Full Surplus Extraction from Samples*

Hu Fu† Nima Haghpanah‡ Jason Hartline§
Robert Kleinberg¶

February 8, 2021

Abstract

We study whether an auctioneer who has only partial knowledge of the distribution of buyers' valuations can extract the full surplus. There is a finite number of possible distributions, and the auctioneer has access to a finite number of samples (independent draws) from the true distribution. Full surplus extraction is possible if the number of samples is at least the difference between the number of distributions and the dimension of the linear space they span, plus one. This bound is tight. The mechanism that extracts the full surplus uses the samples to construct contingent payments, and not for statistical inference.

Keywords: Mechanism design, unknown distributions, samples.

JEL codes: D82, D44.

*We thank Nageeb Ali, David Bindel, Nick Gravin, Bart Lipman, Luca Rigotti, Ron Siegel, Bruno Strulovici, and Gjergji Zaimi, and seminar participants at Stanford, Pennsylvania Economic Theory Conference, SAET, North American Summer Meeting of the Econometric Society, and Erice Workshop on Stochastic Methods in Game Theory for helpful discussions.

†Department of Computer Science, University of British Columbia, hufu@cs.ubc.ca.

‡Corresponding author. Department of Economics, Penn State University, nima.haghpanah@gmail.com.

§EECS Department, Northwestern University, hartline@eecs.northwestern.edu. Supported by NSF CAREER Award CCF-0846113 and NSF CCF-1101717.

¶Department of Computer Science, Cornell University, rdk@cs.cornell.edu. Supported by NSF award AF-0910940, a Microsoft Research New Faculty Fellowship, and a Google Research Grant.

1 Introduction

Cremer and McLean (1988) design an auction that generically extracts the full surplus for selling an indivisible product to one of many potential buyers. The auction given by Crémer and McLean requires full knowledge of the distribution of buyers' valuations, but in reality the auctioneer rarely has access to such detailed information. This paper studies the possibility of full surplus extraction when the auctioneer has only partial knowledge of the distribution.

In our model, the true distribution of buyers' valuations is unknown to the auctioneer and all buyers. The true distribution belongs to a finite commonly-known set of possible distributions. The auctioneer has access to a finite number of independent draws, or *sample*, from the true distribution. Before observing the samples, the auctioneer commits to a mechanism that specifies the allocation and payments based on the bids and the realized samples. After the buyers bid, samples are revealed and the allocation and the payments are specified.

Does there exist a mechanism that extracts the full surplus for each possible distribution? The auctioneer cannot simply ask the buyers to report the true distribution and punish them if they disagree, since the buyers do not know the true distribution either. Thus full surplus extraction is impossible without samples, since the auctions that extract the full surplus for different distributions may be different. Even with samples, it may seem impossible to extract the full surplus since, if samples are used for inference, the auctioneer needs infinitely many samples to infer the true distribution with certainty.

We show that it is in fact possible to extract the full surplus. We identify the number of samples that are necessary and sufficient for doing so. 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.[1] This number is at least one and at most the number of distributions minus one. If the distributions are linearly independent, the number of samples is one. In particular,

[1]It is also necessary that full surplus extraction is possible for every distribution in the set.

for any two distributions, one sample suffices.

The mechanism that extracts the full surplus uses the samples to construct contingent payments, and not for statistical inference. The mechanism is a second price auction plus side payments. The side payment of each buyer depends on others' bids and the realized samples. The second price auction maximizes the total surplus. The surplus is fully extracted if each buyer's interim expected utility is zero for each value of the buyer and each distribution, that is, the expected side payment equals the expected utility from the second price auction.

The problem is thus to verify whether a solution to a system of linear equalities exists, where the variables are the side payments. A solution exists if for each buyer, the rows of the following matrix are linearly independent: The matrix specifies, for each distribution and value of the buyer, indexing the rows, the conditional probability of each profile of others' values and samples, indexing the columns. In the special case where the true distribution is known (the set of distributions is a singleton), the matrix is an extension of the one constructed by Cremer and McLean (1988), where the columns are not only indexed by profiles of others' values, but also samples. In general, our matrix simply stacks such matrices, one for each distribution in the set, on top of each other.

A main technical step in our proof is to identify how many samples are required to guarantee linear independence. The key observation is that for a given distribution and value of a buyer, the profile of others' values and each sample are independently distributed. Thus, each row of the matrix is an outer product of several vectors, i.e., a vector that specifies the conditional probabilities of others' values, and multiple copies of a vector that specifies sample probabilities. We identify exactly how many times the vector of sample probabilities must be multiplied by itself to guarantee linear independence. The fact that this is achieved with relatively few samples is one of the main insights of our paper. Samples are used to create linear independence.

There are two ways to interpret our results. The first interpretation views full surplus extraction as a critique of the features of commonly studied auction models. In this view, our work reexamines these features in order to identify one that is responsible for the unsettling prediction of full surplus extraction. Instead of identifying an assumption that is responsible, we identify one that may be weakened. This assumption is that the true distribution is commonly known. 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), if buyers can collude (Laffont and Martimort, 2000), or if more than one auctioneer compete (Peters, 2001).

The second interpretation of our results views full surplus extraction as a proof of concept. In this view, the possibility of full surplus extraction demonstrates the power of contingent payments to design robust mechanisms. With only partial knowledge of the distribution, one can design mechanisms that perform as well as the mechanisms that have full knowledge of the distribution. The existing literature on contingent payments (Hansen, 1985; DeMarzo et al., 2005) suggests that an auctioneer can increase profit by relating payments to observable events that are correlated with buyers' values (e.g., revenue sharing in auctions for oil lease). In contrast, our work highlights the role of contingent payments to obtain robustness.

1.1 Related Work

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. Similar to ours, these papers allow for uncertainty over the true distributions. Nevertheless, unlike ours, these papers require all buyers to share a common prior distribution over distributions. In contrast, in our settings buyers may have different prior beliefs over the set of possible distributions.

The linear independence property, often referred to as statistical detectability, has also been used in repeated games (Fudenberg et al., 1994), moral hazard in teams (Rahman and Obara, 2010; Rahman, 2012), and mechanism design (Rahman, 2010). However, this condition is often used to design transfer schemes to enforce incentive constraints, and not to resolve uncertainty as used in our setting. That is, in equilibrium, the strategies of all players are known. Thus, the role of transfer is to deter deviations by ensuring that each player prefers the distribution of payoffs it receives by following the equilibrium strategy to what she would receive by deviating. In contrast, the transfers in our setting are designed to ensure robustness of the mechanism with respect to an unknown state.

Other papers have studied the design of mechanisms without full knowledge of the distribution of types, and without access to samples. In Bergemann et al. (2016), the auctioneer knows the distribution of values but not the agent's beliefs. Since no sample is available, the auctioneer cannot extract the full surplus for all possible distributions. As a result, Bergemann et al. (2016) focus on a worst case objective. A common approach in designing mechanisms without full knowledge of the distribution is to use the agents' reported types for inference. For example, in Goldberg et al. (2006) and 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) and Baliga and Vohra (2003), an agent's demand, and consequently her 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), who show that an auction that uses a single sample, namely the VCG auction with a reserve equal to the sample, 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.

2 The Setting

A single indivisible product is to be assigned to at most one of n potential buyers. Each buyer i privately knows her 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.

An 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 probability that i gets the product. Buyers have linear utilities. That is, the utility of buyer i with value $v_i \in V_i$ for receiving the product 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.[2]

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 interpreted as the external information available to the mechanism via market research. We assume that it is commonly known to all buyers that the true distribution belongs to $\mathcal{F}$. Nevertheless, buyers may not agree about the identity of the true distributions (they may have different priors, or may not be equipped with a prior at all). A buyer bids in the mechanism knowing only her own value, and before the signal is revealed. Our goal is to design a mechanism that extracts the full surplus in expectation over any distribution in $\mathcal{F}$, to be

[2] In 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* Θ_i. Each buyer i has a valuation function $v_i : A \times \Theta_i \to \mathbb{R}$. 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 $v_i(a, \theta_i) - p$.

formalized throughout this section. Crucially, since the buyers may disagree about the identity of the true distribution, the mechanism cannot simply elicit the true distribution from agents by asking them to report the true distribution and punish them if they disagree.

We mainly focus on a special case where the signal s consists of independent draws from the distribution of values. In particular, we say that 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^ℓ as a *sample*, and abusing notation, represent $\mathcal{F}$ as $\{D^1, \ldots, D^m\}$. Note that given the independence assumption, for 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 to allocations $x : V_1 \times \ldots \times V_n \times S \to X$. The function p is the payment function mapping actions and signal to payments $p : V_1 \times \ldots \times V_n \times S \to \mathbb{R}^n$.[3]

A mechanism is dominant strategy incentive compatible (DSIC) if for any buyer i, values v_i, v_i', v_{-i}, and signal s,

$$v_i x_i(v_i, v_{-i}, s) - p_i(v_i, v_{-i}, s) \geq v_i x_i(v_i', v_{-i}, s) - p_i(v_i', v_{-i}, s).$$

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

$$\mathbf{E}_{(v,s)\sim F^j}\left[v_i x_i(v, s) - p_i(v, s) | v_i\right] \geq 0. \tag{1}$$

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.[4] Since the participation constraint holds for any buyer i and any distribution in $\mathcal{F}$, participation is an equilibrium as long as $\mathcal{F}$ is commonly known to all buyers,

[3] By risk neutrality there is no loss in focusing on deterministic payment rules.

[4] An alternative notion of incentive compatibility is Bayesian incentive compatibility

regardless of any additional information that the buyers may have about the identity of the true distribution (e.g., in form of heterogeneous prior beliefs over $\mathcal{F}$).

Definition 1. *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*,*

$$\mathbf{E}_{(v,s)\sim F^j}\left[\sum\nolimits_i p_i(v,s)\right] = \mathbf{E}_{(v,s)\sim F^j}\left[\max\nolimits_i v_i\right].$$

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* $\mathcal{F}$*.*[5]

A special case of our model is when the true distribution is known. We say that 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

(BIC), which requires that for all i, v_i, v_i', and distribution $F^j \in \mathcal{F}$,

$$\mathbf{E}_{(v,s)\sim F^j}\left[v_i x_i(v,s) - p_i(v,s)|v_i\right] \geq \mathbf{E}_{(v,s)\sim F^j}\left[v_i x_i(v_i', v_{-i}, s) - p_i(v_i', v_{-i}, s)|v_i\right].$$

BIC is perhaps a more natural notion of incentive compatibility than DSIC in our setting, given that we define the participation constraint also in expectation. DSIC implies BIC. Thus our main result, which is an existence result, holds if we replace DSIC with BIC.

Recall that the allocation is allowed to be randomized. A more restrictive definition than DSIC 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.

[5]If 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 auctioneer's prior belief over $\mathcal{F}$.

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 Crémer-McLean result, we start with some notation for distributions. Consider $j \in \{1, \ldots, m\}$, $i \in \{1, \ldots, n\}$, and $v_i \in V_i$. Let $\vec{D}^j = \Big(D^j(v)\Big)_{v \in V}$ be the distribution D^j over v, represented as a vector of size $|V|$. Let

$$\vec{D}^j_{v_i} = \Big(D^j(v_{-i}|v_i)\Big)_{v_{-i} \in V_{-i}}$$

be the distribution D^j over v_{-i} conditioned on v_i, represented as a vector of size $|V_{-i}|$. A valuation distribution D^j satisfies the *Crémer-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 Crémer-McLean theorem in our setting.

Theorem 0 (Cremer and McLean, 1988)**.** *Assume that the true distribution is known. There exists an auction that extracts full surplus for $\mathcal{F} = \{F^1\}$ if the marginal distribution on value profiles D^1 satisfies the Crémer-McLean condition.*[6]

Let us call 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. Notice that if a mechanism extracts the full surplus on $\mathcal{F}$, then it must also

[6]Further, Cremer and McLean (1988) show that the condition of Theorem 0 is necessary for full surplus extraction if the setting is generalized to the one described in footnote 2.

extract surplus on a singleton set $\{F^j\}$ for all j. We later give conditions on the signal so that full surplus extraction extends to a set of distributions $\mathcal{F}$.

3 The Main Result

Our main result identifies the number of samples that are necessary and sufficient for full surplus extraction. The number of samples is the number of distributions minus the *dimension* of the set of the set of distributions, plus one.

The dimension of the set of distributions is defined as follows. 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. 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. 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.[7] Recall that a set of distributions $\mathcal{F}$ is 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 1. *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 dimension d if and only if $k \geq m - d + 1$.*

To fully extract the surplus on $\mathcal{F}$, 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. 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.

[7]The dimension can not be 1. Otherwise, distributions must be a scaled versions of each other. This is not possible for distinct probability distributions.

The following two subsections prove the sufficient and necessary directions of the main theorem.

3.1 Sufficient Number of Samples

We first define a class of mechanisms that extend the Crémer-McLean construction to our setting, without requiring $\mathcal{F}$ 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. 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 2. *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 product 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 $u_i^{\mathrm{SPA}}(v)$ be the ex post utility of buyer i in the second price auction.[8] For each buyer i with value v_i and

[8] 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.

distribution j, let

$$u_{i,j}^{\text{SPA}}(v_i) = \mathbf{E}_{v\sim D^j}\left[u_i^{\text{SPA}}(v)\right]$$

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

$$\mathbf{E}_{(v,s)\sim F^j}\left[q_i(v,s)|v_i\right] = u_{i,j}^{\text{SPA}}(v_i), \forall i, v_i, j. \tag{2}$$

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 2 exist.

The lemma below specifies conditions for existence of such side payment functions. It provides conditions on $\mathcal{F}$ under which a solution exists for *any* right hand side u^{SPA}, without assuming any structure on u^{SPA}. In particular, it shows that for any i, the set of possible interim expected side payments $\{(\mathbf{E}_{(v,s)\sim F^j}[q_i(v,s)|v_i])_{j,v_i}\}_{q_i}$ is equal to $\mathbb{R}^{m\times|V_i|}$. 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. Consider $j \in \{1,\ldots,m\}$, $i \in \{1,\ldots,n\}$, and $v_i \in V_i$. Let

$$\vec{F}_{v_i}^j = \Big(F^j(v_{-i}, s|v_i)\Big)_{v_{-i}\in V_{-i}, s\in S}$$

be the distribution F^j over (v_{-i}, s) conditioned on v_i, represented as a vector of size $|V_{-i}| \times S$. Notice that 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 1. *Consider a set of distributions $\mathcal{F} = \{F^1,\ldots,F^m\}$. There exists a Crémer-McLean mechanism with samples that extracts the full surplus on $\mathcal{F}$, 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.*

We now state and prove the sufficiency part of Theorem 1.

Proposition 1. *Consider a Crémer-McLean k-sample set of distributions $\mathcal{F}$ of size m and dimension 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 1 is a corollary of Lemma 1 and the lemma below.

Lemma 2. *Consider a Crémer-McLean k-sample set of distributions $\mathcal{F}$ of size m and dimension d. If $k \geq m-d+1$, then for each bidder i, the set of $|V_i| \times m$ vectors $\{\vec{F}^j_{v_i}\}_{v_i \in V_i, j \in \{1,\ldots,m\}}$ are linearly independent.*

We use outer products on vectors to simplify the proof of Lemma 2 notationally. 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.[9] We use the following standard property of outer products.

Lemma 3. *Consider a set of linearly independent vectors $\mathcal{A} = \{A^1, \ldots, A^m\}$ and, for each $j = 1, \ldots, m$, a set $\mathcal{B}_j$ of linearly independent vectors. The set of vectors in the set $\{B \otimes A^j\}_{j \in \{1,\cdots,m\}, B \in \mathcal{B}_j}$ (of size $\sum_j |\mathcal{B}_j|$) 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 4. *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 $k \geq m - d + 1$.*

We next use Lemma 3 and Lemma 4 to prove Lemma 2.

Proof of Lemma 2. Fix a buyer i. Recall the definition of $\vec{F}$, for all j, and v_i,

$$\vec{F}^j_{v_i} = \Big(F^j(v_{-i}, s | v_i) \Big)_{v_{-i} \in V_{-i}, s \in S}.$$

[9] 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 $(3,4) \otimes (1,2) = (3(1,2), 4(1,2)) = (3,6,4,8)$.

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

$$\vec{F}^j_{v_i} = \Big(D^j(v_{-i}|v_i) \times D^j(s^1) \times \ldots \times D^j(s^k)\Big)_{v_{-i},s}.$$

Using the outer product notation, this simplifies to

$$\vec{F}^j_{v_i} = \vec{D}^j_{v_i} \otimes (\otimes \vec{D}^j)^k. \tag{3}$$

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 4, 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 Crémer-McLean condition, the $|V_i|$ vectors in $\{\vec{D}^j_{v_i}\}_{v_i \in V_i}$ are linearly independent. We can then apply Lemma 3, to conclude that the vectors in $\{\vec{F}^j_{v_i}\}_{v_i \in V_i, j \in \{1,\ldots,m\}}$ 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 $\mathcal{B}_j = \{\vec{D}^j_{v_i}\}_{v_i \in V_i}$. Given the Crémer-McLean condition, each $\mathcal{B}_j$ consists of linearly independent vectors. Now Lemma 3 implies that the vectors in the set $\{B \otimes A^j\}_{j \in \{1,\cdots,m\}, B \in \mathcal{B}_j} = \{\vec{D}^j_{v_i} \otimes (\otimes \vec{D}^j)^k\}_{j \in \{1,\cdots,m\}, v_i \in V_i} = \{\vec{F}^j_{v_i}\}_{j \in \{1,\cdots,m\}, v_i \in V_i}$ are also linearly independent. $\square$

3.2 Necessary Number of Samples

We now show that the number of samples in Theorem 1 is necessary.

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

Let us first point out a difficulty. Recall that Proposition 1 was established through Lemma 1 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, 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 $u_{i,3}^{\text{SPA}} = u_{i,1}^{\text{SPA}}/2 + u_{i,2}^{\text{SPA}}/2$. 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 $u_{i,3}^{\text{SPA}} = u_{i,1}^{\text{SPA}}/2 + u_{i,2}^{\text{SPA}}/2$.

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 1. *Buyer 1 has two possible values, $v_1 \in \{2, 3\}$. There are only two profiles of values that other buyers can possibly have, v_{-i}^1 and v_{-i}^2.*[10] *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.*

$$B^1 = \begin{array}{c c} & \begin{array}{cc} v_{-i}^1 & v_{-i}^2 \end{array} \\ \begin{array}{c} 2 \\ 3 \end{array} & \begin{pmatrix} 1/3 & 0 \\ 0 & 2/3 \end{pmatrix} \end{array}, \qquad B^2 = \begin{array}{c c} & \begin{array}{cc} v_{-i}^1 & v_{-i}^2 \end{array} \\ \begin{array}{c} 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 2/3 \\ 1/3 & 0 \end{pmatrix} \end{array}.$$

Consider α_1 to α_m, where $0 \le \alpha_j \le 1$, $\alpha_j \neq 1/2$. Construct each distribution D^j in the set $\mathcal{F}$ as a convex combination of D^1 and D^2 with weight α_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 2. Then buyer 1 must be allocated regardless of the profile of values, since buyer 1 has the highest value. As a result, buyer 1's utility from allocation, ignoring payments, is equal to her value. Therefore, to extract surplus the expected payment of each value must be equal to the value,

$$\mathbf{E}_{v\sim D^j, s\sim(\times D^j)^k}\left[p_1(v,s)|v_1\right] = v_1, \qquad \forall j, v_1.$$

[10] 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.

Since buyer 1 gets the product regardless of her value, incentive compatibility requires that the payment of buyer 1 does not depend on her report. We thus write the payment function of buyer 1 as $p_1(v_{-1}, s)$. The above equality becomes

$$\mathbf{E}_{v \sim D^j, s \sim (\times D^j)^k} \left[p_1(v_{-1}, s) | v_1 \right] = v_1, \qquad \forall j, v_1.$$

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 furthur that $\sum_j \beta_{j,2} \neq 0$. We must have

$$\begin{aligned}
&\sum\nolimits_{j,v_1} \beta_{j,v_1} \mathbf{E}_{v \sim D^j, s \sim (\times D^j)^k} \left[p_1(v_{-1}, s) | v_1 \right] = \sum\nolimits_j \beta_{j,2} \cdot 2 + \sum\nolimits_j \beta_{j,3} \cdot 3 \\
&= (2-3)(\sum_j \beta_{j,2}) \neq 0.
\end{aligned}$$

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 $\beta = (\beta_{j,v_1})_{j,v_1}$ that satisfies (i) $\sum_{j,v_1} \beta_{j,v_1} = 0$ and (ii) $\sum_j \beta_{j,2} \neq 0$, we have

$$\sum\nolimits_{j,v_1} \beta_{j,v_1} \mathbf{E}_{v \sim D^j, s \sim (\times D^j)^k} \left[p_1(v_{-1}, s) | v_1 \right] \neq 0.$$

The next lemma shows that no such function p_1 exists.

Lemma 5. *Consider the set of distributions $\mathcal{F} = \{D^1, \ldots, D^m\}$ defined in Example 1 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*

$$\sum\nolimits_{j,v_1} \beta_{j,v_1} \mathbf{E}_{v \sim D^j, s \sim (\times D^j)^k} \left[p_1(v_{-1}, s) | v_1 \right] = 0.$$

4 Discussion and Conclusions

The surplus extracting auction of Crémer and McLean is often seen as a critique on commonly studied models 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 buyers are risk neutral; second, that the auctioneer has exact knowledge of the underlying distribution of the buyers' 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 the second and third assumptions may not be the main driver of full surplus extraction: The assumption that the auctioneer 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. A main conceptual step in our construction is to define a matrix of conditional probabilities that is common knowledge to all buyers, even though the true distribution is not. We accomplish this by indexing the rows of the conditional probability matrix by each potential distribution.

An important assumption is that the auctioneer can credibly commit to how the mechanism uses the signal. In particular, the signal space must be commonly known. The commitment assumption may be justified in a setting where the signals are publicly verifiable and contractible. For example, in oil-lease auctions, the value of the asset is partially verifiable ex post. Thus it is common to design auctions in which transfers depend on the realization of a state (signaling the value of the asset). If the auctioneer cannot commit, then an inference-based approach may be more effective than our mechanism with contingent payments. With an inference-based approach, the auctioneer observes the signal and then publicly announces an auction that maps bids to allocation and payments. Thus to verify the allocation and payments, there is no need to verify the signal. A natural question is whether such an approach can approximately extract the full surplus. In another manuscript Fu et al. (2020), we provide bounds on the number of samples needed.

Let us comment on how "belief-free" our model is. Note that the expected utility of a buyer does depend on the buyer's belief about the true distribution. However, assuming that all other buyers participate, a buyer expects non-negative utility from participation for any prior belief. In fact, the buyers may not be equipped with a prior at all. For example, the buyer may maximize the minimum expected utility. For all buyers to participate, it only needs to be common knowledge among buyers that the true distribution is in $\mathcal{F}$. Agent 1 may know that the true distribution is in $\mathcal{F}' \subseteq \mathcal{F}$, may know that buyer 2 knows that it is in $\mathcal{F}'' \subseteq \mathcal{F}$, and so on. The assumption that the set of distributions is common knowledge is a weakening, although admittedly a limited one, that the distribution itself is common knowledge.[11]

Another assumption made in this paper is that the set of distributions is finite. This is typically assumed in the literature on auctions with contingent payments. This assumption can also be justified when all buyers have access to an anonymized distribution of values, but it is unknown how the identities in the anonymized distribution maps to the that of the buyers. For instance, with three players, there are six possible ways by which the identities in the distribution can be mapped to the identities of the buyers. Nevertheless, 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, 2010). This seems

[11] Consider an alternative setting in which all the players share a common prior over the set of distributions $\mathcal{F}$. Without samples, the problem reduces to one in which the average distribution is known to all players. It might be possible that the average distribution satisfies the linear independence condition of Crémer and McLean, even though some or all distributions in $\mathcal{F}$ do not. Our approach rules out such a possibility. Conversely, the average distribution may fail to satisfy the linear independence conditions even though all distributions in $\mathcal{F}$ do. We show that it is nevertheless possible to extract the full surplus in such a case.

a prerequisite for possibly extending the approach further to infinite families of distributions.

References

Balcan, M.-F., Blum, A., Hartline, J. D., and Mansour, Y. (2008). Reducing mechanism design to algorithm design via machine learning. *Journal of Computer and System Sciences*, 74(8):1245–1270.

Baliga, S. and Vohra, R. (2003). Market research and market design. *Advances in Theoretical Economics*, 3(1).

Barelli, P. (2009). On the genericity of full surplus extraction in mechanism design. *Journal of Economic Theory*, 144(3):1320–1332.

Bellman, R. (1997). *Introduction to matrix analysis*. SIAM.

Bergemann, D., Brooks, B. A., and Morris, S. (2016). Informationally robust optimal auction design.

Chen, Y.-C. and Xiong, S. (2013). Genericity and robustness of full surplus extraction. *Econometrica*, 81(2):825–847.

Cole, R. and Roughgarden, T. (2014). The sample complexity of revenue maximization. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 243–252. ACM.

Cremer, J. and McLean, R. P. (1988). Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica: Journal of the Econometric Society*, pages 1247–1257.

DeMarzo, P. M., Kremer, I., and Skrzypacz, A. (2005). Bidding with securities: Auctions and security design. *American economic review*, 95(4):936–959.

Dhangwatnotai, P., Roughgarden, T., and Yan, Q. (2010). Revenue maximization with a single sample. In *ACM Conference on Electronic Commerce*, pages 129–138.

Fu, H., Haghpanah, N., Hartline, J., and Kleinberg, R. (2020). Sample complexity of full surplus extraction. *Working paper.*

Fu, H., Hartline, J. D., and Hoy, D. (2013). Prior-independent auctions for risk-averse agents. In *ACM Conference on Electronic Commerce*, pages 471–488.

Fudenberg, D., Levine, D., and Maskin, E. (1994). The folk theorem with imperfect public information. *Econometrica (1986-1998)*, 62(5):997.

Goldberg, A. V., Hartline, J. D., Karlin, A. R., Saks, M., and Wright, A. (2006). Competitive auctions. *Games and Economic Behavior*, 55(2):242–269.

Hansen, R. G. (1985). Auctions with contingent payments. *The American Economic Review*, 75(4):862–865.

Heifetz, A. and Neeman, Z. (2006). On the generic (im) possibility of full surplus extraction in mechanism design. *Econometrica*, 74(1):213–233.

Laffont, J.-J. and Martimort, D. (2000). Mechanism design with collusion and correlation. *Econometrica*, 68(2):309–342.

Lopomo, G., Rigotti, L., and Shannon, C. (2017). Uncertainty in mechanism design.

McAfee, R. P. and Reny, P. J. (1992). Correlated information and mecanism design. *Econometrica*, pages 395–421.

Morgenstern, J. and Roughgarden, T. (2016). Learning simple auctions. In *Conference on Learning Theory*, pages 1298–1318.

Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V. V. (2007). *Algorithmic game theory*, volume 1. Cambridge University Press Cambridge.

Peters, M. (2001). Surplus extraction and competition. *The Review of Economic Studies*, 68(3):613–631.

Rahman, D. (2010). Detecting profitable deviations.

Rahman, D. (2012). But who will monitor the monitor? *American Economic Review*, 102(6):2767–97.

Rahman, D. and Obara, I. (2010). Mediated partnerships. *Econometrica*, 78(1):285–308.

Robert, J. (1991). Continuity in auction design. *Journal of Economic Theory*, 55(1):169–179.

Roughgarden, T. and Talgam-Cohen, I. (2013). Optimal and near-optimal mechanism design with interdependent values. In *ACM Conference on Electronic Commerce*, pages 767–784.

Segal, I. (2003). Optimal pricing mechanisms with unknown demand. *The American economic review*, 93(3):509–529.

A Proofs

A.1 Proofs from Section 3

A.1.1 Proof of Lemma 1

Proof of Lemma 1. Recall that a mechanism that extracts full surplus exists if the following system has a solution

$$\mathbf{E}_{(v,s)\sim F^j}\left[q_i(v,s)|v_i\right] = u_{i,j}^{\mathrm{SPA}}(v_i), \forall i, v_i, j$$

The system has a solution for any right hand side $u_{i,j}^{\mathrm{SPA}}(v_i)$ if 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. $\square$

A.1.2 Proof of Lemma 3

Proof of Lemma 3. Suppose that there is a vector $\vec{\alpha} = (\alpha_{j,B})_{j\in\{1,\cdots,m\},B\in\mathcal{B}_j}$ such that

$$\sum_{j\in\{1,\ldots,m\},B\in\mathcal{B}_j} \alpha_{j,B}(B \otimes A^j) = \vec{0},$$

where $\vec{0}$ is a vector of zeros. We show that $\vec{\alpha}$ must be a vector of zeros. Using the definition of the outer product, and writing $B = (b_1, \ldots, b_\ell)$, we have

$$\sum_{j\in\{1,\ldots,m\},B\in\mathcal{B}_j} \alpha_{j,B}(b_1 A^j, \ldots, b_\ell A^j) = \vec{0}.$$

Therefore,

$$\sum_{j\in\{1,\ldots,m\}} ((\sum_{B\in\mathcal{B}_j} \alpha_{j,B} b_1) A^j, \ldots, (\sum_{B\in\mathcal{B}_j} \alpha_{j,B} b_\ell) A^j) = \vec{0}.$$

We thus have a system of ℓ equalities. For $i \in \{1, \ldots, \ell\}$, the i'th equality is

$$\sum_{j \in \{1,\ldots,m\}} \Big(\sum_{B \in \mathcal{B}_j} \alpha_{j,B} b_i\Big) A^j = 0.$$

Since vectors $A^1, \ldots, A^m$ are linearly independent, we must have

$$\sum_{B \in \mathcal{B}_j} \alpha_{j,B} b_i = 0,$$

for all i and j. Since vectors in $\mathcal{B}_j$ are linearly independent, we must have $\alpha_{j,B} = 0$ for all j and B, completing the proof. $\square$

A.1.3 Proof of Lemma 4

Proof of Lemma 4. 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 $\{B_1, \cdots, B_d\}$ 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,

$$\begin{aligned}
\left(\otimes \vec{D}^j\right)^k &= \left(\otimes \sum_{\ell=1}^{d} \alpha_{j\ell} B_\ell\right)^k \\
&= \sum_{1 \le \ell_1, \ldots, \ell_k \le d} \alpha_{j\ell_1} B_{\ell_1} \otimes \ldots \otimes \alpha_{j\ell_k} B_{\ell_k} \\
&= \sum_{1 \le \ell_1, \ldots, \ell_k \le d} \alpha_{j\ell_1} \ldots \alpha_{j\ell_k} B_{\ell_1} \otimes \ldots \otimes B_{\ell_k}
\end{aligned}$$

where ℓ_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 τ, $1 \leq \tau \leq d$, 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

$$\left(\otimes \vec{D}^j\right)^k = \sum_{\substack{\gamma_1+\cdots+\gamma_d=k,\\ \gamma_1,\cdots,\gamma_d\geq 0}} \alpha_{j1}^{\gamma_1}\alpha_{j2}^{\gamma_2}\ldots\alpha_{jd}^{\gamma_d} \sum_{\substack{\ell_1,\ldots,\ell_k,\\ \gamma_\tau=|\{\ell_j=\tau\}|,\forall\tau}} B_{\ell_1}\otimes\ldots\otimes B_{\ell_k}$$

To simplify notation, define

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

That is, $C_{\gamma_1,\cdots,\gamma_d}$ is the sum of terms that are outer products of $B_1,\cdots,B_d$, such that in each term B_1 appears γ_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

$$\left(\otimes \vec{D}^j\right)^k = \sum_{\substack{\gamma_1+\cdots+\gamma_d=k,\\ \gamma_1,\cdots,\gamma_d\geq 0}} \alpha_{j1}^{\gamma_1}\alpha_{j2}^{\gamma_2}\ldots\alpha_{jd}^{\gamma_d} C_{\gamma_1,\cdots,\gamma_d}.$$

An inductive application of Lemma 3 implies that the set of vectors $\{B_{\ell_1}\otimes\cdots\otimes B_{\ell_k}\}_{\ell_1,\cdots,\ell_k\in[d]}$ are linearly independent. To see this, assume that $\{B_{\ell_1}\otimes\cdots\otimes B_{\ell_{k-1}}\}_{\ell_1,\cdots,\ell_{k-1}\in[d]}$ are linearly independent. Define $\mathcal{A} = \{B_{\ell_1}\otimes\cdots\otimes B_{\ell_{k-1}}\}_{\ell_1,\cdots,\ell_{k-1}\in[d]}$, and for $j\in\{1,\ldots,|\mathcal{A}|\}$, define $\mathcal{B}_j = \{B_1,\ldots,B_d\}$. Now Lemma 3 states that $\{B\otimes B_{\ell_1}\otimes\cdots\otimes B_{\ell_{k-1}}\}_{B\in\{B_1,\ldots,B_d\},\ell_1,\cdots,\ell_{k-1}\in[d]} = \{B_{\ell_1}\otimes\cdots\otimes B_{\ell_k}\}_{\ell_1,\cdots,\ell_k\in[d]}$ are linearly independent. Since each $C_{\gamma_1,\cdots,\gamma_d}$ is a summation over vectors in $\{B_{\ell_1}\otimes\cdots\otimes B_{\ell_k}\}_{\ell_1,\cdots,\ell_k\in[d]}$, the vectors in $\{C_{\gamma_1,\ldots,\gamma_d}\}_{\gamma_1+\ldots+\gamma_d=k}$ are also linearly independent.

Now note that each $(\otimes\vec{D}^j)^k$ is expressed as a linear combination of linearly independent vectors, with the linear coefficient on $C_{\gamma_1,\cdots,\gamma_d}$ being the product $\alpha_{j1}^{\gamma_1} \dots \alpha_{jd}^{\gamma_d}$. To show linear independence of the set of vectors $\{(\otimes \vec{D}^j)^k\}_j$, 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 $\{(\alpha_{j1}^{\gamma_1} \dots \alpha_{jd}^{\gamma_d})_{\gamma_1+\dots+\gamma_d=k}\}_j$ are linearly independent.

The vector $(\alpha_{j1}^{\gamma_1} \dots \alpha_{jd}^{\gamma_d})_{\gamma_1+\dots+\gamma_d=k}$ is the image of the vector $\vec{\alpha}_j = (\alpha_{j1}, \dots, \alpha_{jd})$ under a mapping $\nu : \mathbb{R}^d \to \mathbb{R}^{\binom{d+k-1}{d-1}}$ which evaluates all the k-th degree monomials in $\mathbb{R}[x_1, \dots, x_d]$ at a point in $\mathbb{R}^d$. We now show that these m images $\nu(\vec{\alpha}_1), \dots, \nu(\vec{\alpha}_m)$ are linearly independent when $k = m - d + 1$.

We will show that for every j, there exists a linear form on $\mathbb{R}^{\binom{d+k-1}{d-1}}$ 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 $\{B_1, \cdots, B_d\}$ is a basis of this space, the vectors $\vec{\alpha}_1, \dots, \vec{\alpha}_m$ 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 $f_1 : (y_1, \dots, y_d) \mapsto \beta_1 y_1 + \cdots + \beta_d y_d$ 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 $\vec{\alpha}_{d+1}, \dots, \vec{\alpha}_m$. 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 ℓ, for some ζ. Thus, for each $j' = d+1, \cdots, m$, 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,

$$f = f_1 f_{d+1} \dots f_m.$$

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

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

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. $\square$

A.1.4 Proof of Lemma 5

Before we prove Lemma 5, 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.

$$\bar{V} = \begin{pmatrix} 1 & y_1 & \ldots & y_1^{m-2} & (1 + \frac{3y_1}{1-y_1^2})(1+y_1)^{m-2} \\ 1 & y_2 & \ldots & y_2^{m-2} & (1 + \frac{3y_2}{1-y_2^2})(1+y_2)^{m-2} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & y_m & \ldots & y_m^{m-2} & (1 + \frac{3y_m}{1-y_m^2})(1+y_m)^{m-2} \end{pmatrix}. \tag{4}$$

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

$$V = \begin{pmatrix} 1 & y_1 & \ldots & y_1^{m-2} & y_1^{m-1} \\ 1 & y_2 & \ldots & y_1^{m-2} & y_2^{m-1} \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & y_m & \ldots & y_m^{m-2} & y_m^{m-1} \end{pmatrix}.$$

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 6. *For distinct $y_1, \ldots, y_m$, none equal to ± 1, the m by m matrix $\bar{V}$ defined in Equation 4 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

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

which is equivalent to

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

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.

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

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

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

Index columns from 1 to m. For column ℓ from 1 to $m-1$, the ℓ'th entry is $y^{\ell-1}(1-y)$. Now replace the element in each column ℓ from 2 to $m-1$ with the sum of all elements from 1 to ℓ. 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 ℓ from 1 to $m-2$, multiply the ℓ'th element by $\binom{m-3}{\ell-1}$ and add it to the last element. Since $y(1+y)^{m-3} = \sum_{\ell=0}^{m-3} \binom{m-3}{\ell} y^{\ell+1}$, the result is

$$1-y, 1-y^2, \ldots, 1-y^{m-1}, \sum_{\ell=1}^{m-2} \binom{m-3}{\ell-1}.$$

Divide the last column by $\sum_{\ell=1}^{m-2} \binom{m-3}{\ell-1}$,

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

Mulitply 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 $\bar{V}$. $\square$

We now prove Lemma 5.

Proof of Lemma 5. To prove the lemma, we need to show existence of β of size $m \times |V_1|$ such that $\beta \cdot \vec{1} = \sum_{j,v_1} \beta_{j,v_1} = 0$, $\sum_j \beta_{j,2} \neq 0$, and

$$\sum_{j,v_1} \beta_{j,v_1} \mathbf{E}_{v,s^1,\ldots,s^k \sim D^j} \left[p_1(v_{-1}, s) | v_1 \right] = 0, \forall p_1. \tag{5}$$

Recall that $\vec{F}^j_{v_1}$ 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 $\{\vec{F}^j_{v_1}\}_{j,v_1}$ on top of each other, that is,

$$\begin{array}{cc} & \begin{array}{ccc} \dots & v_{-1}, s & \dots \end{array} \\ F = \begin{array}{c} \vdots \\ j, v_1 \\ \vdots \end{array} & \begin{pmatrix} & & \\ \dots & F^j(v_{-1}, s|v_1) & \dots \\ & & \end{pmatrix}. \end{array}$$

Using this notation we can write

$$\sum_{j,v_1} \beta_{j,v_1} \mathbf{E}_{v,s^1,\dots,s^k \sim D^j} \left[p_1(v_{-1}, s)|v_1\right] = \beta \cdot F \cdot p_1.$$

Therefore, to show Equation 5 it is sufficient to show that $\beta \cdot F = 0$.

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

$$\vec{F}^j_{v_1}(v_{-1}, s) = \vec{D}^j(v_{-1}|v_1)\, \mathbf{Pr}_j\, [s] = \vec{D}^j(v_{-1}|v_1) \cdot \vec{D}^j(s^1) \cdot \ldots \cdot \vec{D}^j(s^k).$$

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 1, $D^j(s^\ell) = \alpha_j$ 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 $\mathbf{Pr}_j[s] = (\alpha_j)^s(1-\alpha_j)^{k-s}$. We therefore simply represent F as a $2m$ by $2(m-1)$ matrix as follows

$$\begin{array}{cc} & \begin{array}{ccc} \dots & v_{-1}, \ell & \dots \end{array} \\ F = \begin{array}{c} \vdots \\ j, v_1 \\ \vdots \end{array} & \begin{pmatrix} & & \\ & \mathbf{Pr}_j[v_{-1}|v_1](\alpha_j)^\ell(1-\alpha_j)^{m-2-\ell} & \\ & & \end{pmatrix}. \end{array}$$

Let V be an m by $m-1$ Vandermonde matrix, that is,

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

Let $\Delta_{v_1,v_{-1}}$ be a m by m diagonal matrix such that $\Delta_{v_1,v_{-1}}(j,j) = \mathbf{Pr}_j[v_{-1}|v_1](1-\alpha_j)^{m-2}$. 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 β is a vector of size $2m$. Let β be composed of two parts β_L and β_H, each of size m. That is, $\beta = (\beta_L, \beta_H)$. Now the equation $\beta \cdot F = 0$ becomes

$$(\beta_L \Delta_{v_1^1,v_{-1}^1} + \beta_H \Delta_{v_1^2,v_{-1}^1})V = 0, \tag{6}$$

$$(\beta_L \Delta_{v_1^1,v_{-1}^2} + \beta_H \Delta_{v_1^2,v_{-1}^2})V = 0. \tag{7}$$

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

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

where $y_j = \alpha_j/(1-\alpha_j)$. By Lemma 6, there exists $\alpha_1, \dots, \alpha_m$ such that matrix

$\bar{V}$ is invertible. In particular, we have

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

Now Lemma 6 applies to imply that $\bar{V}$ has full rank. Therefore, there exists a solution w to the following system $w \cdot \bar{V} = (0, \dots, 0, 1)$ of m equations and m unknowns. The first $m-1$ equations are equivalent to Equation 6 together with Equation 7. Therefore, $\beta \cdot F = 0$ if

$$\begin{aligned} \beta_L \Delta_{v_1^1, v_{-1}^1} + \beta_H \Delta_{v_1^2, v_{-1}^1} &= w, \\ \beta_L \Delta_{v_1^1, v_{-1}^2} + \beta_H \Delta_{v_1^2, v_{-1}^2} &= w. \end{aligned}$$

Solving these two equations gives

$$\begin{aligned} \beta_L &= w(\Delta_{v_1^2, v_{-1}^2} - \Delta_{v_1^2, v_{-1}^1}) \cdot (\Delta_{v_1^1, v_{-1}^1} \cdot \Delta_{v_1^2, v_{-1}^2} - \Delta_{v_1^1, v_{-1}^2} \cdot \Delta_{v_1^2, v_{-1}^1})^{-1}, \\ \beta_H &= w(\Delta_{v_1^1, v_{-1}^2} - \Delta_{v_1^1, v_{-1}^1}) \cdot (\Delta_{v_1^2, v_{-1}^1} \cdot \Delta_{v_1^1, v_{-1}^2} - \Delta_{v_1^2, v_{-1}^2} \cdot \Delta_{v_1^1, v_{-1}^1})^{-1}. \end{aligned}$$

To summarize the arguments so far, if β_L and β_H satisfy the above equations, then $\beta \cdot F = 0$ and therefore $\sum_j \beta_j \, \mathbf{E}_{v_{-1} \sim D^j}[p_1(v_{-1})|v_1] = 0$. We will next show that additionally, $\sum_j \beta_{j,2} \neq 0$.

Multiply Equation by $\vec{1}$, a vector of 1's, to get

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

We next show that

$$(\Delta_{v_1^1, v_{-1}^2} - \Delta_{v_1^1, v_{-1}^1}) \cdot (\Delta_{v_1^2, v_{-1}^1} \cdot \Delta_{v_1^1, v_{-1}^2} - \Delta_{v_1^2, v_{-1}^2} \cdot \Delta_{v_1^1, v_{-1}^1})^{-1} \cdot \vec{1} \tag{8}$$

is the m'th column of matrix $\bar{V}$. Note that since w is the solution to $w \cdot \bar{V} = (0, \ldots, 0, 1)^T$, this implies that

$$\beta_H \cdot \vec{1} = w(\Delta_{v_1^1, v_{-1}^2} - \Delta_{v_1^1, v_{-1}^1}) \cdot (\Delta_{v_1^2, v_{-1}^1} \cdot \Delta_{v_1^1, v_{-1}^2} - \Delta_{v_1^2, v_{-1}^2} \cdot \Delta_{v_1^1, v_{-1}^1})^{-1} \cdot \vec{1} = 1 \neq 0.$$

So to complete the proof, we need to argue that Expression 8 is the m'th column of matrix $\bar{V}$. The j'th element in the column vector is

$$= \frac{(\mathbf{Pr}_j[v_{-1}^2|v_1^1] - \mathbf{Pr}_j[v_{-1}^1|v_1^1])(1-\alpha_j)^{m-2}}{(\mathbf{Pr}_j[v_{-1}^1|v_1^2]\,\mathbf{Pr}_j[v_{-1}^2|v_1^1] - \mathbf{Pr}_j[v_{-1}^2|v_1^2]\,\mathbf{Pr}_j[v_{-1}^1|v_1^1])(1-\alpha_j)^{2(m-2)}}.$$

By construction of Example 1, we have $\mathbf{Pr}_j[v_{-1}^1|v_1^1] = \alpha_j/(\alpha_j + 2(1-\alpha_j))$, $\mathbf{Pr}_j[v_{-1}^2|v_1^1] = 2(1-\alpha_j)/(\alpha_j+2(1-\alpha_j))$, $\mathbf{Pr}_j[v_{-1}^1|v_1^2] = (1-\alpha_j)/((1-\alpha_j)+2\alpha_j)$, and $\mathbf{Pr}_j[v_{-1}^2|v_1^2] = 2\alpha_j/((1-\alpha_j)+2\alpha_j)$. Therefore, the j'th element becomes

$$\begin{aligned}
&= \frac{\frac{2(1-\alpha_i)-\alpha_j}{\alpha_j+2(1-\alpha_j)}}{\frac{1-\alpha_j}{(1-\alpha_j)+2\alpha_j} \cdot \frac{2(1-\alpha_i)}{\alpha_j+2(1-\alpha_j)} - \frac{2\alpha_j}{(1-\alpha_j)+2\alpha_j} \cdot \frac{\alpha_j}{\alpha_j+2(1-\alpha_j)}} \cdot (1-\alpha_j)^{-(m-2)} \\
&= \frac{(2-3\alpha_j)(1+\alpha_j)}{2(1-\alpha_j)^2 - 2\alpha_j^2} \cdot (1-\alpha_j)^{-(m-2)}.
\end{aligned}$$

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

$$\begin{aligned}
&= \frac{\frac{2-y_j}{1+y_j}\frac{1+2y_j}{1+y_j}}{2(\frac{1}{1+y_j})^2 - 2(\frac{y_j}{1+y_j})^2} \cdot (1+y_j)^{(m-2)} \\
&= \frac{(2-y_j)(1+2y_j)}{2(1-y_j^2)} \cdot (1+y_j)^{(m-2)} \\
&= (1 + \frac{3y_j}{1-y_j^2})(1+y_j)^{m-2}.
\end{aligned}$$

We have argued that Expression 8 is the m'th column of matrix $\bar{V}$, which completes the proof. □

A.1.5 Proof of Proposition 2

Proof of Proposition 2. Consider the following extension of Example 1. 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 $\{v_{-1}^1, v_{-1}^2, v_{-1}^{m-d+3}, v_{-1}^{m-d+4}, \ldots, v_{-1}^m\}$. Similarly to Example 1, assume that $\max_{j\neq 1} v_j^1 = \max_{j\neq 1} v_j^2 = 1$. Now consider the following d bases. The two basis B^1 and B^2 are defined as in Example 1, that is

$$B^1 = \begin{array}{c c} & \begin{array}{cc} v_{-i}^1 & v_{-i}^2 \end{array} \\ \begin{array}{c} 2 \\ 3 \end{array} & \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} \end{array}, \qquad B^2 = \begin{array}{c c} & \begin{array}{cc} v_{-i}^1 & v_{-i}^2 \end{array} \\ \begin{array}{c} 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \end{pmatrix} \end{array},$$

with probability zero on every other value profile. For $\ell = m-d+3, \ldots, m$, define a distribution B^ℓ that puts probability 1 on profile (v_1^ℓ, v_{-1}^ℓ), 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 1, $D^j = \alpha_j B^1 + (1-\alpha_j)B^2$. 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 1.

If a surplus extracting mechanisms exists, then we must have

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

If $v_{-1} = v_{-1}^1$ or $v_{-1} = v_{-1}^2$, buyer 1 must win the product. 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

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

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

$$\sum_{j,v_1} \beta_{j,v_1} \, \mathbf{E}_{v,s^1,\ldots,s^k \sim D^j} \left[p_1(v_{-1}, s)|v_1\right] \neq 0.$$

Now Lemma 5 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 5 shows that no function p_1 satisfying the above inequalities exists. $\square$