Mechanisms for Risk Averse Agents, Without Loss

Shaddin Dughmi
Microsoft Research
shaddin@microsoft.com

Yuval Peres
Microsoft Research
peres@microsoft.com

June 15, 2012

Abstract

Auctions in which agents’ payoffs are random variables have received increased attention in recent years. In particular, recent work in algorithmic mechanism design has produced mechanisms employing internal randomization, partly in response to limitations on deterministic mechanisms imposed by computational complexity. For many of these mechanisms, which are often referred to as truthful-in-expectation, incentive compatibility is contingent on the assumption that agents are risk-neutral. These mechanisms have been criticized on the grounds that this assumption is too strong, because “real” agents are typically risk averse, and moreover their precise attitude towards risk is typically unknown a-priori. In response, researchers in algorithmic mechanism design have sought the design of universally-truthful mechanisms — mechanisms for which incentive-compatibility makes no assumptions regarding agents’ attitudes towards risk.

Starting with the observation that universal truthfulness is strictly stronger than incentive compatibility in the presence of risk aversion, we show that any truthful-in-expectation mechanism can be generically transformed into a mechanism that is incentive compatible even when agents are risk averse, without modifying the mechanism’s allocation rule. The transformed mechanism does not require reporting of agents’ risk profiles. Equivalently, our result can be stated as follows: Every (randomized) allocation rule that is implementable in dominant strategies when players are risk neutral is also implementable when players are endowed with an arbitrary and unknown concave utility function for money.

Our result has two main implications: (1) A mechanism designer concerned with an objective which depends only on the allocation rule of the mechanism can feel free to design a truthful-in-expectation mechanism, knowing that the risk-neutrality assumption can be removed by a generic black-box transformation. (2) Studying universally-truthful mechanisms under the pretense of robustness to risk aversion is no longer justified.

1 Introduction

Auctions in which agents’ payoffs are random variables have received increased attention in recent years. We are motivated by recent work in algorithmic mechanism design — particularly in prior-free settings — which has produced mechanisms employing internal randomization. The impetus for that work came from impossibility results for deterministic mechanisms: when constrained to polynomial time, randomized mechanisms provably yield superior approximation guarantees for objectives such as the social welfare than their deterministic counterparts, assuming standard complexity-theoretic conjectures such as $P \neq NP$. But randomization in agents’ payoffs is nothing
new; Bayesian mechanism design has always studied environments where players’ types are drawn from a known prior, and consequently even a deterministic mechanism induces a random payoff for each agent.

In many cases where randomized auctions — or auctions in randomized environments — are studied, incentive compatibility is contingent on the assumption that agents are risk neutral with respect to the random outcome resulting from the auction. In prior-free settings, mechanisms that are incentive-compatible when agents are risk neutral have often been referred to as truthf ul in expectation. Truthful-in-expectation mechanisms have been criticized on the grounds that the risk-neutrality assumption is too strong (e.g. [5, 6, 3, 16]), sideling these mechanisms as objects of mere theoretical interest. Consequently, random combinations of deterministic truthful mechanisms, known as universally-truthful, have received independent attention due to their robustness to agents’ attitudes towards risk (e.g. [6, 16]). This is despite the fact that the space of polynomial-time universally-truthful mechanisms is provably more restrictive, at least for the social welfare objective (see e.g. [2]). Luckily, it is an easy observation that universal truthfulness is a strictly stronger requirement than robustness to risk aversion, at least when adopting the standard approach for modeling risk. This easy observation is the starting point for this note.

We show that, for any truthful-in-expectation mechanism, the assumption of risk neutrality can be removed without altering the mechanism’s allocation rule. In other words, every (randomized) allocation rule that is implementable in dominant strategies when players are risk neutral is also implementable when players are endowed with an arbitrary unknown concave risk profile. A similar result holds for Bayesian settings, where randomness in an agent’s payoff is a result of both internal randomization of the mechanism as well as randomness in other agents’ types. Our result has two main implications: (1) A mechanism designer concerned with an objective which depends only on the allocation rule of the mechanism can feel free to design a truthful-in-expectation mechanism, knowing that the risk-neutrality assumption can be removed by a generic black-box transformation. (2) Studying universally-truthful mechanisms under the pretense of robustness to risk aversion is no longer justified.

We note that the mathematical transformation underlying our result is similar to that used by Esö and Futo [10], who consider a risk averse seller and risk neutral bidders. Nevertheless, we believe that the primary contribution of this note is not the underlying mathematics, but rather a framing much of the research in algorithmic mechanism design in as much as it relates to participants’ attitudes towards risk.

1.1 Results

We show that any truthful-in-expectation mechanism can be converted to a mechanism that is dominant-strategy incentive compatible when players are risk averse, without modifying the mechanism’s allocation rule. The intuition is as follows: a risk averse player facing a random monetary payoff $x$ can contract with a risk-neutral bank. The contract stipulates that the bank will always pay the player $E[x] - x$. This preserves the player’s expected payoff, and removes all associated risk. Such a bank can be simulated by the principal running the auction mechanism, and we show that the simulation preserves incentive compatibility. In essence, the auction mechanism can insure players against risk by correlating monetary transfers with the realized payoffs.

An interpretation of our result is that any allocation rule implementable in dominant strategies when players are risk neutral is also implementable when players are risk averse. The transformation results in no loss in the principal’s utility when the principal is risk neutral, and only increases
agents’ utilities. Moreover, the transformation does not depend on players’ risk profiles, and does not require reporting of said profiles. While our results are motivated by randomized mechanisms in prior-free settings, we note that they immediately apply to Bayesian settings as well. Specifically, in a setting where player types are drawn from a known prior, the principal can remove all risk from the utilities of agents participating the auction by appropriately correlating the payments of an agent with the realized types of his competition in the auction.

One implication of this note is an indictment of the notion of universal truthfulness studied in algorithmic mechanism design. Specifically, mechanisms that are universally truthful are often cited as more desirable, due primarily to their robustness towards players’ risk profiles, whereas truthful-in-expectation mechanisms are often thought of as constrained to quasilinear (risk neutral) utilities. This note demonstrates that, in fact, truthful-in-expectation mechanisms can be made incentive compatible for players with (unknown) risk averse utilities, at no loss in the objective so long as the objective is independent of the higher-order moments of agent payoffs and payments.

1.2 Related Work

Most of the literature in mechanism design assumes risk neutrality, both for the principal (also known as the seller) and the agents (the buyers). However, risk aversion has been studied, and we mention a sample of those works. Maskin and Riley [12] design a revenue-maximizing single-item auction in a Bayesian setting, where buyers are risk averse with given risk profiles. Esö and Futo [10] design a single-item auction that is optimal for a risk-averse seller, in a Bayesian setting where buyers are risk neutral. The mathematics underlying our result is similar to that underlying the mechanism of [10]; their mechanism is built from the revenue-maximizing auction of Myerson [13], and removes all risk from the seller’s revenue by charging each player a random payment equal to that in [13] in expectation, but appropriately correlated with the payments of other buyers such that the resulting revenue is deterministic.

More recent work has considered approximately optimal mechanisms for the seller’s utility when the seller and/or buyers are risk averse. Sundararajan and Yan [15] consider a multi-item Bayesian setting where the seller is risk averse, and buyers are risk neutral. They assume that the seller’s risk profile is unknown, and achieve a constant-factor approximation of the optimal seller utility simultaneously for all risk profiles. Bhalgat et al [1] consider settings with both risk averse buyers and sellers, with known risk profiles, and devise constant-factor approximations for the optimal seller utility.

2 Preliminaries

Let \([n] = \{1, \ldots, n\}\) denote a set of players. Let \(\Omega\) denote a set of feasible outcomes. For each \(i \in [n]\), let \(\mathcal{V}_i\) denote a family of valuation functions for player \(i\) mapping \(\Omega\) to \(\mathbb{R}\), and let \(\mathcal{V} = \mathcal{V}_1 \times \mathcal{V}_2 \times \ldots \times \mathcal{V}_n\). We consider direct revelation mechanisms of the form \((\mathcal{A}, p)\) where \(\mathcal{A} : \mathcal{V} \to \Omega\) is an allocation rule, and \(p : \mathcal{V} \to \mathbb{R}^n\) is a payment rule. We allow the mechanism to be randomized, in which case \(\mathcal{A}(v)\) and \(p(v)\) are random variables.

Such a mechanism is said to be truthful-in-expectation if each player maximizes his expected payoff, defined as his value for the chosen allocation less his payment, by reporting truthfully; precisely, if
\[ \mathbf{E}_i(u_i(A(v)) - p_i(v)) \geq \mathbf{E}_i(u_i(A(v_{-i}, v_i')) - p_i(v_{-i}, v_i')) \] (1)

for all \( i \in [n], v \in \mathcal{V}, v_i' \in \mathcal{V}_i \), where the expectation is taken over the internal randomness of the mechanism \((A, p)\). If a player \( i \) is risk neutral, i.e., his utility is simply his payoff, then reporting truthfully is a dominant strategy in such a mechanism. Truthful-in-expectation mechanisms can also be called dominant strategy incentive compatible for risk neutral players.

We consider risk averse players where risk aversion is modeled as follows. A risk averse player \( i \) with valuation \( v_i \) is also equipped with a non-decreasing, concave function \( u_i : \mathbb{R} \rightarrow \mathbb{R} \) denoting his utility for money. When faced with a random outcome \( \omega \in \Omega \) and random payment \( p_i \), the player’s utility is defined as

\[ u_i(v_i(\omega) - p_i) \] (2)

A mechanism \((A, p)\) is said to be dominant strategy incentive compatible for risk averse players if each player maximizes his expected utility by reporting truthfully; precisely, if

\[ \mathbf{E}_i[u_i(v_i(A(v)) - p_i(v))] \geq \mathbf{E}_i[u_i(v_i(A(v_{-i}, v_i')) - p_i(v_{-i}, v_i'))] \] (3)

for all \( i \in [n], v \in \mathcal{V}, v_i' \in \mathcal{V}_i \), and non-decreasing and concave \( u_i : \mathbb{R} \rightarrow \mathbb{R} \). When participating in such a mechanism, a risk averse player maximizes his utility by bidding truthfully regardless of reports of other players.

### 3 The Transformation

We now show that for every truthful-in-expectation mechanism \((A, p)\), there is a mechanism \((A, p')\) which is dominant-strategy incentive compatible for risk averse players. In other words, every allocation rule implementable for risk-neutral players is also implementable for risk-averse players. We stress that our mechanisms do not require reporting of the risk profiles, and the risk profiles may be arbitrary and unknown concave functions.

**Theorem 3.1.** For every truthful in expectation mechanism \((A, p)\), there is a payment rule \( p' \), correlated with the outcome of \( A \), such that \((A, p')\) is dominant-strategy incentive compatible for risk averse players. Moreover, \( \mathbf{E}[p'_i(v)] = \mathbf{E}[p_i(v)] \) for each player \( i \) and reported valuation profile \( v \).

We define \( p' \) as follows, correlated with the random outcome of the allocation rule \( A \)

\[ p'_i(v) = v_i(A(v)) - \mathbf{E}[v_i(A(v)) - p_i(v)]. \] (4)

Two claims are self-evident:

**Claim 3.2.** \( \mathbf{E}[p'_i(v)] = \mathbf{E}[p_i(v)] \) for each \( i \) and \( v \in \mathcal{V} \).

**Claim 3.3.** \( v_i(A(v)) - p'_i(v) \) is equal to its expectation, which is \( \mathbf{E}[v_i(A(v)) - p_i(v)] \), for every realization of the mechanism’s coins.

Claim 3.2 implies that \((A, p')\) is also truthful-in-expectation. Claim 3.3 implies that a player who bids truthfully faces a deterministic monetary payoff. Now, invoking Jensen’s inequality completes the proof that \((A, p')\) is incentive compatible for risk averse players.
Equality (5) follows from claim 3.3, inequality (6) follows from the assumption that $(A, p)$ is truthful in expectation and from monotonicity of $u_i$, equality (7) follows from claim 3.2, and (8) follows from Jensen’s inequality for concave functions.

4 Additional Discussion

The main takeaway from this note is the following: Once a mechanism designer has committed to an allocation rule, the risk neutrality assumption is without loss. Specifically, the assumption of risk neutrality can be relaxed to risk aversion generically and without modifying the allocation rule.

There are additional issues worth mentioning. First, while we posed our technical statements for prior free settings, a similar result holds in Bayesian environments. Second, this note weakens the most commonly invoked justification for the study of universally-truthful mechanisms, which have been a mainstay of algorithmic mechanism design for mainly historical reasons. Third, while our transformation can easily be implemented in polynomial time when an agent’s expected payoff can be calculated efficiently, there are computational considerations in mechanisms where the expectation is unknown and can only be sampled. Fourth, there are valid critiques of our result, including the fact that transfers in the resulting mechanisms may flow in both directions — from seller to buyer and vice versa — even when only one direction is natural, and the fact that the transformation may be unjustified when the choice of allocation rule may itself be contingent on the risk neutrality assumption. We discuss these four issues below.

4.1 Extension to Bayesian Settings

In Bayesian mechanism design, the valuation profile $v$ is drawn from a prior distribution $D$ which is common knowledge. In this context, incentive-compatibility is defined in reference to the prior $D$. Assuming risk neutrality, a mechanism $(A, p)$ is Bayesian Incentive Compatible if each agent $i$ with valuation $v_i$ maximizes his expected payoff by reporting truthfully, where the expectation is over the randomness is over draws of other agents’ types, as well as the internal randomness of the mechanism. In other words, Bayesian incentive compatibility is equivalent to truth-telling being a Bayes-Nash equilibrium of the resulting game.

As in the prior-free setting discussed in the rest of this note, risk neutrality can be relaxed to risk aversion in Bayesian settings. Specifically, observe that our transformation in Section 3 was agnostic to the source of randomness in the agent’s payoff, and so was its proof of correctness. This implies the following analogous result.

**Theorem 4.1.** For every Bayesian incentive compatible mechanism $(A, p)$ in a setting with risk neutral players, there is a payment rule $p'$, correlated with the outcome of $A$ and the types of the agents, such that $(A, p')$ is Bayesian incentive compatible for risk averse players. Moreover, for
each agent $i$ with fixed report $v_i$, we have that $E[p'_i(v_i, v_{-i})] = E[p_i(v_i, v_{-i})]$, where the randomness is over the internal random coins of the payment rules as well as the types $v_{-i}$ of players other than $i$.

4.2 Implications for Universal Truthfulness

Much of the literature on algorithmic mechanism design concerns universally-truthful mechanisms. We posit three main reasons for this: (1) Due to technical barriers, impossibility results were constrained to universally-truthful mechanisms in the early days of the field (e.g. [4, 14]), (2) truthful-in-expectation mechanisms surpassing the guarantees of their universally-truthful counterparts were few and far between until recently, and (3) recently-discovered truthful-in-expectation mechanisms appear impractical at first glance, particularly because they exploit the risk neutrality assumption to excess; for example by allocating the player many items with some probability, and nothing the rest of the time (e.g. [2]).

The study of universally-truthful mechanisms has been justified on the grounds of robustness to risk aversion (e.g. (e.g. [5, 6, 16])), an appealing argument in light of point (3) above. However, this justification misses an important distinction: universal truthfulness is a provably stronger assumption than robustness to risk aversion, specifically when risk aversion is modeled by a concave utility function for money. In fact, in many settings universally-truthful mechanisms are provably weaker, in terms of approximation the social welfare objective in polynomial time, than truthful-in-expectation mechanisms. Yet, as we show in this note, truthful-in-expectation mechanisms can be generically converted to mechanisms that are incentive-compatible for risk averse players without changing the allocation rule. This weakens the argument for the study of universally-truthful mechanisms, unless one is concerned with risk-seeking behavior.

4.3 Computational Considerations

It is easy to see that the only nontrivial requirement for efficient implementation of the transformation of Section 3 is that the expected payoff of each agent be calculable efficiently. This is indeed the case for some truthful-in-expectation mechanisms in the literature (e.g. [2, 11]). However, other recent examples do not admit a simple closed form for the expected payoff ([9, 7]). In such cases, it is not clear whether incentive compatibility in the presence of risk aversion can be recovered, even in an approximate sense, in a generic way. However, we argue that multiplicative approximations of the expected payoff suffices for approximate incentive compatibility in the presence of risk aversion, and show an example where this is possible.

To make this discussion concrete, we examine the mechanism of Dughmi, Roughgarden and Yan [9] for combinatorial auctions as an illustrative example. In combinatorial auctions, the outcomes $\Omega$ are the set of partial allocations of items $[m] = \{1, \ldots, m\}$ among the players $1, \ldots, n$. We represent an allocation as a set of disjoint bundles $(S_1, \ldots, S_n)$, where each $S_i$ is a subset of the items $[m]$. Each player $i$ is equipped with a valuation function $v_i : 2^{[m]} \to \mathbb{R}_+$, and player $i$’s payoff from an allocation $(S_1, \ldots, S_n)$ is simply $v_i(S_i)$. One class of valuation functions tackled in [9], which we will focus on here, is succinct coverage valuations. Specifically, $v_i$ is represented explicitly as a pair $(U^i, X^i)$, where $U^i$ is a set and $X^i$ is a family of $m$ subsets of $U^i$, indexed by the items $[m]$. The valuation function $v_i$ is then defined as follows: $v_i(S) = |\bigcup_{j \in S} X_j^i|$.

---

1 We note that, as defined here, elements of $U^i$ are unweighted. This is merely to simplify our presentation; our arguments apply equally well when each element of $U^i$ is equipped with a weight.
The mechanism of [9] uses a randomized allocation rule \( \mathcal{A} \), defined as follows. First, let the set \( \mathcal{P} \) be the family of fractional allocations; specifically, \( \mathcal{P} = \{ x \in \mathbb{R}^{n \times m}_+ : \forall j \in [m], \sum_{i=1}^n x_{ij} \leq 1 \} \). Then, they associate each \( x \in \mathcal{P} \) with a distribution \( D_x \) over allocations; namely, the allocation assigning item \( j \) to each player \( i \) with probability \( 1 - e^{-x_{ij}} \), independently for all items. The range \( \mathcal{R} \) of the allocation rule is then defined as \( \{ D_x : x \in \mathcal{P} \} \). Fortuitously, for each profile \( v \) of coverage valuations the expected social surplus of \( D_x \) is a concave function of \( x \). Therefore computing the \( x^* = x^*(v) \in \mathcal{P} \) associated with the welfare-maximizing distribution \( D_{x^*} \in \mathcal{R} \) is a convex optimization problem. Letting \( \mathcal{A}(v) \) be a sample from \( D_{x^*}(v) \), and charging each player a payment \( p(v) \) with expectation equal to his externality with respect to \( \mathcal{A}(v) \) gives a randomized member of the VCG family, and is therefore truthful in expectation.

Implementing this mechanism raises an important complication: \( x^* \) can not, in general, be computed explicitly. In fact this is manifestly impossible, as \( x^* \) may be irrational. This obstacle is overcome in [9] by exploiting the fact that \( D_{x^*} \) need only be sampled, rather than written down explicitly. Therefore, by interleaving the sampling process with an approximation procedure for \( x^* \) (using, e.g. the ellipsoid method), the distribution \( D_{x^*} \) can be sampled in expected polynomial time as needed for implementing the mechanism.

Combining this idea with the reduction we propose in this note does not appear possible in general. Specifically, our reduction requires explicit computation of each player’s expected payoff from the mechanism, whereas the trick just described can merely sample it. Even resorting to approximation via random sampling does not appear to yield approximate incentive compatibility for risk averse players. Specifically, random sampling can obtain an additive \( O(\epsilon v_i([m])) \)-approximation for player \( i \)’s payoff with high probability. However, due to concavity of player \( i \)’s utility, this does not lead to an \( O(\epsilon u_i([m])) \) additive approximation for player \( i \)’s utility.

Suppose instead that we are able to compute a multiplicative \( (1 - \epsilon) \)-approximation for player \( i \)’s payoff from the mechanism. Jensen’s inequality now implies that this yields a \( (1 - \epsilon) \)-approximation for player \( i \)’s utility. Plugging this into our reduction yields a \( (1 - \epsilon) \)-approximately incentive compatible mechanism \( (\mathcal{A}, \hat{p}) \) for risk averse players, in the following sense:

\[
E[u_i(v_i(\mathcal{A}(v)) - \hat{p}_i(v))] \geq (1 - \epsilon) E[u_i(v_i(\mathcal{A}(v_{-i}, v'_i) - \hat{p}_i(v_{-i}, v'_i))]
\]

Unfortunately, such multiplicative approximation of agent payoffs from a mechanism is not possible in general, and does not appear possible for the mechanism of [9], specifically for players whose payoffs are exponentially smaller than their maximum possible value for an allocation. Nevertheless, in this case an approximately-truthful (in the multiplicative sense of [9]) modification of the mechanism of [9], presented in [8], allows such an approximation and therefore yields an approximately-incentive compatible mechanism for risk averse players through our reduction. We omit the details. We expect that, in most cases, such approximations will be possible.

4.4 Critiques

The main critique of our reduction, in our opinion, can equally be levied against many critiques of truthful-in-expectation mechanisms for optimizing social welfare. It is best described as a hypothetical. Suppose, for example, that a principal designs an allocation rule \( \mathcal{A} \) that (approximately) maximizes the social welfare in the presence of risk neutral players, and a payment rule \( p \) rendering \( (\mathcal{A}, p) \) truthful in expectation. Our reduction yields \( \hat{p} \) such that \( (\mathcal{A}, \hat{p}) \) is incentive compatible even

\[\text{An agent } i \text{’s externality with respect to an allocation rule } \mathcal{A} \text{ is } E[\sum_{j \neq i} v_j(\mathcal{A}(v_{-i}, 0)) - \sum_{j \neq i} v_j(\mathcal{A}(v))]\]
when players are risk averse. However, in the presence of risk aversion, \(A\) is no longer a desirable allocation rule because it is *optimizing the wrong objective!* Specifically, in the presence of risk averse bidders, the social welfare is no longer the sum of bidders' values for the allocation, but rather the sum of their utilities plus the revenue of the principal (or the utility of the principal, when the principal is not risk neutral).

In other words, the mere adoption of the sum of agent values as the objective implicitly assumes risk neutrality. Therefore, a truthful in expectation mechanism is already justified in light of this choice, and our transformation does not buy us anything unless we concede that the objective being optimized is the wrong one. Nevertheless, even then our reduction does no harm, so long as we are unconcerned with the riskiness of the principal's revenue. Moreover, in situations where the choice of allocation rule is not influenced by the risk profiles of the agents, or is otherwise constrained to be independent of said profiles due to practical reasons or informational constraints, our reduction seems particularly appropriate.

Another deficiency of our reduction, discussed in the previous section, is that it is not efficiently implementable in general, in particular for mechanisms not admitting a simple closed form for the distribution of their random allocation rule. Nevertheless, as described there, this can often be remedied by resorting to approximation.

The final critique we mention concerns the direction of payments: since our payments are designed to insure an agent against risk, they may flow both from the principal to an agent and vice versa. This is true even in settings, such as auctions, where only one direction may be natural or practical.

References


