Maximum Entropy Prediction Markets

JACOB ABERNETHY, University of Michigan, Ann Arbor
SINDHU KUTTY, University of Michigan, Ann Arbor
SÉBASTIEN LAHAIE, Microsoft Research, New York City
RAHUL SAMI, University of Michigan, Ann Arbor

In this paper, we draw connections between the aggregation performed by learning algorithms and the information aggregation done in prediction markets. We show that, under reasonable conditions, the behavior of rational traders can be understood as the result of performing a learning algorithm on their private data. Similarly, the market state can be interpreted as a distribution over the outcome space. In particular, we show that a proper scoring rule can be derived from maximum entropy distributions. This scoring rule can be used as a general form of LMSR in prediction markets with over continuous outcome spaces. In order to provide insight on the behavior of rational traders in the market, we use the concept of exponential utility. We show that the traders' behavior can be understood as updating his belief using a Bayesian process and updating the market state in accordance with this utility function. These maxent prediction markets can also be used to design markets that are robust against adversarial traders. In fact, when traders are required to report their budgets and their beliefs, we can show that an informative trader eventually makes money and damaging traders eventually have limited influence in the market. Using ideas from convex analysis and the properties of the prediction market, we analyze the properties of the maxent market maker thus providing insight into the information content of the prediction market.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics; I.2.6 [**Artificial Intelligence**]: Learning

General Terms: Prediction Market Design, Machine Learning Algorithms

Additional Key Words and Phrases: Exponential Families, Bayesian Learning, Maximum Entropy Distributions, generalized LMSR

ACM Reference Format:
Jacob Abernethy, Sindhu Kutty, Sébastien Lahaie, and Rahul Sami, 2014. Maximum Entropy Prediction Markets *ACM* 9, 4, Article 39 (June 2013), 21 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Prediction markets are aggregation mechanisms that allow market prices to be interpreted as predictive probabilities on an event. Each trader in the market is assumed to have some private information that he uses to make a prediction on the outcome of the event. Traders are allowed to report their beliefs on the outcome of the event

DOI:http://dx.doi.org/10.1145/0000000.0000000

by allowing them to buy and sell securities whose value depends on the outcome of this future event. This will effect the state of the market, thus updating the predictive probabilities for the event. Further, since the trades are done sequentially, the trader is allowed to observe all past trades in the market and update his private information based on this information. Traders can see the past history of trades, so the price at which a current trader is willing to buy and sell these securities can be interpreted as an aggregate "consensus probability forecast" of a particular candidate winning the election.

(JAKE: I want to use this paragraph for a general overview of mm frameworks) One popular form of prediction markets is the market scoring rule [Hanson 2003]. A market scoring rule considers all trades as a single chronological sequence. Traders earn rewards proportional to the incremental reduction in prediction loss caused by their trades in comparison to the previous trade. In other words, their rewards depend on the change in market probabilities caused by their trade, as well as on the eventual outcome. Thus, each trader has an incentive to minimize the prediction loss. In this format, the *market maker* who runs the market can suffer an overall loss, but Hanson [Hanson 2003] showed that, for market scoring rules on finite outcome spaces, the loss of the market maker can be bounded.

Much the work on prediction market frameworks has focused primarily on structural properties of the mechanism: incentive compatibility, the market maker loss, the available liquidity, the fluctuations of the prices as a function of the trading volume, to name a few. Absent from much of the literature is a corresponding *semantics* of the market behavior or the observed prices. That is, how can we interpret the equilibrium market state when we have a number of traders with diverse beliefs on the underlying state of the world? In what sense is the market an aggregation mechanism? Do price changes relate to our usual Bayesian notion of information incorporation via posterior updating?

In the present work we show that a number of classical statistical tools can be leveraged to design a prediction market framework in the mold of an *exponential family distributions* that possesses a number of attractive properties. Common concepts in statistics *entropy maximization*, *log loss*, and *bayesian inference*, relate to natural aspects of our class of mechanisms. In particular, the central objects in our market framework can be interpreted as elements of exponential families:

— the *payoff function* of the market corresponds to the *sufficient statistics* of the probabiilty distribution;

— the vector of *oustanding shares* in the market corresponds to the *natural parameter vector* of the distribution;

— the market prices correspond to *mean parameters*;

— the profit potential for a trader corresponds to a *Kullback-Leibler divergence* between the trader's belief and that of the market.

In addition to showing this syntactic relationship between exponential families and prediction markets, we also explore the semantic implications as well. In particular, we show that our formulation allows us to analyze the evolution of the market under various models of trader behavior:

— We consider *budget-limited traders* who are constrained in how much they influence the market. We analyze the market under these circumstances; we are able to show that traders with good information can expect to profit and their influence over the market state increases over time whereas malicious traders have limited impact on the market.

— We consider when our agents are risk-averse in an interestins special case, that is under the assumption they utilize *exponential utility* to optimize their bets. In this case we can characterize precisely how a single trader interacts with the market as well as the equilibrium reached given multiple traders. The eventual market state turns out to be a weighted combination of trader beliefs and initial market state; the weights depend on the risk aversion parameter of the individual traders.

— In addition, we observe that trader behavior varies depending on whether they assimilate information as Bayesians or as frequentists. Interestingly, we show that although these two kinds of traders pick the eventual market state as a convex combination of their private belief and current market state, they do so in dual spaces.

1.1. Related Work

Designing prediction markets to handle a large outcome space is an active area of research. In [Chen et al. 2008], the authors use a restricted betting language to design efficient markets for a combinatorial outcome space. This technique is generalized by [Pennock and Xia 2011]. [Gao et al. 2009] consider extending various automated market makers to an infinite outcome space. For the logarithmic market scoring rule they show that unbounded market maker loss can result in this setting. [Abernethy et al. 2011] and [Othman and Sandholm 2011] specify frameworks under which they design cost function based markets that satisfy the desirable property of bounded market maker loss even in infinite outcome spaces.

The connection between machine learning and prediction markets has been studied previously. [Chen and Vaughan 2010] and [Abernethy et al. 2011] have previously explored the connection to learning algorithms to inform the design and understanding of prediction markets. In particular, [Chen and Vaughan 2010] consider the correspondence between prediction markets with market scoring rules and the Follow the Regularized Leader algorithm proposed by [Kalai and Vempala 2005] and thus provide insight into the aggregation mechanism of a prediction market.

Independently of this work, Beygelzimer et al. [Beygelzimer et al. 2012] have shown that, for a particular form of binary prediction markets and traders with log-utility, the long-run dynamics of trading activity and budgets over many prediction markets lead the markets to satisfy a bounded regret property with respect to the best single trader. Our results here, and the future work we have suggested, form a program to prove bounded regret properties of budget-limited prediction markets under much more general conditions.

2. GENERALIZED LOG SCORING RULES

We consider a measurable space that consists of a set of outcomes $\mathcal{X}$ together with a σ-algebra $\mathcal{F}$. An agent or expert has a *belief* over potential outcomes taking the form of a probability measure absolutely continuous with respect to a base measure ν.[1] Throughout we represent the belief as the corresponding density p with respect to ν.

[1]Recall that a measure P is absolutely continuous with respect to ν if $P(A) = 0$ for every $A \in \mathcal{F}$ for which $\nu(A) = 0$. In essence the base measure ν restricts the support of P. In our examples ν will typically be a restriction of the Lebesgue measure for continuous outcomes or the counting measure for discrete outcomes.

Let $\mathcal{P}$ denote the set of all such probability densities, and let $\mathcal{D} \subseteq \mathcal{P}$ be a subset from which the beliefs are drawn.

We are interested in eliciting information about the agent's belief, in particular expectation information. Let $\phi : \mathcal{X} \to \mathbf{R}^d$ be a vector-valued random variable or *statistic*, where d is finite. The aim is to elicit $\mu = \mathbf{E}_p[\phi(x)]$ where x is the random outcome. A *scoring rule* is a device for this purpose. Let

$$\mathcal{M} = \left\{\mu \in \mathbf{R}^d : \mathbf{E}_p[\phi(x)] = \mu, \text{ for some } p \in \mathcal{P}\right\}$$

be the set of realizable statistic expectations. A scoring rule $S : \mathcal{M} \times \mathcal{X} \to \mathbf{R} \cup \{-\infty\}$ pays the agent $S(\hat{\mu}, x)$ according to how well its report $\hat{\mu} \in \mathcal{M}$ agrees with the eventual outcome $x \in \mathcal{X}$.

Definition 2.1. A scoring rule S is *proper over domain* $\mathcal{D}$ for statistic ϕ if for each $\mu \in \mathcal{M}$ and $p \in \mathcal{D}$ with expected statistic μ, we have

$$\mathbf{E}_p[S(\mu, x)] \geq \mathbf{E}_p[S(\hat{\mu}, x)] \tag{1}$$

for all alternative $\hat{\mu}$. If the domain is $\mathcal{D} = \mathcal{P}$, the set of all possible densities, then we simply say the scoring rule is *proper*.

Note that given a proper scoring rule S any affine transformation $\tilde{S}(\mu, x) = aS(\mu, x) + b(x)$ of the rule, with $a > 0$ and b an arbitrary real-valued function of the outcomes, again yields a proper scoring rule termed *equivalent*. Throughout we will often apply such affine transformations to obtain the clearest version of the scoring rule. We will also focus on scoring rules where inequality (1) is strict to avoid trivial cases such as constant scoring rules.

Our definition is more general than classical scoring rules in two respects. Classically, scoring rules take in the entire density p rather than just some statistic, and incentive compatibility must hold over all of $\mathcal{P}$. When the outcome space is large or infinite, it is not feasible to directly communicate p, so our definition allows for summary information of the belief. If there is a bijection between the sets $\mathcal{D}$ and $\mathcal{M}$, then we say that $\mathcal{M}$ parametrizes $\mathcal{D}$ and write $p(\cdot\,; \mu)$ for the density mapping to μ.

Note that this definition of a scoring rule places only mild information requirements on the part of the agent to ensure truthful reporting. Because condition (1) holds for all p consistent with expectation μ, it is enough for the agent to simply know the latter and not the complete density to be properly incentivized. However, the agent must also agree that the density is actually drawn from domain $\mathcal{D}$; if the scoring rule is proper, it is enough that the agent agree with the support of the density as implicitly defined by base measure ν.

When the outcome space is finite we can recover classical scoring rules from the definition by taking $\mathcal{D} = \mathcal{P}$ and using the statistic $\phi : \mathcal{X} \to \{0,1\}^{\mathcal{X}}$ that maps an outcome x to a unit vector with a 1 in the component corresponding to x. The expectation of ϕ is then exactly the probability mass function.

2.1. Proper Scoring from Consistency

Our starting point for designing proper scoring rules is the classic logarithmic scoring rule for eliciting probabilities in the case of finite outcomes. This rule is simply $S(p, x) = \log p(x)$ and there are several ways to understand why it is proper, with different implications for generalization.

First observe that the scoring rule compensates the agent with the log likelihood it assigned to the outcome, so the agent is maximizing the expected log likelihood via its report. Now it is a fundamental result in statistics that, under a variety of different sufficient conditions, the maximum likelihood estimator is statistically consistent; see [] for an overview of such conditions. This means that in the limit (as the size of the i.i.d. empirical sample increases) the log likelihood is maximized by the true parameters. By this reasoning, the following result follows immediately from the definition of consistency and the law of large numbers.

THEOREM 2.2. *Suppose that $\mathcal{M}$ parametrizes $\mathcal{D}$. The logarithmic scoring rule defined by*

$$S(\mu, x) = \log p(x; \mu), \tag{2}$$

where $\mu \in \mathcal{M}$, is proper over $\mathcal{D}$ if and only if the maximum likelihood estimator for μ is consistent.

We note that although our focus is on parametrizations by expectations, the preceding theorem in fact holds for more generic parameter spaces. The range of scoring rules it provides is broader than those characterized by [], because it allows for a restriction over the set of beliefs. On the other hand, its applicability is limited to those domains $\mathcal{D}$ that can be parametrized by the relevant statistics. The following provide concrete examples of these points.

Example 2.3. Suppose the outcomes are supported on $[1, +\infty)$ and follow a Pareto distribution with density $f(x; \alpha) = \alpha / x^{\alpha+1}$ parametrized by an index $\alpha > 0$. The mean μ is related to the index via the one-to-one mapping $\mu = \frac{\alpha}{\alpha - 1}$, so the density can alternatively be parametrized by the mean. Theorem 2.2 gives the following proper scoring rule for the mean of a Pareto distribution with support on $[1, +\infty)$:

$$S(\mu, x) = \log \frac{\mu}{\mu - 1} - \left(\frac{\mu}{\mu - 1} + 1 \right) \log x. \tag{3}$$

We stress that the rule can only elicit the mean assuming the agent knows its belief is a Pareto distribution over $[1, +\infty]$. It does not elicit the mean of other families of densities parametrized by the mean (e.g., the exponential distribution).

Example 2.4. Suppose the outcomes are supported on $(-\infty, +\infty)$ and follow a Cauchy distribution with density $f(x; m) = 1/\pi[1 + (x - m)^2]$, parametrized by the median m. Theorem 2.2 leads to the following proper scoring rule for the median:

$$S(m, x) = -\log[1 + (x - m)^2]. \tag{4}$$

Note that the median cannot be obtained as the expectation of any statistic. This highlights the range of parameters than may be elicited by the log score in different circumstances.

2.2. Proper Scoring from Maximum Entropy

We now turn to scoring rules that are proper over the entire domain of densities $\mathcal{P}$. Our construction is the same as in the previous section: we take the log likelihood of a density parametrized by the relevant statistics. If the family of densities is chosen appropriately, the resulting rule will in fact hold over all $\mathcal{P}$. To achieve this we draw on a well-known duality between maximum likelihood and maximum entropy.

Exponential Families. We let $p(x;\mu)$ be the maximum entropy distribution with expected statistic μ. Specifically, it is the solution to the following mathematical program:[2]

$$\min_{p\in\mathcal{P}} F(p) \quad \text{s.t.} \quad \mathbf{E}_p[\phi(x)] = \mu, \tag{5}$$

where the objective function is the negative entropy of the distribution, namely

$$F(p) = \int_{x\in\mathcal{X}} p(x)\log p(x)\, d\nu(x).$$

Note that the explicit set of constraints in (5) are linear, and to stress this fact we find it helpful to re-write them as $A_\phi p = \mu$, where A_ϕ is the expectation operator of statistics ϕ. We let $G : \mathcal{M} \to \mathbf{R}$ be the optimal value function of (5), meaning $G(\mu)$ is the (negative) entropy of the maximum entropy distribution with expected statistics μ.

It is well-known that solutions to (5) are *exponential family* distributions, whose densities with respect to ν take the form

$$p(x;\theta) = \exp(\langle\theta, \phi(x)\rangle - T(\theta)). \tag{6}$$

The density is stated here in terms of its *natural* parametrization $\theta \in \mathbf{R}^d$, where θ arises as the Lagrange multiplier associated with linear constraints $A_\phi p = \mu$. The term $T(\theta)$ essentially arises as the multiplier for the normalization constraint (the density must integrate to 1), and so ensures that (6) is normalized:

$$T(\theta) = \log \int_{\mathcal{X}} \exp\langle\theta, \phi(x)\rangle\, d\nu(x). \tag{7}$$

The function T is known as the *log-partition* or *cumulant* function corresponding to the exponential family. Its domain is $\Theta = \{\theta \in \mathbf{R}^d : T(\theta) < +\infty\}$, called the natural parameter space. The exponential family is *regular* if Θ is open—almost all exponential families of interest, and all those we consider in this work, are regular. The family is *minimal* if there is no $\alpha \in \Theta$ such that $\langle\alpha, \phi(x)\rangle$ is a constant over $\mathcal{X}$ (ν-almost everywhere); minimality is a property of the associated statistic ϕ, usually called the *sufficient statistic* in the literature.

The following proposition collects the relevant results on regular exponential families and their associated parameter domains and functions. A convex function T is of *Legendre type* if it is proper, closed, strictly convex and differentiable on the interior of its domain, and $\lim_{\theta\to\bar\theta} ||\nabla T(\theta)|| = +\infty$ when $\bar\theta$ lies on the boundary of the domain.

PROPOSITION 2.1. *Consider a regular exponential family with minimal sufficient statistic. The following properties hold:*

(1) *T and G are of Legendre type, and $T = G^*$ (equivalently $G = T^*$).*

(2) *The gradient map ∇T is one-to-one and onto the interior of $\mathcal{M}$. Its inverse is ∇G which is one-to-one and onto the interior of Θ.*

(3) *The exponential family distribution with natural parameter $\theta \in \Theta$ has expected statistic $\mu = \mathbf{E}_p[\phi(x)] = \nabla T(\theta)$.*

[2]We assume throughout that the minimum is finite and achieved for all $\mu \in \mathcal{M}$. Some care is needed to ensure this holds for specific statistics and outcome spaces. For example, taking outcomes to be the real numbers, there is no maximum entropy distribution with a given mean μ (one can take densities tending towards the uniform distribution over the reals), but there is always a solution if we constrain both the mean and variance.

(4) *The maximum entropy distribution with expected statistic μ is the exponential family distribution with natural parameter $\theta = \nabla G(\mu)$.*

In the above T^* denotes as usual the convex conjugate of T, which here can be evaluated as $T^*(\mu) = \sup_{\theta \in \Theta} \langle \theta, \mu \rangle - T(\theta)$. Similarly, $G^*(\theta) = \sup_{\mu \in \mathcal{M}} \langle \theta, \mu \rangle - G(\mu)$.

Proper Log Scoring. We are now in a position to analyze the log scoring rule under exponential family distributions. From our discussion so far, we have that an exponential family density can be parametrized either by the natural parameter θ, or by the mean parameter μ, and that the two are related by the invertible gradient map $\mu = \nabla T(\theta)$. We will write $p(x; \theta)$ or $p(x; \mu)$ given the parametrization used, which should be clear from context.

The following observation is crucial. Let $\tilde{p} \in \mathcal{P}$ be a density (not necessarily from an exponential family) with expected statistic μ, let $p(\cdot\,; \mu)$ be the exponential family with the same expected statistic, and let $\mu' \in \mathcal{M}$ be an alternative report. Then note from (6) that

$$\mathbf{E}_{\tilde{p}}[\log p(x; \mu')] = \mathbf{E}_{p(\cdot;\mu)}[\log p(x; \mu')] = \langle \theta', \mu \rangle - T(\theta'), \tag{8}$$

where $\theta' = \nabla G(\mu')$ is the natural parameter for the exponential family with statistic μ'. We see from this that the expected log score only depends on the expectation μ of the underlying density, not the full density, which is how we can achieve proper scoring according to Definition 2.1.

THEOREM 2.5. *Consider the logarithmic scoring rule $S(\mu, x) = \log p(x; \mu)$ defined over a set of densities $\mathcal{D}$ parametrized by $\mathcal{M}$. The scoring rule is proper over the entire domain of densities $\mathcal{P}$ if and only if $\mathcal{D}$ is the exponential family with statistic ϕ.*

PROOF. Let $\mu, \mu' \in \mathcal{M}$ be the agent's true belief and an alternative report, and let $p, p' \in \mathcal{P}$ be densities consistent with each belief. Let $\theta = \nabla G(\mu)$ and $\theta' = \nabla G(\mu')$, and note that $\mu = \nabla T(\theta)$. We have

$$\begin{aligned} & \mathbf{E}_p[\log p(x; \mu)] - \mathbf{E}_p[\log p(x; \mu')] \\ =\; & \langle \theta, \mu \rangle - T(\theta) - \langle \theta', \mu \rangle + T(\theta') \\ =\; & T(\theta') - T(\theta) - \langle \theta' - \theta, \mu \rangle \\ =\; & T(\theta') - T(\theta) - \langle \theta' - \theta, \nabla T(\theta) \rangle. \end{aligned} \tag{9}$$

The latter is positive by the strict convexity of T, which shows that the log score is proper. Now assume the defined log score is proper. By the Savage characterization of proper scoring rules for expectations (see []), we must have

$$S(\mu, x) = G(\mu) - \langle \nabla G(\mu), \mu - \phi(x) \rangle$$

for some strictly convex function G. Let $T = G^*$, so that $\nabla G = \nabla T^{-1}$, and let $\theta = \nabla G(\mu)$. Then the above can be written as

$$\begin{aligned} \log p(x; \mu) &= G(\mu) - \langle \nabla G(\mu), \mu - \phi(x) \rangle \\ &= \langle \theta, \mu \rangle - T(\theta) - \langle \theta, \mu - \phi(x) \rangle \\ &= \langle \theta, \phi(x) \rangle - T(\theta), \end{aligned}$$

which shows that $p(x; \mu)$ takes the form of an exponential family. $\square$

As further intuition for the result, note that (9) is the definition of the so-called Bregman divergence with respect to strictly convex function T. Therefore we have

$$\mathbf{E}_p[\log p(x; \mu)] - \mathbf{E}_p[\log p(x; \mu')] = D_T(\theta', \theta) = D_G(\mu, \mu'),$$

where the last equality is a well-known identity relating the Bregman divergences of T and $T^* = G$ []. The equation states that the agent's regret from misreporting its mean parameter does not depend on the full density p, only the mean μ.

Theorem 2.5 leads to a straightforward procedure for constructing score rules for expectations: define the relevant statistic, and consider the maximum entropy (equivalently, exponential family) distribution consistent with the agent's reported mean μ. The scoring rule compensates the agent according to the log likelihood of the eventual outcome according to the latter. The interpretation is that the agent is only providing partial information about the underlying density, so the principal first infers a full density according to the principle of maximum entropy, and then scores the agent using the usual log score.

An advantage of this generalization of the log score is that, for many expectations of interest, it leads to simple *closed-form* formulas for the scoring rule (including for multi-dimensional outcomes in some cases). This covers for instance several different combinations of the arithmetic, geometric, and harmonic means, as well as higher order moments. The following examples illustrate the construction.

Example 2.6. As base measure we take the Lebesgue restricted to $[0, +\infty)$, and we consider the statistic $\phi(x) = x$ so that we are simply eliciting the mean. The maximum entropy distribution with a given mean μ is the exponential distribution, and taking its log density gives the scoring rule

$$S(\mu, x) = -\frac{x}{\mu} - \log \mu. \tag{10}$$

We stress that although this rule is derived from the exponential distribution, Theorem 2.5 implies that it elicits the mean of any distribution supported on the nonnegative reals (e.g., Pareto, lognormal), in contrast to (3). Indeed, it is easy to see that the expected score (10) depends only on the mean of the agent's belief because it is linear in x. As a generalization of this example, the maximum entropy distribution for the k-th moment $\phi(x) = x^k$ with respect to the same base measure is the Weibull distribution. Taking its log density leads to the following equivalent scoring rule:

$$S(\mu, x) = (k-1)\log x - k \log \mu - \Gamma\left(1 + \frac{1}{k}\right)^k \left(\frac{x}{\mu}\right)^k,$$

where Γ denotes the usual gamma function (the extension of the factorial to the reals).

Example 2.7. As a base measure we take the Lebesgue over the real numbers $\mathbf{R}$. We are interested in eliciting the mean μ and variance σ^2, so as a statistic we take $\phi(x) = (x, x^2)$ for which $\mathbf{E}_p[\phi(x)] = (\mu, \mu^2 + \sigma^2)$. The maximum entropy distribution for a given mean and variance is the normal distribution, and taking its log density gives the scoring rule

$$S((\mu, \sigma^2), x) = -\frac{(x-\mu)^2}{\sigma^2} - \log \sigma^2. \tag{11}$$

Again, we stress that this scoring rule elicits the mean and variance of any density over the real numbers, not just those of a normal distribution. The construction easily generalizes to a multi-dimensional outcome space by taking the log density of the multivariate normal:

$$S((\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{x}) = -(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) - \log |\boldsymbol{\Sigma}|. \tag{12}$$

Here the statistics being elicited are the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$. These scoring rules have been studied by Dawid and Sebastiani [] as rules that only depend on the mean and variance of the reported density. They note that these rules are weakly proper (because they do not distinguish between densities with the same first and second moments), but do not make the point that knowledge of the full density is not necessary on the part of the agent.

In the above, Example 2.7 illustrates an important point about parametrizations of the elicited expectations. The variance σ^2 cannot be written as $\mathbf{E}[\phi(x)]$ for any ϕ, because the mean μ enters the definition of σ^2 but is not available when ϕ is defined (indeed it is elicited in tandem with the variance).[3] Instead one must use the first two *uncentered* moments $\mathbf{E}[x]$ and $\mathbf{E}[x^2]$. These are in bijection with μ and σ^2, so the resulting scoring rule can be re-written in terms of the latter. Therefore, it is possible to elicit not just expectations but also bijective transformations of expectations.

3. MAXIMUM ENTROPY MARKET MAKING

In a single-agent setting, a scoring rule is used to *elicit* the agent's belief. In a multi-agent setting, an information market can be used to *aggregate* the beliefs of the agents. In his seminal paper introduced the idea of a market scoring rule, which inherits the appealing elicitation and aggregation properties of both in order to perform well in thin or thick markets. In this section, we adapt the generalized log scoring rule to a market scoring rule which leads to markets with simple closed-form cost functions for many statistics of interest. The exponential families framework also allows us to move beyond risk-neutral agents to understand the aggregation properties of information markets under alternative behaviors, such as Bayesian updating and risk aversion.

3.1. Information Market

In an information market an agent's expected belief μ is elicited indirectly through the purchase and sale of contingent claim securities. Under this approach, each component i of the statistic ϕ is interpreted as the payoff function of a security; that is, a single share of security i pays off $\phi_i(x)$ when $x \in \mathcal{X}$ occurs. Thus if the portfolio of shares held by the agent is $\delta \in \mathbf{R}^d$, where entry δ_i corresponds to the number of shares of security i, then the payoff to the agent when x occurs is evaluated by taking the inner product $\langle \delta, \phi(x) \rangle$.

As a concrete example, recall that in the classic finite-outcome case the statistic has a component for each outcome x such that $\phi_x(x') = 1$ if $x' = x$ and 0 otherwise. Therefore the corresponding security pays 1 dollar if outcome x occurs. (These are known as Arrow-Debreu securities.) In Example 2.6 the one-dimensional statistic is $\phi(x) = x$, corresponding to a security whose payoff is linear in the outcome $x \in \mathbf{R}_+$. (This amounts to a futures contract.)

The standard way to implement an information market in the literature, due to , is via a centralized market maker. The market maker maintains a convex, differentiable cost function $C : \mathbf{R}^d \to (-\infty, +\infty]$, where $C(\theta)$ records the revenue collected when the vector of outstanding shares is θ. The cost to an agent of purchasing portfolio δ under a market state of θ is $C(\theta + \delta) - C(\theta)$, and therefore the instantaneous prices of the securities are given by the gradient $\nabla C(\theta)$.

A risk-neutral agent will choose to acquire shares up to the point where, for each share, expected payoff equals marginal price. Formally, if the agent acquires portfolio

[3]This is an intuitive but far from formal explanation for the fact that the dimension of the message space, or *elicitation complexity*, for eliciting the variance is at least 2 [].

δ, moving the market state vector to $\theta' = \theta + \delta$, then we must have

$$\mathbf{E}_p[\phi(x)] = \nabla C(\theta'). \tag{13}$$

In this way, by its choice of δ, the agent reveals that its expected belief is $\mu = \nabla C(\theta')$. We stress that this observation relies on the assumptions that 1) the agent is risk-neutral, 2) the agent does not incorporate the market's information into its own beliefs, and 3) the agent is not budget constrained. We will examine relaxations of each assumption in later sections.

3.2. Information-Theoretic Interpretation

In the remainder of this paper we focus on the following cost function, which arises from the "generalized" logarithmic market scoring rule (LMSR):

$$C(\theta) = \log \int_{x \in \mathcal{X}} \exp\left[\langle \theta, \phi(x) \rangle\right] \nu(dx). \tag{14}$$

This is of course exactly the log-partition function (7) for the exponential family with sufficient statistic ϕ, and we recover the classic LMSR using outcome indicator vectors as statistics. Because an agent would never select a portfolio with infinite cost, the effective domain (i.e., the possible vectors of outstanding shares) of C is $\Theta = \{\theta \in \mathbf{R}^d : C(\theta) < +\infty\}$, which gives an economic interpretation to the natural parameter space of an exponential family.

The correspondence between the cost function (14) and the log-partition function (7) suggests the following interpretation. The market maker maintains an exponential family distribution over the state space $\mathcal{X}$ parametrized by share vectors that lie in Θ. When an agent buys shares, it moves the distribution's natural parameter so that the market prices matches its beliefs, or in other words the market's mean parametrization matches the agent's expectation.

There is a well-known duality between scoring rules and cost-function based markets. To see this in our context, recall from (8) that

$$\mathbf{E}_{\tilde{p}}[\log p(x; \mu')] = \langle \theta', \mu \rangle - T(\theta')$$

where $\tilde{p}$ is the agent's belief and μ' the agent's report. The expected log score from reporting μ' is exactly the same as the expected payoff from buying portfolio of shares $\theta' = \nabla C(\mu')$ (assuming an initial market state of 0), as $\langle \theta', \mu \rangle$ is the expected revenue and $T(\theta')$ is the cost. As in Section 2 this reasoning relies on the assumption of risk-neutrality (and no budget constraints), not on any specific form for the agent's belief.

The agent's expected profit from moving the share vector from θ to θ' is

$$\begin{aligned} & \langle \theta' - \theta, \mu \rangle - C(\theta') + C(\theta) \\ = \ & C(\theta) - C(\theta') - \langle \theta - \theta', \nabla C(\theta) \rangle \\ = \ & D_C(\theta, \theta') = D_{C^*}(\mu', \mu), \end{aligned}$$

recalling (9). Now have observed (among others) that the Kullback-Leibler divergence between two exponential family distributions is the Bregman divergence, with repect to the log-partition function, between their natural parameters. The agent's expected profit is therefore the KL divergence between the market's implied expectation and the exponential family corresponding to the agent's expectation, a well-known property from the classical LMSR [].

3.3. Examples: Real Line and the Sphere

Let us now revisit our scoring rules examples from Section 2 in the context of prediction markets. The relevant entities now are the payoff function, the effective domain of shares, and the cost function.

Example 3.1. We consider outcomes over the positive reals bR_+ and set up a market for the expected outcome, consisting of a single security that pays off $\phi(x) = x$. The log partition funciton of the exponential distribution leads to the following cost function:

$$C(\theta) = -\log(-\theta).$$

The effective domain is $\Theta = \{\theta \in \mathbf{R} : \theta < 0\}$. This means the market must start with a negative number of outstanding shares for the security, and the number shares must stay negative. The market maker need not explicitly enforce this, because by the Legendre property of C the cost tends to $+\infty$ as the outstanding shares approach the boundary, which is straightforward to see in this example.

Example 3.2. We consider outcomes over the real line $\mathbf{R}$ and set up a market with securities corresponding to the first two uncentered moments (i.e, agents are betting on the return and volatility). The securities are defined by the payoffs $\phi(x) = (x, x^2)$. The log partition function of the normal distribution, under its natural parametrization, leads to the following cost function:

$$C(\theta) = -\frac{\theta_1^2}{4\theta_2} - \frac{1}{2}\log(-2\theta_2).$$

The effective domain is $\Theta = \{(\theta_1, \theta_2) \in \mathbf{R}^2 : \theta_2 < 0\}$. Again, we have here an instance where it is not possible for the number of outstanding shares of the second security to exceed 0. However, an arbitrary amount of the securities can be sold short, which corresponds to increasing the variance of the market's estimate.

Example 3.3. As another example let the outcome space be the d-dimensional unit sphere. This setting was considered by who provide a cost function implicitly defined through a variational characterization. The maximum entropy approach leads to another alternative. We have a security for each of the d dimensions, and security i simply pays off $\phi_i(x) = x_i$, where $x \in \mathbf{R}^d$ is the unit-norm outcome. The maximum entropy distribution over the sphere with such sufficient statistics is the von Mises-Fisher distribution. The log partition function corresponds to the cost function

$$C(\theta) = I_{\frac{d}{2}-1}(||\theta||) - \left(\frac{d}{2} - 1\right)\log||\theta||,$$

where I_r refers to the modified Bessel function of first kind and order r. The effective domain of θ is the positive orthant in $\mathbf{R}^d$. The mean parametrization of the von Mises-Fisher distribution gives a generalized log scoring rule for the expected outcome components, but it is unwieldy and involves several special functions.

4. THE EXPONENTIAL FAMILY MARKET MECHANISM WITH BUDGETS

In the previous section, we saw that we can define a cost-function based prediction market so that the aggregated belief of the traders represents the maximum likelihood estimate of the natural parameters of the true exponential family distribution.

In this section, we consider the the prediction market setup with traders that may be either informative or malicious. The malicious traders may want to inject faulty information into the market. The informative traders on the other hand receive points drawn from the true distribution on which they base their beliefs.

We will show that if we are able to impose finite initial budgets on the traders and control the market prices based on these budgets, then it possible to set up the market so that it is prohibitive for damaging traders to participate in the market. Further, the informative traders can be shown to have expected growth in budget so that they are eventually able to move the market prices without restriction.

In this section, we assume that the traders have exponential family beliefs. The cost function has the same form as the log partition function T of the exponential family and the payoff is determined by the sufficient statistics of the data $\phi(x)$.

4.1. Budget-limited Aggregation

Imposing budget limits on the traders will allow us to control the amount of influence any one trader can have on moving the market prices. We will also satisfy an additional requirement that no trader has negative budget at any point of participation in the market. This is achieved by restricting the movement of the market and hence influencing the cost incurred by the trader. Recall that the payoff in this market is non-negative and hence the only adverse influence on a trader's budget is the cost of movement of the market state.

In this section, we assume that the budget of each trader is known to the market maker, and that the market maker can directly limit the allowed trades based on a trader's budget. Let α be the budget of a trader in the market. Suppose with infinite budget, the trader would have moved the market state from θ to $\hat{\theta}$, where $\hat{\theta}$ represents his true belief. Let C be the cost function. Now suppose further that $\alpha < C(\hat{\theta}) - C(\theta)$; that is the trader's budget does not allow for purchasing enough shares to move the market state to his belief. In this case, we want to budget-limit the trader's influence on the market state.

We define the budget-limited final market state as θ'. Here, we consider a specific functional form of θ':

$$\theta' = \lambda\hat{\theta} + (1-\lambda)\theta$$

where

$$\lambda = \min\left(1, \frac{\alpha}{C(\hat{\theta}) - C(\theta)}\right)$$

First, we show that this trade is feasible given the trader's budget:

THEOREM 4.1. *Let the current market state be given by θ. Let the final market state $\theta' = \lambda\hat{\theta}+(1-\lambda)\theta$ where $\lambda = \min\left(1, \frac{\alpha}{C(\hat{\theta})-C(\theta)}\right)$. The cost to the trader to move the market state from θ to θ' is at most his budget α.*

PROOF. From the convexity of C, we have

$$C(\theta') \leq (1-\lambda)C(\theta) + \lambda C(\hat{\theta})$$

Now

$$\begin{aligned} C(\theta') - C(\theta) &\leq (1-\lambda)C(\theta) \\ &\quad +\lambda C(\hat{\theta}) - C(\theta) \\ &= \lambda \left(C(\hat{\theta}) - C(\theta) \right) \end{aligned}$$

Thus, $C(\theta') - C(\theta) \leq \alpha$. □

We note that moving to θ' as defined may not be the optimal trade for a rational trader maximizing her expected profit. In general, the inequality above is strict, and so a trader does not fully exhaust her budget by moving to θ. Our results below will continue to hold in the case that strategic informative traders move to a position closer to their beliefs $\hat{\theta}$.

4.2. Damage Bound

In this section, we will quantify the error in prediction that the market maker might have to endure as a result of malicious entities entering the market. We assume that these malicious entities trade in multiple instances of the market; thus the exposure of the market maker is over several *rounds*. We use the standard log loss to measure this error in terms of the initial budget of traders.

We define the loss function for θ shares held:

$$L(\theta, x) = -\log(P_\theta(x)) = \log \int \exp\{\theta^T \phi(x)\}\, dx - \theta^T \phi(x) = C(\theta) - \theta^T \phi(x)$$

Suppose the prediction market runs over multiple rounds t. Let θ_0^t be the initial number of shares of each security that are held. Let $\hat{\theta}_k^t$ be the final values corresponding to the market state after the traders have made their reports. Let us assume that at this point the outcome is revealed; that is, we receive the value of the random variable x^t.

Over multiple instances of the prediction market, we can track the change in budget of each trader. Let the budget at rounds t and $t-1$ be α^t and α^{t-1} respectively. The change in budget for the trader is

$$\begin{aligned} \alpha^t - \alpha^{t-1} &= C(\theta^t) - C(\theta'^t) - (\theta^t - \theta'^t)^T \phi(x^t) \\ &= L(\theta, x^t) - L(\theta', x^t) \end{aligned}$$

Define the myopic impact of a trader i in segment t as

$$\Delta_i^t := L(\hat{\theta}_{i-1}^t, x^t) - L(\hat{\theta}_i^t, x^t)$$

Thus, the myopic impact captures incremental gain in prediction due to the trader in a round. Note that the myopic impact caused by trader i at round t is equal to the change in his budget in that round.

The total myopic impact due to all k active traders is given by

$$\Delta^t = L(\theta_0^t, x^t) - L(\hat{\theta}_k^t, x^t)$$

Thus $-\Delta^t$ captures the incremental loss of the market prediction after aggregation of all k traders.

THEOREM 4.2. *A coalition of b malicious traders can at most cause loss bounded by their initial budgets.*

PROOF. Consider the myopic impact of a single trader i after participating in the market T times. Since the market evolves so that the budget of any trader never falls below zero, the total myopic impact in T rounds caused due to trader i is:

$$\Delta_i := \sum_{t=1}^{T} \Delta_i^t = \sum_{t=1}^{T} (\alpha_i^t - \alpha_i^{t-1}) = \alpha_i^T - \alpha_i^0 \geq -\alpha_i^0$$

Thus, any coalition of b adversaries $\{1, \ldots, b\}$ can cause at most $\sum_{i=1}^{b} \alpha_i^0$ damage. □

This means that if it can be made prohibitively expensive for an attacker to generate clones, we can set up the prediction market with mostly informative traders.

In Section 4.3 we show that for an informative trader in every round, his budget increases in expectation. The intuition behind this is that a trader's prediction moves the input moves the market probability closer to the true probability distribution resulting in net expected profit.

4.3. Budget of Informative Traders

Given the information-theoretic interpretation of the cost-function based prediction market, we now show that the informative trader in the prediction market defined above increases his budget in a round in expectation *under his own belief distribution*.

We now characterize the expected change in budget for an informative trader. The following result holds for any round t; for simplicity, we have therefore dropped the superscript from the notation.

THEOREM 4.3. *Let θ be the current market state in the exponential family prediction market. Suppose that an informative trader with belief distribution parametrized by θ' moves the market state to the budget-limited state $\hat{\theta} = \lambda\theta' + (1-\lambda)\theta$. Then, the expectation (over the trader's belief) of the trader's profit is greater than zero whenever his budget is positive and his belief differs from the previous market position θ.*

PROOF. Let the cost function C be equal to the log partition function T of the belief distribution. The payoff is given by the sufficient statistics $\phi(x)$. Then, the trader's expected net payoff is given by

$$\begin{aligned}
& \mathbf{E}_{x \sim P_{\hat{\theta}}}[C(\theta) - C(\theta') - (\theta - \theta')\phi(x)] \\
= \ & T(\theta) - \theta \mathbf{E}_{x \sim P_{\hat{\theta}}}[\phi(x)] - (T(\theta') - \theta' \mathbf{E}_{x \sim P_{\hat{\theta}}}[\phi(x)]) \\
= \ & T(\theta) - \theta \nabla T(\hat{\theta}) - (T(\theta') - \theta' \nabla T(\hat{\theta})) \\
= \ & T(\theta) - T(\hat{\theta}) - \nabla T(\hat{\theta})(\theta - \hat{\theta}) - (T(\theta') - T(\hat{\theta}) - \nabla T(\hat{\theta})(\theta' - \hat{\theta})) \\
= \ & D_T(\theta, \hat{\theta}) - D_T(\theta', \hat{\theta}) \\
\geq \ & \lambda D_T(\theta, \hat{\theta}) \geq 0
\end{aligned}$$

The second to last inequality holds since $D_T(\theta', \hat{\theta})$ is convex in θ' and we have:

$$\begin{aligned}
D_T(\theta', \hat{\theta}) &= D_T\left(\lambda\hat{\theta} + (1-\lambda)\theta, \hat{\theta}\right) \\
&\leq \lambda D_T(\hat{\theta}, \hat{\theta}) + (1-\lambda) D_T(\theta, \hat{\theta}) \\
&= (1-\lambda) D_T(\theta, \hat{\theta})
\end{aligned}$$

Thus, a trader who moves the market state can expect his profit to be positive and at least $\lambda D_T(\theta, \hat{\theta})$. $\square$

For continuous distributions with a density, the probability that a trader with private information will form exactly the same beliefs as the current market position is 0, and thus, each trader will have positive expected profit on almost all sequences of observed samples and beliefs. This result suggests that, eventually, every informative trader will have the ability to influence the market state in accordance with his beliefs, without being budget limited.

Notice that Theorem 4.3 only required that the market state to which the trader moves, be representable as a convex combination of the current market state and his belief. This means that the result holds for exponential utility traders aiming to maximize their utility. In this case, the trader who moves the market state can expect his profit to be positive and at least $\frac{1}{a} D_T(\theta, \hat{\theta})$ where a is the exponential utility parameter.

We note one important aspect of Theorem 4.3: The expectation is taken with respect to each trader's belief at the time of trade, rather than with respect to the true distribution. This is needed because we have made no assumptions about the optimality of the traders' belief updating procedure. If we assume that the traders' belief formation is optimal, then this growth result will extend to the true distribution as well.

5. BAYESIAN TRADERS WITH LINEAR UTILITY

In this section, we lay the foundation for interpreting trader behavior in terms of Bayesian updates.

As before, we are interested in eliciting the sufficient statistics of the data. We assume that the outcome is drawn from an exponential family distribution; the prediction market is setup as before with the cost function corresponding to the log partition function and the payoff function corresponding to the sufficient statistics. Thus, the market state provides an estimate on the natural parameter of the distribution from which the outcome is drawn. Additionally, we assume that the market also makes public the total number of traders that have traded in the market.

The goal is to aggregate information from risk neutral agents who have a belief distribution over the natural parameters. This prior distribution is updated by the agents based on the current market state, They also each have to access to the empirical mean of sufficient statistics based on a fixed number m of data points. Assuming a conjugate prior, both the prior and posterior belief distributions on the natural parameters are also an exponential family distributions.

Now we will define the setup more formally.

Let the data distribution be given by $p(x;\beta) = \exp\{\beta \cdot \phi(x) - T(\beta)\}$ where $T()$ is the log partition function and $\phi()$ are the sufficient statistics. Then the conjugate prior parametrized by $b_0 = (n\nu, n)$ is given by $p(\beta; b_0) = \exp\{b_0 \cdot (\beta, \psi(\beta)) - \psi(b_0)\}$where $\psi()$ is the corresponding log partition function. The posterior distribution on the natural parameters is given by $p(\beta; b)$ where $b = (n\nu + m\hat{\mu}, n+m)$ where $\hat{\mu}$ is the empirical mean of the sufficient statistics of the m data points drawn from the data distribution; that is, $\hat{\mu} = \sum_{i=1}^{m} \phi(x_i)$. It turns out that

$$\mathbf{E}_{\beta \sim b_0} \mathbf{E}_{x \sim \beta}[\phi(x)] = \nu \tag{15}$$

Thus, it is helpful to think of the prior as being based on a 'phantom' sample of size n and mean ν. Thus the posterior mean is a convex combination of the prior and posterior means, and their relative weights depend on the phantom and empirical sample sizes.

Suppose the current market state is θ and i traders have traded in the market when trader $i+1$ with prior belief distribution $p(\beta; b_0^i)$ enters the market. Here $b_0^i = (n\nu_i, n)$. This trader also has access to private information in the form of empirical sufficient statistics $\hat{\mu}_i$ from m data points. Recall from Proposition 2.1 that natural parameter θ corresponds to expected statistics $\nabla T(\theta)$. Thus, he updates his belief as $p(\beta; b_i)$ where $b = (m\hat{\mu} + mi\nabla T(\theta) + n\nu_i, n + mi + m)$.

Suppose the trader wishes to maximize his expected payoff. Then the number of shares δ_i that he purchases when the current market state is θ is given by

$$\arg\max_{\delta_i} \mathbf{E}_{\beta\sim b_i} \mathbf{E}_{x\sim\beta} \left[\delta_i \phi(x) - T(\delta_i + \theta) + T(\theta)\right]$$

But, from Equation 15 we have $\mathbf{E}_{\beta\sim b}\mathbf{E}_{x\sim\beta}[\phi(x)] = \frac{n\nu + m\hat{\mu} + mi\nabla T(\theta)}{n+m(i+1)}$. To obtain the maximum, we set the gradient of the above expression with respect to δ_i to 0. Thus, we have for the optimal number of shares δ_i^*

$$\nabla T(\delta_i^* + \theta) = \frac{n\nu + m\hat{\mu} + mi\nabla T(\theta)}{n + m(i+1)}$$

Thus, from Proposition 2.1 we have that for an exponential family prediction market, the final market state is given by

$$\nabla G\left(\frac{n\nu + m\hat{\mu} + mi\nabla T(\theta)}{n + m(i+1)}\right)$$

where G is the convex conjugate of T. Thus, the final market state is a convex combination of prior, posterior and market means.

6. RISK AVERSE TRADERS WITH EXPONENTIAL UTILITY

In this section we relax that standard assumption that agents in the market are risk-neutral. We show that, with sufficient extra structure to the agents' beliefs and utilities, the market performs a clean aggregation of the agents' expectations in the form of a simple weighted averages. Assume that the agent has an exponential utility function for money w:

$$U_a(w) = -\frac{1}{a}\exp(-aw). \tag{16}$$

Here a controls the level of risk aversion: the agent is more risk averse as a increases, and as a tends to 0 we approach linear utility (risk-neutrality). Specifically, a is the Arrow-Pratt coefficient of absolute risk aversion, and exponential utilities of the form (16) are the unique utilities that exhibit constant absolute risk aversion (CARA).

If wealth is distributed according to a probability measure P, then the *certainty equivalent* of a random amount of wealth is defined as

$$CE(w) = U_a^{-1}\mathbf{E}_P\left[U_a(w)\right].$$

Suppose as before that the agent's belief over outcomes takes the form of a density p with respect to base measure ν. There is a close relationship between the log-partition function and the certainty equivalent under exponential utility.

LEMMA 6.1. *The certainty equivalent of the agent's expected profit, under exponential utility, when acquiring shares δ under a market state of θ is*

$$\log a - T_p(-a\delta) - aC(\theta + \delta) + aC(\theta), \tag{17}$$

where T_p is the log partition function (7) with a base measure of $p\,d\nu$. Furthermore, if the agent's belief is an exponential family with parameter $\hat{\theta}$, we have

$$T_p(\delta) = T(\hat{\theta} + \delta) - T(\hat{\theta}),$$

where T is the usual log partition function with base measure ν.

PROOF. Explicitly, the certainty equivalent of the profit is

$$\begin{aligned}
& CE(\,\langle \delta, \phi(x)\rangle - [C(\theta+\delta) - C(\theta)]\,) \\
= \; & -\log \int_{\mathcal{X}} \frac{1}{a} \exp\left(\langle -a\delta, \phi(x)\rangle + a[C(\theta+\delta) - C(\theta)]\right) p(x) d\nu(x) \\
= \; & \log a - a[C(\theta+\delta) - C(\theta)] - \log \int_{\mathcal{X}} \exp\left(\langle -a\delta, \phi(x)\rangle\right) p(x) d\nu(x) \\
= \; & \log a - a[C(\theta+\delta) - C(\theta)] - T_p(-a\delta).
\end{aligned}$$

For the second part of the result, we have

$$\begin{aligned}
T_p(\delta) \; &= \; \log \int_{\mathcal{X}} \exp(\langle \delta, \phi(x)\rangle)\, p(x;\hat{\theta})\, d\nu(x) \\
&= \; \log \int_{\mathcal{X}} \exp(\langle \delta + \hat{\theta}, \phi(x)\rangle - T(\hat{\theta}))\, d\nu(x) \\
&= \; T(\hat{\theta} + \delta) - T(\theta) + \log \int_{\mathcal{X}} \exp(\langle \delta + \hat{\theta}, \phi(x)\rangle - T(\hat{\theta} + \delta))\, d\nu(x) \\
&= \; T(\hat{\theta} + \delta) - T(\theta) + \log \int_{\mathcal{X}} p(x; \hat{\theta} + \delta)\, d\nu(x) \\
&= \; T(\hat{\theta} + \delta) - T(\theta),
\end{aligned}$$

where the last line follows from the fact that density $p(x; \hat{\theta} + \delta)$ integrates to 1. □

Recall that for the generalized LMSR, the cost function C is exactly the log partition function T. We are therefore lead to the following clear understanding of a risk-averse agent's behavior in such a market.

THEOREM 6.2. *Suppose an agent has exponential utility with coefficient a and exponential family beliefs with natural parameter $\hat{\theta}$. In the generalized LMSR market with current market state θ, the agent's optimal trade δ moves the state vector to*

$$\theta + \delta = \frac{1}{1+a}\hat{\theta} + \frac{a}{1+a}\theta. \tag{18}$$

PROOF. The agent's optimal trade maximizes its expected utility, or equivalently the certainty equivalent. From Lemma 6.1 and the fact that $T = C$, the agent therefore maximizes

$$\log a - T(\hat{\theta} - a\delta) + T(\hat{\theta}) - aT(\theta + \delta) + aT(\theta).$$

This objective is strictly concave, from the strict convexity of T. The optimum is therefore characterized by the first-order conditions:

$$\nabla T(\hat{\theta} - a\delta) = \nabla T(\theta + \delta).$$

As the gradient map ∇T is one-to-one, this is solved by equating the arguments, which leads to $\delta = (\hat{\theta} - \theta)/(1 + a)$ and (18). $\square$

Note that as, a tends to 0, we approach risk neutrality and the agent moves the share vector all the way to its private estimate $\hat{\theta}$. As a grows larger (agent grows more risk averse) the agent makes smaller trades to reduce it exposure, and the final state stays closer to the current state θ. Update (18) implies that under the conditions of the theorem, a market that receives a sequence of myopic traders aggregates their natural parameters in the form of an exponentially weighted moving average. The final market estimates (i.e., prices) are obtained by applying ∇T to this average.

7. REINTERPRETING REPEATED TRADES

In previous sections we have analyzed trader behavior as if it is his first entry into the market. In this section we will quantify how prior exposure in the market affects a trader's choices. In particular, we will show that a trader who has previously purchased shares in a market will, on subsequent entry, behave as if this purchase has updated his private belief. This result implies that any financial trade made in a market is equivalent to changing the trader's effective belief.

As before, we consider the exponential family prediction market where traders have exponential belief. Let us also suppose that traders in this market have exponential utility

$$U(w) = -\frac{1}{a}\exp(-aw).$$

Here a is the coefficient of risk aversion (higher means more risk averse, and the utility function is more concave).

Suppose an agent has exponential family belief parametrized by natural parameter $\hat{\theta}$. Based on this belief, let δ_1^* be the optimal vector of shares the agent decides to trade on first entering the market. Thus, his belief distribution is given by the density

$$p(x; \hat{\theta}) = \exp(\langle \hat{\theta}, \phi(x) \rangle - T(\hat{\theta}))$$

where $T(\hat{\theta}) = \int_{\mathcal{X}} \exp\{\langle \hat{\theta}, \phi(x) \rangle\}\, dx$ is the log partition function and $\phi(x)$ are the sufficient statistics of the outcome $x \in \mathcal{X}$. On a subsequent entry into this market with market state θ', his optimal purchase δ_2^* is given by the solution of

$$\begin{aligned}
& \arg\max_{\delta_2} \mathbf{E}_{x \sim p(x;\hat{\theta})} U\left[(\delta_1^* + \delta_2)\phi(x) - C(\delta_1^* + \theta) + C(\theta) - C(\delta_2 + \theta') + C(\theta')\right] \\
= \;& \arg\max_{\delta_2} \int_{\mathcal{X}} -\frac{1}{a}[\exp(-a(\delta_1^* + \delta_2)\phi(x) \\
& + aC(\delta_1^* + \theta) - aC(\theta) + aC(\delta_2 + \theta') - aC(\theta'))\exp\{\hat{\theta}\phi(x) - T(\hat{\theta})\}]\, dx \\
= \;& \arg\max_{\delta_2} \int_{\mathcal{X}} -\frac{1}{a}[\exp\{-a\delta_2\phi(x) + aC(\delta_2 + \theta') - aC(\theta') \\
& + aC(\delta_1^* + \theta) - aC(\theta)\}\exp\{(\hat{\theta} - a\delta_1^*)\phi(x) - T(\hat{\theta})\}]\, dx \\
= \;& \arg\max_{\delta_2} \int_{\mathcal{X}} U[N(\delta_2, \theta')]\exp\{(\hat{\theta} - a\delta_1^*)\phi(x) - T(\hat{\theta}) + aC(\delta_1^* + \theta) - aC(\theta)\}]\, dx
\end{aligned}$$

Here, the first equality follows from the fact that we are taking expectation over the agent's belief parameter $\hat{\theta}$ and the second equality follows simply from rearranging the factors of $\phi(x)$. And lastly we have written $-\frac{1}{a}[\exp\{-a\delta_2\phi(x) + aC(\delta_2 + \theta') - aC(\theta')]$ as $U[N(\delta_2, \theta')]$ where $N(\delta_2, \theta')$ is the net payoff when δ_2 shares are purchased when the current market state is θ'.

Note that since $C = T$, $-T(\hat{\theta}) + aC(\delta_1^* + \theta) - aC(\theta)$ is proportional to $T(\hat{\theta} - a\delta_1^*)$, it follows that $\int_{\mathcal{X}} \exp\{(\hat{\theta} - a\delta_1^*)\phi(x) - T(\hat{\theta}) + aC(\delta_1^* + \theta) - aC(\theta)\}]\, dx$ is proportional to the density with parameter $\hat{\theta} - a\delta_1^*$. Thus, picking the optimal share vector is equivalent to maximizing expected utility of $N(\delta_2, \theta')$, where expectation is taken with respect to an exponential family distribution over $\mathcal{X}$ parametrized by $\hat{\theta} - a\delta_1^*$.

Let $\Theta \stackrel{\text{def}}{=} \hat{\theta} - a\delta_1^*$ be the effective belief. Thus we have that the trader chooses his share vector as follows.

$$\begin{aligned}
&\arg\max_{\delta_2} \int_{\mathcal{X}} U[N(\delta_2, \theta')] \exp\{(\hat{\theta} - a\delta_1^*)\phi(x) - T(\hat{\theta}) + aC(\delta_1^* + \theta) - aC(\theta)\}\, dx \\
=\ &\arg\max_{\delta_2} \int_{\mathcal{X}} U[N(\delta_2, \theta')] \exp\{(\hat{\theta} - a\delta_1^*)\phi(x) - T(\hat{\theta} - a\delta_1^*)\}]\, dx \\
=\ &\arg\max_{\delta_2} \int_{\mathcal{X}} -\frac{1}{a} \exp\{-a\delta_2\phi(x) + aC(\delta_2 + \theta') - aC(\theta')\} \exp\{\Theta \cdot \phi(x) - T(\Theta)\}\, dx \\
=\ &\arg\max_{\delta_2} U\left[-C(\delta_2 + \theta') + C(\theta') - \frac{1}{a}T(\Theta - a\delta_2) + \frac{1}{a}T(\Theta)\right]
\end{aligned}$$

This follows from Proposition **??**. The maximizer is $\delta_2^* = (\Theta - \theta')/(1 + a)$, which moves the share vector to $\theta' + \delta_2^* = \frac{1}{1+a}\Theta + \frac{a}{1+a}\theta'$ which is a convex combination of the effective belief and the current market state.

In other words, an exponential utility maximizing trader who has belief $\hat{\theta}$ with prior exposure δ in a market will behave identically to an exponential utility maximizing trader with belief $\hat{\theta} - a\delta$ and no prior exposure in the market. Here a is the utility parameter. This means that financial exposure can be equivalently understood as changing the privately held beliefs.

8. EQUILIBRIUM MARKET STATE FOR EXPONENTIAL UTILITY AGENTS

We have shown that every exponential-utility maximizing trader picks the share vector δ so that the eventual market state can be represented as a convex combination of the initial market state and the natural parameter of his (exponential family) belief distribution. In this section we will compute the equilibrium state in an exponential family market with multiple exponential-utility maximizing traders.

Recall the following result from game theory.

THEOREM 8.1. *Let $U_i(\vec{\delta})$ be the utility function of the i^{th} trader given strategies $\vec{\delta} \stackrel{\text{def}}{=} \delta_1, \ldots, \delta_i, \ldots, \delta_n$. If there exists a potential function $g(\vec{\delta})$ such that*

$$U_i(\vec{\delta}) - U_i(\vec{\delta}_{-i}, \delta_i') = g(\vec{\delta}) - g(\vec{\delta}_{-i}, \delta_i')$$

then when $g(\vec{\delta})$ is maximized, $\vec{\delta}$ is an equilibrium.

In the exponential family market, the cost function C is identical to the log partition function T. Let $\vec{\delta}$ be the vector of vectors of shares purchased by every trader in the market when the market has reached equilibrium, θ the initial market state, $\hat{\theta}_i$ the natural parameter of trader i's belief distribution and a_i his utility parameter.

Define a potential function as

$$g(\vec{\delta}) \stackrel{\text{def}}{=} T(\theta + \sum_i \delta_i) + \sum_i \frac{1}{a_i} T(\hat{\theta}_i - a_i\delta_i)$$

Now the utility of trader i is $U_i(\vec{\delta}) = -T(\theta + \sum_j \delta_j) + T(\theta + \sum_{j\neq i} \delta_j) - \frac{1}{a_i}T(\hat{\theta}_i - a_i\delta_i) + \frac{1}{a_i}T(\hat{\theta}_i)$. Thus, Theorem 8.1 applies and we can find the equilibrium market state by maximizing $g(\vec{\delta})$ for each δ_i.

$$\begin{aligned}\nabla_{\delta_i} g(\vec{\delta}) &= \nabla T(\theta + \sum_{j=1}^{n} \delta_i) - \nabla T(\hat{\theta}_i - a_i\delta_i) \\ &= 0\end{aligned}$$

This can be achieved by equating the arguments. That is, for each trader i,

$$\hat{\theta}_i - a_i\delta_i = \theta + \sum_{j=1}^{n} \delta_j \tag{19}$$

Rewriting, we have for each trader i,

$$\frac{\hat{\theta}_i}{a_i} - \delta_i = \frac{1}{a_i}(\theta + \sum_{j=1}^{n} \delta_j)$$

Thus,

$$\sum_{i=1}^{n} \left(\frac{\hat{\theta}_i}{a_i}\right) - \sum_{i=1}^{n} \delta_i = \left(\theta + \sum_{j=1}^{n} \delta_j\right) \sum_{i=1}^{n} \frac{1}{a_i}$$

And

$$\sum_{j=1}^{n} \delta_j = \frac{\sum_{i=1}^{n} \left(\frac{\hat{\theta}_i}{a_i}\right) - \theta \sum_{i=1}^{n} \left(\frac{1}{a_i}\right)}{1 + \sum_{i=1}^{n} \frac{1}{a_i}}$$

Substituting in Equation 19 we have the following expression for the final market state.

$$\theta + \sum_{j=1}^{n} \delta_j = \frac{\theta + \sum_{i=1}^{n} \left(\frac{\hat{\theta}_i}{a_i}\right)}{1 + \sum_{i=1}^{n} \frac{1}{a_i}}$$

Thus the equilibrium state is a convex combination of trader beliefs and initial market state.

9. CONCLUSIONS

REFERENCES

ABERNETHY, J., CHEN, Y., AND WORTMAN VAUGHAN, J. 2011. An optimization-based framework for automated market-making. In *Proceedings of the 12th ACM conference on Electronic commerce*. EC '11. ACM, New York, NY, USA, 297–306.

BEYGELZIMER, A., LANGFORD, J., AND PENNOCK, D. M. 2012. Learning performance of prediction markets with kelly bettors. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 3*. AAMAS '12. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1317–1318.

CHEN, Y., GOEL, S., AND PENNOCK, D. M. 2008. Pricing combinatorial markets for tournaments. In *Proceedings of the 40th annual ACM symposium on Theory of computing*. STOC '08. ACM, New York, NY, USA, 305–314.

CHEN, Y. AND VAUGHAN, J. W. 2010. A new understanding of prediction markets via no-regret learning. In *Proceedings of the 11th ACM conference on Electronic commerce*. EC '10. ACM, New York, NY, USA, 189–198.

GAO, X., CHEN, Y., AND PENNOCK, D. M. 2009. Betting on the real line. In *Proceedings of the 5th International Workshop on Internet and Network Economics*. WINE '09. Springer-Verlag, Berlin, Heidelberg, 553–560.

HANSON, R. 2003. Combinatorial information market design. *Information Systems Frontiers 5,* 1, 107–119.

KALAI, A. AND VEMPALA, S. 2005. Efficient algorithms for online decision problems. *J. Comput. Syst. Sci. 71*, 291–307.

OTHMAN, A. AND SANDHOLM, T. 2011. Automated market makers that enable new settings: Extending constant-utility cost functions. In *Proceedings of the Second Conference on Auctions, Market Mechanisms and Their Applications (AMMA)*. 19–30.

PENNOCK, D. AND XIA, L. 2011. Price updating in combinatorial prediction markets with bayesian networks. In *Proceedings of the Twenty-Seventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-11)*. AUAI Press, Corvallis, Oregon, 581–588.