Submitted to *Operations Research*
manuscript OPRE-2018-01-735

Mechanism Design for Correlated Valuations: Efficient Methods for Revenue Maximization

Michael Albert
Darden School of Business, University of Virginia, Charlottesville, VA 22903, albertm@darden.virginia.edu

Vincent Conitzer
Department of Computer Science, Duke University, Durham, NC 27708, conitzer@cs.duke.edu

Giuseppe Lopomo
Fuqua School of Business, Duke University, Durham, NC 27708, glopomo@duke.edu

Peter Stone
Department of Computer Science, University of Texas at Austin, Austin, TX 78712, pstone@cs.utexas.edu

Traditionally, the mechanism design literature has been primarily focused on settings where the bidders' valuations are independent. However, in settings where valuations are *correlated*, much stronger results are possible. For example, the entire surplus of efficient allocations can be extracted as revenue. These stronger results are true, in theory, under generic conditions on parameter values. But in practice, they are rarely, if ever, implementable due to the stringent requirement that the mechanism designer knows the distribution of the bidders types exactly. In this work, we provide a computationally efficient and sample efficient method for designing mechanisms that can robustly handle imprecise estimates of the distribution over bidder valuations. This method guarantees that the selected mechanism will perform at least as well as any ex-post mechanism with high probability. The mechanism also performs nearly optimally with sufficient information and correlation. Further, we show that when the distribution is not known, and must be estimated from samples from the true distribution, a sufficiently high degree of correlation is essential to implement optimal mechanisms. Finally, we demonstrate through simulations that this new mechanism design paradigm generates mechanisms that perform significantly better than traditional mechanism design techniques given sufficient samples.

Key words: mechanism design; robust optimization; revenue maximization; correlated valuations

History: The definition and computational properties of ϵ-robust mechanisms in Section 4 originally appeared in the proceedings of AAAI-17 as Albert et al. (2017a). The infinite sample complexity for arbitrary correlation result in Section 6 and Appendix C originally appeared in the proceedings of AAMAS-17 as Albert et al. (2017b).

1. Introduction

Auctions are one of the fundamental tools for allocating resources in the modern economy. They are used to allocate online ad space (Interactive Advertising Bureau (IAB) 2017), offshore oil drilling rights, famous artwork, small and medium lift capacity to planetary orbit (NASA 2014), government supply contracts (Government Accountability Office 2013), FCC spectrum licenses (FCC 2017), and almost limitless numbers of other things, large and small. Given the economic magnitude and breadth of these markets, it is crucial that these auctions are implemented optimally; even small deviations from optimality can lead to millions of dollars worth of lost revenue, inefficiencies in resource allocation, and overspending.

The application of auction design to online advertising has been a particularly fruitful area of research, both academically and practically. For example, in 2017 $88 billion dollars of online ad revenue was generated through automated auctions (Interactive Advertising Bureau (IAB) 2017). Moreover, these online advertising markets are representative of a new class of auction platforms characterized by the frequency with which they repeat identical auctions and the granularity of the data that the platforms collect. Real time bidding for Amazon Web Services EC2 "Spot Instances" is another such platform.

These rapidly repeated, data-rich auction platforms provide unique opportunities for the auction (or mechanism) designer. This is due to well-established results indicating that revenue optimal mechanisms are *prior-dependent* (also known as *Bayesian*) mechanisms (Myerson 1981, Cremer and McLean 1985, 1988, Lopomo 2001), i.e., mechanisms that assume some knowledge of the kinds of bidders that are likely to participate. These repeated settings provide the mechanism designer the opportunity to learn the distribution of the bidders and achieve the optimal mechanism.

This dependence on distributional information is particularly strong in the setting of correlated valuations, i.e the setting where bidders' valuations are correlated. Moreover, the combination of correlated valuations and Bayesian mechanisms allows for the strongest possible result in revenue maximizing mechanism design, that of full surplus extraction as revenue for the seller (Cremer and

McLean 1985, 1988, Albert et al. 2016). With a small degree of correlation (i.e., the Cremer-McLean (CM) condition), the seller can, in expectation, generate as much revenue as if she knew the bidders' true valuations. Further, correlated valuations are likely to be the norm, not the exception, in many mechanism design settings of interest, e.g. any setting with a common value component. For online advertising, the bidders (i.e., the advertisers) generally receive a profile of the user who will be shown the advertisement. The bidders then use sophisticated valuation algorithms to convert this profile into a valuation of the user viewing the advertisement. For two bidders that are advertising similar products, they are naturally going to have correlated valuations for the ad impression due to similar valuation algorithms.

Therefore, if a mechanism designer intends to maximally exploit a correlated valuations setting, she must use information about the distribution. Unfortunately, traditional mechanism design paradigms assume either perfect information about the distribution (Bayesian mechanism design) or use very little information about the distribution (ex-post mechanism design). If the mechanism designer tries to naïvely use an estimate of the distribution, the mechanism is unlikely to be incentive compatible or individually rational, leading to mechanisms that are hard to reason about and may perform very poorly (Lopomo et al. 2009, Albert et al. 2015).

As our primary contribution, we develop and study a new class of mechanisms, *ϵ-robust mechanisms*, that extend *robust mechanisms* (Bergemann and Morris 2005) by allowing for a small probability of violation of incentive compatibility and individual rationality when there is uncertainty over the distribution of bidders. This new class of mechanisms permits the design of mechanisms in settings where the bidder distribution is estimated. Specifically, we assume that there is a set of distributions, i.e., the *ϵ-consistent set*, with the true bidder distribution being within the set with high probability (probability of $1-\epsilon$). We also assume that we have a point estimate of the true distribution. This is consistent with a procedure that estimates the true distribution and provides a "confidence interval" around that estimate. This new class of mechanisms is then guaranteed to be incentive compatible and individually rational for *all* distributions within the ϵ-consistent set.

Using this new class of mechanisms, we show the following:

(Informal) Corollary 2 *An optimal ϵ-robust mechanism can be computed in polynomial time.*

Furthermore, we demonstrate that achieving approximately optimal revenue, i.e., nearly full surplus extraction, is not only possible for this class of mechanisms, for distributions that are "sufficiently" correlated, it can be done efficiently. We do this using a non-parametric estimation procedure for the distribution combined with carefully choosing ϵ. This reinterprets ϵ as a regularization parameter for Bayesian mechanism design that parameterizes the space between ex-post and Bayesian mechanisms. The optimal ϵ balances the robustness against the specificity of the mechanism, which leads to our main result:

(Informal) Theorem 3 *For "sufficiently" correlated distributions, an ϵ-robust mechanism that approximately extracts full surplus as revenue can be learned using a polynomial number of samples from the underlying bidder distribution.*

Though Theorem 3 states that an approximately revenue optimal mechanism can be learned using an efficient number of samples, we do require that there is "sufficient" correlation. One naturally wonders whether the sufficient correlation condition is strictly necessary. To address this, we show the following:

(Informal) Corollary 3 *For "insufficiently" correlated distributions, there is no mechanism design procedure that uses a finite number of samples and guarantees approximately optimal revenue.*

2. Related Work

Much of the focus of the mechanism design community has been on the approximate optimality of *simple* mechanisms (Bulow and Klemperer 1996, Hartline and Roughgarden 2009, Roughgarden and Talgam-Cohen 2013, Morgenstern and Roughgarden 2015), that is, mechanisms that are either prior-independent or weakly prior-dependent (this distinction will be made clear in Section 3.4). This focus is due to two factors. First, prior-dependent mechanisms can be very brittle to misspecified priors (Lopomo 2001, Albert et al. 2015). That is to say, if a prior-dependent mechanism is constructed using an incorrect prior it can perform much worse than simple mechanisms (Hartline and Roughgarden 2009). Second, competition can be an effective substitute for knowledge of the distribution (Bulow and Klemperer 1996), at least for independent and identically distributed bidders. Therefore, instead of implementing prior-dependent mechanisms, practitioners generally implement prior-independent mechanisms under the assumption that there are many bidders, or that it will somehow be possible to *acquire* new bidders at a reasonable cost.

Unfortunately, in many auctions there is no feasible way to acquire more bidders. In online advertising, auctions are often very thin due to the platform giving preference to more relevant advertisements to the user of the platform. For example, when a user searches for the keyword "Nike" on Google, Nike appears, anecdotally, to be the only company that can win the first ad slot, since Google knows the user is likely interested in Nike products. Additionally, some auctions, such as Amazon Web Services EC2 "Spot Instances," may be thick on both sides, i.e., the demand is nearly equal to or less than the supply, leading to effectively thin auctions. Given this thin auctions problem, a recent literature (Pardoe et al. 2010, Fu et al. 2014, Kanoria and Nazerzadeh 2014, Mohri and Medina 2014, Blum et al. 2015, Mohri and Medina 2015, Morgenstern and Roughgarden 2015) has begun developing techniques to learn the optimal mechanism given samples from the true distribution. However, the literature has, primarily, focused on the restrictive case of independent private value distributions, where each bidder's valuation is independent of all other bidders, and the restrictive class of ex-post individually rational and incentive compatible mechanisms.

The most closely related paper is Fu et al. (2014), which explores the sample complexity of optimal mechanism design with correlated valuations. They are able to show that if there is a finite set of distributions from which the true distribution will be drawn, then the sample complexity is of the same order as the number of possible distributions. However, the results are in a sense *too* strong. Specifically, their findings suggest that maximizing revenue in settings with correlated distributions with finite types is trivial from a sample complexity standpoint, at least if the set of possible distributions is known. Moreover, outside of a very weak/modest condition (effectively stating that there *is* correlation), the degree of correlation does not play a role in the ability to implement the mechanism, an intuitively strange result. The key to reconciling this strangeness is realizing that there is something fundamentally distinct between infinite sets of distributions and finite sets, as we demonstrate in Section 6.

In the traditional mechanism design literature, this work is closely related to work on *robust mechanism design* (Bergemann and Morris 2005, Lopomo et al. 2009). This line of literature assumes that the bidders in the mechanism have a belief over other bidders, but that the belief for each agent is not known by the mechanism designer, similar to our notion of uncertainty over the distribution. Instead, the belief of each bidder over other bidders becomes part of the "type" of the agent, following Harsanyi (1967). This is, in some ways, more flexible than our approach; we define a bidder's type as his *payoff* type, while his belief is unknown but is from a known set. Since we are primarily interested in uncertain, but well-defined, beliefs, our notation will be sufficiently flexible for this work, and it allows us to more easily go beyond well defined beliefs. Our results differ from this previous work in two main respects. First, we develop and analyze a computational framework for computing a new class of mechanisms that satisfy the properties of Bergemann and Morris (2005) and Lopomo et al. (2009), whereas previous work is primarily interested in the theoretical limitations of such mechanisms. Second, we extend their definition of uncertainty in that we allow for probabilistic violation of the standard constraints in mechanism design.

From a computational standpoint, our work uses techniques both from the automated mechanism design literature (Conitzer and Sandholm 2002, 2004, Guo and Conitzer 2010, Sandholm and Likhodedov 2015) and the literature on robust optimization (Bertsimas and Sim 2004, Aghassi and Bertsimas 2006). However, the combination of these techniques is unique as far as we are aware.

3. ϵ-Robust Mechanisms

In this section, we introduce the problem setting and notation. Then we define a novel notion of incentive compatibility and individual rationality that allows for a relaxation of the traditional assumptions about the mechanism designers knowledge of the bidder distribution. Finally, we show how our new notions of incentive compatibility and individual rationality are connected to previous results in the literature.

3.1. Preliminaries

We consider a single monopolistic seller auctioning one object, which the seller values at zero, to a single bidder whose valuation is correlated with an external signal. The special case of a single bidder and an externally verifiable signal captures many of the important aspects of this problem while increasing ease of exposition relative to the case of many bidders, and this setting has been used in the literature on correlated mechanism design (McAfee and Reny 1992, Albert et al. 2015) for this purpose. The external signal can, but does not necessarily, represent other bidders' bids. For example, in online advertising for search results (so called "sponsored search auctions"), the external signal may be bids on a different but related keyword. E.g., if the target auction is for the keyword "Nike" the external signal may be bids on the auction for the keyword "shoes".

The bidder has a *valuation* type θ drawn from a finite set of discrete types $\Theta = \{1,...,|\Theta|\}$. Further, the bidder has a valuation function $v : \Theta \to \mathbb{R}^+$ that maps types to valuations for the object. Assume, without loss of generality, that for all $\theta, \theta' \in \Theta$, if $\theta > \theta'$ then $v(\theta) \geq v(\theta')$, and $v(1) > 0$. The discrete external signal is denoted by $\omega \in \Omega = \{1, 2, \ldots, |\Omega|\}$. Throughout the paper, we will denote vectors, matrices, and tensors as bold symbols.

There is a probability distribution, $\boldsymbol{\pi}$, over the types of the bidder and external signal where the probability of type and signal (θ, ω) is $\pi(\theta, \omega)$. Note that in contrast to much of the literature on mechanism design, we do not require that the bidder type be distributed independently of the external signal.

The distribution over the external signal ω given θ will be denoted by the $|\Omega|$ dimensional vector $\boldsymbol{\pi}(\cdot|\theta)$. We will, in many cases, be primarily interested in the conditional distribution over the external signal given the bidder's type, $\boldsymbol{\pi}(\cdot|\theta)$. We will, alternatively, represent the full distribution as a marginal distribution over Θ, $\boldsymbol{\pi}_\Theta$, and a set of conditional distributions over Ω, $\boldsymbol{\pi}(\cdot|\cdot) = \{\boldsymbol{\pi}(\cdot|1), \boldsymbol{\pi}(\cdot|2), ..., \boldsymbol{\pi}(\cdot||\Theta|)\}$. Therefore, if the true distribution is $\boldsymbol{\pi}$, we will alternatively represent it as $\{\boldsymbol{\pi}_\Theta, \boldsymbol{\pi}(\cdot|\cdot)\}$. If for all $\theta, \theta' \in \Theta$, $\boldsymbol{\pi}(\cdot|\theta) = \boldsymbol{\pi}(\cdot|\theta')$, it as an independent private values (IPV) setting and the optimal mechanism is a *reserve price mechanism*, a mechanism where the seller makes a take it or leave it offer at the *reserve price* (Myerson 1981).

3.2. Uncertainty Over the Distribution

Traditionally, the mechanism design literature has assumed that the mechanism designer knows the true distribution, $\boldsymbol{\pi}$, with arbitrary precision. A more realistic assumption is that the distribution is not perfectly known, but instead estimated, i.e., the seller estimates the distribution $\boldsymbol{\pi}$ as $\hat{\boldsymbol{\pi}}$. A reasonable estimation procedure would provide both a point estimate of the distribution, as well as a confidence interval around the estimate. This is the information structure we assume for this work, and we formalize this in the following definition.

Definition 1 (Set of ϵ-Consistent Distributions). Let $P(A)$ be the set of probability distributions over a set A. A subset $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}) \subseteq P(A)$ is an *ϵ-consistent set of distributions* for the estimated distribution $\hat{\boldsymbol{\pi}}$ if the true distribution, $\boldsymbol{\pi}$, is in $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ with probability $1-\epsilon$.

While an ϵ-consistent set is a set of joint distributions over both Θ and Ω, it will be useful to refer to the set of ϵ-consistent conditional distributions for $\theta \in \Theta$ as $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$, i.e., the set of true conditional distributions $\boldsymbol{\pi}(\cdot|\cdot)$ is such that $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$ for all $\theta \in \Theta$ with probability $1-\epsilon$. Similarly, the set of consistent marginal distributions over Θ will be referred to as $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_\Theta)$. Clearly, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ completely identifies $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$ and $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_\Theta)$.

While, this is similar to the notion of beliefs being part of the type space introduced in Bergemann and Morris (2005) for robust mechanisms, it differs in two important areas. First, we explicitly keep the uncertainty in the distribution separate from concerns over uncertainty in the type, whereas Bergemann and Morris (2005) combines the notion of bidder type and distribution uncertainty of bidder type into a single meta-type. Secondly, our notion of ϵ-consistent sets differs from that of robust mechanism design in Bergemann and Morris (2005) by allowing the true distribution to not be in the consistent set, with probability ϵ. Though if $\epsilon = 0$ our definition corresponds to theirs in this way, and we will refer to this special case by dropping the ϵ subscript, i.e., $\mathcal{P}(\hat{\boldsymbol{\pi}})$ implies $\epsilon = 0$.

While it is completely valid to view the mechanism designer as having access to a single consistent set $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ that is provided a-priori, it will also be useful to consider a situation in which ϵ is best viewed as a parameter. Specifically, a general estimation procedure for the distribution will allow for a range of ϵ to be used when computing the ϵ-consistent set. This is analogous to a statistician choosing the size of the confidence interval. In Section 5, we will show how, by the optimal choice of ϵ, the mechanism designer can trade off robustness of the mechanism with the expected revenue of the mechanism to lead to a good outcome with high probability.

3.3. ϵ-Robust Individual Rationality and ϵ-Robust Incentive Compatibility

A (direct) revelation mechanism is defined by, given the bidder type and external signal (θ, ω), 1) a probability that the seller allocates the item to the bidder and 2) a monetary transfer from the bidder to the seller. We will denote the probability of allocating the item to the bidder as $p(\theta, \omega)$, which is a value between zero and one, and the transfer from the bidder to the seller as $x(\theta, \omega)$, where a positive value denotes a payment to the seller and a negative value a payment from the seller to the bidder. We will denote a mechanism as $(\boldsymbol{p}, \boldsymbol{x})$.

Definition 2 (Bidder's Utility). Given a reported type $\theta' \in \Theta$, true type $\theta \in \Theta$, and external signal $\omega \in \Omega$, the bidder's realized utility under mechanism $(\boldsymbol{p}, \boldsymbol{x})$ is *quasi-linear*, i.e., the *bidder's utility* is $v(\theta)p(\theta', \omega) - x(\theta', \omega)$:

Due to the well-known revelation principle (e.g., Gibbons (1992)), the seller can restrict her attention to incentive compatible (IC) mechanisms, i.e., mechanisms where it is always optimal for the bidder to truthfully report his valuation. While it is theoretically possible to allow bidders to report both their valuations and their beliefs, and design optimal mechanisms given this joint report, standard automated mechanism design techniques require finitely specified input, and we are explicitly interested in infinite sets of distributions. Additionally, just as it is unlikely that the mechanism designer will know the true distribution over the bidder's valuation and external signal, it is equally unlikely that the bidder will know the true distribution with arbitrary precision. Asking the bidder to report the true distribution seems unreasonable. Therefore, we will simplify the mechanism design process by only considering mechanisms for which the payments, $\boldsymbol{x}$, and probabilities of allocations, $\boldsymbol{p}$, depend only on the reported bidder types and the realization of the external signal. While this is not without loss of generality, it will be sufficient to achieve our goals of better than ex-post performance while allowing for the possibility of nearly optimal performance.

With the defined mechanism $(\boldsymbol{p},\boldsymbol{x})$ and an ϵ-consistent set $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$, we can introduce our novel notion of ϵ-robust incentive compatibility.

Definition 3 (ϵ-Robust Incentive Compatibility). A mechanism is *ϵ-robust incentive compatible* (IC) for the ϵ-consistent set of distributions $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ if for all $\theta,\theta' \in \Theta$ and $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$,

$$\sum_{\omega\in\Omega} \pi(\omega|\theta)\left(v(\theta)p(\theta,\omega) - x(\theta,\omega)\right) \geq \sum_{\omega\in\Omega} \pi(\omega|\theta)\left(v(\theta)p(\theta',\omega) - x(\theta',\omega)\right)$$

ϵ-Robust incentive compatibility is a statement about the *beliefs* of the bidder over the external signal, $\boldsymbol{\pi}(\omega|\theta)$. Specifically, it allows the seller to determine payments by *lottery*. The lottery that the bidder faces can be dependent on his valuation, but the lottery itself is over the external signal. Additionally, the lotteries may have negative payments, i.e., the seller pays the bidder, for some outcomes of the external signal. An ϵ-robust incentive compatible mechanism guarantees with probability $1-\epsilon$ that the bidder with type θ has a higher expected value from reporting his type truthfully than he would from lying about his type. Stated another way, if the true distribution is in the ϵ-consistent set, the bidder weakly prefers to tell the truth about his type.

In addition to incentive compatibility, we are interested in mechanisms that are *individually rational* (IR), i.e., it is rational for a bidder to participate in the mechanism. We similarly define ϵ-robust individual rationality as a mechanism that is individually rational when the true distribution is in the ϵ-consistent set.

Definition 4 (ϵ-Robust Individual Rationality). A mechanism is *ϵ-robust individually rational* (IR) for the ϵ-consistent set of distributions $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ if for all $\theta \in \Theta$ and $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$,

$$\sum_{\omega\in\Omega} \pi(\omega|\theta)\left(v(\theta)p(\theta,\omega) - x(\theta,\omega)\right) \geq 0$$

Since we are interested in maximizing revenue, we will define an optimal mechanism as one that maximizes the expected revenue given the point estimate while satisfying ϵ-robust incentive compatibility and ϵ-robust individual rationality. Maximizing revenue with respect to the point estimate will maximize expected revenue, assuming that the point estimate is an unbiased estimate of the true distribution.

Definition 5 (Optimal ϵ-Robust Mechanism). A mechanism $(\boldsymbol{p}, \boldsymbol{x})$, given an estimated distribution $\hat{\boldsymbol{\pi}}$ and an ϵ-consistent set $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$, is a *revenue optimal ϵ-robust mechanism* if under the constraint of ϵ-robust individual rationality and ϵ-incentive compatibility it maximizes the following:

$$\sum_{\theta,\omega} x(\theta,\omega)\hat{\pi}(\theta,\omega)$$

3.4. Relation to Bayesian and Ex-Post Mechanisms

It is informative to consider two special cases of the ϵ-consistent set. The first case, which we will refer to as the *Bayesian* case for reasons that will become apparent, is when $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}) = \{\hat{\boldsymbol{\pi}}\}$, i.e., it is a singleton consisting of just the point estimate. The second case, which we will refer to as the *ex-post* case, is when $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}) = P(\Theta \times \Omega)$, i.e., the set of consistent distributions is all possible distributions. The Bayesian case corresponds to two possible interpretations, either the mechanism designer has perfect knowledge over the distribution, or ϵ is very large, $\epsilon \approx 1$, and the mechanism has a high probability of failing to be IC or IR. Alternatively, the ex-post case corresponds to no information about the distribution except for the point estimate, or a very small ϵ, $\epsilon \approx 0$, and the subsequent mechanism is guaranteed to be both IC and IR.

It is straightforward to see that the definition of ϵ-robust individual rationality and incentive compatibility reduce to the traditional ex-interim IR and Bayesian IC definitions in the Bayesian case. Additionally, it would be expected that when we have no useful information about the distribution, the optimal ϵ-robust mechanism should do no better than a mechanism that satisfies the traditional notion of ex-post IC and ex-post IR, since ex-post IR and IC are independent of the distribution. The following corollary shows that this is indeed the case.

Corollary 1. *If for all $\theta \in \Theta$, and $\omega \in \Omega$, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ is such that there exists a distribution $\boldsymbol{\pi'} \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ where $\pi'(\omega'|\theta) = 1$ if $\omega' = \omega$ and 0 otherwise, then an optimal ϵ-robust mechanism is an optimal ex-post mechanism for the distribution $\hat{\boldsymbol{\pi}}$.*

Proof. If for all $\theta \in \Theta$ and $\omega \in \Omega$, the distribution such that $\pi(\omega'|\theta) = 1$ if $\omega' = \omega$ and 0 otherwise is in $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$, the robust IR constraints contain the following set of constraints

$$v(\theta)p(\theta,\omega) - x(\theta,\omega) \geq 0 \qquad \forall \quad \omega \in \Omega, \theta \in \Theta$$

which implies ex-post IR. Conversely, ex-post IR implies robust IR.

By an identical argument, the ex-post IC constraints can be recovered. □

Bayesian mechanisms are what we have been referring to as prior-dependent mechanisms, while ex-post is weakly prior-dependent, i.e., only the objective function depends on the distribution, not the constraints over incentive compatibility and individual rationality. If, instead of precise knowledge of the distribution of bidder types and external signals, the mechanism designer has an imprecise estimate of the distribution, the prior-dependent, or Bayesian, mechanism can fail to be both incentive compatible and individually rational. This failure can be a significant problem for two reasons. First, if the mechanism is not individually rational bidders will not participate in the mechanism. In markets with few bidders, the loss of even a single bidder can lead to significant decreases in expected revenue, even relative to simple mechanisms (Bulow and Klemperer 1996). Second, if the mechanism is not incentive compatible, the bidder may optimally choose to mis-report his true valuation, leading both to biases in future estimates of the distribution and difficulty in reasoning about the performance of the mechanism, since it is unclear a priori how the bidder will report.

It is in this sense that Bayesian mechanisms are, in general, strongly prior-dependent. The mechanism depends not only on the seller's estimate of the distribution, but also the bidder's belief over the distribution. The consequences of these being mis-aligned is not just slightly lower expected revenue, as would be the case for weakly prior-dependent mechanisms such as a second price auction with a reserve price; it is a failure of the mechanism to maintain fundamental characteristics (Hartline 2014, Albert et al. 2015). Therefore, unless the seller has perfect knowledge of the bidder's beliefs, standard mechanism design techniques will leave only the option of using sub-optimal, weakly prior-dependent mechanisms.

Using these sub-optimal mechanisms can have significant implications for the achievable revenue. Note that given individual rationality constraints, the maximum revenue that any mechanism can possibly achieve, in expectation, is the expected bidder valuation. When a mechanism achieves this, we say that the mechanism *extracts the full social surplus as revenue*. However, due to the incentive compatibility constraint, mechanisms in IPV settings generally fail to extract the full surplus as revenue due to the necessity of providing incentives for the bidder to reveal his true type, which is private information. In contrast, in correlated settings, much more can be done. Moreover, there does not need to be much correlation; the following definition is sufficient.

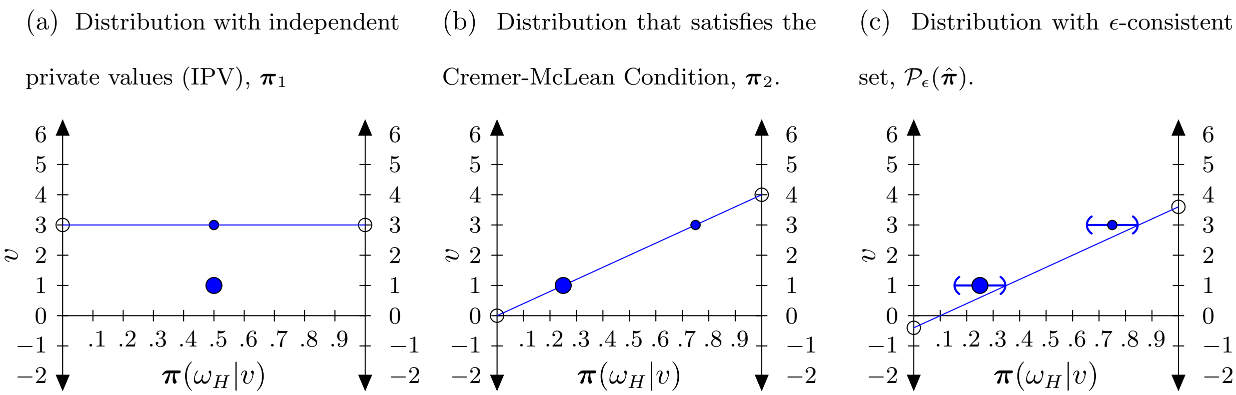
DEFINITION 6 (CREMER-MCLEAN CONDITION). The distribution over types $\boldsymbol{\pi}$, is said to satisfy the *Cremer-McLean (CM) condition* if the set of conditional distributions with respect to bidder type, $\{\boldsymbol{\pi}(\cdot|\theta)\}_{\theta\in\Theta}$, are convex independent. Equivalently, for all $\theta \in \Theta$, there does not exist $\rho(\theta') \geq 0$ for $\theta' \in \Theta/\{\theta\}$ such that $\sum_{\theta'\in\Theta/\{\theta\}} \rho(\theta') = 1$ and $\boldsymbol{\pi}(\cdot|\theta) = \sum_{\theta'\in\Theta/\{\theta\}} \rho(\theta')\boldsymbol{\pi}(\cdot|\theta')$:

While the CM condition is not specifically about Pearson correlation, it is a statement about dependence between the external signal and the bidder type, the intuitive notion of correlation. Specifically, it says that every distinct bidder type contains information about the probability of the external signal. Moreover, it is a *generic* condition, i.e., a condition that holds with probability one for a random distribution.

THEOREM 1 **(Cremer and McLean (1988))**. *If the Cremer-McLean condition is satisfied by the distribution* $\boldsymbol{\pi}$*, then there exists an ex-interim IR and Bayesian IC mechanism that extracts the full social surplus as revenue.*

By Theorem 1, the mechanism designer can generate as much revenue in expectation as if she knew the bidder's valuation. She can do this by offering lotteries over the external signal. This allows expected payments to depend on the bidder's conditional distribution, which the bidder cannot affect. Therefore, with correlation, private information has no value for the bidder. This is in sharp contrast to the IPV setting where a consequence of the bidder possessing private information is that the seller must share some of the expected social surplus with the bidder.

Figure 1 The points represent the bidder type, where the position along the x-axis is the probability that the external signal is high, and the relative size of the point represents the marginal probability of that bidder type. The lines represent lotteries offered in the mechanism, with the payment for the lottery if ω_H is observed being the intersection with the right vertical axis, and the payment if ω_L is observed is the intersection with the left axis. The height of the line at each point is the expected payment for that lottery for the conditional distribution represented by the point on the horizontal axis. The bidder accepts a lottery if and only if the expected payment is less than or equal to his valuation (IR) and chooses the lottery with the lowest expected payment (IC). For these mechanisms, if a bidder accepts a lottery, the item is allocated with probability 1.



Therefore, ϵ-robust mechanisms allow for a middle ground between the distribution independence of ex-post mechanisms and the optimality of Bayesian mechanisms. If ϵ is small, this will give us strong guarantees that our mechanism will likely do at least as well as an ex-post mechanism while allowing for the mechanism to perform nearly optimally if the ϵ-consistent set is sufficiently small. Additionally, expected revenue (i.e., a smaller ϵ-consistent set) can be traded off against robustness (a smaller ϵ). Indeed, in Section 5 we demonstrate that ϵ can be set such that the expected revenue of the ϵ-robust mechanism is nearly optimal. Example 1 demonstrates how the Bayesian mechanism maximizes revenue when the CM condition is satisfied and how the ϵ-robust mechanism trades off revenue for robustness.

Example 1. Suppose that there is a single bidder and an external signal that is correlated with the bidder's valuation. Both the bidder valuations and the external signal are binary, and we will denote the bidder valuations by $v \in \{1,3\}$ and the possible values of the external signal by $\omega \in \{\omega_L, \omega_H\}$. Denote the distribution of the bidder's valuations and the external signal by

$$\boldsymbol{\pi}_1(v,\omega) = \begin{bmatrix} 1/3 \; 1/3 \\ \\ 1/6 \; 1/6 \end{bmatrix} \qquad\qquad \boldsymbol{\pi}_2(v,\omega) = \begin{bmatrix} 1/2 \;\; 1/6 \\ \\ 1/12 \; 1/4 \end{bmatrix}$$

where the indices are ordered such that $\pi_2(v=3,\omega=\omega_L) = 1/12$. Note that the marginal distributions over v are identical for $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$. It is clear that in $\boldsymbol{\pi}_1$ the bidder's valuation and the external signal are uncorrelated, implying that the optimal mechanism is a reserve price mechanism (Myerson 1981), shown in Figure 1a, with a price of 3 and an expected revenue of 1. However, $\boldsymbol{\pi}_2$ satisfies the Cremer-McLean condition, and therefore, the seller can extract full surplus as revenue (the full $\frac{5}{3}$), as in Figure 1b. She can do this by offering a lottery such that if $\omega = \omega_H$ the bidder pays 4 and if $\omega = \omega_L$ the bidder pays 0.

Now suppose that instead of precise knowledge of the distribution, the mechanism designer has a point estimate $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}_2$ and an ϵ-consistent set of distributions, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$, such that $\pi(\omega_H|1) \in [.15,.35]$ and $\pi(\omega_H|1) \in [.65,.85]$ with probability $1-\epsilon$. The robust mechanism is shown in Figure 1c. Note that the expected payments for each type is reduced by $\frac{2}{5}$ relative to the known distribution case. Therefore, the expected revenue of the mechanism is $\frac{19}{15}$, lower than the mechanism in Figure 1b. However, for all distributions in the ϵ-consistent set, the mechanism is IC and IR.

4. Computing ϵ-Robust Mechanisms

In this section, we will introduce a new mechanism design technique that can efficiently compute ϵ-robust mechanisms. Specifically, we will combine techniques from automated mechanism design and robust convex optimization to automate the design of ϵ-robust mechanisms.

The definition of an ϵ-robust optimal mechanism, Definition 5, combined with the constraints for ϵ-robust individual rationality, ϵ-robust incentive compatibility, and feasibility (i.e., that the item can be allocated at most once) define a linear optimization problem. For either the ex-post or Bayesian case, an optimal mechanism can be efficiently computed as a solution to this linear program (Guo and Conitzer 2010). However, for an ϵ-robust mechanism, this linear program generally contains an infinite number of constraints over a, potentially, non-convex set. As a consequence, designing an ϵ-robust mechanism may be computationally intractable. However, the following assumption allows computational tractability.

ASSUMPTION 1. *For all $\theta \in \Theta$, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$ is a convex n-polyhedron where n is polynomial in the number of bidder types.*

Assumption 1 includes very reasonable cases such as the case where for all $\theta \in \Theta$ and $\omega \in \Omega$, $\pi(\omega|\theta) \in [\underline{\pi}(\omega|\theta), \overline{\pi}(\omega|\theta)]$ for some fixed values $\underline{\pi}(\omega|\theta) \leq \overline{\pi}(\omega|\theta)$. Further, any set that does not satisfy Assumption 1 can be contained in a set that does. Therefore, we can always make the assumption hold by using a larger ϵ-consistent set.

THEOREM 2. *Let $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ satisfy Assumption 1. Then for a given $(\boldsymbol{p}, \boldsymbol{x})$ there exists a polynomial time algorithm that determines whether there exists a $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$ such that ϵ-robust individual rationality or ϵ-robust incentive compatibility is violated.*

Proof. For each $\theta \in \Theta$, solve the following linear program

$$\min_{\boldsymbol{\pi}(\cdot|\theta)} \sum_{\omega} \pi(\omega|\theta)(v(\theta)p(\omega,\theta) - x(\omega,\theta))$$

$$\textit{subject to } \boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$$

Note that in the above program, $(\boldsymbol{p}, \boldsymbol{x})$ are no longer variables but coefficients. If the objective value is less than 0, then the ϵ-robust IR constraint with distribution $\boldsymbol{\pi}(\cdot|\theta)$ is violated. If the objective value is at least 0, none of the ϵ-robust IR constraints are violated for the given θ. There are $|\Theta|$ linear programs that must be solved, each with a polynomial number of variables and constraints, due to Assumption 1.

Similarly for ϵ-robust incentive compatibility, the following program, for all $\theta, \theta' \in \Theta$, finds violated constraints:

$$\min_{\boldsymbol{\pi}(\cdot|\theta)} \sum_{\omega} \pi(\omega|\theta)\left(v(\theta)p(\omega,\theta) - x(\omega,\theta) - (v(\theta)p(\omega,\theta') - x(\omega,\theta'))\right)$$

$$\textit{subject to } \boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta)) \quad \square$$

COROLLARY 2. *Let $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ satisfy Assumption 1. Then an optimal ϵ-robust mechanism can be computed with time complexity that is polynomial in the number of types of the bidder, $|\Theta|$ and external signal, $|\Omega|$.*

Proof. First, we can compute the Bayesian mechanism using the point estimate $\hat{\boldsymbol{\pi}}$ with a linear program. By Theorem 2, we can then determine whether or not an ϵ-robust IR or ϵ-robust IC constraint is violated in polynomial time for this Bayesian mechanism. Additionally, the argmax of the linear programs is a conditional distribution for a bidder type θ for which the constraint is violated. We can then add in an additional Bayesian IC or ex-interim IR constraint, using the identified conditional distribution, to the original Bayesian mechanism linear program and compute a new mechanism. Repeat this process until there are no conditional distributions in $\mathcal{P}(\hat{\boldsymbol{\pi}})$ that violate ϵ-robust IC or ϵ-robust IR. Given that there are $2|\Theta||\Omega|$ variables in the linear program in Definition 5, and there are $2|\Theta||\Omega|$ non-IC and IR constraints, this constraint generation procedure is guaranteed to terminate in a polynomial number of iterations (Kozlov et al. 1980). □

Note that Corollary 2 states that the optimal ϵ-robust mechanism is polynomial in the *number of bidder types*, not the number of bidders. In general, mechanism design is NP-hard in the number of bidders (Guo and Conitzer 2010). Given that ϵ-robust mechanisms are a generalization of traditional mechanism design, an extension of this mechanism design procedure to multiple bidders will also be NP-hard in the number of bidders. However, much of the advantage of prior-dependent mechanisms will be in thin auctions, so we do not view this as a significant weakness of this approach.

5. Sample Complexity of ϵ-Robust Mechanisms

In this section, we show, using a *non-parametric* estimate of the true distribution, that we can compute an optimal ϵ-robust mechanism that converges to the optimal revenue achievable if the true distribution is known, assuming that the true underlying distribution satisfies the CM condition (Definition 6). Moreover, we show that we require only a polynomial number of samples from the underlying distribution to achieve an additive k-approximation, where the sample complexity is over the number of bidder types, the number of external signals, the largest valuation $v(|\Theta|)$, and the amount of correlation (a concept that will be made precise). We will achieve this by carefully balancing the robustness of the mechanism against the expected revenue.

We will restrict this analysis to the case where the true underlying distribution satisfies the CM condition. This is due to the absence of a general characterization of optimal mechanisms in correlated valuation settings when the CM condition does not hold. However, we expect some elements of the sample complexity results presented here to carry over to bounding the general approximation error if and when a full characterization of the optimal mechanism is given when the CM condition does not hold. We will require that the true distribution not only satisfies the CM condition, but that it satisfies the CM condition by a sufficient margin. Intuitively, the true distribution must be "sufficiently" correlated, where the definition of correlation is the CM condition. The following definition formalizes this:

Definition 7 (γ-Separated Sets). Let $\{A_i\}_{i\in K}$ be a set of sets of real valued vectors, i.e., for all $\boldsymbol{x}\in\bigcup_{i\in K}A_i$, $\boldsymbol{x}\in\mathbb{R}^n$. For any set A, let $Conv\,(A)$ be the convex hull of the set A. Then $\{A_i\}_{i\in K}$ is said to be *γ-separated* if there exists $\gamma>0$ such that for all $i\in K$:

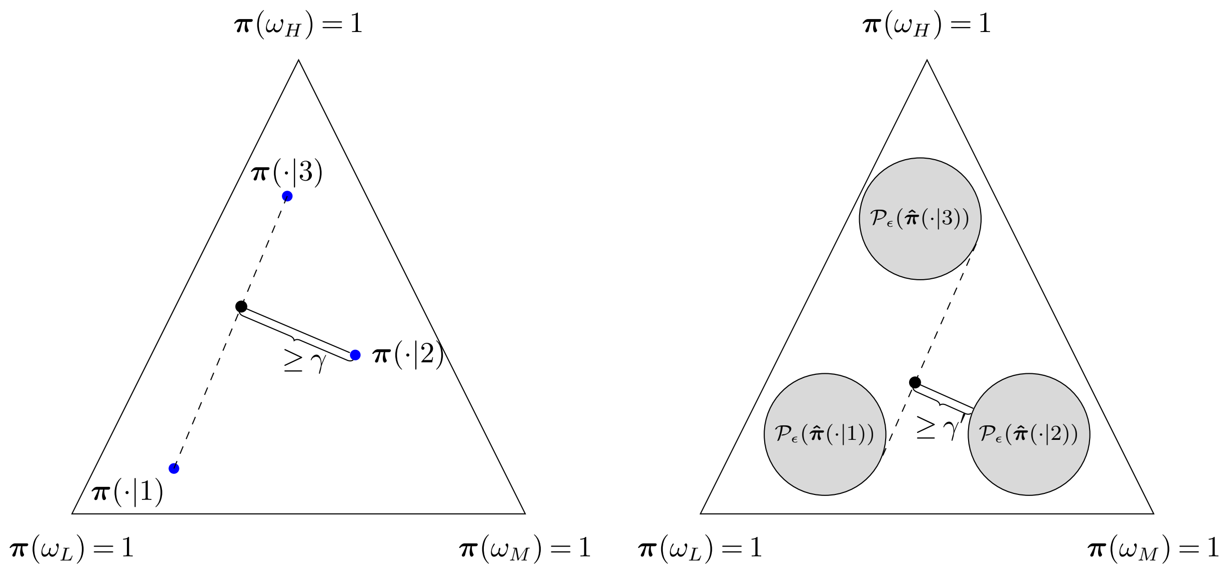
$$\gamma \leq \min_{\boldsymbol{x}\in Conv\left(\bigcup_{j\in K/i}\{A_j\}\right),\,\boldsymbol{x'}\in Conv(A_i)} ||\boldsymbol{x}-\boldsymbol{x'}||. \tag{1}$$

We will refer to a distribution, $\boldsymbol{\pi}$, as a *γ-separated distribution* if the set of singleton sets $\{\{\boldsymbol{\pi}(\cdot|\theta)\}\}_{\theta\in\Theta}$ is γ-separated. Additionally, we will refer to an ϵ-consistent set, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$, as a *$\gamma$-separated ϵ-consistent set* if $\{\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))\}_{\theta\in\Theta}$ is γ-separated.

A γ-separated distribution is, intuitively, a distribution for which the set of conditional distributions are convex independent (and satisfies the CM condition) by at least γ (see Figure 2). Moreover, if an ϵ-consistent set is γ-separated, then every distribution in the ϵ-consistent set is *at least* γ-separated. Therefore, if the true distribution is in the ϵ-consistent set, then it is at least γ-separated. Conversely, if the true distribution is γ-separated, then a γ'-separated ϵ-consistent set that contains the true distribution must have $\gamma'\leq\gamma$, i.e., the γ-separation of a distribution is an

Figure 2 Illustration of a γ-separated distribution, $\boldsymbol{\pi}$, and ϵ-consistent set, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$. Illustrated below is a 2-simplex on which points represent distributions over three outcomes. The external space $\Omega = \{\omega_L, \omega_M, \omega_H\}$, and the bidder type space $\Theta = \{1, 2, 3\}$.

(a) Illustration of a γ-separated distribution, $\boldsymbol{\pi}$ (b) Illustration of a γ'-separated ϵ-consistent set $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$



upper bound on the γ'-separation of an ϵ-consistent set that contains that distribution. Throughout this section we will use γ for γ-separated distributions and γ' for γ'-separated ϵ-consistent sets.

We assume that we have a set of samples $X = \{(\theta_1, \omega_1), (\theta_2, \omega_2), ..., (\theta_{|X|}, \omega_{|X|})\}$ from the true γ-separated distribution, $\boldsymbol{\pi}$. The mechanism designer does not know γ. For a given number of samples $|X|$, there will be a minimum $\gamma_{\min}$ such that the true distribution $\boldsymbol{\pi}$ must be at least $\gamma_{\min}$-separated in order to bound the additive approximation. We will use $\hat{\boldsymbol{\pi}}_X$ to denote the empirical distribution function of the set of samples X; $\hat{\pi}_X(\theta, \omega) = \frac{1}{|X|} \sum_{x \in X} \mathbb{1}_{\{x = (\theta, \omega)\}}$. We will refer to the subset of samples in which the bidder type is $\theta \in \Theta$ as X_θ. Additionally, for the sake of exposition, we will assume that the marginal probability for any bidder type $\theta \in \Theta$ is bounded below by $\underline{\pi}_\Theta$. The sample complexity result we present in this section will be a function of $\underline{\pi}_\Theta$. This will allow us to bound the probability of seeing a sufficient number of samples of each bidder type. Given that, in general, a bidder type may have a marginal probability of 0, this is a significant assumption. However, we relax this assumption in Appendix B.1, and a modified polynomial sample complexity result holds independent of the minimum marginal probability.

The mechanism designer will use the following procedure to implement the mechanism. First, the mechanism designer computes the empirical distribution $\hat{\boldsymbol{\pi}}_X$. Second, the ϵ-consistent set will be computed as $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta)) = \{\boldsymbol{\pi} \in P(\Omega) | \sum_{\omega\in\Omega} |\hat{\pi}_X(\omega|\theta) - \pi(\omega)| \leq \delta_\theta\}$ for all $\theta \in \Theta$ and, similarly, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X) = \{\boldsymbol{\pi} \in P(\Theta\times\Omega) | \sum_{(\theta,\omega)\in\Theta\times\Omega} |\hat{\pi}_X(\theta,\omega) - \pi(\theta,\omega)| \leq \delta\}$ for carefully chosen δ, $\{\delta_\theta\}_{\theta\in\Theta}$, and ϵ. The choice of δ, $\{\delta_\theta\}_{\theta\in\Theta}$, and ϵ will be made to balance the conflicting needs of a small ϵ-consistent set and a small probability of the true distribution not being in the ϵ-consistent set. Then, the mechanism designer will compute an optimal ϵ-robust mechanism, $(\boldsymbol{p}, \boldsymbol{x})$, with the additional constraints:

$$-\frac{2v(|\Theta|)}{\gamma_{\min}} \leq x(\theta,\omega) \leq \frac{4v(|\Theta|)}{\gamma_{\min}} \qquad \forall\,\theta\in\Theta,\, \omega\in\Omega. \tag{2}$$

Equation (2) ensures that the payment is always bounded, which will allow the mechanism designer to bound the worst and best case outcomes. We will show that the constraints (2) do not bind for a "sufficiently" γ'-separated consistent set $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$. Specifically, if $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is at least $\frac{\gamma_{\min}}{2}$-separated, then the constraints (2) will not affect the revenue of an optimal ϵ-robust mechanism.

We will predicate all of our analysis of the expected revenue of the ϵ-robust mechanism on the assumption that $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is at least $\frac{\gamma_{\min}}{2}$-separated. However, if $\delta_\theta \leq \frac{\gamma_{\min}}{4}$ for all $\theta\in\Theta$, it is easy to see that if $\boldsymbol{\pi}\in\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ must be $\frac{\gamma_{\min}}{2}$-separated given that we assume that $\boldsymbol{\pi}$ is at least $\gamma_{\min}$-separated. Therefore, ϵ is not just an upper bound on the probability that the mechanism is not IR or IC, it is also an upper bound on the probability that $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is not $\frac{\gamma_{\min}}{2}$-separated.

In order to show that the above procedure generates nearly optimal revenue in expectation, we will need to bound three sources of loss. First, we must bound the loss if $\boldsymbol{\pi}\notin\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ or $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is not $\frac{\gamma_{\min}}{2}$-separated. The probability of this is bounded by ϵ, assuming that for all $\theta\in\Theta$, $\delta_\theta \leq \frac{\gamma_{\min}}{2}$. As a worst case assumption, we will assume that we have constructed a mechanism where the bidder should face a payment of $\frac{4v(|\Theta|)}{\gamma_{\min}}$ (an upper bound on the payments by (2)) but, instead, faces a payment of $-\frac{2v(|\Theta|)}{\gamma_{\min}}$ (the minimum possible payment), for a loss in expected revenue of $\frac{6v(|\Theta|)}{\gamma_{\min}}$. Second, we must bound the loss when $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is $\frac{\gamma_{\min}}{2}$-separated and $\boldsymbol{\pi}\in\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$. This loss is due to two factors; the optimal ϵ-robust mechanism must satisfy constraints over a set of distributions instead of just a single distribution (namely ϵ-robust IR and IC), and the distribution used to optimize the ϵ-robust mechanism is an estimate of the true distribution. We will show that for a $\frac{\gamma_{\min}}{2}$-separated "sufficiently" small ϵ-consistent set, an ϵ-robust mechanism performs nearly optimally. Third and finally, we must ensure that we have a sufficient number of samples for each bidder type $\theta \in \Theta$, i.e., $|X_\theta|$ is sufficiently large, to learn a sufficiently small ϵ-consistent set.

First, we will identify the relationship between ϵ, δ, $\{\delta_\theta\}_{\theta\in\Theta}$, and $|X|$ in Lemmas 2 and 3. To do so, we will rely on the following concentration inequality.

Lemma 1 **(Devroye (1983))**. *Let $(\theta_1, ..., \theta_k)$ be a multinomial $(n, p_1, ..., p_k)$ random vector. For all $\delta \in (0,1)$ and all k satisfying $\frac{k}{n} \le \frac{\delta^2}{20}$, we have*

$$P\left(\sum_{i=1}^{k} |\theta_i - E(\theta_i)| > n\delta\right) \le 3e^{\frac{-n\delta^2}{25}} \tag{3}$$

Lemma 2. *Let $\boldsymbol{\pi}$ be the true distribution over $\Theta \times \Omega$, and let X be a set of independent samples from $\boldsymbol{\pi}$. For $\theta \in \Theta$, if $|X_\theta| \ge 25|\Omega| \ln\left(\frac{1}{\epsilon_\theta}\right)\left(\frac{1}{\delta_\theta}\right)^2$, then $\sum_{\omega\in\Omega} |\hat{\pi}_X(\omega|\theta) - \pi(\omega|\theta)| \le \delta_\theta$ with probability $1-\epsilon_\theta$.*

Proof. By Devroye's Lemma (Lemma 1), with $\epsilon_\theta = 3e^{\frac{-|X_\theta|\delta_\theta^2}{25}}$, the number of samples necessary is:

$$|X_\theta| \ge 25 \ln\left(\frac{1}{\epsilon_\theta}\right)\left(\frac{1}{\delta_\theta}\right)^2.$$

However, $\frac{|\Omega|}{|X_\theta|} \le \frac{\delta_\theta^2}{20}$ implying:

$$|X_\theta| \ge 25|\Omega| \ln\left(\frac{1}{\epsilon_\theta}\right)\left(\frac{1}{\delta_\theta}\right)^2 \ge \max\left\{25\ln\left(\frac{1}{\epsilon_\theta}\right)\left(\frac{1}{\delta_\theta}\right)^2, \frac{20|\Omega|}{\delta_\theta^2}\right\} \tag{4}$$

is sufficient. □

Lemma 3. *Let $\boldsymbol{\pi}$ be the true distribution over $\Theta \times \Omega$, and let X be a set of independent samples from $\boldsymbol{\pi}$. Then, if $|X| \ge 25|\Omega||\Theta| \ln\left(\frac{1}{\epsilon_1}\right)\left(\frac{1}{\delta_1}\right)^2$, $\sum_{(\theta,\omega)\in\Theta\times\Omega} |\hat{\pi}_X(\theta,\omega) - \pi(\theta,\omega)| \le \delta_1$ with probability $1-\epsilon_1$.*

Proof. Proof is identical to Lemma 2. □

Now, we will bound the number of samples $|X|$ necessary to ensure that we see a sufficient number of samples from each bidder type $\theta \in \Theta$.

LEMMA 4 **(Mitzenmacher and Upfal (2005))**. *Let θ be a binomial (n,p) random variable. For all $\delta \in (0,1)$, we have*

$$P(\theta < (1-\delta)np) \leq e^{-\frac{\delta^2 np}{2}} \tag{5}$$

LEMMA 5. *Let $\boldsymbol{\pi}$ be the true distribution over $\Theta \times \Omega$, and let X be a set of independent samples from $\boldsymbol{\pi}$. Then, if $|X| \geq \frac{8M}{\underline{\pi}_\Theta} \ln\left(\frac{|\Theta|}{\epsilon_2}\right)$, for all $\theta \in \Theta$, $|X_\theta| \geq M$ with probability $1-\epsilon_2$.*

Proof. We can model the probability of any sample from X being of bidder type $\theta \in \Theta$ as a binomial distribution with success probability $\pi(\theta) \geq \underline{\pi}_\Theta$. Since we need every bidder type $\theta \in \Theta$ to have a sufficient number of samples with probability $1-\epsilon_2$, we require that any specific bidder type have a sufficient number of samples with probability at least $1-\frac{\epsilon_2}{|\Theta|}$. Then, by a union bound, the probability that there exists $\theta \in \Theta$ such that $|X_\theta| < M$ will be bounded by ϵ_2. By Lemma 4 with $\delta = \frac{1}{2}$, we can bound the number of samples necessary to ensure that $|X_\theta| \geq M$ with probability at least $1-\frac{\epsilon_2}{|\Theta|}$ by

$$|X| \geq \frac{8M}{\underline{\pi}_\Theta} \ln\left(\frac{|\Theta|}{\epsilon_2}\right) \geq \max\left\{\frac{8}{\underline{\pi}_\Theta} \ln\left(\frac{|\Theta|}{\epsilon_2}\right), \frac{2M}{\underline{\pi}_\Theta}\right\}.$$

□

In order to bound the potential loss in expected revenue due to the possibility that $\boldsymbol{\pi} \notin \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$ we need to bound the worst case outcome. We have done this by constraining the payments with the constraints in (2). In Lemma 6, we show that for a sufficiently γ'-separated ϵ-consistent set, this does not affect the revenue of an optimal ϵ-robust mechanism.

LEMMA 6. *If the ϵ-consistent set, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}})$, is γ'-separated for $\gamma' > 0$, then there exists a revenue optimal ϵ-robust mechanism $(\boldsymbol{p}, \boldsymbol{x})$ such that for all $\theta \in \Theta$ and $\omega \in \Omega$,*

$$-\frac{v(|\Theta|)}{\gamma'} \leq x(\theta,\omega) \leq \frac{v(|\Theta|)}{\gamma'} + v(|\Theta|) \leq \frac{2v(|\Theta|)}{\gamma'}. \tag{6}$$

The proof is in Appendix B.

Finally, we must bound the loss in revenue due to using a robust mechanism. This loss will be due to both the robust constraints as well as the estimated objective. The following lemma demonstrates that when the consistent set is small enough, the loss due to using a robust mechanism is arbitrarily small.

LEMMA 7. *Let $\mathcal{P}(\hat{\boldsymbol{\pi}})$ be a γ'-separated consistent set, i.e., $\epsilon = 0$, such that for all $\theta \in \Theta$ there exists a $\delta > 0$ such that $\delta \geq \max_{\boldsymbol{\pi},\boldsymbol{\pi}' \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))} ||\boldsymbol{\pi} - \boldsymbol{\pi}'||$, $\delta \geq \max_{\boldsymbol{\pi},\boldsymbol{\pi}' \in \mathcal{P}(\hat{\boldsymbol{\pi}})} ||\boldsymbol{\pi} - \boldsymbol{\pi}'||$, and $\delta \leq \frac{k\gamma'}{6v(|\Theta|)|\Omega|}$. Let $\boldsymbol{\pi}^* \in \mathcal{P}(\hat{\boldsymbol{\pi}})$ be a distribution with optimal revenue R when the distribution is known. Then an optimal robust mechanism achieves at least $R - k$ in expected revenue for any $\boldsymbol{\pi} \in \mathcal{P}(\hat{\boldsymbol{\pi}})$.*

The proof is in Appendix B.

Combining the above results, we can achieve our main result, a polynomial bound on the number of samples necessary to achieve a k-additive approximation to the optimal expected revenue when the distribution is exactly known.

THEOREM 3. *Let the true distribution, $\boldsymbol{\pi}$, be a γ-separated distribution that satisfies the CM condition with an optimal revenue of R. Assume that $\gamma \geq \gamma_{\min}$. Let X be independent samples from $\boldsymbol{\pi}$. If $|X| \geq \frac{200|\Omega|^3|\Theta|}{\underline{\pi}_\Theta} \ln^2\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)}{k\gamma_{\min}}\right)^2$, then an optimal $\frac{k\gamma_{\min}}{18v(|\Theta|)}$-robust mechanism with payments bounded by (2) has an expected revenue of at least $R - k$.*

Proof. We will start with the assumption that δ and $\{\delta_\theta\}_{\theta\in\Theta}$ are less than $\frac{\gamma_{\min}}{4}$. We will bound the loss due to this not being the case later in the proof. If $\boldsymbol{\pi} \notin \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$, then the loss of revenue is bounded by $\frac{6v(|\Theta|)}{\gamma_{\min}}$ due to the assumption that payments are bounded by (2). Therefore, if $\epsilon \leq \frac{k\gamma_{\min}}{18v(|\Theta|)}$, the revenue loss due to $\boldsymbol{\pi} \notin \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is at most $\frac{k}{3}$.

If $\boldsymbol{\pi} \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ and $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is $\frac{\gamma_{\min}}{2}$-separated, we can bound the loss due to robustness using Lemma 7. If $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X) \subseteq \left\{\boldsymbol{\pi} \,\middle|\, ||\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}_X|| \leq \frac{k\gamma_{\min}}{72v(|\Theta|)|\Omega|}\right\}$, and similarly for $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$, then the expected revenue of the mechanism is at least $R - \frac{k}{3}$.

To ensure that $\boldsymbol{\pi} \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ and for all $\theta \in \Theta$, $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$ with probability at least $\epsilon = \frac{k\gamma_{\min}}{18v(|\Theta|)}$, set $\epsilon_\theta = \epsilon_1 = \frac{k\gamma_{\min}}{18(|\Theta|+1)v(|\Theta|)}$, using the notation of Lemmas 2 and 3. By Lemma 2, if

$\delta_\theta \le \frac{k\gamma_{\min}}{72v(|\Theta|)|\Omega|} \le \frac{\gamma_{\min}}{4}$, it must be that $|X_\theta| \ge 25|\Omega| \ln\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2$. By Lemma 3, if $\delta_1 \le \frac{k\gamma_{\min}}{72v(|\Theta|)|\Omega|}$, it must be that $|X| \ge 25|\Theta||\Omega| \ln\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2$.

Finally, we must bound the loss due to the probability that for some $\theta \in \Theta$, $|X_\theta| < 25|\Omega| \ln\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2$. If there does exist $\theta \in \Theta$ such that $|X_\theta| < 25|\Omega| \ln\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2$, then we will again assume that we lose $\frac{6v(|\Theta|)}{\gamma_{\min}}$. Using the notation of Lemma 5, if $\epsilon_2 \le \frac{k\gamma_{\min}}{18v(|\Theta|)}$ this revenue loss is bounded by $\frac{k}{3}$. Therefore, by Lemma 5,

$$|X| \ge \frac{200|\Omega|^3}{\underline{\pi}_\Theta} \ln\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right) \ln\left(\frac{18v(|\Theta|)|\Theta|}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)}{k\gamma_{\min}}\right)^2. \tag{7}$$

Putting this all together, if $|X| \ge \frac{200|\Omega|^3|\Theta|}{\underline{\pi}_\Theta} \ln^2\left(\frac{18(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{72v(|\Theta|)}{k\gamma_{\min}}\right)^2$ all conditions are satisfied. $\square$

Note that, as is common in sample complexity results, Theorem 3 gives a very loose upper bound on the number of samples necessary. Many of the simplifying steps in the proof increased the number of samples by significant factors, and a more complicated analysis would likely significantly reduce this bound. We show in Section 7, with a naïve estimation procedure, we can generate nearly optimal revenue with far fewer samples than Theorem 3 suggests. However, Theorem 3 formally demonstrates that this estimation problem, even in the worst case, requires only a polynomial number of samples.

6. Necessity of γ-Separation

Suppose that we start with some consistent set $\mathcal{P}(\hat{\boldsymbol{\pi}})$ that does not satisfy Definition 7 for any value of γ, and we have access to a finite number of samples from the true distribution $\boldsymbol{\pi}$, can we guarantee nearly optimal revenue for the true distribution by using the samples in the mechanism design process? Put another way, we show in Section 5 that γ-separation of the true distribution, Definition 7, is sufficient, but is it necessary? In the setting where $\mathcal{P}(\hat{\boldsymbol{\pi}})$ is finite, Fu et al. (2014) showed that with relatively few samples, full surplus extraction is possible, but does this extend to an infinite set of possible distributions?

In our setting, and likely in practice, the true distribution lies in a continuous space and any reasonable distribution estimation procedure will return a continuous set of distributions that are consistent with the observed samples. In this section we demonstrate that in the worst case, there is *no mechanism that uses a finite number of samples from the true distribution that can approximate the optimal revenue*, implying that the assumption of a γ-separated distribution is both necessary and sufficient.

If we are to state that γ-separation is a necessary condition for learning the optimal mechanism, we must show that if γ-separation does not hold, the optimal revenue cannot be achieved *by any mechanism design procedure that uses a finite number of samples.* However, for any given distribution that satisfies the CM condition, γ-separation will hold for some γ, and therefore can be learned with a finite number of samples, by Theorem 3, using an ϵ-robust mechanism. Therefore, we consider the case where we have an ϵ-consistent set of distributions where $\epsilon = 0$ for the set. Then, we will argue that even if every distribution in the consistent set satisfies the CM condition, but the set of distributions converges to a distribution that does not satisfy the CM condition, then there is no mechanism design procedure that uses a finite number of samples that can guarantee nearly optimal revenue. Our definition of convergence of distributions is given below.

Definition 8 (Converging Distributions). A countably infinite sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ is said to be *converging to the distribution* $\boldsymbol{\pi}^*$, the *convergence point*, if for all $\theta \in \Theta$ and for an arbitrary $\epsilon > 0$, there exists a $T \in \mathbb{N}$ such that for all $i \geq T$, $||\boldsymbol{\pi}_i(\cdot|\theta) - \boldsymbol{\pi}^*(\cdot|\theta)|| < \epsilon$.

It is trivial to show that if a sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ converges to a distribution $\boldsymbol{\pi}^*$ that does not satisfy the CM condition, then there does not exist $\gamma > 0$ such that the sequence of distributions is γ-separated. Moreover, if for any consistent set of distributions $\mathcal{P}(\hat{\boldsymbol{\pi}})$ where for any fixed $\gamma' > 0$, there exists an infinite set of distributions $\mathcal{P}' \subset \mathcal{P}(\hat{\boldsymbol{\pi}})$ such that for all distributions $\boldsymbol{\pi'} \in \mathcal{P}'$, $\boldsymbol{\pi'}$ is not γ'-separated, then there exists a sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ that converges to a distribution that does not satisfy the CM condition. This will be true for any continuous consistent set of distributions $\mathcal{P}(\hat{\boldsymbol{\pi}})$ that is not γ-separated.

Additionally, we will require that the sequence of distributions converges to a point in the interior of the convex hull of the sequence. This is not without loss of generality, but given that the convergence point must be in the closure of the convex hull, the set of convergence points we rule out is a measure zero set, so we feel Assumption 2 is reasonable. For a full discussion and intuition for Assumption 2, see Appendix C. We leave it to future work to relax Assumption 2.

ASSUMPTION 2. *For the sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ converging to $\boldsymbol{\pi}^*$ and for any $\theta' \in \Theta$, there exists a subset of distributions of size $|\Omega|$ from the set $\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ that is affinely independent and the distribution $\boldsymbol{\pi}^*(\cdot|\theta')$ is a strictly convex combination of the elements of the subset. I.e., there exists $\{\alpha_k\}_{k=1}^{|\Omega|}$, $\alpha_k \in (0,1)$ and $\{\boldsymbol{\pi}_k(\cdot|\theta_k)\}_{k=1}^{|\Omega|}$, where every $\boldsymbol{\pi}_k(\cdot|\theta_k) \in \{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$, such that $\boldsymbol{\pi}^*(\cdot|\theta') = \sum_{k=1}^{|\Omega|} \alpha_k \boldsymbol{\pi}_k(\cdot|\theta_k)$.*

With Assumption 2 and Definition 8, we can state the main result.

COROLLARY 3. *Let $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ be a sequence of distributions that satisfies Assumption 2 and converges to $\boldsymbol{\pi}^*$. Let the true (unknown) distribution $\boldsymbol{\pi'} \in \{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$. There does not exist a mechanism design procedure using a finite number of independent samples from $\boldsymbol{\pi'}$ that guarantees a constant approximation to the optimal revenue.*

See Appendix C for a full discussion and proof of Corollary 3. The intuition is that any mechanism with positive expected revenue must have a bounded minimum payment by Assumption 2. Therefore, any mechanism that satisfies IR must have a bounded maximum payment. These bounded payments imply that the achievable revenue for each distribution in the sequence, even distributions that satisfy the CM condition, must converge smoothly to the achievable revenue at the convergence distribution, which can be an IPV distribution and arbitrarily low.

7. Experimental Results

While we have demonstrated that sufficient correlation is both necessary and sufficient, it is difficult to gain an intuitive appreciation for ϵ-robust mechansims with purely theoretical results. In this section, we demonstrate the improvement ϵ-robust mechanisms provide experimentally. Throughout the experiments, we have a single bidder with type $\theta \in \{1, 2, ..., 10\}$ and valuation $v(\theta) = \theta$. The external signal is $\omega \in \{1, 2, ..., 10\}$. We model the true distribution as a categorical distribution with 10×10 elements, with each element corresponding to a tuple (θ, ω).

To the best of our knowledge, there are no standard distributions to test correlated mechanism design procedures available, so we use a discretized bi-variate normal distribution. Specifically, we discretize the area under the bi-variate standard normal distribution between $[-1.96, 1.96]$ in both dimensions as a 10×10 grid and normalize. We chose the bi-variate normal distribution for its broad relevance to many empirically observed distributions and the ability to easily vary the correlation. Note that the bi-variate normal distribution always satisfies the Cremer-McLean condition if the correlation is positive.

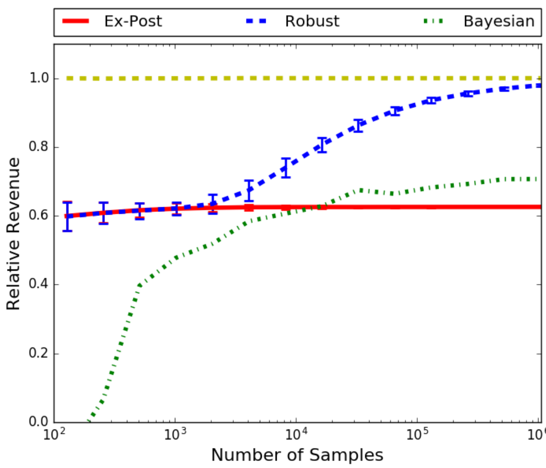
To estimate the distribution, we sample from the true distribution and use Bayesian updating with a maximally uninformative Dirichlet prior ($\alpha = [1, ..., 1]$) to arrive at a Dirichlet posterior over the distribution of bidder types and external signals. We then calculate empirical confidence intervals by sampling from the Dirichlet posterior and observing the $\epsilon/(2 \times 10 \times 10)$ and $(1 - \epsilon/(2 \times 10 \times 10))$ quantiles for each element of the conditional distributions $\boldsymbol{\pi}(\omega|\theta)$ and use the quantiles as the ϵ-consistent set. Note that we do not simply use the $\epsilon/2$ and $(1 - \epsilon/2)$ quantiles due to jointly estimating confidence intervals for 100 variables and applying a union bound.

For our experiments, we solve for the optimal ex-post, ϵ-robust, and Bayesian mechanisms given our estimated distribution $\hat{\boldsymbol{\pi}}$ and our ϵ-consistent set. Given that both the optimal ϵ-robust and Bayesian mechanisms can fail to be incentive compatible and/or individually rational due to the difference between the estimated and true distribution, we compute the optimal action for the bidder: either report truthfully, strategically misreport, or do not participate. We then calculate the revenue accordingly.

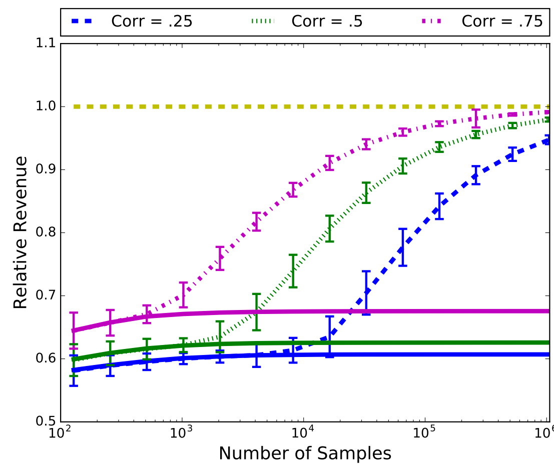
In Figure 3a, we show the performance of the optimal ex-post, robust, and Bayesian mechanisms using our estimated distribution as we increase the number of samples. We report confidence intervals for both the ex-post mechanisms and the robust mechanisms; however for the Bayesian mechanisms, the confidence intervals were off the chart. Figure 3a demonstrates how badly the Bayesian mechanism performs when the distribution is not exactly known. Even after $10,000$ samples from the true distribution, the Bayesian mechanism fails to outperform the ex-post mechanism.

Figure 3 The performance of the ex-post, ϵ-robust, and Bayesian mechanisms using the estimated distribution. All revenue is scaled by the full social surplus, denoted as 1. Note that the Number of Samples is in log scale. The parameters used, unless explicitly given, were as follows: Correlation = .5, ϵ = .05. Each experiment was repeated 200 times, and the 95% confidence interval is included for the ϵ-robust and ex-post mechanisms. The Bayesian mechanism confidence interval is off the plot.

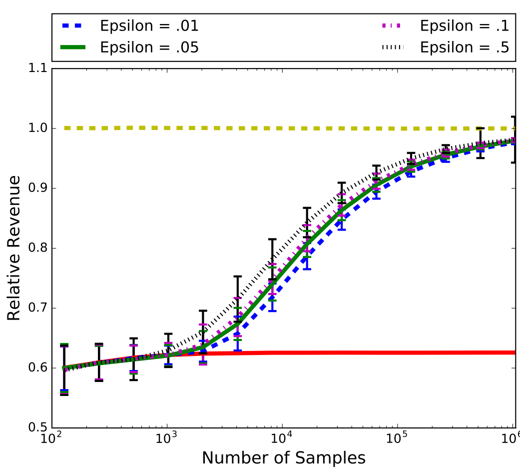
(a) Ex-post, ϵ-robust, and Bayesian mechanisms.



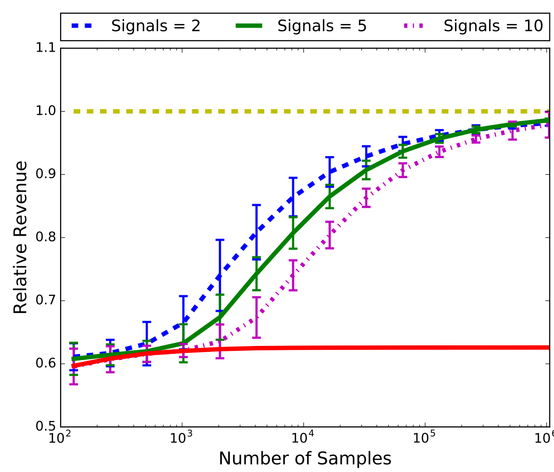
(b) Number of Samples versus Correlation.



(c) Number of Samples versus ϵ.



(d) Number of Samples versus Signal Space.



For few samples, the expected revenue is even negative for the Bayesian mechanism due to negative payments in the lotteries offered. By contrast, the optimal ϵ-robust mechanism generates revenue indistinguishable from the ex-post mechanism for low numbers of samples, while significantly outperforming the ex-post mechanism starting at about 10,000 samples.

In Figures 3b and 3c, we vary correlation and ϵ with increasing numbers of samples. As the bidder type and external signal are more highly correlated, the ϵ-robust mechanism requires fewer samples to perform well (Figure 3b). Also, we see that the ϵ-robust mechanism is not very sensitive to the choice of ϵ (Figure 3c) a fact that we attribute to being overly cautious in requiring all elements of the distribution to be in the bounded intervals.

In Figure 3d, we bin some of the external signals together in order to explore the trade-off between estimating a lower dimensional distribution and constructing a mechanism over the full information. Specifically, for the Signals $= 2$ case we put all of $\omega = \{1, ..., 5\}$ into one bin and $\omega = \{6, ..., 10\}$ to a second bin. Note the true signal still has 10 values, we are just binning the observed signal. We find that for a low number of samples, we do much better by binning the external signal, but, while difficult to see on the plot, at higher numbers of samples, it is better to use the full distribution.

Note that we consider the results here to be lower bounds on the performance of optimal ϵ-robust mechanisms. We assume a completely uninformative prior, increasing the required sample size. Further, we have used a naïve distribution estimation procedure, so there is likely significant room to improve upon the estimation.

8. Conclusion

In this work, we have presented a new mechanism design paradigm, ϵ-robust mechanisms, that takes a non-trivial step away from traditional mechanism design paradigms. Specifically, we allow for the traditional constraints of incentive compatibility and individual rationality to be probabilistically violated, and in return, we are able to compute these mechanisms efficiently and learn nearly optimal mechanisms using a polynomial number of samples from the true distribution, at least when the true distribution is γ-separated. More generally, this class of mechanisms naturally spans the distance between traditional ex-post mechanisms, a setting we can replicate with $\epsilon = 0$, and Bayesian mechanisms, $\epsilon = 1$. Therefore, we effectively parameterize the design of Bayesian mechanisms in settings with distributional uncertainty.

We have also bounded the complexity of learning mechanisms for settings with correlated bidder distributions. Corollary 3 suggests that learning the optimal mechanism is doomed in the worst case, putting a floor on what is possible. Our positive result (Theorem 3) provides a ceiling on how difficult it is. While we do not claim that our procedure is either the most sample efficient or the most computationally efficient, it does provide a benchmark and puts limits on what is possible. We leave it to future research to improve both the upper and lower bounds.

While the sample complexity is polynomially bounded, it is still relatively large, at least in our experiments. However, settings with frequently repeated auctions, such as online ad auctions or the real time auction of Amazon Web Services (AWS) EC2 "Spot Instances," are likely to have a sufficiently large set of historic bids to improve relative to ex-post mechanisms.

We do not analyze the convergence of this ϵ-robust mechanism design procedure in settings where full surplus extraction as revenue is not possible because this seems to require a better understanding of the optimal mechanism under full knowledge of the distribution, currently an open question. We hope recent results characterizing necessary and sufficient conditions for full surplus extraction (Albert et al. 2016) may point to a better understanding of the optimal mechanism when the conditions fail. Specifically, it remains an open question as to whether the optimal mechanism is *deterministic* (i.e., the allocation probability is always 0 or 1), and if there exists a way to extend the virtual valuation function approach of Myerson (1981) to a correlated valuation setting that would allow for the simple computation of the allocation function. It seems likely that a similar polynomial bound on the sample complexity of learning optimal mechanisms using ϵ-robust mechanisms will hold for the general case.

An area of particular interest for future research is applying these techniques to the problem of budget balanced, socially efficient mechanisms. As the well known Myerson-Satterthwaite impossibility result (Myerson and Satterthwaite 1983) states, it is generally impossible to have strong budget balanced, socially efficient mechanisms. However, in a correlated valuation setting, there is a generic condition which states that strong budget balanced, socially efficient mechanisms are possible (Kosenok and Severinov 2008), but these mechanisms are, much like revenue maximizing mechanisms, highly dependent on the mechanism designer's precise knowledge of the true distribution and a common prior among bidders. However, the results from this work suggest that there are likely to be computationally feasible robust mechanisms that approximately achieve budget balance and social efficiency. This would lead to new applications of incentive compatible distributed systems, such as federated server farms, where a group shares resources in an efficient manner without any money being transferred out of the system. We are currently exploring questions in this direction.

Appendix A: Type Dependent Robustness

While in Section 3.3 we present ϵ-robust mechanisms in the context of a single ϵ-consistent set, it is not strictly necessary. In the definition of ϵ-robust IC (Definition 3) and ϵ-robust IR (Definition 4) for any individual $\theta \in \Theta$ the constraints are only over the conditional distributions in $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))$. Moreover, the ϵ-consistent set only affects the mechanism to the extent that it affects the ϵ-robust IR and ϵ-robust IC constraints.

Therefore, for any given $\theta \in \Theta$ we can use a distinct ϵ_θ-consistent set of conditional distributions. Moreover, the set does not have to be the same for both IC and IR, i.e., an $\epsilon_{\theta,IC}$ and $\epsilon_{\theta,IR}$. The interpretation is that given a $\theta \in \Theta$, the probability that the mechanism fails to be IC is at most $\epsilon_{\theta,IC}$ and the probability that it fails to be IR is at most $\epsilon_{\theta,IR}$. With this more flexible view, we can introduce the more flexible notion of $\{\epsilon_{\theta,IC}\}_{\theta\in\Theta}$-robust incentive compatibility.

Definition 9 ($\{\epsilon_{\theta,IC}\}_{\theta\in\Theta}$-Robust Incentive Compatibility). A mechanism is $\{\epsilon_{\theta,IC}\}_{\theta\in\Theta}$-*robust incentive compatible* (IC) if for all $\theta \in \Theta$, given an $\epsilon_{\theta,IC}$-consistent set of distributions $\mathcal{P}_{\epsilon_{\theta,IC}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$, for all $\theta, \theta' \in \Theta$ and $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_{\epsilon_{\theta,IC}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$,

$$\sum_{\omega\in\Omega} \pi(\omega|\theta)\left(v(\theta)p(\theta,\omega) - x(\theta,\omega)\right) \geq \sum_{\omega\in\Omega} \pi(\omega|\theta)\left(v(\theta)p(\theta',\omega) - x(\theta',\omega)\right)$$

We can similarly define the more flexible notion of $\{\epsilon_{\theta,IR}\}_{\theta\in\Theta}$-robust individual rationality.

Definition 10 ($\{\epsilon_{\theta,IR}\}_{\theta\in\Theta}$-Robust Individual Rationality). A mechanism is $\{\epsilon_{\theta,IR}\}_{\theta\in\Theta}$-*robust individually rational* (IR) if for all $\theta \in \Theta$, given an $\epsilon_{\theta,IR}$-consistent set of distributions $\mathcal{P}_{\epsilon_{\theta,IR}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$, for all $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_{\epsilon_{\theta,IR}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$,

$$\sum_{\omega\in\Omega} \pi(\omega|\theta)\left(v(\theta)p(\theta,\omega) - x(\theta,\omega)\right) \geq 0$$

We can now extend the definition of an ϵ-robust mechanism to include the above extensions of IC and IR.

Definition 11 (Optimal $\{\epsilon_{\theta,IR}, \epsilon_{\theta,IC}\}_{\theta\in\Theta}$-Robust Mechanism). A mechanism $(\boldsymbol{p}, \boldsymbol{x})$ is a *revenue optimal* $\{\epsilon_{\theta,IR}, \epsilon_{\theta,IC}\}_{\theta\in\Theta}$*-robust mechanism*, given an estimated distribution $\hat{\boldsymbol{\pi}}$, if under the constraint of $\{\epsilon_{\theta,IR}\}_{\theta\in\Theta}$-robust individual rationality for $\{\epsilon_{\theta,IR}\}_{\theta\in\Theta}$-consistent sets of distributions $\{\mathcal{P}_{\epsilon_{\theta,IR}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))\}_{\theta\in\Theta}$ and $\{\epsilon_{\theta,IC}\}_{\theta\in\Theta}$-incentive compatibility for $\{\epsilon_{\theta,IC}\}_{\theta\in\Theta}$-consistent sets of distributions $\{\mathcal{P}_{\epsilon_{\theta,IC}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))\}_{\theta\in\Theta}$ it maximizes the following:

$$\sum_{\theta,\omega} x(\theta,\omega)\hat{\pi}(\theta,\omega)$$

This gives the mechanism designer the ability to allow for less robust mechanisms for certain bidder types. This may be useful for bidder types that have a very low marginal probability, for example. If the seller sees type θ with only a 1% chance, it may be okay if that type fails to find the mechanism incentive compatible or individually rational if it allows for higher revenue from higher marginal probability types. We will take advantage of this in Appendix B.1 in order to bound the sample complexity of learning an $\{\epsilon_{\theta,IR}, \epsilon_{\theta,IC}\}_{\theta\in\Theta}$-robust mechanism that performs nearly optimally with respect to the true distribution. We will be able to do this without considering the minimum marginal probability of a type, unlike in Theorem 3.

Note that while the optimal $\{\epsilon_{\theta,IR}, \epsilon_{\theta,IC}\}_{\theta\in\Theta}$-robust mechanism in Definition 11 is perfectly well defined, it is significantly more difficult to reason about. Now, there is not just a single parameter that dictates the likelihood of the mechanism being IC and IR, there are $2|\Theta|$ such parameters. However, it is trivial to show that Corollary 2 still applies, and optimal $\{\epsilon_{\theta,IR}, \epsilon_{\theta,IC}\}_{\theta\in\Theta}$-robust mechanisms can be computed efficiently. Moreover, the negative result in Corollary 3 also applies since that result relied on a consistent set with $\epsilon = 0$, so this extended definition also does not allow us to learn mechanisms for non-γ-separated distributions.

Appendix B: Additional Proofs and Discussion for Section 5

In order to demonstrate that, using a sufficient number of samples from a γ-separated true distribution, an ϵ-robust mechanism can be computed that is nearly optimal for a γ-separated true distribution, it will be useful to first guarantee that a mechanisms exists that is nearly optimal for all distributions in a small enough consistent set. The following lemma guarantees this.

Lemma 8. *For any distribution $\boldsymbol{\pi}^*$ that satisfies the CM condition and given any positive constant $k > 0$, there exists $\delta > 0$ and a mechanism such that for all distributions, $\boldsymbol{\pi}'$, for which for all $\theta \in \Theta$, $||\boldsymbol{\pi}^*(\cdot|\theta) - \boldsymbol{\pi}'(\cdot|\theta)|| < \delta$, the revenue generated by the mechanism is greater than or equal to $R - k$, where R is the optimal revenue for distribution $\boldsymbol{\pi}'$.*

Proof. By the assumption that there exists a mechanism that extracts full surplus for the distribution $\boldsymbol{\pi}^*$, there must be a mechanism that always allocates the item and leaves the bidder with an expected utility of 0. Let this mechanism be denoted by $(\boldsymbol{p}^*, \boldsymbol{x}^*)$. The lottery denoted by the payments $\boldsymbol{x}^*(\theta, \cdot)$ can be viewed as a hyperplane over the $|\Omega|$-simplex of distributions over Ω. The height of the hyperplane at any point over the simplex corresponds to the expected payment associated with the distribution at that point (see Figure 1 for intuition). Let C be the value for the largest gradient of any lottery in the mechanism (e.g., the slope of the line in Figure 1). Choose $\delta = k/(2C)$. Then the expected utility for any distribution $\boldsymbol{\pi}'(\cdot|\theta)$ with $||\boldsymbol{\pi}^*(\cdot|\theta) - \boldsymbol{\pi}'(\cdot|\theta)|| < \delta$ when optimally reporting $\theta' \in \Theta$ (note that this is, potentially, an optimal *mis*-report) is bounded by:

$$-C\delta \leq \sum_{\omega} \pi'(\omega|\theta)\left(v(\theta)p^*(\theta', \omega) - x^*(\theta', \omega)\right) \leq C\delta$$

Construct a new mechanism (not necessarily truthful) where all payments $x'(\theta, \omega) = x^*(\theta, \omega) - C\delta$ and set $p'(\theta, \omega) = p^*(\theta, \omega) = 1$. Then, the utility of the bidder for optimally misreporting is:

$$0 \leq \sum_{\omega} \pi'(\omega|\theta)\left(v(\theta)p'(\theta', \omega) - x'(\theta', \omega)\right) \leq 2C\delta$$

which implies that the bidder always participates. Since the item is always allocated, the loss in revenue is equivalent to the gain in utility for the bidder. Therefore, the mechanism $(\boldsymbol{p}', \boldsymbol{x}')$ always guarantees revenue within $2C\delta = k$ of the optimal mechanism for any θ, so the expected revenue of the mechanism is greater than or equal to $R - k$. $\square$

Note that the proof of Lemma 8 is constructive, but the mechanism is not necessarily incentive compatible. Therefore, the constructed mechanism is not an ϵ-robust mechanism. Lemma 8 is intuitively very reasonable, and likely what one would expect a priori. For a class of distributions that are sufficiently close, there should be a mechanism that does about as well on all of them.

Proof of Lemma 6. Note that the payments varying by $\omega \in \Omega$ only affects the ϵ-robust IC constraint (Definition 3). This is because the objective is only affected by the expected payment given $\theta \in \Theta$ and the marginal probability, $\hat{\pi}_{\Theta}(\theta)$. Moreover, the ϵ-robust IR constraint (Definition 4), similarly, only depends on the expected payment for any given conditional distribution.

Therefore, the only role of conditioning payments on ω is to separate types to ensure that the bidder reports his type truthfully, so it will be the IC constraints that determine the necessary range of payments. Let $\Theta' \subseteq \Theta$ be such that $|\Theta'| = \max\{|\Theta|, |\Omega|\}$ Then for all $\theta \in \Theta'$ and for all $\boldsymbol{\pi}'(\cdot|\theta) \in \mathcal{P}_{\epsilon}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$, there exists a hyperplane through the points $\{(\boldsymbol{\pi}'(\cdot|\theta), v(\theta))\}_{\theta \in \Theta'}$, by the assumption of γ'-separation. Moreover, there exists a hyperplane such that the gradient is bounded by $\frac{|v(|\Theta|) - v(1)|}{\gamma'} < \frac{v(|\Theta|)}{\gamma'}$, again by the definition of γ'-separation. Therefore, a lottery defined by a hyperplane with a gradient of $\frac{v(|\Theta|)}{\gamma'}$ is sufficient to separate any subset of types.

Moreover, if for some $\theta \in \Theta$ the expected payment for a lottery $\boldsymbol{x}(\theta,\cdot)$ is negative for every possible distribution $\boldsymbol{\pi} \in P(\Omega)$, then the expected revenue for the mechanism can be increased by raising the payments for type θ until there is some distribution over Ω such that the expected payment is non-negative. Therefore, there exists an optimal mechanism, $(\boldsymbol{p},\boldsymbol{x})$, such that for every lottery defining hyperplane, the hyperplane's gradient is bounded by $v(|\Theta|)/\gamma'$, and the value of the hyperplane is at least 0 somewhere in the set of distributions over Ω. Similarly, the value of the hyperplane must be less than or equal to $v(|\Theta|)$ for some distribution over Ω or the lottery is not individually rational for any bidder type. Therefore for all $\theta \in \Theta$ and $\omega \in \Omega$, the maximum payment is $-\frac{v(|\Theta|)}{\gamma'} \leq x(\theta,\omega) \leq \frac{v(|\Theta|)}{\gamma'} + v(|\Theta|) \leq \frac{2v(|\Theta|)}{\gamma'}$, since $\gamma' \leq 1$. □

Proof of Lemma 7. Construct a mechanism, $(\boldsymbol{p}',\boldsymbol{x}')$, as in the proof of Lemma 8, for the distribution $\boldsymbol{\pi}^*$ and δ. By Lemma 6 and Lemma 8, this mechanism guarantees that the revenue for any distribution $\boldsymbol{\pi} \in \mathcal{P}(\hat{\boldsymbol{\pi}})$ is within $\frac{k}{3|\Omega|}$ of the optimal revenue, and the mechanism is robust IR. However, the mechanism does not necessarily satisfy robust incentive compatibility. Therefore, we must adjust payments to ensure that the mechanism is robust IC.

Choose type $\theta \in \Theta$, let $\Theta'(\theta) \subset \Theta$ be the set of types that maximizes utility for some $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$, i.e.,

$$\Theta'(\theta) = \left\{ \theta' \in \Theta \;\middle|\; \exists\, \boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))\; s.t. \forall \hat{\theta} \in \Theta, \right.$$
$$\left. \sum_{\omega} \pi(\omega|\theta)\left(v(\theta)p'(\theta',\omega) - x'(\theta',\omega)\right) \geq \sum_{\omega} \pi(\omega|\theta)\left(v(\theta)p'(\hat{\theta},\omega) - x'(\hat{\theta},\omega)\right) \right\}.$$

Stated differently, $\Theta'(\theta)$ is the set of optimal misreports for bidder type θ for at least one $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$.

Construct a new set of payments $\boldsymbol{x}_1^*$ such that 1) for all $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}(\cdot|\theta)$ and $\theta' \in \Theta'(\theta)$, $\sum_\omega \pi(\omega|\theta)x_1^*(\theta,\omega) \leq \sum_\omega \pi(\omega|\theta)x'(\theta',\omega)$, and 2) minimizes

$$\beta = \max_{\{\pi_\omega \mid \sum_\omega \pi_\omega = 1\}} \min_{\theta' \in \Theta'} \sum_{\omega} \pi_\omega (x'(\theta',\omega) - x_1^*(\theta,\omega)).$$

Notice that β must be positive by the first condition. The first condition also ensures that it is indeed incentive compatible for the bidder to report θ, and the second condition ensures that it is the largest set of payments such that this is true. Note that for each $\theta' \in \Theta'(\theta)$ there is a hyperplane that is defined by $x'(\theta',\cdot)$, and this set of hyperplanes intersects inside the set $Conv(\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta)))$, the convex hull of $\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$. Given that the minimum of an intersecting set of hyperplanes defines a concave function, one can always construct a hyperplane that is less than or equal to the concave function for a given set, $\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$, and the difference between the concave function and the hyperplane is non-increasing outside of $Conv(\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta)))$. Therefore, the maximum occurs within

$Conv(\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta)))$, and the value of the hyperplane at the boundaries of $Conv(\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta)))$ are greater than or equal to $v(\theta) - \frac{k}{3|\Omega|}$, by the construction of $\boldsymbol{x'}$, and the fact that $\boldsymbol{x}_1^*$ minimizes β. A geometric interpretation is that $\boldsymbol{x}_1^*(\theta,\cdot)$ is the highest hyperplane that is strictly lower than any other hyperplane defining lottery in $\boldsymbol{x'}$ in the region of the consistent set $\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$. Similarly, this implies that the utility of bidder type θ with conditional distribution $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$ facing the lottery $\boldsymbol{x}_1^*(\theta,\cdot)$ is at most $\frac{k}{3|\Omega|}$.

However, there may be another $\theta^* \in \Theta/\Theta'(\theta)$ for some $\boldsymbol{\pi}(\cdot|\theta^*) \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta^*))$ that now chooses to report θ in order to have payments $x_1^*(\theta,\omega)$. Construct new payments $x_2^*(\theta^*,\omega) = x_2^*(\theta,\omega) = x_1^*(\theta,\omega) - \frac{k}{3|\Omega|}$ for $\omega \in \Omega$ and $x_2^* = x_1^*(\theta',\omega)$ for $\theta' \neq \theta$. This sets the lottery defined for type θ^* to be equal to the lottery for θ, and it ensures that for every conditional distribution within $\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta))$ to also choose this lottery. Also, given that the expected payments for $\boldsymbol{x}_1^*$ are greater than or equal to the expected payments of existing lotteries in $\boldsymbol{x'}$ for distributions outside of $Conv(\mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta)))$, the expected payment of bidder type θ^* facing $\boldsymbol{x'}$ is greater than or equal to $v(\theta^*) - \frac{k}{3|\Omega|}$ by construction of $\boldsymbol{x'}$, and the fact that $\mathcal{P}(\hat{\boldsymbol{\pi}})$ is γ-separated, the new expected payment for θ^* under $\boldsymbol{x}_2^*$ is at least $v(\theta^*) - \frac{2k}{3|\Omega|}$. Additionally, because under $\boldsymbol{x'}$ the utility for type θ^* was at most $\frac{k}{3|\Omega|}$ and payments for $\boldsymbol{x}_1^*$ have been lowered by $\frac{k}{3|\Omega|}$, for all $\boldsymbol{\pi}(\cdot|\theta^*) \in \mathcal{P}(\hat{\boldsymbol{\pi}}(\cdot|\theta^*))$, bidder type θ^* will optimally report θ^* under mechanism $\boldsymbol{x}_2^*$. Now, check to see if there exists another type θ^{**} that now prefers to misreport type θ. If so, repeat the above procedure. Note that this can happen for at most $|\Omega| - 1$ types, by the definition of γ'-separation and the necessity for all types that would choose to misreport to be near the hyperplane defined by $\boldsymbol{x}^*(\theta,\cdot)$. Every type that faces this set of payments, has a utility less than $k\left(\frac{|\Omega|}{3|\Omega|}\right) = \frac{k}{3}$.

Now choose another $\theta'' \in \Theta$ that is not yet IC and repeat. However, all types that would have optimally misreported as θ in the previous round, will never choose to misreport as θ'', since they must not be near the hyperplane that the payments defined for θ'', otherwise θ'' would have chosen to misreport as θ, and the above would have already ensured the mechanism $\boldsymbol{x}_2^*$ would already be robust IC for type θ''. Continue this process until there are no $\theta \in \Theta$ for which the mechanism is not robust IC. Denote the final set of payments as $\boldsymbol{x}$.

Therefore, there exist a set of payments $\boldsymbol{x}$ such that the utility for all types is less than $\frac{k}{3}$ and the item is always allocated. This implies the revenue generated by the mechanism at $\boldsymbol{\pi}^*$ is at within $\frac{k}{3}$ of the optimal revenue, R. Moreover, using this mechanism, the revenue achievable for any $\boldsymbol{\pi} \in \mathcal{P}(\hat{\boldsymbol{\pi}})$, including $\hat{\boldsymbol{\pi}}$, is bounded by shifting all probability mass from the largest payment to the smallest payment, a loss of revenue of less than $\delta\left(\frac{3v(|\Theta|)}{\gamma'}\right) \leq \frac{k}{2|\Omega|}$ by Lemma 6. Thus, the optimal robust mechanism at $\mathcal{P}(\hat{\boldsymbol{\pi}})$ has an objective value of at least $R - \frac{k}{3} - \frac{k}{2|\Omega|}$.

The last piece of the proof is to show that for any mechanism that does nearly optimally for $\hat{\boldsymbol{\pi}}$ will also do nearly optimally for all $\boldsymbol{\pi} \in \mathcal{P}(\hat{\boldsymbol{\pi}})$. However, the worst that could happen is, again, all probability mass could shift to the worst possible payment from the best possible payment, which the difference between the two is bounded by $\frac{3v(|\Theta|)}{\gamma'}$ by Lemma 6. Therefore, the maximum loss due to mis-specifying the objective is $\delta\left(\frac{3v(|\Theta|)}{\gamma'}\right) \leq \frac{k}{2|\Omega|}$. Therefore, for all $\boldsymbol{\pi} \in \mathcal{P}(\hat{\boldsymbol{\pi}})$, the optimal robust mechanism for $\mathcal{P}(\hat{\boldsymbol{\pi}})$ achieves a revenue of $R - \frac{k}{3} - \frac{k}{2|\Omega|} - \frac{k}{2|\Omega|} \geq R - k$ on $\boldsymbol{\pi}$, since $|\Omega| \geq 2$.

□

B.1. Sample Complexity for Learning Optimal Mechanisms with Unbounded Marginal Probabilities

In Section 5, we showed that a nearly optimal mechanism could be learned using a polynomial number of samples from the true distribution, as long as the true distribution was γ-separated, Theorem 3. However, this result depended on a lower bound for the marginal probability of any type $\theta \in \Theta$, $\underline{\pi}_\Theta$. Given that there is no guarantee that the marginal probability will be even non-zero for all $\theta \in \Theta$, this is not an ideal bound. Since we do not put a lower bound on the marginal probability of a bidder type θ, there will be, in general, no finite number of samples sufficient to ensure that the number of samples with bidder type θ is sufficient to estimate the conditional distribution. Therefore, we bound the probability mass of the bidder types for which we will not see a sufficient number of samples with Lemma 9.

In this appendix, we relax this condition by using the extension of ϵ-robust mechanisms to type dependent robustness described in Appendix A. First, assuming that the set $\{\mathcal{P}_{\epsilon_{\theta,IR}}(\hat{\boldsymbol{\pi}}(\cdot|\theta)) \cup \mathcal{P}_{\epsilon_{\theta,IC}}(\hat{\boldsymbol{\pi}}(\cdot|\theta))\}_{\theta\in\Theta}$ is γ'-separated, Lemmas 6 and 7 apply directly since the proofs do not rely on a specific value of ϵ, nor do they rely on ϵ being uniform. Therefore, much of the proof structure will remain the same.

However, we will have to modify the mechanism design procedure. Specifically, we will use two values for the robustness parameter, ϵ and 1. There will be a threshold for the number of samples necessary to sufficiently accurately estimate the ϵ-consistent set $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$. Specifically, the set $\Theta_{\underline{X}} \subset \Theta$ will be the set of $\theta \in \Theta$ such that there are an insufficient number of samples. For $\theta \in \Theta/\Theta_{\underline{X}}$, $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta)) = \{\boldsymbol{\pi} \in P(\Omega) | \sum_{\omega\in\Omega} |\hat{\pi}_X(\omega|\theta) - \pi(\omega)| \leq \delta_\theta\}$ where the relationship between ϵ, $|X_\theta|$, and δ_θ is determined as in Lemma 2. For $\theta \in \Theta_{\underline{X}}$, we will use a robustness parameter of 1 for both IC and IR and set $\mathcal{P}_1(\hat{\boldsymbol{\pi}}(\cdot|\theta)) = \{\hat{\boldsymbol{\pi}}(\cdot|\theta)\}$. The set $\{\hat{\boldsymbol{\pi}}(\cdot|\theta)\}_{\theta\in\Theta_{\underline{X}}}$ will be chosen to maximize the γ'-separation for the set $\{\{\hat{\boldsymbol{\pi}}(\cdot|\theta)\}_{\theta\in\Theta_{\underline{X}}}, \{\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))\}_{\theta\in\Theta/\Theta_{\underline{X}}}\}$. Note that this set can be computed efficiently using a linear program. Additionally, if $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\cdot|\theta)$ for all $\theta \in \Theta/\Theta_{\underline{X}}$, $\boldsymbol{\pi}$ is γ-separated with $\gamma \geq \gamma_{\min}$, and $\delta_\theta \leq \frac{\gamma_{\min}}{4}$, it is easy to verify that $\{\{\hat{\boldsymbol{\pi}}(\cdot|\theta)\}_{\theta\in\Theta_{\underline{X}}}, \{\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}(\cdot|\theta))\}_{\theta\in\Theta/\Theta_{\underline{X}}}\}$ will be $\frac{\gamma_{\min}}{2}$-separated.

This will imply that the mechanism will fail to be IC and IR for $\theta \in \Theta_{\underline{X}}$, and we will have to assume a worst case outcome for them, specifically the seller will lose $\frac{6v(|\Theta|)}{\gamma_{\min}}$. However, if we can ensure that the probability mass of the bidders such that this happens is sufficiently small, then the mechanism can still achieve nearly optimal expected revenue. For types for which we have a sufficient number of samples, the proof will be nearly identical to Theorem 3.

LEMMA 9. *Let $\boldsymbol{\pi}$ be the true distribution over $\Theta \times \Omega$, and let X be a set of independent samples from $\boldsymbol{\pi}$. Let $\Theta' = \{\theta | \theta \in \Theta, |X_\theta| \geq M\}$. Then, if $|X| \geq 8M|\Theta| \ln\left(\frac{|\Theta|}{\epsilon_3}\right)\left(\frac{1}{\delta_3}\right)$, $\sum_{\theta \in \Theta'} \pi(\theta) > 1 - \delta_3$ with probability $1 - \epsilon_3$.*

Proof. For any sample, we can model the probability of that sample being of bidder type $\theta \in \Theta$ as a binomial distribution with success probability of $\pi(\theta)$. In order to ensure that $\sum_{\theta \in \Theta'} \pi(\theta) > 1 - \delta_3$ with probability $1 - \epsilon_3$, we need to cover any bidder with type $\frac{\delta_3}{|\Theta|}$ marginal probability with probability at least $1 - \frac{\epsilon_3}{|\Theta|}$. If we do this, then there is at most a set of bidders with collective marginal probability of less than δ_3 that we do not cover, with probability of at least $1 - \epsilon_3$. The number of samples we need to ensure that we cover a bidder with marginal probability of $\frac{\delta_3}{|\Theta|}$ with probability $1 - \frac{\epsilon_3}{|\Theta|}$ is given by Lemma 4. Set $\delta = \frac{1}{2}$ in Lemma 4, then $\frac{\epsilon_3}{|\Theta|} = e^{-\frac{|X|\delta_3}{8|\Theta|}}$:

$$|X| \geq 8|\Theta| \ln\left(\frac{|\Theta|}{\epsilon_3}\right)\left(\frac{1}{\delta_3}\right) \tag{8}$$

and $M \leq \frac{|X|\delta_3}{2|\Theta|}$ which implies:

$$|X| \geq 8M|\Theta| \ln\left(\frac{|\Theta|}{\epsilon_3}\right)\left(\frac{1}{\delta_3}\right) \geq \max\left\{8|\Theta| \ln\left(\frac{|\Theta|}{\epsilon_3}\right)\left(\frac{1}{\delta_3}\right), \frac{2M|\Theta|}{\delta_3}\right\}. \tag{9}$$

□

Now all of the pieces are in place to prove the desired corollary.

COROLLARY 4. *Let the true distribution, $\boldsymbol{\pi}$, be a γ-separated distribution that satisfies the CM condition with an optimal revenue of R. Assume that $\gamma \geq \gamma_{\min}$. Let X be independent samples from $\boldsymbol{\pi}$. Let $\Theta_{\underline{X}} = \{\theta | \theta \in \Theta, |X_\theta| < 25|\Omega| \ln\left(\frac{24(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{96v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2\}$ If $|X| \geq 200|\Theta||\Omega|^3 \ln^2\left(\frac{24(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{96v(|\Theta|)}{k\gamma_{\min}}\right)^3$, then an $\{\{\frac{k\gamma_{\min}}{24v(|\Theta|)}\}_{\theta \in \Theta/\Theta_{\underline{X}}, IC, IR}, \{1\}_{\theta \in \Theta_{\underline{X}}, IC, IR}\}$-robust mechanism with payments bounded by (2) has an expected revenue of at least $R - k$.*

Proof. The proof of Corollary 4 proceeds nearly identically to Theorem 3. However, since there is an additional source of loss (the bidder types without sufficient samples), the bound on the other sources of loss must be a little tighter, $\frac{k}{4}$ versus $\frac{k}{3}$.

Let $\theta \in \Theta/\Theta_{\underline{X}}$, then if $\boldsymbol{\pi}(\cdot|\theta) \notin \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$, then the loss of revenue is bounded by $\frac{6v(|\Theta|)}{\gamma_{\min}}$. Therefore, if $\epsilon \leq \frac{k\gamma_{\min}}{24v(|\Theta|)}$, the revenue loss due to $\boldsymbol{\pi}(\cdot|\theta) \notin \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$ is at most $\frac{k}{4}$.

Similarly, for all $\theta \in \Theta/\Theta_{\underline{X}}$, if $\boldsymbol{\pi} \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ and $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ is $\frac{\gamma_{\min}}{2}$-separated, we can bound the loss due to robustness using Lemma 7. If $\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X) \subseteq \left\{\boldsymbol{\pi} \,\middle|\, ||\boldsymbol{\pi} - \hat{\boldsymbol{\pi}}_X|| \leq \frac{k\gamma_{\min}}{96v(|\Theta|)|\Omega|}\right\}$, and similarly for

$\mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$, then the expected revenue of the mechanism is within $\frac{k}{4}$ of the optimal revenue for type $\theta \in \Theta$.

To ensure that $\boldsymbol{\pi} \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X)$ and for all $\theta \in \Theta/\Theta_{\underline{X}}$, $\boldsymbol{\pi}(\cdot|\theta) \in \mathcal{P}_\epsilon(\hat{\boldsymbol{\pi}}_X(\cdot|\theta))$ with probability at least $\epsilon = \frac{k\gamma_{\min}}{24v(|\Theta|)}$, set $\epsilon_\theta = \epsilon_1 = \frac{k\gamma_{\min}}{24(|\Theta|+1)v(|\Theta|)}$, using the notation of Lemmas 2 and 3. By Lemma 2, if $\delta_\theta \le \frac{k\gamma_{\min}}{96v(|\Theta|)|\Omega|} \le \frac{\gamma_{\min}}{4}$, it must be that $|X_\theta| \ge 25|\Omega| \ln\left(\frac{24(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{96v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2$. By Lemma 3, if $\delta_1 \le \frac{k\gamma_{\min}}{96v(|\Theta|)|\Omega|}$, it must be that $|X| \ge 25|\Theta||\Omega| \ln\left(\frac{24(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{96v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2$.

For all $\theta \in \Theta_{\underline{X}}$, we assume that we lose the maximum possible revenue, $\frac{6v(|\Theta|)}{\gamma_{\min}}$. Therefore, if we ensure that the probability mass of bidder types in $\Theta_{\underline{X}}$ is less than $\frac{k\gamma_{\min}}{24v(|\Theta|)}$ with high probability, we can bound the loss. Then if the probability mass is less than $\frac{k\gamma_{\min}}{24v(|\Theta|)}$, we lose at most $\frac{k}{4}$ in revenue. Further, we must bound the probability that the probability mass is larger than $\frac{k\gamma_{\min}}{24v(|\Theta|)}$ by, again, $\frac{k\gamma_{\min}}{24v(|\Theta|)}$. If the probability mass is more than $\frac{k\gamma_{\min}}{24v(|\Theta|)}$, we again assume a worst case in that we lose the maximum possible amount, $\frac{6v(|\Theta|)}{\gamma_{\min}}$. This implies that the loss in revenue in expectation is less than $\frac{k}{4}$ for the case where there are too many bidder types for which the estimate of the conditional distribution is insufficient. To do this, by Lemma 9 with $\epsilon_3 = \frac{k\gamma_{\min}}{24v(|\Theta|)}$ and $\delta_3 \le \frac{k\gamma_{\min}}{24v(|\Theta|)}$, we must have

$$|X| \ge 200|\Theta||\Omega| \ln\left(\frac{24(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{96v(|\Theta|)|\Omega|}{k\gamma_{\min}}\right)^2 \ln\left(\frac{24v(|\Theta|)|\Theta|}{k\gamma_{\min}}\right)\left(\frac{24v(|\Theta|)}{k\gamma_{\min}}\right)$$

If the number of samples is sufficient to ensure that the above conditions hold, then the expected revenue from the mechanism is greater than or equal to $R-k$.

Therefore, if $|X| \ge 200|\Theta||\Omega|^3 \ln^2\left(\frac{24(|\Theta|+1)v(|\Theta|)}{k\gamma_{\min}}\right)\left(\frac{96v(|\Theta|)}{k\gamma_{\min}}\right)^3$, all conditions are satisfied. □

Appendix C: Additional Proofs and Discussion for Section 6

In this appendix, we discuss and rigorously prove Corollary 3 in Section 6. Note that in Definition 8, we do not explicitly assume that the elements of the sequence satisfy the CM condition, nor do we assume that the distribution to which the sequence is converging is an IPV distribution. However, it is straightforward to construct examples of converging sequences such that every element of the sequence satisfies CM but the limit is IPV. Figure 4a demonstrates one such set. We will make use of the following standard definition.

Definition 12 (Affine Independence). A set of vectors $\{\boldsymbol{v}_i\}_{i=1}^m$ over $\mathbb{R}^n$ are affinely independent if for $\{\alpha_i\}_{i=1}^m$, $\sum_i \alpha_i \boldsymbol{v}_i = \mathbf{0}$ and $\sum_i \alpha_i = 0$ implies $\alpha_i = 0$ for all $i \in \{1, ..., m\}$.

The set of distributions over Ω are the points on a $|\Omega|$-simplex where the vertices of the simplex are denoted by the set of distributions such that $\pi(\omega) = 1$ for all $\omega \in \Omega$ (see Figure 4b). Further, any set of distributions over Ω of size $|\Omega|$ that are affinely independent must span the $|\Omega|$-simplex with affine combinations. If the set $\{\boldsymbol{\pi}_i\}_{i=1}^{|\Omega|}$ is affinely independent, then for any distribution $\boldsymbol{\pi'}$ over Ω, there must exist $\{\alpha_i\}_{i=1}^{|\Omega|}$ where $\sum_i \alpha_i = 1$ and $\boldsymbol{\pi'} = \sum_i \alpha_i \boldsymbol{\pi}_i$.

We can assume, without loss of generality, that for any sequence of distributions we consider, $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$, there must exist a subset of $\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ of size $|\Omega|$ that is affinely independent. If not, the affine combination of vectors $\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ spans a lower dimensional simplex, and we can reduce the dimensionality of Ω until an affinely independent subset exists. Note that this relies on the assumption that the bidder is risk neutral. Specifically, a risk neutral bidder is indifferent between a payment for an outcome of the external signal, $p(\theta', \boldsymbol{\pi}_{i'}, \omega)$, and a lottery over multiple values of the external signal with the same expected payoff. Therefore, if there is not a subset of $\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ of size $|\Omega|$ that is affinely independent we can always replace the true signal with a lower dimensional set of lotteries over the external signal without affecting the expected utility of the bidder.

We will further simplify this setting by assuming that we know the marginal distribution, $\boldsymbol{\pi}_\Theta$, perfectly, and we must only estimate the conditional distributions $\{\boldsymbol{\pi}(\cdot|\theta)\}_{\theta\in\Theta}$. Since we will show that this more restrictive set is sufficient for our impossibility results, the results will naturally extend to the more permissive set.

In addition to Definition 8, we will require Assumption 2, stated in Section 6 and restated here for ease of reference.

Assumption 2. *For the sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ converging to $\boldsymbol{\pi}^*$ and for any $\theta' \in \Theta$, there exists a subset of distributions of size $|\Omega|$ from the set $\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ that is affinely independent and the distribution $\boldsymbol{\pi}^*(\cdot|\theta')$ is a strictly convex combination of the elements of the subset. I.e., there exists $\{\alpha_k\}_{k=1}^{|\Omega|}$, $\alpha_k \in (0,1)$ and $\{\boldsymbol{\pi}_k(\cdot|\theta_k)\}_{k=1}^{|\Omega|}$, where every $\boldsymbol{\pi}_k(\cdot|\theta_k) \in \{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$, such that $\boldsymbol{\pi}^*(\cdot|\theta') = \sum_{k=1}^{|\Omega|} \alpha_k \boldsymbol{\pi}_k(\cdot|\theta_k)$.*

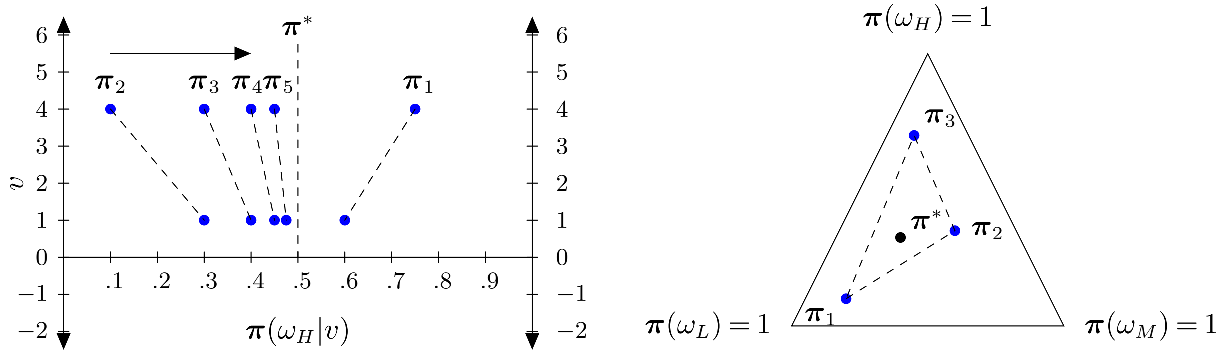
Assumption 2 states that the sequence of distributions is converging to a distribution that is in the interior of the sequence. This is not without loss of generality, but it is only violated for a measure zero set of distributions, the distributions at the boundary of the space of the strictly convex combinations of the distributions within the sequence of distributions. Specifically, Assumption 2 is a statement about the *conditional distributions*, and particularly that all conditional distributions of the convergence point, $\boldsymbol{\pi}^*(\cdot|\theta)$ for all θ, is in some sense in the interior of some other estimate (see Figures 4b for a graphical depiction of this statement). Moreover, the distributions that "enclose" the convergence point do not have to have the same θ, i.e., any conditional distributions for any θ in the set $\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ can be the distributions that "enclose" the convergence point. Further, if the full set of all potential distributions is a continuous closed set, then there will be an infinite number of sequences that satisfy this assumption.

With these definitions, we are able to introduce our main impossibility results, Theorem 4 and Corollary 5.

Figure 4 Figure 4a demonstrates a converging sequence as in Definition 8. Each point represents a conditional distribution, and conditional distributions linked by a dashed line both belong to the same full distribution. Specifically, the conditional distributions are all converging to $\boldsymbol{\pi}(\omega_H|v) = 1/2$, i.e., a distribution where the bidder's value and the external signal are uncorrelated. However, all of the distributions in the sequence satisfy the Cremer-McLean condition. Figure 4b demonstrates distributions that satisfy Assumption 2 in a set of distributions over three possible external signals $\{\omega_L, \omega_M, \omega_H\}$ as points in a 2-simplex. Specifically, $\boldsymbol{\pi}^*$ is a strictly convex combination of $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, and $\boldsymbol{\pi}_3$. The sequence of distributions in Figure 4a also satisfies Assumption 2 due to $\boldsymbol{\pi}_1$, but if $\boldsymbol{\pi}_1$ was excluded from the sequence, it would not.

(a) Sequence of converging distributions with a binary signal.

(b) Sequence elements that satisfy Assumption 2.



Theorem 4. *Let $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ be a sequence of distributions converging to $\boldsymbol{\pi}^*$ that satisfies Assumption 2. Denote the revenue of the optimal mechanism for the distribution $\boldsymbol{\pi}^*$ by R. For any $k > 0$ and for any mechanism, there exists a $T \in \mathbb{N}$ such that for all $\boldsymbol{\pi}_{i'} \in \{\boldsymbol{\pi}_i\}_{i=T}^{\infty}$, the expected revenue is less than $R + k$.*

Theorem 4, whose proof we shall defer to the end of this section, states that *no mechanism* can guarantee revenue better than the optimal revenue achievable at the convergence point for all distributions in the sequence. Namely, if the sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ satisfy the CM condition, but the convergence point is IPV, then *no mechanism* can always do better than the optimal mechanism for the IPV point (in our setting, a reserve price mechanism (Myerson 1981)).

It may not seem surprising that we cannot construct mechanisms that do well on large sets of distributions. However, the following corollary indicates that *we cannot learn* a mechanism that always does well either.

Corollary 5. *Let $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ be a sequence of distributions converging to $\boldsymbol{\pi}^*$ that satisfies Assumption 2. Denote the revenue of the optimal mechanism for the distribution $\boldsymbol{\pi}^*$ by R. For any $k > 0$*

and for any mechanism that uses a finite number of independent samples from the underlying distribution, there exists a $T \in \mathbb{N}$ *such that for all* $\boldsymbol{\pi}_{i'} \in \{\boldsymbol{\pi}_i\}_{i=T}^{\infty}$*, the expected revenue is less than* $R + k$.

We must be careful in interpreting Theorem 4 and Corollary 5; they are both statements about distributions close to the convergence point. They do not provide a bound for distributions that are far from the convergence point. Therefore, even if the convergence point is an IPV distribution, it is still potentially possible to generate near optimal revenue for some distributions in the sequence. However, even with sampling, mechanisms cannot generate significantly higher revenue than the optimal IPV mechanism for distributions sufficiently close to IPV, though sampling may still substantially increase the expected revenue for some subset of the sequence of distributions.

These results indicate that the setting where the bidder may have a distribution from an infinite set is fundamentally different from the setting where the bidder's distribution is one of a finite set (as in Fu et al. (2014)). Note that the set of all mechanisms includes mechanisms that first applies some procedure to reduce the infinite set to a finite set.

In the remainder of this section, we prove Theorem 4 and Corollary 5. The strategy that we will use to prove the above results relies on bounding the maximum possible payments for any mechanism. Specifically, the revelation principle (Gibbons 1992) ensures that the revenue achievable by any mechanism can be achieved by a mechanism that not only truthfully elicits the bidder's valuation, but also truthfully elicits the distribution of the bidder. We will show that Assumption 2 implies that any mechanism with payments too large (either from or to the bidder), will create an incentive for some bidder type to lie either about his valuation or his distribution, violating the revelation principle. Once we show that payments are bounded, we can use a standard continuity result in linear programming to show that the expected revenue of the mechanism must converge to something less than or equal to the optimal revenue achievable at the convergence point.

To bound payments, we will require that for distributions "sufficiently close" to the convergence point, we can always find another distribution that is a finite step in any direction. This is what Assumption 2 provides (see Figure 4b for intuition), as the following lemma formally demonstrates.

Lemma 10. *Let* $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ *be a sequence of distributions converging to* $\boldsymbol{\pi}^*$ *that satisfies Assumption 2. There exists an* $\epsilon_{\min} > 0$*, such that for all distributions* $\boldsymbol{\pi}$ *over* Ω *where* $||\boldsymbol{\pi} - \boldsymbol{\pi}^*(\cdot|\theta)|| < \epsilon_{\min}$ *for some* $\theta \in \Theta$*, and all unit vectors* $\boldsymbol{z} \in \mathbb{R}^{|\Omega|}$ *where* $\sum_{\omega} z(\omega) = 0$*, there exists a* $\boldsymbol{\pi}_j(\cdot|\theta_j) \in \{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i,\theta}$ *such that* $(\boldsymbol{\pi} - \boldsymbol{\pi}_j(\cdot|\theta_j)) \bullet \boldsymbol{z} \geq \epsilon_{\min}$.

Proof. First, note that by Assumption 2, for all $\theta \in \Theta$, there exists $\{\alpha_k\}_{k=1}^{|\Omega|}$ and an affinely independent set of vectors $\{\boldsymbol{\pi}_k(\cdot|\theta_k)\}_{k=1}^{|\Omega|}$, where $\alpha_k \in (0,1)$ and $\boldsymbol{\pi}^*(\cdot|\theta) = \sum_{k=1}^{|\Omega|} \alpha_k \boldsymbol{\pi}_k(\cdot|\theta_k)$. The set of affinely independent points $\{\boldsymbol{\pi}_k(\cdot|\theta_k)\}_k$ define a simplex in $\mathbb{R}^{|\Omega|}$, and the lth face of the simplex, where $l \in \{1, ..., |\Omega|\}$, is the set of points denoted by $\sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k)$ such that $\sum_{k \neq l} \alpha'_k = 1$ and $\alpha'_k \in [0, 1]$. The distance from the distribution $\boldsymbol{\pi}^*(\cdot|\theta)$ to any point on the lth face is:

$$\begin{aligned}\min_{\alpha'_k} ||\boldsymbol{\pi}^*(\cdot|\theta) - \sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k)|| &= \min_{\alpha'_k} ||\sum_{k=1}^{|\Omega|} \alpha_k \boldsymbol{\pi}_k(\cdot|\theta_k) - \sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k)|| \\ &= \min_{\alpha'_k} ||\alpha_l \boldsymbol{\pi}_l(\cdot|\theta_l) - \sum_{k \neq l} (\alpha_k - \alpha'_k) \boldsymbol{\pi}_k(\cdot|\theta_k)|| > 0.\end{aligned}$$

The last inequality is due to $\alpha_l \neq 0$ (Assumption 2) and affine independence. Let $\epsilon' > 0$ be the minimum such distance for all $\theta \in \Theta$ and all faces of the simplex.

Define $\epsilon_{\min} = \frac{\epsilon'}{2}$. Let $\boldsymbol{\pi}$ be a distribution over Ω where $||\boldsymbol{\pi} - \boldsymbol{\pi}^*(\cdot|\theta)|| < \epsilon_{\min}$, for some $\theta \in \Theta$. Let $\{\boldsymbol{\pi}_k(\cdot|\theta_k)\}_k$ define the simplex that contains $\boldsymbol{\pi}^*(\cdot|\theta)$. Therefore, the distance from $\boldsymbol{\pi}$ to any face of the simplex is at least $\epsilon_{\min}$ by an application of the triangle inequality. Let $\boldsymbol{z} \in \mathbb{R}^{|\Omega|}$ be a unit vector such that $\sum_\omega z(\omega) = 0$. Since a simplex is a closed and bounded set, there exists some face of the simplex which we will denote as the lth face, such that for some $\{\alpha'_k\}_{k \neq j}$ and some $\epsilon \geq \epsilon_{\min} > 0$:

$$\boldsymbol{\pi} - \epsilon \boldsymbol{z} = \sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k).$$

Let $\boldsymbol{\pi}_j(\cdot|\theta_j)$ be a vertex of that face such that

$$\left(\boldsymbol{\pi}_j(\cdot|\theta_j) - \sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k)\right) \bullet \boldsymbol{z} \leq 0.$$

This must exist by virtue of the face being a segment of a hyper-plane. Then:

$$\begin{aligned}\epsilon_{\min} \leq \epsilon = \epsilon \boldsymbol{z} \bullet \boldsymbol{z} &= \left(\boldsymbol{\pi}(\cdot) - \sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k)\right) \bullet \boldsymbol{z} \\ &= \left(\boldsymbol{\pi}_j(\cdot|\theta_j) - \sum_{k \neq l} \alpha'_k \boldsymbol{\pi}_k(\cdot|\theta_k)\right) \bullet \boldsymbol{z} + (\boldsymbol{\pi} - \boldsymbol{\pi}_j(\cdot|\theta_j)) \bullet \boldsymbol{z} \\ &\leq (\boldsymbol{\pi} - \boldsymbol{\pi}_j(\cdot|\theta_j)) \bullet \boldsymbol{z} \quad \square\end{aligned}$$

The payments in a Bayesian mechanism are a lottery over the external signal. A lottery over the external signal can be viewed as a linear function (or a hyper-plane) whose domain is the Ω-simplex of distributions and whose value is the expected payment for the lottery. Lemma 10 ensures that for points close enough to the convergence point, there exists a distribution in the sequence that is in the "opposite direction" of the gradient of the hyperplane that defines the lottery. For any possible lottery with a gradient of magnitude K, there exists a distribution for which the expected payment for the lottery is at least $\epsilon_{\min} K$ less than for any distributions "sufficiently close" to the convergence point. Therefore, if payments are too large (either from or to the bidder) for some distribution $\boldsymbol{\pi}'$ and θ', there is another distribution and type $\boldsymbol{\pi}''$ and θ'' that will find reporting $\boldsymbol{\pi}'$ and θ' irresistible. The following lemma formalizes this argument.

LEMMA 11. *Let $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$ be a sequence of distributions converging to $\boldsymbol{\pi}^*$ that satisfies Assumption 2. For any mechanism $(\boldsymbol{p},\boldsymbol{x})$ that is incentive compatible and individually rational and guarantees non-negative revenue in expectation for all distributions in $\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}$, there exists some $M>0$ such that for all $\boldsymbol{\pi}_{i'}\in\{\boldsymbol{\pi}_i\}_{i=1}^{\infty}, \theta\in\Theta$, and $\omega\in\Omega$:*

$$|x(\theta,\boldsymbol{\pi}_{i'},\omega)|\leq M$$

Proof of Lemma 11. Let $\epsilon_{\min}>0$ be defined as in Lemma 3. By the definition of converging sequences of distributions (Definition 8), there exists a $T\in\mathbb{N}$ such that for all $\theta\in\Theta$ and $\boldsymbol{\pi}_{i^*}(\cdot|\theta)\in\{\boldsymbol{\pi}_i(\cdot|\theta)\}_{i=T}^{\infty}$, $||\boldsymbol{\pi}_{i^*}(\cdot|\theta)-\boldsymbol{\pi}^*(\cdot|\theta)||\leq\epsilon_{\min}$. Since there are a finite number of distributions such that $i^*<T$, choose $M_{i^*<T}=\max_{i^*<T,\theta,\omega}|x(\theta,\boldsymbol{\pi}_{i^*},\omega)|$.

Therefore if payments are not bounded, for any $M'>0$, there must exist some $\boldsymbol{\pi}_{i'}\in\{\boldsymbol{\pi}_i\}_{i=T}^{\infty}, \theta'\in\Theta$, and $\omega'\in\Omega$ such that $x(\theta',\boldsymbol{\pi}_{i'},\omega')>M'$ or $x(\theta',\boldsymbol{\pi}_{i'},\omega')<-M'$.

First, we will consider the case where $x(\boldsymbol{\pi}_{i'},\theta',\omega')<-M'$. Denote the type $\theta\in\Theta$ with the minimum marginal probability by π_{min}, i.e., $\pi_{min}=\min_\theta\{\pi_\theta\}$. Without loss of generality, we can assume that $\pi_{min}>0$. Note that the expected revenue generated for any type θ must be bounded from below by $-v(|\Theta|)/\pi_{\min}$ if the mechanism guarantees non-zero expected revenue. This is because the maximum amount of expected revenue for any type can be at most $v(|\Theta|)$ or individual rationality will not be satisfied, and if expected revenue for any type is less than $-v(|\Theta|)/\pi_{\min}$, it is not possible to make up the revenue from other types. Further, this implies that in order for the mechanism to generate non-negative revenue, the bidder's expected utility for any type must be less than $v(|\Theta|)+v(|\Theta|)/\pi_{\min}$. Therefore, set

$$M'=\frac{v(|\Theta|)}{\pi_{\min}}+\frac{(1-\epsilon_{\min})(2v(|\Theta|)+\frac{v(|\Theta|)}{\pi_{\min}}+1)}{\epsilon_{\min}}$$

Then, the magnitude of the gradient of the hyper-plane defined by the affine combination of $x(\theta',\boldsymbol{\pi}_{i'},\omega)$ for all $\omega\in\Omega$ must be at least:

$$||\nabla\boldsymbol{x}(\theta',\boldsymbol{\pi}_{i'},\cdot)||\geq\frac{(-\frac{v(|\Theta|)}{\pi_{\min}}+M')}{(1-\epsilon_{\min})}=\frac{2v(|\Theta|)+\frac{v(|\Theta|)}{\pi_{\min}}+1}{\epsilon_{\min}}$$

Let $\boldsymbol{z}\in\mathbb{R}^{|\Omega|}$ with $\sum_\omega z(\omega)=0$ be the direction of the gradient of the hyper-plane defined by the lottery in the plane of the $|\Omega|$-simplex. Then by Lemma 10, there exists a $\boldsymbol{\pi}_j(\cdot|\theta_j)$ such that $(\boldsymbol{\pi}_{i'}(\cdot|\theta')-\boldsymbol{\pi}_j(\cdot|\theta_j))\bullet\boldsymbol{z}>\epsilon_{\min}$. Then:

$$\begin{aligned}\sum_\omega\pi_j(\omega|\theta_j)\left(v(\theta_j)p(\theta_j,\boldsymbol{\pi}_j,\omega)-x(\theta_j,\boldsymbol{\pi}_j,\omega)\right)&\geq\sum_\omega\pi_j(\omega|\theta_j)\left(v(\theta_j)p(\theta',\boldsymbol{\pi}_{i'},\omega)-x(\theta',\boldsymbol{\pi}_{i'},\omega)\right)\quad\text{(by IC)}\\&\geq\sum_\omega\pi_j(\omega|\theta_j)\left(v(\theta_j)p(\theta',\boldsymbol{\pi}_{i'},\omega)-x(\theta',\boldsymbol{\pi}_{i'},\omega)\right)\end{aligned}$$

$$
\begin{aligned}
&- \sum_{\omega} \pi_{i'}(\omega|\theta')\left(v(\theta')p(\theta', \boldsymbol{\pi}_{i'}, \omega) - x(\theta', \boldsymbol{\pi}_{i'}, \omega)\right) \quad \text{(by IR)} \\
&\geq \sum_{\omega} (\pi_{i'}(\omega|\theta') - \pi_j(\omega|\theta_j)) x(\theta', \boldsymbol{\pi}_{i'}, \omega) - v(|\Theta|) \\
&= (\boldsymbol{\pi}_{i'}(\cdot|\theta') - \boldsymbol{\pi}_j(\cdot|\theta_j)) \cdot \boldsymbol{z} ||\nabla \boldsymbol{x}(\theta', \boldsymbol{\pi}_{i'}, \cdot)|| - v(|\Theta|) \\
&\geq \epsilon_{\min} ||\nabla \boldsymbol{x}(\theta', \boldsymbol{\pi}_{i'}, \cdot)|| - v(|\Theta|) \\
&\geq v(|\Theta|) + v(|\Theta|)/\pi_{\min} + \epsilon_{\min}
\end{aligned}
$$

Therefore, the seller cannot earn non-negative expected revenue for distribution $\boldsymbol{\pi}_j$, a contradiction.

It is straightforward to show that the combination of individual rationality and payments being bounded from below by $-\max\{M_{i^*<T}, M'\}$ implies that all payments must be bounded from above. We omit the details. Denote this upper bound by M''.

Therefore, let $M = \max\{M_{i^*<T}, M', M''\}$, and all payments are bounded by M. □

With payments bounded, the final necessary result is the following lemma stating that for any linear program where the variables for the set of optimal solutions is bounded, the corresponding sequence of linear programs is upper semi-continuous.

LEMMA 12 **(Martin (1975))**. *Let $\boldsymbol{a}(\boldsymbol{t}), \boldsymbol{b}(\boldsymbol{t}), \boldsymbol{c}(\boldsymbol{t})$, and $\boldsymbol{d}(\boldsymbol{t})$ be vectors parameterized by the parameter vector $\boldsymbol{t} \in \mathcal{Q}$. Assume that $\boldsymbol{a}(\boldsymbol{t}), \boldsymbol{b}(\boldsymbol{t}), \boldsymbol{c}(\boldsymbol{t})$, and $\boldsymbol{d}(\boldsymbol{t})$ converge continuously to $\boldsymbol{a}(\boldsymbol{0}), \boldsymbol{b}(\boldsymbol{0}), \boldsymbol{c}(\boldsymbol{0})$, and $\boldsymbol{d}(\boldsymbol{0})$ as $\boldsymbol{t} \to \boldsymbol{0}$. Similarly, $\boldsymbol{A}(\boldsymbol{t})$, $\boldsymbol{B}(\boldsymbol{t})$, $\boldsymbol{C}(\boldsymbol{t})$, and $\boldsymbol{D}(\boldsymbol{t})$ are matrices that converge continuously to $\boldsymbol{A}(\boldsymbol{0})$, $\boldsymbol{B}(\boldsymbol{0})$, $\boldsymbol{C}(\boldsymbol{0})$, and $\boldsymbol{D}(\boldsymbol{0})$.*

Define the parameterized linear program $LP(\boldsymbol{t})$ as:

$$
\begin{aligned}
&\max_{\boldsymbol{x},\boldsymbol{q}} \quad \boldsymbol{c}'(\boldsymbol{t})\boldsymbol{x} + \boldsymbol{d}'(\boldsymbol{t})\boldsymbol{q} \\
&\textit{subject to} \\
&\boldsymbol{A}(\boldsymbol{t})\boldsymbol{x} + \boldsymbol{B}(\boldsymbol{t})\boldsymbol{q} = \boldsymbol{a}(\boldsymbol{t}) \\
&\boldsymbol{C}(\boldsymbol{t})\boldsymbol{x} + \boldsymbol{D}(\boldsymbol{t})\boldsymbol{q} \leq \boldsymbol{b}(\boldsymbol{t}) \\
&\boldsymbol{q} \geq \boldsymbol{0}
\end{aligned}
$$

If the set of optimal solutions of $LP(\boldsymbol{0})$, $\{(\boldsymbol{x}, \boldsymbol{q}) : (\boldsymbol{x}, \boldsymbol{q}) \in \arg\max(LP(\boldsymbol{0}))\}$, is bounded, then the objective value of $LP(\boldsymbol{t})$ is upper semi-continuous at $\boldsymbol{t} = \boldsymbol{0}$.

Proof of Theorem 4. Note that the maximum revenue achievable for any given $\boldsymbol{\pi}_{i'}$ can be bounded from above by the following linear program:

$$
\begin{aligned}
&\max_{\boldsymbol{p},\boldsymbol{x}} \sum_{\theta} \sum_{\omega} \pi_{i'}(\theta, \omega) x(\theta, \boldsymbol{\pi}_{i'}, \omega) \\
&\textit{subject to}
\end{aligned}
$$

$$\sum_{\omega} \pi_{i'}(\omega|\theta)\left(v(\theta)p(\theta,\boldsymbol{\pi}_{i'},\omega) - x(\theta,\boldsymbol{\pi}_{i'},\omega)\right) \geq 0 \qquad \forall \quad \theta \in \Theta$$

$$\sum_{\omega} \pi_{i'}(\omega|\theta)\left(v(\theta)p(\theta,\boldsymbol{\pi}_{i'},\omega) - x(\theta,\boldsymbol{\pi}_{i'},\omega)\right)$$

$$\geq \sum_{\omega} \pi_{i'}(\omega|\theta)\left(v(\theta)p(\theta',\boldsymbol{\pi}_{i'},\omega) - x(\theta',\boldsymbol{\pi}_{i'},\omega)\right) \quad \forall \quad \theta,\theta' \in \Theta$$

$$0 \leq p(\theta,\boldsymbol{\pi}_{i'},\omega) \leq 1 \qquad \forall \quad \theta \in \Theta, \omega \in \Omega$$

$$-M \leq x(\theta,\boldsymbol{\pi}_{i'},\omega) \leq M \qquad \forall \quad \theta \in \Theta, \omega \in \Omega$$

where the last constraint is a consequence of Lemma 11. Therefore, by Lemma 12, the objective of this program is upper semi-continuous at $\boldsymbol{\pi}^*$, and the result follows immediately. □

Corollary 5 directly follows. The key insight is that any finite number of samples from the underlying distribution can be viewed as one signal from a more complicated distribution, and that this distribution still converges to a convergence point that will be IPV if the original convergence point is IPV.

Proof of Corollary 5 Let $\{(\theta_j,\omega_j)\}_{j=1}^N$ be a finite number of independent samples from the true distribution $\boldsymbol{\pi}_i$. Note the true distribution can be written as $\boldsymbol{\pi}_i = \boldsymbol{\pi}^* + \boldsymbol{\epsilon}_{\theta,i}$ for some $\boldsymbol{\epsilon}_{\theta,i} \in \mathbb{R}^{|\Omega|}$. Therefore, the probability of seeing samples $\{(\theta_j,\omega_j)\}_{j=1}^N$ and external signal ω is:

$$\boldsymbol{\pi}_i(\{(\theta_j,\omega_j)\}_{j=1}^N,\omega|\theta) = \pi_i(\omega|\theta)\prod_{j=1}^N \pi_i(\omega_j|\theta_j)\pi(\theta_j)$$

$$= (\pi^*(\omega|\theta) + \epsilon_{\theta,i}(\omega))\prod_{j=1}^N (\pi^*(\omega_j|\theta_j) + \epsilon_{\theta_j,i}(\omega_j))\pi(\theta_j)$$

which converges to $\boldsymbol{\pi}^*(\{(\theta_j,\omega_j)\}_{j=1}^N,\omega|\theta)$ as $\boldsymbol{\pi}_i$ converges to $\boldsymbol{\pi}^*$. Moreover, the samples $\{(\theta_j,\omega_j)\}_{j=1}^N$ are independent of the final round's bidder type, so the optimal mechanism over the distribution $\boldsymbol{\pi}^*(\{(\theta_j,\omega_j)\}_{j=1}^N,\omega|\theta)$ is revenue equivalent to the optimal mechanism over $\boldsymbol{\pi}^*$. Therefore, a finite number of samples is equivalent to a higher dimensional signal, and Theorem 4 applies directly. □

C.1. Unbounding the Approximation Ratio on a Converging Sequence of Distributions

While Corollary 5 states that we can't learn a mechanism that guarantees optimal revenue, it leaves open the possibility that we can learn a mechanism that achieves nearly optimal revenue. However, as the following example and Lemma shows, there exists a sequence of distributions that all satisfy the CM condition and whose full surplus revenue grows without bound in the number of bidder types. However, the sequence converges to an IPV distribution that has constant revenue in the number of bidder types.

EXAMPLE 2. Let the marginal distribution over the type of the bidder be given by $\pi(\theta) = 1/2^\theta$ for $\theta \in \{1, ..., |\Theta| - 1\}$ and $\pi(|\Theta|) = 1/2^{|\Theta|-1}$. Further let the value of the bidder for the item be $v(\theta) = 2^\theta$. Therefore, the expected value of the bidder's valuation is

$$\sum_{\theta=1}^{|\Theta|-1} \left(\frac{1}{2^\theta}\right) 2^\theta + \left(\frac{1}{2}\right)^{|\Theta|-1} 2^{|\Theta|} = |\Theta| + 1$$

Assume that the external signal set is $\Omega = \{1, ..., |\Theta| - 1\}$. Note that for a reserve price mechanism with a reserve price of $2^{|\Theta|}$, the expected revenue is 2. Further, if the distribution is IPV, this is the optimal mechanism (Myerson 1981).

LEMMA 13. *In the setting of Example 2, there exists a sequence of distributions $\{\boldsymbol{\pi}_i\}_{i=1}^\infty$ that converges to an IPV distribution and satisfies Assumption 2 such that for each distribution $\boldsymbol{\pi}_i$, their exists a mechanism $(\boldsymbol{p}_i, \boldsymbol{x}_i)$ whose expected revenue is $|\Theta| + 1$.*

Proof. Let, for all $k \in \{1, ..., |\Theta| - 1\}$, $\pi_i(\omega = k|\theta = k) = \frac{1}{|\Omega|} + \frac{1}{|\Omega|i}$ and $\pi_i(\omega \neq k|\theta = k) = \frac{1}{|\Omega|} - \frac{1}{(|\Omega|-1)|\Omega|i}$. Note that for all i, $\boldsymbol{\pi}_i$ satisfies the CM condition, and by Theorem 1, there exists a mechanism such that the expected revenue is $|\Theta| + 1$. Furthermore, $\boldsymbol{\pi}_i(\cdot|\cdot)$ converges to $\boldsymbol{\pi}^*(\cdot|\cdot) = 1/|\Omega|$, an IPV distribution. Finally, $\boldsymbol{\pi}^*(\cdot|\cdot) = 1/|\Omega|$ is in the interior of the convex hull of $\{\boldsymbol{\pi}_i\}_{i=1}^\infty$, so the sequence satisfies Assumption 2. □

Corollary 3 follows immediately from Lemma 13 and Corollary 5.

Acknowledgments

Albert and Conitzer are thankful for support from NSF under awards IIS-1527434, IIS-1814056, and CCF-1337215, ARO under grants W911NF-12-1-0550 and W911NF-11-1-0332, and a Guggenheim Fellowship.

A portion of this work has taken place in the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CPS-1739964, IIS-1724157, NRI-1925082), ONR (N00014-18-2243), FLI (RFP2-000), ARO (W911NF-19-2-0333), DARPA, Lockheed Martin, GM, and Bosch. Peter Stone serves as the Executive Director of Sony AI America and receives financial compensation for this work. The terms of this arrangement have been reviewed and approved by the University of Texas at Austin in accordance with its policy on objectivity in research.

References

Aghassi M, Bertsimas D (2006) Robust game theory. *Mathematical Programming* 107(1):231–273, ISSN 1436-4646, URL `http://dx.doi.org/10.1007/s10107-005-0686-0`.

Albert M, Conitzer V, Lopomo G (2015) Assessing the robustness of Cremer-McLean with automated mechanism design. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 763–769, AAAI '15 (Austin, TX, USA: AAAI Press), URL http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9296/9367.

Albert M, Conitzer V, Lopomo G (2016) Maximizing revenue with limited correlation: The cost of ex-post incentive compatibility. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)*.

Albert M, Conitzer V, Stone P (2017a) Automated design of robust mechanisms. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*.

Albert M, Conitzer V, Stone P (2017b) Mechanism design with unknown correlated distributions: Can we learn optimal mechanisms? *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 69–77, AAMAS '17 (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), URL http://dl.acm.org/citation.cfm?id=3091125.3091141.

Bergemann D, Morris S (2005) Robust mechanism design. *Econometrica* 73(6):1771–1813, ISSN 00129682, 14680262, URL http://www.jstor.org/stable/3598751.

Bertsimas D, Sim M (2004) The price of robustness. *Operations Research* 52(1):35–53, URL http://dx.doi.org/10.1287/opre.1030.0065.

Blum A, Mansour Y, Morgenstern J (2015) Learning valuation distributions from partial observation. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 798–805, AAAI '15 (Austin, TX, USA: AAAI Press), URL http://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/view/9550/9375.

Bulow J, Klemperer P (1996) Auctions versus negotiations. *The American Economic Review* 86(1):pp. 180–194, ISSN 00028282, URL http://www.jstor.org/stable/2118262.

Conitzer V, Sandholm T (2002) Complexity of mechanism design. *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence*, 103–110, UAI '02 (Alberta, Canada: Morgan Kaufmann), URL http://dl.acm.org/citation.cfm?id=2073889.

Conitzer V, Sandholm T (2004) Self-interested automated mechanism design and implications for optimal combinatorial auctions. *Proceedings of the 5th ACM Conference on Electronic Commerce*, 132–141, EC '04 (New York, NY, USA: ACM), ISBN 1-58113-771-0, URL http://dx.doi.org/10.1145/988772.988793.

Cremer J, McLean RP (1985) Optimal selling strategies under uncertainty for a discriminating monopolist when demands are interdependent. *Econometrica* 53(2):pp. 345–361, ISSN 00129682, URL http://www.jstor.org/stable/1911240.

Cremer J, McLean RP (1988) Full extraction of the surplus in Bayesian and dominant strategy auctions. *Econometrica* 56(6):pp. 1247–1257, ISSN 00129682, URL http://www.jstor.org/stable/1913096.

Devroye L (1983) The equivalence of weak, strong and complete convergence in l¡sub¿1¡/sub¿ for kernel density estimates. *The Annals of Statistics* 11(3):896–904, ISSN 00905364, URL http://www.jstor.org/stable/2240651.

FCC (2017) Broadcast incentive auction and post-auction transition. https://www.fcc.gov/about-fcc/fcc-initiatives/incentive-auctions, URL https://www.fcc.gov/about-fcc/fcc-initiatives/incentive-auctions.

Fu H, Haghpanah N, Hartline J, Kleinberg R (2014) Optimal auctions for correlated buyers with sampling. *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, 23–36, EC '14 (New York, NY, USA: ACM), ISBN 978-1-4503-2565-3, URL http://dx.doi.org/10.1145/2600057.2602895.

Gibbons R (1992) *Game Theory for Applied Economists* (Princeton University Press).

Government Accountability Office (2013) Reverse auctions: Guidance is needed to maximize competition and achieve cost savings. Technical report, Government Accountability Office, URL http://www.gao.gov/products/GAO-14-108.

Guo M, Conitzer V (2010) Computationally feasible automated mechanism design: General approach and case studies. *Proceedings of the Twentieth AAAI Conference on Artificial Intelligence*, 1676–1679, AAAI '10 (Atlanta, GA, USA: AAAI Press), URL http://cs.adelaide.edu.au/~mingyu/publication/nectar2010.pdf, nECTAR track.

Harsanyi JC (1967) Games with incomplete information played by "bayesian" players, i-iii: Part i. the basic model. *Management Science* 14(3):159–182, URL http://dx.doi.org/10.1287/mnsc.1040.0270.

Hartline JD (2014) Mechanism design and approximation, URL http://jasonhartline.com/MDnA/, http://jasonhartline.com/MDnA/.

Hartline JD, Roughgarden T (2009) Simple versus optimal mechanisms. *Proceedings of the 10th ACM Conference on Electronic Commerce*, 225–234, EC '09 (New York, NY, USA: ACM), ISBN 978-1-60558-458-4, URL http://dx.doi.org/10.1145/1566374.1566407.

Interactive Advertising Bureau (IAB) (2017) IAB programmatic revenue report 2016 results. https://www.iab.com/wp-content/uploads/2016/04/IAB_Internet_Advertising_Revenue_Report_FY_2016.pdf, URL https://www.iab.com/wp-content/uploads/2016/04/IAB_Internet_Advertising_Revenue_Report_FY_2016.pdf.

Kanoria Y, Nazerzadeh H (2014) Dynamic reserve prices for repeated auctions: Learning from bids, URL http://dx.doi.org/10.2139/ssrn.2444495, sSRN Paper. Suggested citation: Kanoria, Yash and Nazerzadeh, Hamid, Dynamic Reserve Prices for Repeated Auctions: Learning from Bids (July 1, 2014). Available at SSRN: http://ssrn.com/abstract=2444495 or http://dx.doi.org/10.2139/ssrn.2444495.

Kosenok G, Severinov S (2008) Individually rational, budget-balanced mechanisms and allocation of surplus. *Journal of Economic Theory* 140(1):126 – 161, ISSN 0022-0531, URL http://dx.doi.org/http://dx.doi.org/10.1016/j.jet.2007.07.002.

Kozlov M, Tarasov S, Khachiyan L (1980) The polynomial solvability of convex quadratic programming. *USSR Computational Mathematics and Mathematical Physics* 20(5):223 – 228, ISSN 0041-5553, URL http://dx.doi.org/http://dx.doi.org/10.1016/0041-5553(80)90098-1.

Lopomo G (2001) Optimality and robustness of the English auction. *Games and Economic Behavior* 36(2):219–240, ISSN 0899-8256, URL http://www.sciencedirect.com/science/article/pii/S0899825600908170.

Lopomo G, Rigotti L, Shannon C (2009) Uncertainty in mechanism design, working Paper.

Martin DH (1975) On the continuity of the maximum in parametric linear programming. *Journal of Optimization Theory and Applications* 17(3):205–210, ISSN 1573-2878, URL http://dx.doi.org/10.1007/BF00933875.

McAfee RP, Reny PJ (1992) Correlated information and mechanism design. *Econometrica* 60(2):pp. 395–421, ISSN 00129682, URL http://www.jstor.org/stable/2951601.

Mitzenmacher M, Upfal E (2005) *Probability and Computing: Randomized Algorithms and Probabilistic Analysis* (Cambridge University Press), ISBN 9780521835404, URL https://books.google.com/books?id=0bAYl6d7hvkC.

Mohri M, Medina AM (2014) Optimal regret minimization in posted-price auctions with strategic buyers. Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger K, eds., *Advances in Neural Information Processing Systems 27*, 1871–1879 (Curran Associates, Inc.), URL http://papers.nips.cc/paper/5438-optimal-regret-minimization-in-posted-price-auctions-with-strategic-buyers.pdf.

Mohri M, Medina AM (2015) Learning algorithms for second-price auctions with reserve. *Journal of Machine Learning Research* 3, URL http://www.cs.nyu.edu/~mohri/pub/jres.pdf.

Morgenstern J, Roughgarden T (2015) The pseudo-dimension of nearly-optimal auctions. *NIPS*, Forthcoming, URL http://arxiv.org/abs/1506.03684.

Myerson RB (1981) Optimal auction design. *Mathematics of Operations Research* 6(1):58–73, URL http://dx.doi.org/10.1287/moor.6.1.58.

Myerson RB, Satterthwaite MA (1983) Efficient mechanisms for bilateral trading. *Journal of Economic Theory* 29(2):265 – 281, ISSN 0022-0531, URL http://dx.doi.org/http://dx.doi.org/10.1016/0022-0531(83)90048-0.

NASA (2014) NASA chooses american companies to transport U.S. astronauts to international space station. https://www.nasa.gov/, URL NASAChoosesAmericanCompaniestoTransportU.S.AstronautstoInternationalSpaceStation.

Pardoe D, Stone P, Saar-Tsechansky M, Keskin T, Tomak K (2010) Adaptive auction mechanism design and the incorporation of prior knowledge. *Informs Journal on Computing* 22(3):353–370, URL http://joc.journal.informs.org/cgi/content/abstract/ijoc.1090.0353v1.

Roughgarden T, Talgam-Cohen I (2013) Optimal and near-optimal mechanism design with interdependent values. *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, 767–784, EC '13 (New York, NY, USA: ACM), ISBN 978-1-4503-1962-1, URL http://dx.doi.org/10.1145/2482540.2482606.

Sandholm T, Likhodedov A (2015) Automated design of revenue-maximizing combinatorial auctions. *Operations Research* 63(5):1000–1025, URL `http://dx.doi.org/10.1287/opre.2015.1398`.