over distribution $D^0$, is at most $2$.

**Proof.** For each player $i$, let $x_i(v_i) = E_{v_{-i} \sim D^0}[x_i(v)]$ be the *interim* allocation probability of player $i$ with value $v_i$. By the discrete analogue of Myerson’s characterization [Myerson, 1981],

$$E_{v \sim D^0} \left[ p_i(v) \right] \leq E_{v_{-i} \sim D^0} \left[ p_i(1, v_{-i}) \right] + \sum_{v_i} x_i(v_i)(G_i(v_i)v_i - G_i(v_{i+1})v_{i+1}),$$

where $G_i(v_i) = \sum_{v_j \geq v_i} D^0(v_j)$. Note however that $G_i(v_i) = 1/v_i$, and therefore $G_i(v_i)v_i - G_i(v_{i+1})v_{i+1} = 0$. As a result, the above inequality becomes,

$$E_{v \sim D^0} \left[ p_i(v) \right] \leq E_{v_{-i} \sim D^0} \left[ p_i(1, v_{-i}) \right].$$

The IR constraint for $v_i = 1$ implies that the right hand side of the above inequality is at most $1$. We conclude that the expected payment of each player is at most $1$, and that the expected revenue from both players is at most $2$. 

\[ \square \]
The next lemma follows from the fact that the probability of any profile of values \( v \) in \( D^1 \) and \( D^2 \) is at most \( 1 + \epsilon \) times the probability of \( v \) in \( D^0 \).

**Lemma 13.** In Example 1, any \( F \)-feasible auction \((x, p)\) satisfies

\[
E_{D^j} \left[ \sum_i p_i^B(v) \right] \leq 2(1 + \epsilon), \forall j \in \{1, 2\}.
\]

**Proof.** Note from definition 7 that \( D^j(v) \leq (1 + \epsilon)D^0(v) \). Since \( p_i^B \geq 0 \),

\[
E_{D^j} \left[ \sum_i p_i^B(v) \right] \leq (1 + \epsilon)E_{D^0} \left[ \sum_i p_i^B(v) \right].
\]

Note that the auction with allocation \( x \) and payment \( p^B \) (i.e., without the part \( p^A \)) is DSIC and ex post IR. By Lemma 12, the expected revenue of any such auction for distribution \( D^0 \) is at most 2. Therefore, we conclude that \( E_{D^j} \left[ \sum_i p_i^B(v) \right] \leq 2(1 + \epsilon) \).

By combining Lemma 11 and Lemma 13 we conclude that no \( F \)-feasible auction can achieve large revenue in expectation over both distributions.

**Lemma 14.** In Example 1, any \( F \)-feasible auction \((x, p)\) satisfies

\[
E_{D^1} \left[ \sum_i p_i(v) \right] + E_{D^2} \left[ \sum_i p_i(v) \right] \leq 8 + 4\epsilon.
\]

**Proof.** By Lemma 10, the revenue of any auction \((x, p)\) can be broken down to \( p^A \) and \( p^B \) as in 12. By applying Lemma 11 and Lemma 13 in turn, we have

\[
E_{D^1} \left[ \sum_i p_i(v) \right] + E_{D^2} \left[ \sum_i p_i(v) \right]
= E_{D^1} \left[ \sum_i p_i^A(v) + p_i^B(v) \right] + E_{D^2} \left[ \sum_i p_i^A(v) + p_i^B(v) \right]
\leq 4 + E_{D^1} \left[ \sum_i p_i^B(v) \right] + E_{D^2} \left[ \sum_i p_i^B(v) \right]
\leq 4 + 2(1 + \epsilon) + 2(1 + \epsilon) = 8 + 4\epsilon.
\]
We can now combine Lemma 1 and Lemma 14 to prove Theorem 1. By Lemma 14, since no \( \mathcal{F} \)-feasible auction can perform well on both distributions, the learning rule must select an auction that performs well for the true distribution with high probability. Lemma 1 specifies the number of samples that are required for any given probability of identifying the true distribution.

**Proof of Theorem 1.** For \( j \in \{1, 2\} \), let \( M_j \subseteq M_{SI}(\mathcal{F}) \) be the subset of all \( \mathcal{F} \)-feasible auctions \((x, p)\) that achieve a revenue of at least \( 4 + 2\epsilon \) in expectation over distribution \( D_j \), that is

\[
E_{v \sim D_j} \left[ \sum_i p_i(v) \right] \geq 4 + 2\epsilon. \tag{16}
\]

By Lemma 14, any auction in \( M_{SI}(\mathcal{F}) \) that satisfies the above inequality for \( j \) must violate it for \(-j\), that is, \( M_1 \cap M_2 = \emptyset \). Now consider any signal-feasible mechanism \((x, p)\) identified by a learning rule \( L \). By definition of \( M_j \), for any given \( s \),

\[
E_{v \sim D_j} \left[ \sum_i p_i(v, s) \right] \leq 4 + 2\epsilon, \quad \text{if } L(s) \notin M_j,
\]

and by IIR,

\[
E_{v \sim D_j} \left[ \sum_i p_i(v, s) \right] \leq \text{FS}(H), \quad \text{if } L(s) \in M_j.
\]

(Recall that \( \text{FS}(H) \) is the full surplus of either distribution in Example 1.) Therefore we can write

\[
E_{v \sim D_j, s \sim (\times D_j)^k} \left[ \sum_i p_i(v, s) \right] \\
\leq \Pr_{s \sim (\times D_j)^k} [L(s) \in M_j] \text{FS}(H) + (1 - \Pr_{s \sim (\times D_j)^k} [L(s) \in M_j])(4 + 2\epsilon) \\
= \Pr_{s \sim (\times D_j)^k} [L(s) \in M_j] \left(\text{FS}(H) - (4 + 2\epsilon)\right) + (4 - 2\epsilon).
\]
By the assumption of the theorem, for all \( j \),

\[
E_{v \sim D^j, s \sim (x D^j)^k} \left[ \sum_i p_i(v, s) \right] \geq (1 - \delta)FS(H).
\]

We therefore must have

\[
Pr_{s \sim (x D^j)^k} \left[ L(s) \in M_j \right] \left( FS(H) - (4 + 2\epsilon) \right) + (4 - 2\epsilon) \geq (1 - \delta)FS(H)
\]

and therefore for all \( j \),

\[
Pr_{s \sim (x D^j)^k} \left[ L(s) \in M_j \right] \geq \frac{(1 - \delta)FS(H) - (4 + 2\epsilon)}{FS(H) - (4 + 2\epsilon)} = 1 - \frac{\delta FS(H)}{FS(H) - (4 + 2\epsilon)}.
\]

Define the function \( h \) such that \( h(s) = 1 \) if \( L(s) \in M_1 \) and \( h(s) = 2 \) if \( L(s) \in M_2 \) (and define \( h \) arbitrarily elsewhere). The above equation becomes,

\[
Pr_{s \sim (x D^j)^k} \left[ h(s) = j \right] \geq 1 - \frac{\delta FS(H)}{FS(H) - (4 + 2\epsilon)}.
\]

The theorem follows from the above inequality and Lemma 1.

\[\square\]

### A.2 Proofs from Section 4

#### A.2.1 Proof of Lemma 3

**Proof of Lemma 3** Recall that a mechanism that extracts full surplus exists if the following system has a solution

\[
E_{(v,s) \sim F^j} \left[ q_i(v, s) \right] = u_{i,j}^{SPA}(v_i), \forall i, v, j
\]

The system has a solution for any right hand side \( u_{i,j}^{SPA}(v_i) \) if the set of \( |V_i| \times m \) vectors \( \{F_{i,j}^v\}_{v_i \in V_i, j \in \{1, ..., m\}} \) are linearly independent.

\[\square\]

#### A.2.2 Proof of Proposition 2

**Proof of Proposition 2** For a buyer \( i \), consider a distribution \( \hat{D}^i \) that draws \( \nu_1 \) to \( \nu_n \) independently, and then assigns the highest value to buyer \( i \) with probability 1 (and other
values uniformly at random). Note that $D^i$ is a convex combination of $\hat{D}^i$ and an independent distribution. Therefore it is sufficient to show that the distributions $\hat{D}^i$ are linearly independent.

To see that $\hat{D}^i$ are linearly independent, consider a profile of values such that only the value of one buyer, buyer $i$, is 2. If the true distribution is $i$, then the probability of such a profile is some positive number $c$. Otherwise, the probability of this profile is 0. Since any convex combination of zeros is equal to zero, no convex combination of other distributions can be equal to $\hat{D}^i$.

Note that with probability $(1 - \epsilon)$, the random variables $\nu_1, \ldots, \nu_n$ are directly assigned to $v_1, \ldots, v_n$. In that case, the values do not contain any information about the true distribution, and therefore the probability of error is $1 - 1/n$. Even if a mechanism identifies the true distribution with probability 1 if random variables are not assigned directly, the probability of an error is at least $(1 - \epsilon)(1 - 1/n)$.

\section*{A.2.3 Proof of Proposition 3}

\textbf{Proof of Proposition 3} Immediately from Lemma 3 and Lemma 4.

\section*{A.2.4 Proof of Lemma 6}

\textbf{Proof of Lemma 6} Recall the assumption that $d$ is the dimension of the linear space spanned by the $m$ vectors in $\{\vec{D}^1, \ldots, \vec{D}^m\}$. Let $\{B_1, \ldots, B_d\}$ be a basis. Then for each $j$, we can write $D^j$ as a linear sum of these vectors: $\vec{D}^j = \sum_{\ell=1}^d \alpha_{j\ell} B_{\ell}$.

We consider the outer product $(\otimes \vec{D}^j)^k$. By bilinearity,

\[
(\otimes \vec{D}^j)^k = \left( \otimes \sum_{\ell=1}^d \alpha_{j\ell} B_{\ell} \right)^k = \sum_{1 \leq \ell_1, \ldots, \ell_k \leq d} \alpha_{j\ell_1} B_{\ell_1} \otimes \ldots \otimes \alpha_{j\ell_k} B_{\ell_k} = \sum_{1 \leq \ell_1, \ldots, \ell_k \leq d} \alpha_{j\ell_1} \ldots \alpha_{j\ell_k} B_{\ell_1} \otimes \ldots \otimes B_{\ell_k}
\]
where \( \ell_j \) identifies the term selected from the \( j \)'th multiplier. We next factor out the terms with the same scalar multiplier. That is, for any \( \tau, 1 \leq \tau \leq d \), let \( \gamma_\tau = |\{\ell_j = \tau\}| \) be the number of times that \( \alpha_{j\tau} \) appears in the multiplication. Note that \( \gamma_1 + \ldots + \gamma_d = k \).

Factoring out the terms with the same scalar multiplier, we have

\[
\left( \otimes \vec{D} \right)^k = \sum_{\gamma_1 + \ldots + \gamma_d = k, \gamma_1, \ldots, \gamma_d \geq 0} \alpha_{j_1}^{\gamma_1} \alpha_{j_2}^{\gamma_2} \ldots \alpha_{j_d}^{\gamma_d} \sum_{\gamma_\tau = |\{\ell_j = \tau\}|, \forall \tau} B_{\ell_1} \otimes \ldots \otimes B_{\ell_k}
\]

To simplify notation, define

\[
C_{\gamma_1, \ldots, \gamma_d} = \sum_{\ell_1, \ldots, \ell_k, \gamma_\tau = |\{\ell_j = \tau\}|, \forall \tau} B_{\ell_1} \otimes \ldots \otimes B_{\ell_k}.
\]

That is, \( C_{\gamma_1, \ldots, \gamma_d} \) is the sum of terms that are outer products of \( B_1, \ldots, B_d \), such that in each term \( B_1 \) appears \( \gamma_1 \) times, and so on. Since outer product is not commutative, these products do not have to be the same. For instance, when \( d = 2 \), \( C_{1,2} = B_1 \otimes B_2 \otimes B_2 + B_2 \otimes B_1 \otimes B_2 + B_2 \otimes B_2 \otimes B_1 \). We have

\[
\left( \otimes \vec{D} \right)^k = \sum_{\gamma_1 + \ldots + \gamma_d = k, \gamma_1, \ldots, \gamma_d \geq 0} \alpha_{j_1}^{\gamma_1} \alpha_{j_2}^{\gamma_2} \ldots \alpha_{j_d}^{\gamma_d} C_{\gamma_1, \ldots, \gamma_d}.
\]

An inductive application of [Lemma 5](#) implies that the set of vectors \( \{B_{\ell_1} \otimes \cdots \otimes B_{\ell_k}\}_{\ell_1, \ldots, \ell_k \in [d]} \) are linearly independent. To see this, assume that \( \{B_{\ell_1} \otimes \cdots \otimes B_{\ell_{k-1}}\}_{\ell_1, \ldots, \ell_{k-1} \in [d]} \) are linearly independent. Define \( A = \{B_{\ell_1} \otimes \cdots \otimes B_{\ell_{k-1}}\}_{\ell_1, \ldots, \ell_{k-1} \in [d]} \), and for \( j \in \{1, \ldots, |A|\} \), define \( B_j = \{B_1, \ldots, B_d\} \). Now [Lemma 5](#) states that \( \{B \otimes B_{\ell_1} \otimes \cdots \otimes B_{\ell_{k-1}}\}_{B \in \{B_1, \ldots, B_d\}, \ell_1, \ldots, \ell_{k-1} \in [d]} = \{B_{\ell_1} \otimes \cdots \otimes B_{\ell_k}\}_{\ell_1, \ldots, \ell_k \in [d]} \) are linearly independent. Since each \( C_{\gamma_1, \ldots, \gamma_d} \) is a summation over vectors in \( \{B_{\ell_1} \otimes \cdots \otimes B_{\ell_k}\}_{\ell_1, \ldots, \ell_k \in [d]} \), the vectors in \( \{C_{\gamma_1, \ldots, \gamma_d}\}_{\gamma_1 + \ldots + \gamma_d = k} \) are also linearly independent.

Now note that each \( (\otimes \vec{D})^k \) is expressed as a linear combination of linearly independent vectors, with the linear coefficient on \( C_{\gamma_1, \ldots, \gamma_d} \) being the product \( \alpha_{j_1}^{\gamma_1} \ldots \alpha_{j_d}^{\gamma_d} \). To show linear independence of the set of vectors \( \{((\otimes \vec{D})^k)_j\} \), we only need to show that the set of \( m \) linear coefficients as vectors are linearly independent. More specifically, we show that the \( m \) vectors in the set \( \{(\alpha_{j_1}^{\gamma_1} \ldots \alpha_{j_d}^{\gamma_d})_{\gamma_1 + \ldots + \gamma_d = k}\}_j \) are linearly independent.

The vector \( (\alpha_{j_1}^{\gamma_1} \ldots \alpha_{j_d}^{\gamma_d})_{\gamma_1 + \ldots + \gamma_d = k} \) is the image of the vector \( \vec{\alpha}_{j} = (\alpha_{j_1}, \ldots, \alpha_{j_d}) \) under a mapping \( \nu : \mathbb{R}^d \to \mathbb{R}^{(d+k-1)} \) which evaluates all the \( k \)-th degree monomials in \( \mathbb{R}[x_1, \ldots, x_d] \) at
a point in \( \mathbb{R}^d \). We now show that these \( m \) images \( \nu(\vec{\alpha}_1), \ldots, \nu(\vec{\alpha}_m) \) are linearly independent when \( k = m - d + 1 \).

We will show that for every \( j \), there exists a linear form on \( \mathbb{R}^{(d+k-1)} \) that vanishes at \( \nu(\vec{\alpha}_{j'}) \) for all \( j' \neq j \) and does not vanish at \( \nu(\vec{\alpha}_j) \). This will show that there cannot be any linear dependence among the \( m \) points \( \nu(\vec{\alpha}_j) \).

Since \( \{\vec{D}_j\}_j \) spans a linear space of dimension \( d \), and since \( \{B_1, \ldots, B_d\} \) is a basis of this space, the vectors \( \vec{\alpha}_1, \ldots, \vec{\alpha}_m \) span a \( d \)-dimensional linear space. Without loss of generality, consider \( \vec{\alpha}_1 \), we can find \( d - 1 \) other vectors that are linearly independent with \( \vec{\alpha}_1 \). Therefore, we can find a linear form \( f_1 : (y_1, \ldots, y_d) \mapsto \beta_1 y_1 + \cdots + \beta_d y_d \) which vanishes at all these \( d - 1 \) vectors but does not vanish at \( \vec{\alpha}_j \). Without loss of generality, let the remaining \( m - d \) vectors be \( \vec{\alpha}_{d+1}, \ldots, \vec{\alpha}_m \). Note that since each \( \vec{D}_j \) represents a probability distribution, its entries sum to one. Therefore, no two \( \vec{\alpha}_j \) and \( \vec{\alpha}_{j'} \) are scalar copies of each other, i.e., there are no \( j \neq j' \) such that \( \alpha_{j\ell} = \zeta \alpha_{j'\ell} \) for each \( \ell \), for some \( \zeta \). Thus, for each \( j' = d + 1, \ldots, m \), we can find a linear form \( f_{j'} \) such that \( f_{j'} \) vanishes at \( \vec{\alpha}_{j'} \) but does not vanish at \( \vec{\alpha}_j \). Now consider the product of these \( m - d + 1 \) linear forms, 

\[
  f = f_1 f_{d+1} \cdots f_m.
\]

If we take \( k \) to be \( m - d + 1 \), \( f \) itself is a linear form on \( \mathbb{R}^{(d+k-1)} \), and can be evaluated at \( \nu(\vec{\alpha}_1), \ldots, \nu(\vec{\alpha}_m) \), and

\[
  f(\nu(\vec{\alpha})) = f_1(\vec{\alpha}) f_{d+1}(\vec{\alpha}) \cdots f_m(\vec{\alpha}), \quad \forall \vec{\alpha} \in \mathbb{R}^d.
\]

By construction, \( f(\nu(\vec{\alpha}_j)) = 0 \) for all \( j \neq 1 \) and \( f(\nu(\vec{\alpha}_1)) \neq 0 \). Since the choice of \( \vec{\alpha}_1 \) was arbitrary, the construction works for arbitrary \( \vec{\alpha}_j \), and so \( \nu(\vec{\alpha}_1), \ldots, \nu(\vec{\alpha}_m) \) are linearly independent for \( k = m - d + 1 \). This completes the proof.
A.2.5 Proof of Lemma 7

Before we prove Lemma 7, we need some definitions and a technical lemma. Given \( y_1, \ldots, y_m \in \mathbb{R} \), consider an \( m \) by \( m \) matrix \( \bar{V} \) defined as follows.

\[
\bar{V} = \begin{pmatrix}
1 & y_1 & \cdots & y_1^{m-2} & (1 + \frac{3y_1}{1-y_1})(1 + y_1)^{m-2} \\
1 & y_2 & \cdots & y_2^{m-2} & (1 + \frac{3y_2}{1-y_2})(1 + y_2)^{m-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & y_m & \cdots & y_m^{m-2} & (1 + \frac{3y_m}{1-y_m})(1 + y_m)^{m-2}
\end{pmatrix}.
\] (17)

Matrix \( \bar{V} \) is closely related to an \( m \) by \( m \) Vandermonde matrix \( V \).

\[
V = \begin{pmatrix}
1 & y_1 & \cdots & y_1^{m-2} & y_1^{m-1} \\
1 & y_2 & \cdots & y_2^{m-2} & y_2^{m-1} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & y_m & \cdots & y_m^{m-2} & y_m^{m-1}
\end{pmatrix}.
\]

The Vandermonde matrix \( V \) has full rank (Bellman, 1997). In the following lemma, we use this fact to show that the matrix \( \bar{V} \) also has full rank.

**Lemma 15.** For distinct \( y_1, \ldots, y_m \), none equal to \( \pm 1 \), the \( m \) by \( m \) matrix \( \bar{V} \) defined in Equation 17 has rank \( m \).

**Proof.** The proof strategy is to convert the matrix \( \bar{V} \) to the Vandermonde matrix using operations that preserve the rank. Note that

\[
(1 + y)^{m-2} = \sum_{i=0}^{m-2} \binom{m-2}{i} y^i.
\]

Therefore, by multiplying each column \( i \) by \( \binom{m-2}{i} \) and subtracting it from the last column,
we can convert the matrix to

$$
\begin{pmatrix}
1 & y_1 & \ldots & y_1^{m-2} & \frac{3y_1}{1-y_1}(1+y_1)^{m-2} \\
1 & y_2 & \ldots & y_2^{m-2} & \frac{3y_2}{1-y_2}(1+y_2)^{m-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & y_m & \ldots & y_m^{m-2} & \frac{3y_m}{1-y_m}(1+y_m)^{m-2} \\
\end{pmatrix},
$$

which is equivalent to

$$
\begin{pmatrix}
1 & y_1 & \ldots & y_1^{m-2} & \frac{3y_1}{1-y_1}(1+y_1)^{m-3} \\
1 & y_2 & \ldots & y_2^{m-2} & \frac{3y_2}{1-y_2}(1+y_2)^{m-3} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & y_m & \ldots & y_m^{m-2} & \frac{3y_m}{1-y_m}(1+y_m)^{m-3} \\
\end{pmatrix}.
$$

Now divide the last column by 3, and multiply each row $j$ of $\bar{V}$ by $1 - y_j$. The result is the following matrix.

$$
\begin{pmatrix}
1(1-y_1) & y_1(1-y_1) & \ldots & y_1^{m-2}(1-y_1) & y_1(1+y_1)^{m-3} \\
1(1-y_2) & y_2(1-y_2) & \ldots & y_2^{m-2}(1-y_2) & y_2(1+y_2)^{m-3} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1(1-y_2) & y_m(1-y_2) & \ldots & y_m^{m-2}(1-y_2) & y_m(1+y_m)^{m-3} \\
\end{pmatrix}.
$$

The remaining operations are on columns. So we focus on a fixed row and drop the index $j$ for simplicity. A row is

$$1(1-y), y(1-y), \ldots, y^{m-2}(1-y), y(1+y)^{m-3}.$$

Index columns from 1 to $m$. For column $\ell$ from 1 to $m - 1$, the $\ell$’th entry is $y^{\ell-1}(1-y)$. Now replace the element in each column $\ell$ from 2 to $m - 1$ with the sum of all elements from 1 to $\ell$. Note that the sum $\sum_{i=1}^{\ell} y^{i-1}(1-y)$ is equal to $1 - y^\ell$. The result is

$$1 - y, 1 - y^2, \ldots, 1 - y^{m-1}, y(1+y)^{m-3}.$$
Now for each $\ell$ from 1 to $m - 2$, multiply the $\ell$'th element by $\left(\begin{array}{c} m-3 \\ \ell-1 \end{array}\right)$ and add it to the last element. Since $y(1 + y)^{m-3} = \sum_{\ell=0}^{m-3} \left(\begin{array}{c} m-3 \\ \ell \end{array}\right) y^{\ell+1}$, the result is

$$1 - y, 1 - y^2, \ldots, 1 - y^{m-1}, \sum_{\ell=1}^{m-2} \left(\begin{array}{c} m - 3 \\ \ell - 1 \end{array}\right).$$

Divide the last column by $\sum_{\ell=1}^{m-2} \left(\begin{array}{c} m - 3 \\ \ell - 1 \end{array}\right)$,

$$1 - y, 1 - y^2, \ldots, 1 - y^{m-1}, 1.$$

Multiply the first $m - 1$ columns by $-1$, and subtract the last column from it to obtain

$$y, y^2, \ldots, y^{m-1}, 1.$$

This is the row of the Vandermonde matrix (permuted such that the first column appears last). Since the Vandermonde matrix has rank $m$, we conclude that so should the matrix $\bar{V}$. 

We now prove Lemma 7.

**Proof of Lemma 7.** To prove the lemma, we need to show existence of $\beta$ of size $m \times |V_1|$ such that $\beta \cdot \bar{1} = \sum_{j,v_1} \beta_{j,v_1} = 0, \sum_j \beta_{j,2} \neq 0$, and

$$\sum_{j,v_1} \beta_{j,v_1} E_{v,s_1,\ldots,s_k \sim D^j} [p_1(v_{-1}, s)|v_1] = 0, \forall p_1. \quad (18)$$

Recall that $\bar{F}^j_{v_1}$ is the vector representation of probability of $(v_{-i}, s)$ conditioned on $v_1$ in distribution $j$. In this proof it is convenient to represent $F$ as a $|V_1| \cdot m$ by $|V_{-i}| \cdot |S|$ matrix that stacks the vectors $\{\bar{F}^j_{v_1}\}_{j,v_1}$ on top of each other, that is,

$$F = \begin{pmatrix} \ldots & v_{-1}, s & \ldots \\ \vdots & \vdots & \vdots \\ j, v_1 \begin{pmatrix} \ldots & F^j(v_{-1}, s|v_1) & \ldots \end{pmatrix} \end{pmatrix}.$$
Using this notation we can write

\[ \sum_{j,v_1} \beta_{j,v_1} E_{v_1,s_1,...,s_k} [p_1(v_{-1}, s)|v_1] = \beta \cdot F \cdot p_1. \]

Therefore, to show Equation 18 it is sufficient to show that \( \beta \cdot F = 0 \).

For any \( j \), since the samples in \( s \) are drawn independently,

\[ \vec{F}_j(v_{-1}, s) = \vec{D}_j^{(v_{-1}|v_1)} \cdot \vec{D}_j^{(s_1)} \cdot \ldots \cdot \vec{D}_j^{(s_k)}. \]

Recall the assumption of the lemma that \( k \leq m - 2 \). Let \( v_{-1}^1 = 2 \) and \( v_{-1}^2 = 3 \). By construction of Example 5, \( D_j(s_\ell) = \alpha_j \) if the sample is a “match”, that is, \( s_\ell = (v_1^1, v_{-1}^1) \) or \( s_\ell = (v_1^2, v_{-1}^2) \), and otherwise \( D_j(s_\ell) = 1 - \alpha_j \). Therefore, to abbreviate notation we simply assume that \( s \in [0, m - 2] \) encodes the number of matches, and thus \( \Pr_j[s] = (\alpha_j)^s (1 - \alpha_j)^{m-2-s} \). We therefore simply represent \( F \) as a \( 2m \) by \( 2(m-1) \) matrix as follows

\[
F = \begin{pmatrix}
\vdots & \ldots & v_{-1}, \ell & \ldots
\vdots
\end{pmatrix}
\begin{pmatrix}
\Pr_j[v_{-1}|v_1](\alpha_j)^\ell (1 - \alpha_j)^{m-2-\ell}
\end{pmatrix}.
\]

Let \( V \) be an \( m \) by \( m-1 \) Vandermonde matrix, that is,

\[
V = \begin{pmatrix}
1 & (\alpha_1/(1 - \alpha_1)) & \ldots & (\alpha_1/(1 - \alpha_1))^{m-2} \\
1 & (\alpha_2/(1 - \alpha_2)) & \ldots & (\alpha_2/(1 - \alpha_2))^{m-2} \\
\vdots & \vdots & \vdots & \vdots \\
1 & (\alpha_m/(1 - \alpha_m)) & \ldots & (\alpha_m/(1 - \alpha_m))^{m-2}
\end{pmatrix}.
\]

Let \( \Delta_{v_1,v_{-1}} \) be an \( m \) by \( m \) diagonal matrix such that \( \Delta_{v_1,v_{-1}}(j,j) = \Pr_j[v_{-1}|v_1](1 - \alpha_j)^{m-2} \).

Using this notation, rewrite \( F \) as

\[
F = \begin{pmatrix}
\Delta_{v_1^1,v_{-1}^1} \cdot V & \Delta_{v_1^1,v_{-1}^2} \cdot V \\
\Delta_{v_1^2,v_{-1}^1} \cdot V & \Delta_{v_1^2,v_{-1}^2} \cdot V
\end{pmatrix}.
\]

Recall that \( \beta \) is a vector of size \( 2m \). Let \( \beta \) be composed of two parts \( \beta_L \) and \( \beta_H \), each of size
m. That is, \( \beta = (\beta_L, \beta_H) \). Now the equation \( \beta \cdot F = 0 \) becomes

\[
(\beta_L \Delta_{v_1,v_{1-1}} + \beta_H \Delta_{v_2,v_{1-1}}) V = 0, \tag{19}
\]

\[
(\beta_L \Delta_{v_1,v_{2-1}} + \beta_H \Delta_{v_2,v_{2-1}}) V = 0. \tag{20}
\]

For a reason to become clear shortly, consider adding an extra column as the \( m \)'th column to matrix \( V \). In particular, consider an \( m \) by \( m \) matrix \( \bar{V} \), whose first \( m - 1 \) columns are identical to that of \( V \), and the entry in row \( j \) and column \( m \) is

\[
(1 + \frac{3y_j}{1 - y_j})(1 + y_j)^{m-2},
\]

where \( y_j = \alpha_j/(1 - \alpha_j) \). By Lemma 15, there exists \( \alpha_1, \ldots, \alpha_m \) such that matrix \( \bar{V} \) is invertible. In particular, we have

\[
\bar{V} = \begin{pmatrix}
1 & y_1 & \ldots & y_1^{m-2} & (1 + \frac{3y_1}{1 - y_1})(1 + y_1)^{m-2} \\
1 & y_2 & \ldots & y_2^{m-2} & (1 + \frac{3y_2}{1 - y_2})(1 + y_2)^{m-2} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & y_m & \ldots & y_m^{m-2} & (1 + \frac{3y_m}{1 - y_m})(1 + y_m)^{m-2}
\end{pmatrix}.
\]

Now Lemma 15 applies to imply that \( \bar{V} \) has full rank. Therefore, there exists a solution \( w \) to the following system \( w \cdot \bar{V} = (0, \ldots, 0, 1) \) of \( m \) equations and \( m \) unknowns. The first \( m - 1 \) equations are equivalent to Equation 19 together with Equation 20. Therefore, \( \beta \cdot F = 0 \) if

\[
\beta_L \Delta_{v_1,v_{1-1}} + \beta_H \Delta_{v_2,v_{1-1}} = w,
\]

\[
\beta_L \Delta_{v_1,v_{2-1}} + \beta_H \Delta_{v_2,v_{2-1}} = w.
\]

Solving these two equations gives

\[
\beta_L = w(\Delta_{v_2,v_{2-1}} - \Delta_{v_2,v_{1-1}}) \cdot (\Delta_{v_1,v_1} \cdot \Delta_{v_2,v_{2-1}} - \Delta_{v_1,v_{2-1}} \cdot \Delta_{v_2,v_{2-1}})^{-1},
\]

\[
\beta_H = w(\Delta_{v_1,v_{2-1}} - \Delta_{v_1,v_{1-1}}) \cdot (\Delta_{v_1,v_1} \cdot \Delta_{v_1,v_{2-1}} - \Delta_{v_1,v_{2-1}} \cdot \Delta_{v_1,v_{1-1}})^{-1}.
\]

To summarize the arguments so far, if \( \beta_L \) and \( \beta_H \) satisfy the above equations, then \( \beta \cdot F = 0 \) and therefore \( \sum_j \beta_j E_{v_{-1} \sim D_{-1}}[p_1(v_{-1})|v_1] = 0 \). We will next show that additionally, \( \sum_j \beta_{j,2} \neq 0 \).
We next show that this implies that is the \( m \)th column of matrix \( \bar{V} \). By construction of Example 5, we have \( \Pr \bar{V} = \begin{pmatrix} \alpha_j & \alpha_j & \cdots & \alpha_j \end{pmatrix} \). So to complete the proof, we need to argue that Expression 21 is the \( m \)th column of matrix \( \bar{V} \). The \( j \)th element in the column vector is

\[
\beta_H \cdot \bar{1} = \frac{(\Pr_j[v^2_{-1}|v_1^1] - \Pr_j[v^1_{-1}|v_1^1])(1 - \alpha_j)^{m-2}}{(\Pr_j[v^1_{-1}|v_1^1] - \Pr_j[v^2_{-1}|v_1^1])}(1 - \alpha_j)^{2(m-2)}.
\]

By construction of Example 5, we have \( \Pr_j[v^1_{-1}|v_1^1] = \alpha_j/(\alpha_j + 2(1 - \alpha_j)) \), \( \Pr_j[v^2_{-1}|v_1^1] = 2(1 - \alpha_j)/(\alpha_j + 2(1 - \alpha_j)) \), \( \Pr_j[v^1_{-1}|v_1^2] = (1 - \alpha_j)/((1 - \alpha_j) + 2\alpha_j) \), and \( \Pr_j[v^2_{-1}|v_1^2] = 2\alpha_j/((1 - \alpha_j) + 2\alpha_j) \). Therefore, the \( j \)th element becomes

\[
\beta_H \cdot \bar{1} = \frac{2(1-\alpha_j)-\alpha_j}{\alpha_j+2(1-\alpha_j)} \cdot (1 - \alpha_j)^{-(m-2)}
\]

Substituting \( \alpha_j = y_j/(1 + y_j) \),

\[
\beta_H \cdot \bar{1} = \frac{2 - y_j}{2(1 + y_j)^2} \cdot (1 + y_j)^{(m-2)}
\]

So to complete the proof, we need to argue that Expression 21 is the \( m \)th column of matrix \( \bar{V} \). The \( j \)th element in the column vector is

\[
\beta_H \cdot \bar{1} = w(\Delta_v^1, v^2_{-1} - \Delta_v^1, v^1_{-1}) \cdot (\Delta_v^2, v^1_{-1} \cdot \Delta_v^1, v^2_{-1} - \Delta_v^1, v^2_{-1} \cdot \Delta_v^1, v^1_{-1})^{-1} \cdot \bar{1}.
\]
We have argued that Expression $21$ is the $m$'th column of matrix $\bar{V}$, which completes the proof.

\section*{A.2.6 Proof of \textbf{Proposition 4}}

\textit{Proof of Proposition 4} Consider the following extension of Example 5. The set of values of buyer 1 is $V_1 = \{v_1^1 = 2, v_1^2 = 3, v_1^{m-d+3}, v_1^{m-d+4}, \ldots, v_1^m\}$. The set of possible profiles of other buyers is $\{v_{-1}^1, v_{-1}^{m-d+3}, v_{-1}^{m-d+4}, \ldots, v_{-1}^m\}$. Similarly to Example 5, assume that $\max_{j \neq 1} v_1^j = \max_{j \neq 1} v_2^j = 1$. Now consider the following $d$ bases. The two basis $B^1$ and $B^2$ are defined as in Example 5 that is

$$B^1 = \begin{pmatrix} v_{-1}^1 & v_{-1}^2 \\ 1/2 & 0 \\ 3 & 1/2 \end{pmatrix}, \quad B^2 = \begin{pmatrix} v_{-1}^1 & v_{-1}^2 \\ 0 & 1/2 \\ 3 & 0 \end{pmatrix},$$

with probability zero on every other value profile. For $\ell = m - d + 3, \ldots, m$, define a distribution $B^\ell$ that puts probability 1 on profile $(v_1^\ell, v_{-1}^\ell)$, and probability zero everywhere else. Note that defined bases are linearly independent, and that the number of bases is $d$. Now define the $m$ distributions as follows. For $j = 1, \ldots, m - d + 2$, define $D^j$ similar to Example 5 $D^j = \alpha_j B^1 + (1 - \alpha_j) B^2$. For $j = m - d + 3, \ldots, m$, define $D^j = B^j$. Note that the number of distributions is $m$ and that the dimension of the linear space spanned by them is $d$. We show that no mechanism can extract full surplus with $k \leq m - d$ samples. The argument parallels the argument following Example 5.

If a surplus extracting mechanisms exists, then we must have

$$\mathbb{E}_{v,s^1,\ldots,s^k \sim D^j} [p_1(v,s)|v_1] = v_1, \quad \forall j \in \{1, \ldots, m - d + 2\}, v_1 \in \{2, 3\}.$$ 

If $v_{-1} = v_{-1}^1$ or $v_{-1} = v_{-1}^2$, buyer 1 must win the item. Incentive compatibility implies that in this case, the payment of agent 1 does not depend in its own report. Therefore we write the payment as $p_1(v_{-1,s})$, and must have

$$\mathbb{E}_{v,s^1,\ldots,s^k \sim D^j} [p_1(v_{-1,s})|v_1] = v_1, \quad \forall j \in \{1, \ldots, m - d + 2\}, v_1 \in \{2, 3\}.$$ 

45
The argument following Example 5 implies that for the function \( p_1 : V_{-1} \times S \to \mathbb{R} \), and for any profile \( \beta = (\beta_{j,v_1})_{j,v_1} \) satisfying i) \( \sum_{j,v_1} \beta_{j,v_1} = 0 \) and ii) \( \sum_j \beta_{j,2} \neq 0 \), we have

\[
\sum_{j,v_1} \beta_{j,v_1} E_{v_1,v_2,\ldots,s_k \sim D_j} [p_1(v_{-1}, s)|v_1] \neq 0.
\]

Now Lemma 7 can be applied to show that no such function \( p_1 \) exists. In particular, consider the set of \( m' = m - d + 2 \) distributions \( \{D_1, \ldots, D^{m-d+2}\} \), and \( k \leq m' - 2 = m - d \). Lemma 7 shows that no function \( p_1 \) satisfying the above inequalities exists. \( \square \)