X

Exponential Family Markets for Information Aggregation

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

[JA: I will take care of this.] 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: Algorithms, Economics

Additional Key Words and Phrases: logarithmic score, exponential family, maximum entropy, risk aversion, budget constraints

ACM Reference Format:
Jacob Abernethy, Sindhu Kutty, Sébastien Lahaie, and Rahul Sami, 2014. Maximum Entropy Prediction Markets *ACM* X, X, Article X (February 2014), 19 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

[JA: This is a comment from jake] [SK: This is a comment from sindhu] [SL: This is a comment from sebastien] 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

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

prediction on the outcome of the event. Traders are allowed to report their beliefs on the outcome of the event 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 event occurring.

[JA: 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 of 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, e.g. *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 via conceptual objects used to define 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.

We begin with a discussion of scoring rules in Section 2 for exponential family distributions, and in particular we show that the logarithmic scoring rule is the only score that is *proper* for the class of exponential families. We turn our attention to market design in Section 3 and give a full description of our proposed mechanisms. In addition to showing the 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:

— Trader behavior varies depending on how they assimilate information; for example, should we consider our agents as Bayesians or frequentists. In Section 4 we consider traders that use a conjugate prior to update their beliefs, and we study how their trades would affect the market state. [JA: Do we show the following?] 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.

— In Section 5 we consider when our agents are risk-averse in an interesting special case, that is under the assumption they utilize *exponential utility* to optimize their bets. We show how this dovetails nicely with the notion of a *certainty equivalent* which is often used in risk-aversion settings. In this case we can characterize precisely how a single trader interacts with the market as well as the equilibrium reached given multiple traders; this result is achieved via using a *potential game* argument. 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 Section 6 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.

[JA: I want to add more to this list. It's ok but doesn't totally reflect what's in the last few sections.]

1.1. Related Work

[JA: I'll add more here. Sebastien: do you have anything to put in here regarding the exponential utility stuff?] Prediction markets have been known to perform well in practice (see, for instance, [Pennock and Sami 2007; Wolfers and Zitzewitz 2004; Pennock et al. 2001]). However, a sound theory for interpreting trader behavior and market prices is an ongoing area of study [Wolfers and Zitzewitz 2006].

Scoring rules are a measure of merit of a predictor [Gneiting and Raftery 2007]. [Hanson 2003] adapted scoring rules to markets with sequential participation of multiple traders. A cost function based formulation of markets was first introduced by [Chen and Pennock 2007].

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. 2013] 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.

In this paper, we use exponential family distributions to construct our cost function. For a comprehensive introduction to the class of exponential family distributions see

[Wainwright and Jordan 2008]. [Azoury and Warmuth 2001] consider learning exponential family distributions in a traditional online learning model and provide worst case loss bounds relative to using an offline algorithm. [Storkey 2011] constructs a prediction market where each participating agent is a learning algorithm with varying utility functions.

[SL: Need to talk about certainty equivalents somewhere]

2. GENERALIZED LOG SCORING RULES

We consider a measurable space consisting 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. Let $\mathcal{P}$ denote the set of all such probability densities.

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 $x \in \mathcal{X}$. The following definition is due to Lambert et al. [2008].

Definition 2.1. A scoring rule S is *proper* for statistic ϕ if for each $\mu \in \mathcal{M}$ and $p \in \mathcal{P}$ 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 $\hat{\mu} \neq \mu$.

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* [Dawid 1998; Gneiting and Raftery 2007]. Throughout we will implicitly 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.

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 the definition allows for summary information of the belief.

Note that Definition 2.1 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 with the support of the density as implicitly defined by base measure 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.

When the outcome space is finite we can recover classical scoring rules from the definition by 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 Maximum Entropy

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 $S(p, x) = \log p(x)$, namely we take the log likelihood of the reported density at the eventual outcome. To generalize the rule to expected statistics rather than full densities, we consider a subset of densities $\mathcal{D} \subseteq \mathcal{P}$. 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. Given such a family parametrized by the relevant statistics, the generalized log scoring rule is then

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

Even though the log score is only applied to densities from $\mathcal{D}$, according to Definition 2.1 it must work over all densities in $\mathcal{P}$. It turns out this is possible if $\mathcal{D}$ is chosen appropriately, drawing on a well-known duality between maximum likelihood and maximum entropy [Grünwald and Dawid 2004].

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]

$$\min_{p \in \mathcal{P}} F(p) \quad \text{s.t.} \quad \mathbf{E}_p[\phi(x)] = \mu, \tag{3}$$

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 (3) are linear, whereas the objective is convex. We let $G : \mathcal{M} \to \mathbf{R}$ be the optimal value function of (3), meaning $G(\mu)$ is the negative entropy of the maximum entropy distribution with expected statistics μ.

It is well-known that solutions to (3) are *exponential family* distributions, whose densities with respect to ν take the form

$$p(x; \theta) = \exp(\langle \theta, \phi(x) \rangle - T(\theta)). \tag{4}$$

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 the linear constraints in (3). The term $T(\theta)$ essentially arises as the multiplier for the normalization constraint (the density must integrate to 1), and so ensures that (4) is normalized:

$$T(\theta) = \log \int_{\mathcal{X}} \exp\langle \theta, \phi(x) \rangle \, d\nu(x). \tag{5}$$

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

[2]We assume 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.

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; proofs may be found in Wainwright and Jordan [2008, Prop. 3.1–3.2, Thm. 3.3–3.4] and see also Banerjee et al. [2005a, Lem. 1, Thm. 2]. 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 $\mu = \mathbf{E}_p[\phi(x)] = \nabla T(\theta)$.*

(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 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.

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 $\hat{\mu} \in \mathcal{M}$ be an alternative report. Then note from (4) that

$$\mathbf{E}_{\tilde{p}}[\log p(x; \hat{\mu})] = \mathbf{E}_{p(\cdot;\mu)}[\log p(x; \hat{\mu})] = \langle \hat{\theta}, \mu \rangle - T(\hat{\theta}), \tag{6}$$

where $\hat{\theta} = \nabla G(\hat{\mu})$ is the natural parameter for the exponential family with statistic $\hat{\mu}$. 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 2.2. *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 if and only if $\mathcal{D}$ is the exponential family with statistic ϕ.*

PROOF. Let $\mu, \hat{\mu} \in \mathcal{M}$ be the agent's true belief and an alternative report, and let $p \in \mathcal{P}$ be a density consistent with μ. Let $\theta = \nabla G(\mu)$ and $\hat{\theta} = \nabla G(\hat{\mu})$, and note that

$\mu = \nabla T(\theta)$. We have

$$\begin{aligned} & \mathbf{E}_p[\log p(x;\mu)] - \mathbf{E}_p[\log p(x;\hat{\mu})] \\ = & \ \langle \theta, \mu \rangle - T(\theta) - \langle \hat{\theta}, \mu \rangle + T(\hat{\theta}) \\ = & \ T(\hat{\theta}) - T(\theta) - \langle \hat{\theta} - \theta, \mu \rangle \\ = & \ T(\hat{\theta}) - T(\theta) - \langle \hat{\theta} - \theta, \nabla T(\theta) \rangle. \end{aligned} \tag{7}$$

The latter is positive by the strict convexity of T, which shows that the log score is proper. For the converse, assume the defined log score is proper. By the Savage characterization of proper scoring rules for expectations (see Gneiting and Raftery [2007, Thm. 1] and Savage [1971]), 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. □

As further intuition for the result, note that (7) is the definition of the 'Bregman divergence' with respect to strictly convex function T, written D_T. Therefore we have

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

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.

2.2. Examples: Moments over the Real Line

Theorem 2.2 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 this distribution. 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 domains (multi-dimensional included) and expectations of interest, it leads to novel closed-form scoring rules. By examining the log densities of various exponential families, we can for instance obtain scoring rules for several different combinations of the arithmetic, geometric, and harmonic means, as well as higher order moments. The following examples illustrate the construction.

Example 2.3. 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

$$S(\mu, x) = -\frac{x}{\mu} - \log \mu. \tag{8}$$

We stress that although this rule is derived from the exponential distribution, Theorem 2.2 implies that it elicits the mean of any distribution supported on the non-negative reals (e.g., Pareto, lognormal). Indeed, it is easy to see that the expected score (8) 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:

$$S(\mu, x) = (k-1)\log x - k \log \mu - \Gamma\left(1 + \frac{1}{k}\right)^k \left(\frac{x}{\mu}\right)^k, \tag{9}$$

where Γ denotes the gamma function (the extension of the factorial to the reals). We have not found either scoring rule (8) or (9) in the literature.

Example 2.4. 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

$$S((\mu, \sigma^2), x) = -\frac{(x-\mu)^2}{\sigma^2} - \log \sigma^2. \tag{10}$$

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:

$$S((\mu, \Sigma), x) = -(x-\mu)'\Sigma^{-1}(x-\mu) - \log|\Sigma|. \tag{11}$$

Here the statistics being elicited are the mean vector μ and the covariance matrix Σ. These scoring rules have been studied by Dawid and Sebastiani [1999] 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 2.4 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 PREDICTION MARKETS VIA EXPONENTIAL FAMILIES

In a single-agent setting, a scoring rule is used to *elicit* the agent's belief. In a multi-agent setting, a prediction market can be used to *aggregate* the beliefs of the agents. In

[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 [Lambert et al. 2008].

his seminal paper Hanson [2003] 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.

3.1. Prediction Market

In a prediction 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 2.3 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 a prediction market in the literature, due to Chen and Pennock [2007], 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 δ, moving the market state vector to $\theta' = \theta + \delta$, then we must have

$$\mathbf{E}_p[\phi(x)] = \nabla C(\theta'). \tag{12}$$

[SK: sebastien/jake, should this be $\mathbf{E}_p(x;\theta')[\phi(x)]$?] 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):

$$C(\theta) = \log \int_{x \in \mathcal{X}} \exp\langle \theta, \phi(x) \rangle \, d\nu(x). \tag{13}$$

This is of course exactly the log-partition function (5) 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 (13) and the log-partition function (5) 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 [Abernethy et al. 2013; Hanson 2003]. To see this in our context, recall from (6) that

$$\mathbf{E}_{\tilde{p}}[\log p(x;\hat{\mu})] = \langle \hat{\theta}, \mu \rangle - T(\hat{\theta})$$

where $\tilde{p}$ is the agent's belief and $\hat{\mu}$ the agent's report. The expected log score from reporting $\hat{\mu}$ is exactly the same as the expected payoff from buying portfolio of shares $\hat{\theta} = \nabla C(\hat{\mu})$ (assuming an initial market state of 0), as $\langle \hat{\theta}, \mu \rangle$ is the expected revenue and $T(\hat{\theta})$ is the cost. As in Section 2 this reasoning relies on the assumption of risk-neutrality, 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 (7). Now Banerjee et al. [2005b] 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 [Hanson 2007].

3.3. Examples: Real Line and the Sphere

Let us now revisit our scoring rule 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 $\mathbf{R}_+$ and set up a market for the expected outcome, consisting of a single security that pays off $\phi(x) = x$. The log partition function 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 of 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 Abernethy et al. [2013] 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; see Banerjee et al. [2005b] for an explanation of these quantities. 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. BAYESIAN TRADERS WITH LINEAR UTILITY

In the standard model of cost-function based prediction markets, a sequence of myopic, risk-neutral agents arrive and trade in the market [Chen and Pennock 2007; Chen and Vaughan 2010]. As we saw in Section 3.1, such a trader moves the prices to its own expectation μ. However, this means that the market does not perform any meaningful aggregation of the agent's belief, as the final prices are simply the final agent's expectation.

In this section we examine the aggregation behavior of the market when agents are Bayesian and take into account the current market state when forming their beliefs. This requires more structure to their beliefs. For this section and the remainder of the paper, we will assume that agents have *exponential family beliefs*.

The exponential families framework is well-suited to reasoning about Bayesian updates. As before let the data distribution be given by $p(x;\theta) = \exp(\langle\theta, \phi(x)\rangle - T(\theta))$ where T is the log partition function and ϕ are the sufficient statistics. Instead of direct beliefs about the data distribution the agent maintains a conjugate prior over the parameters θ. Every exponential family admits a conjugate prior of the form

$$p(\theta; b_0) = \exp(\langle n\nu, \theta\rangle + nT(\theta) - \psi(\nu, n)).$$

Note that this is also an exponential family with natural parameter $b_0 = (n\nu, n)$ where $\nu \in \mathbf{R}^d$ and n is a positive integer. The sufficient statistic maps θ to $(\theta, T(\theta))$, and the log partition function ψ is defined as the normalizer as usual. For a complete treatment of exponential families conjugate priors, see for instance Barndorff-Nielsen

[1978]. Now Diaconis and Ylvisaker [1979, Thm. 2] and Jewell [1974] have shown that

$$\mathbf{E}_{\theta\sim b_0}\mathbf{E}_{x\sim\theta}[\phi(x)] = \nu, \tag{14}$$

[SK: Maybe this is nitpicking but $\mathbf{E}_{\theta\sim p(\theta;b_0)}\mathbf{E}_{x\sim p(x;\theta)}[\phi(x)]$?] meaning that $\nu = n\nu/n$ is the posterior mean. Thus, it is helpful to think of the prior as being based on a 'phantom' sample of size n and mean ν. Suppose now that the agent observes an empirical sample with mean $\hat{\mu}$ and size m. By a standard derivation [see Diaconis and Ylvisaker 1979], the posterior conjugate prior parameters become $n\nu \leftarrow n\nu + m\hat{\mu}$ and $n \leftarrow n + m$, and the posterior expectation (14) evaluates to

$$\frac{n\nu + m\hat{\mu}}{n+m}. \tag{15}$$

Thus the posterior mean is a convex combination of the prior and empirical means, and their relative weights depend on the phantom and empirical sample sizes.

Consider Bayesian agents maintaining an exponential family conjugate prior over the data model's natural parameters (equivalently, the expected security payoffs). Each agent has access to a private sample of the data of size m with mean statistic $\hat{\mu}$. If n agents have arrived before to trade, then the current market prices μ correspond to the phantom sample, and the phantom sample size is nm. After forming the posterior (15) with these substitutions, the (risk-neutral) agent purchases shares δ to move the current market share vector to

$$\nabla C(\theta + \delta) = \frac{n\nu + \hat{\mu}}{n+1}.$$

As a result, the final market prices under this behavior are a simple average of the agent's mean parameters and the initial market prices. We note that to facilitate such belief updating, the market should post the number of trades since initialization.

5. RISK-AVERSE TRADERS WITH EXPONENTIAL UTILITY

[SK: Sebastien, do you think we could cite [Norstad 1999] for instance?]

In this section we relax the 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 average. Assume that the agent has an exponential utility function for wealth 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 [see Ben-Tal and Teboulle 2007].

LEMMA 5.1. *The certainty equivalent of the agent's expected profit, with 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 (5) with a base measure of $p\,d\nu$. Furthermore, if the agent's belief is an exponential family with natural 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

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

For the second part of the result, we have

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

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 understanding of a risk-averse agent's behavior in such a market.

THEOREM 5.2. *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 5.1 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 (the 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.

Liquidity Adjustment. In practice the centralized market maker allows itself some control over the *liquidity* in the market, which captures how responsive prices are to trades. To adjust liquidity we consider the parametrized cost $C_\lambda(\theta) = \frac{1}{\lambda} C(\lambda\theta)$. Here λ is construed as the *inverse liquidity*, or price responsiveness. A larger setting of λ means fewer shares need to be bought to reach the same prices.[4]

In the context of the generalized LMSR we write T rather than C where T is the log partition function, with liquidity-adjusted version T_λ. Let $\hat{\mu}$ be the agent's mean belief with corresponding natural parameter $\hat{\theta} = \nabla T^{-1}(\hat{\mu})$. Recall that a risk-neutral agent moves the share vector so that the prices match its mean parameter. Therefore, define the *target shares* as $\tilde{\theta} = \nabla T_\lambda^{-1}(\hat{\mu})$. The target shares $\hat{\theta}$ and natural parameter $\tilde{\theta}$ are related by

$$\nabla T(\hat{\theta}) = \nabla T_\lambda(\tilde{\theta}) = \hat{\mu}.$$

It is straightforward to check that $\nabla T_\lambda(\tilde{\theta}) = \nabla T(\lambda\tilde{\theta})$ so we have

$$\tilde{\theta} = \hat{\theta}/\lambda. \tag{19}$$

Higher price responsiveness means fewer shares must be bought to make the market prices match the agent's expectation, so the natural parameter is scaled down accordingly. With a liquidity adjustment, the analysis of Theorem 5.2 can be extended and yields the following result.

COROLLARY 5.3. *[JA: μ without a hat is never defined... (this is for sebastien)] Under the conditions of Theorem 5.2 and an inverse liquidity of λ, the agent's optimal trade δ moves the state vector to*

$$\theta + \delta \;=\; \frac{\lambda}{\lambda + a}\tilde{\theta} + \frac{a}{\lambda + a}\theta \;=\; \frac{\lambda}{\lambda + a}\nabla T_\lambda^{-1}(\hat{\mu}) + \frac{a}{\lambda + a}\nabla T_\lambda^{-1}(\mu). \tag{20}$$

According to (20), as λ grows large the agent moves the market state closer to the *target shares*, rather than its true natural parameter. Note that the target shares themselves depend on λ by (19), but the update can be directly written in terms of the agent's beliefs as in the right-hand side of (20).

5.1. Repeated Trading and the Effective Belief

In previous sections we have analyzed trader behavior as if it is his first entry into the market. We now pose the question: how will a trader reason about a possible future

[4]The liquidity adjustment to the cost function takes the same form as the risk-aversion adjustment to the exponential utility in (16). In convex analysis, this transformation is known as the perspective function [Hiriart-Urruty and Lemaréchal 2000, p. 90].

investment when the trader holds an *existing portfolio?* When the agent is exposed to the market given a prior investment this should clearly affect the agent's preference for additional risk. In the context of a trader possessing an exponential family belief and with exponential utility we show that we can analyze explicitly how an agent incorporates an existing portfolio. Indeed, the key conclusion we make in this section is that a trader will reason about a future investment simply *as though* he had simply updated his belief and he had no prior investment.

Suppose an exponential utility agent has exponential family belief parametrized by natural parameter $\hat{\theta}$. Based on this belief, let δ_1 be the vector of shares the agent has purchased on first entry in 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

$$\arg\max_{\delta_2} \mathbf{E}_{x\sim p(x;\hat{\theta})} U\left[(\delta_1+\delta_2)\phi(x) - C(\delta_1+\theta) + C(\theta) - C(\delta_2+\theta') + C(\theta')\right]$$

Then if $\Theta \stackrel{\text{def}}{=} \hat{\theta} - a\delta_1$ is the effective belief, the trader optimal purchase is given by $\delta_2^* = (\Theta - \theta')/(1+a)$, 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.

THEOREM 5.4. *Suppose an exponential utility maximizing trader with utility parameter a who has belief $\hat{\theta}$ makes a purchase δ in a market. On subsequently re-entering this market, he will behave identically to an exponential utility maximizing trader with belief $\hat{\theta} - a\delta$ and no prior exposure in the market.*

Theorem 5.4 implies that financial exposure can be equivalently understood as changing the privately held beliefs.

5.2. 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.

We now review a well-known result from game theory regarding the class of *potential games* [Monderer and Shapley 1996]. Note that we say a function $f(\vec{x})$ is at a *local optimum* if changing any coordinate of $\vec{x}$ does not increase the value of f.

THEOREM 5.5. *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 $\Phi(\vec{\delta})$ such that*

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

then $\vec{\delta}$ is a Nash equilibrium if and only if $\Phi(\vec{\delta})$ is at a local optimum.

In the exponential family market, the cost function C is identical to the log partition function T defined in (5). Let $\vec{\delta}$ be the matrix of share vectors 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 $\Phi(\vec{\delta}) \stackrel{\text{def}}{=} T\left(\theta + \sum_i \delta_i\right) + \sum_i \frac{1}{a_i} T(\hat{\theta}_i - a_i\delta_i)$.

Rather than work directly with the utilities of every trader, we will work with the log of their utility values[5]. Now the log-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 5.5 applies and we can find the equilibrium market state by maximizing $\Phi(\vec{\delta})$ for each δ_i.

$$\nabla_{\delta_i} \Phi(\vec{\delta}) \;=\; \nabla T\left(\theta + \sum_{j=1}^{n} \delta_i\right) - \nabla T(\hat{\theta}_i - a_i\delta_i) = 0$$

And we have the following expression for the final market state.

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

The equilibrium state is a convex combination of all beliefs and initial market state.

6. BUDGET-LIMITED AGGREGATION

In this section, we consider the evolution of the market state when traders are budget-limited and their budgets are known to the market maker. We consider a cost-function based market with cost function C and payoff $\phi(x)$ with non-negative payoff. We assume that the traders trade in multiple instances of the market. As before, the market price is interpreted as a probability density over the outcome space by interpreting the share vector as the natural parameter of an exponential family distribution. We measure the error in prediction using the standard log loss.

We show that traders with faulty information can only induce a limited amount of additional loss to the market's prediction. Further, since informative traders experience an expected increase in budget, they will eventually be unconstrained and allowed to carry out unrestricted trades. Taken together, this means that while the market suffers limited damage from ill-informed traders, it is also able to make use of all the information from informative traders in the long run.

Budget-limited trades. 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. 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.

Definition 6.1. 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 and is called his *budget-limited trade*.

[5]It is important to note that the potential function analysis still applies for any monotonically increasing transformation of the traders' utility functions

Limited Damage. We will now quantify the error in prediction that the market maker might have to endure as a result of ill-informed entities entering the market. We assume that these entities trade in multiple instances of the market; thus the exposure of the market maker is over several *rounds*. We use the log loss to measure this error in terms of the initial budget of traders; the loss function for θ shares held is defined as: $L(\theta, x) = -\log(p(x;\theta)) = C(\theta) - \theta^T \phi(x)$.

LEMMA 6.2. *The loss induced on the market by an uninformative trader is bounded by his initial budget.*

PROOF. First consider the change in budget of a trader i over multiple rounds of the prediction market. Let his budget at rounds t and $t-1$ be α_i^t and α_i^{t-1} respectively. The change in budget for trader i moving the market state from θ to θ' with outcome x^t is

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

Where Δ_i^t is called the myopic impact of a trader i in round t. Thus, the myopic impact captures incremental gain in prediction due to the trader in a round and is equal to the change in his budget in that round.

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:

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

□

Budget of Informative Traders. We now characterize the expected change in budget for an informative trader.

LEMMA 6.3. *Let θ be the current market state. 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

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

where $D_T(\cdot, \cdot)$ is the Bregman divergence based on T. 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 $\lambda D_T(\theta, \hat{\theta})$. $\square$

We note one important aspect of Lemma 6.3: 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.

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 Lemma 6.3 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. When the cost function is adjusted with a liquidity parameter as $C_l(\theta) = (1/l)C(l\theta)$ as described in Section 5, the trader receives an expected payoff of at least $\frac{1}{a} D_T(l\theta', \hat{\theta})$.

REFERENCES

ABERNETHY, J., CHEN, Y., AND VAUGHAN, J. W. 2013. Efficient market making via convex optimization, and a connection to online learning. *ACM Transactions on Economics and Computation 1,* 2, 12:1–12:39.

AZOURY, K. S. AND WARMUTH, M. K. 2001. Relative loss bounds for on-line density estimation with the exponential family of distributions. *Machine Learning 43*, 211–246.

BANERJEE, A., DHILLON, I. S., AND GHOSH, J. 2005a. Clustering with Bregman divergences. *Journal of Machine Learning Research 6*, 1705–1749.

BANERJEE, A., DHILLON, I. S., GHOSH, J., AND SRA, S. 2005b. Clustering on the unit hypersphere using von Mises-Fisher distributions. *Journal of Machine Learning Research 6*, 1345–1382.

BARNDORFF-NIELSEN, O. 1978. *Information and Exponential Families in Statistical Theory*. Wiley Publishers.

BEN-TAL, A. AND TEBOULLE, M. 2007. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance 17,* 3, 449–476.

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)*. 305–314.

CHEN, Y. AND PENNOCK, D. M. 2007. A utility framework for bounded-loss market makers. In *In Proceedings of the 23rd conference on uncertainty in artificial intelligence (UAI)*. 49–56.

CHEN, Y. AND VAUGHAN, J. 2010. A new understanding of prediction markets via no-regret learning. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*. 189–198.

DAWID, A. P. 1998. Coherent measures of discrepancy, uncertainty and dependence with applications to Bayesian predictive experimental design. Tech. Rep. 139, University College London, Dept. of Statistical Science.

DAWID, A. P. AND SEBASTIANI, P. 1999. Coherent dispersion criteria for optimal experimental design. *Annals of Statistics 27*, 65–81.

DIACONIS, P. AND YLVISAKER, D. 1979. Conjugate priors for exponential families. *Annals of Statistics 7,* 2, 269–281.

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)*. 553–560.

GNEITING, T. AND RAFTERY, A. 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102,* 477, 359–378.

GRÜNWALD, P. AND DAWID, A. P. 2004. Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory. *Annals of Statistics 32,* 4, 1367–1433.

HANSON, R. 2003. Combinatorial information market design. *Information Systems Frontiers 5,* 1, 105–119.

HANSON, R. 2007. Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets 1,* 1, 3–15.

HIRIART-URRUTY, J.-B. AND LEMARÉCHAL, C. 2000. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer.

JEWELL, W. S. 1974. Credible means are exact Bayesian for exponential families. *Astin Bulletin 8,* 1, 77–90.

LAMBERT, N. S., PENNOCK, D. M., AND SHOHAM, Y. 2008. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*. 129–138.

MONDERER, D. AND SHAPLEY, L. S. 1996. Potential games. *Games and Economic Behavior 14,* 1, 124–143.

NORSTAD, J. 1999. An introduction to utility theory. *Unpublished manuscript at http://homepage. mac. com/j. norstad*.

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 27th annual Conference on Uncertainty in Artificial Intelligence (UAI)*. 581–588.

PENNOCK, D. M., LAWRENCE, S., GILES, C. L., NIELSEN, F. A., ET AL. 2001. The real power of artificial markets. *Science 291,* 5506, 987–988.

PENNOCK, D. M. AND SAMI, R. 2007. Computational aspects of prediction markets. In *Algorithmic Game Theory*, N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds. Cambridge University Press, Chapter 26.

SAVAGE, L. J. 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association 66*, 783–801.

STORKEY, A. J. 2011. Machine learning markets. In *Proceedings of AI and Statistics (AISTATS)*. 716–724.

WAINWRIGHT, M. J. AND JORDAN, M. I. 2008. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning 1*, 1–305.

WOLFERS, J. AND ZITZEWITZ, E. 2004. Prediction markets. *Journal of Economic Perspectives 18,* 2, 107–126.

WOLFERS, J. AND ZITZEWITZ, E. 2006. Interpreting prediction market prices as probabilities. Tech. rep., National Bureau of Economic Research.