Exponential Family Prediction Markets

December 12, 2013

1 Paper Outline

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 maxent 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 market maker loss.

2 Generalized Log Scoring Rule

Suppose that an agent holds a belief in the expected value of the moments of a random variable x. The maximum entropy distribution that is consistent with these beliefs is an exponential family distribution.

Definition. For an outcome x with predicted probability p, the generalized log scoring rule is defined as

$$S(p,x) \stackrel{\text{def}}{=} a + b \log p(x)$$

where a and b are constants that we will set to 0 and 1 respectively in these notes.

Theorem 2.1. *The generalized log scoring rule for the maximum entropy distribution is proper.*

Proof. Recall that a proper scoring rule $\mathbf{S}$ satisfies

$$\mathbb{E}_{x\sim p}[\mathbf{S}(p,x)] \geq \mathbb{E}_{x\sim p}[\mathbf{S}(p',x)]$$

Note that the maximum entropy distribution is an exponential family distribution. Let $p(x) = \exp\{\beta\cdot\phi(x) - \psi(\beta)\}$ be the true distribution and $p'(x) = \exp\{\beta'\cdot\phi(x) - \psi(\beta')\}$ be the predicted distribution on the event x.
We need to show that

$$\mathbb{E}_{x\sim p}[\mathbf{S}(p,x)] \overset{?}{\geq} \mathbb{E}_{x\sim p}[\mathbf{S}(p',x)]$$

In other words, we need

$$\mathbb{E}_{x\sim p}[\log p(x)] \overset{?}{\geq} \mathbb{E}_{x\sim p}[\log p'(x)]$$

Equivalently,

$$\mathbb{E}_{x\sim p}[\beta\cdot\phi(x) - \psi(\beta)] \overset{?}{\geq} \mathbb{E}_{x\sim p}[\beta'\cdot\phi(x) - \psi(\beta')]$$

Or

$$\beta\cdot\mathbb{E}_{x\sim p}[\phi(x)] - \psi(\beta) \overset{?}{\geq} \beta'\cdot\mathbb{E}_{x\sim p}[\phi(x)] - \psi(\beta') \tag{1}$$

For an exponential family distribution, $\mathbb{E}_{x\sim p}[\phi(x)] = \nabla\psi(\beta)$. Substituting in (1), we need to show

$$\beta\cdot\nabla\psi(\beta) - \psi(\beta) \overset{?}{\geq} \beta'\cdot\nabla\psi(\beta) - \psi(\beta')$$

Rearranging the terms,

$$(\beta-\beta')\cdot\nabla\psi(\beta) \overset{?}{\geq} \psi(\beta) - \psi(\beta')$$

which is the first order condition on the convexity of $\psi(\cdot)$. □

Remark. The distance between the scores of the true and any predictive exponential family distribution is, in fact, the Bregman divergence (based on the log partition function) between the corresponding natural parameters. That is,

$$\mathbb{E}_{x\sim p}[\mathbf{S}(p,x)] - \mathbb{E}_{x\sim p}[\mathbf{S}(p',x)] = D_\psi(\beta,\beta')$$

3 Generalized Log Market Scoring Rule

Consider the dual of the cost function, $C^*(\mu)$ defined as

$$C^*(\mu) = \sup_\beta \beta\cdot\mu - C(\beta)$$

This supremum is obtained at the value of β for which $\mu = \nabla C(\beta)$; that is the natural parameter β for which $\mu = \mathbb{E}_{p_\beta}[\phi(x)]$ is the mean parameter. Rewriting,

$$C^*(\mathbb{E}_{p_\beta}[\phi(x)]) = \beta \cdot \mathbb{E}_{p_\beta}[\phi(x)] - C(\beta)$$

Thus, the mean parameters are the dual variables to the natural parameters.

Example 3.1. *We will now derive an expression for the dual of the log partition function for the Gaussian distribution. In this case, β is a 2-dimensional vector. Let $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ and $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$. So*

$$\begin{aligned} \boldsymbol{\mu} &= \nabla\psi(\boldsymbol{\beta}) \\ &= \nabla\left(-\frac{\beta_1^2}{4\beta_2} - \frac{1}{2}\log(-2\beta_2)\right) \\ &= \begin{pmatrix} -\frac{\beta_1}{2\beta_2} \\ \frac{\beta_1^2}{4\beta_2^2} - \frac{1}{2\beta_2} \end{pmatrix} \end{aligned}$$

Thus, $\boldsymbol{\beta}$ can be written in terms of $\boldsymbol{\mu}$ as follows: $\beta_1 = \frac{\mu_1}{\mu_2 - \mu_1^2}$ and $\beta_2 = \frac{1}{\mu_1^2 - \mu_2}$. This leads to the following closed form expression for the dual of the log partition function:

$$\psi^*(\mu) = -\frac{1}{2} - \frac{1}{2}\log(\mu_2 - \mu_1^2)$$

Theorem 3.2. *The value of the dual of the cost function is the negative entropy of the exponential family distribution obtained from backward mapping the mean parameters.*

Proof. To see this, we note that for $p(x) = \exp\{\beta \cdot \phi(x) - C(\beta)\}$

$$\begin{aligned} -H(p) &= \int_x p(x)\log p(x) \\ &= \int_x p(x)[\beta \cdot \phi(x) - C(\beta)] \\ &= \beta \cdot \int_x p(x)[\phi(x)] - \int_x p(x)C(\beta) \\ &= \beta \cdot \mathbb{E}_p[\phi(x)] - C(\beta) \end{aligned}$$

□

Remark. This result shows a nice parallel to LMSR, since the dual of the cost function of the LMSR is the negative entropy.

Example 3.3. *We will now derive the expression for the differential entropy for the normal distribution. Recall that for the normal distribution*

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$$

$$\begin{aligned}
H(p) &= -\int_x p(x) \log p(x) dx \\
&= -\int_x p(x) \left[-\log(2\pi) - \log\sigma + \frac{-(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \log(2\pi) + \log\sigma + \frac{1}{2\sigma^2} \int_x p(x)(x-\mu)^2] dx \\
&= \log(2\pi) + \log\sigma + \frac{1}{2\sigma^2} \mathbb{E}_{p(x)}[(x-\mu)^2] dx \\
&= \log(2\pi) + \log\sigma + \frac{1}{2\sigma^2} \sigma^2 dx \\
&= \log(2\pi) + \frac{1}{2} + \log\sigma
\end{aligned}$$

Recall that the value of the dual of the log partition function is the value of the negative entropy of the distribution for an exponential family distribution. Thus when the variance of the normal distribution $\sigma^2 = 0$, the value of the dual of the cost function goes to ∞ and is thus unbounded.

Example 3.4. *For the normal distribution, the natural parameters $\boldsymbol{\beta} = (\beta_1, \beta_2)$ can be written in terms of μ and σ as*

$$\beta_1 = \frac{\mu}{\sigma^2}, \quad \beta_2 = \frac{-1}{2\sigma^2}$$

and the log partition function is

$$\psi(\boldsymbol{\beta}) = -\frac{\beta_1^2}{4\beta_2} - \frac{1}{2}\log(-2\beta_2)$$

Alternately, the log partition function can be written in terms of its variance and mean as $\frac{\mu^2}{2\sigma^2} + \log\sigma$. Also note that the mean parameters are μ and $\mu^2 + \sigma^2$.

Remark. The LMSR is essentially a special case applied to a multinomial distribution. The LMSR is known to have bounded ($\log n$) market maker loss. Thus, while in general the exponential family LMSR does not guarantee bounded market maker loss, for some special cases it can.

Theorem 3.5. *Negative differential entropy $-H(p) \stackrel{\text{def}}{=} \int_x p(x) \log p(x) dx$ is unbounded from above for an exponential family distribution.*

Proof. Note that, for an exponential family distribution

$$\begin{aligned} H(p) &= -\int_x p(x)\log p(x)dx \\ &= -\int_x p(x)[\beta\cdot\phi(x)-\psi(\beta)]dx \\ &= -\beta\cdot\mathbb{E}[\phi(x)]+\psi(\beta) \end{aligned}$$

If the range of β is unbounded, the negative differential entropy is unbounded as well. □

3.1 Loss of the market maker

First we derive an expression for the conjugate dual C^* in terms of the primal variables. Recall the definition of the conjugate dual:

$$C^*(\mu) = \sup_q q\cdot\mu - C(q)$$

The supremum is achieved at q' such that $\nabla C(q') = \mu$. So we may rewrite:

$$C^*(\nabla C(q')) = \sup_q q\cdot\nabla C(q') - C(q)$$

This supremum is achieved at q such that $\nabla C(q') = \nabla C(q)$. One such value of q is q'. So we have

$$C^*(\nabla C(q')) = q'\cdot\nabla C(q') - C(q')$$

Also, recall that $\nabla C^*(\nabla C(q)) = q$ for the exponential family market. (I think there are some restrictions on q.)
Let q_0 be the initial and q_f the final market state. Then if μ is the expected value of the outcome sufficient statistics (*i.e.* the mean parameter) under the true distribution, the loss of the exponential family market maker can be written as:

$$\begin{aligned} \phi(x)(q_f-q_0)-C(q_f)+C(q_0) &= \phi(x)(q_f-q_0)-q_f\nabla C(q_f)+C^*(\nabla C(q_f))+q_0\nabla C(q_0)-C^*(\nabla C(q_0)) \\ &= q_f(\phi(x)-\nabla C(q_f))+C^*(\nabla C(q_f))-q_0(\phi(x)-\nabla C(q_0))-C^*(\nabla C(q_0)) \\ &= -C^*(\nabla C(q_0))+C^*(\phi(x))-q_0(\phi(x)-\nabla C(q_0)) \\ &\quad +C^*(\nabla C(q_f))-C^*(\phi(x))+q_f(\phi(x)-\nabla C(q_f)) \\ &= C^*(\phi(x))-C^*(\nabla C(q_0))-\nabla C^*(\nabla C(q_0))(\phi(x)-\nabla C(q_0)) \\ &\quad -\Big[C^*(\phi(x))-C^*(\nabla C(q_f))-\nabla C^*(\nabla C(q_f))(\phi(x)-\nabla C(q_f))\Big] \\ &= D_{C^*}[\phi(x),\nabla C(q_0))]-D_{C^*}[\phi(x),\nabla C(q_f))] \end{aligned}$$

Let's now work out the loss for the Gaussian market maker. First note that for a Gaussian market for mean parameters $\begin{pmatrix} \mu_1 \\ \mu_1^2 + \sigma_1^2 \end{pmatrix}$ and $\begin{pmatrix} \mu_2 \\ \mu_2^2 + \sigma_2^2 \end{pmatrix}$, we have

$$
\begin{aligned}
D_{C^*}\left(\begin{pmatrix} \mu_1 \\ \mu_1^2 + \sigma_1^2 \end{pmatrix}, \begin{pmatrix} \mu_2 \\ \mu_2^2 + \sigma_2^2 \end{pmatrix}\right) &= C^*\left(\begin{pmatrix} \mu_1 \\ \mu_1^2 + \sigma_1^2 \end{pmatrix}\right) - C^*\left(\begin{pmatrix} \mu_1 \\ \mu_1^2 + \sigma_1^2 \end{pmatrix}\right) - \nabla C^*\left(\begin{pmatrix} \mu_2 \\ \mu_2^2 + \sigma_2^2 \end{pmatrix}\right)\begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1^2 + \sigma_1^2 - \mu_2^2 - \sigma_2^2 \end{pmatrix} \\
&= -\log \frac{\sigma_1}{\sigma_2} - \begin{pmatrix} \frac{\mu_2}{\sigma_2^2} \\ -\frac{1}{2\sigma_2^2} \end{pmatrix} \cdot \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1^2 + \sigma_1^2 - \mu_2^2 - \sigma_2^2 \end{pmatrix} \\
&= -\log \frac{\sigma_1}{\sigma_2} - \frac{\mu_2(\mu_1 - \mu_2) - \frac{1}{2}(\mu_1^2 - \mu_2^2 + \sigma_1^2 - \sigma_2^2)}{\sigma_2^2} \\
&= \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_2^2 - \sigma_1^2}{2\sigma_2^2} + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}
\end{aligned}
$$

So the Gaussian market maker loss is

$$
\begin{aligned}
D_{C^*}[\phi(x), \nabla C(q_0))] - D_{C^*}[\phi(x), \nabla C(q_f))] &= \log \frac{\sigma_0}{\sigma_x} + \frac{\sigma_0^2 - \sigma_x^2}{2\sigma_0^2} + \frac{(\mu_x - \mu_0)^2}{2\sigma_0^2} \\
&\quad - \left[\log \frac{\sigma_f}{\sigma_x} + \frac{\sigma_f^2 - \sigma_x^2}{2\sigma_f^2} + \frac{(\mu_x - \mu_f)^2}{2\sigma_f^2}\right] \\
&= \log \frac{\sigma_0}{\sigma_f} + \frac{(x - \mu_0)^2}{2\sigma_0^2} - \frac{(x - \mu_f)^2}{2\sigma_f^2}
\end{aligned}
$$

Here we have used the fact that $\mu_x = x$ and $\sigma_x = 0$.

Remark. In statistics, the standard score (aka z-score), $(x - \mu)/\sigma$, is the (signed) number of standard deviations an observation is above the mean.

4 Some Interesting Asides

Proposition 4.1. *Let $f(x) : S \to \mathbb{R}$ be a convex function. Then*

$$\max_{x \in S} f(x) = \max_{x \in \mathrm{conv}(S)} f(x)$$

Proof. Note that it is trivially true that $\max_{x\in S} f(x) \le \max_{x\in \mathrm{conv}(S)} f(x)$ since every point in S must also be in $\mathrm{conv}(S)$.
Thus, it remains to show that $\max_{x\in S} f(x) \ge \max_{x\in \mathrm{conv}(S)} f(x)$.

Recall,

$$\operatorname{conv}(S) \stackrel{\text{def}}{=} \left\{ \sum_i a_i x_i : x_i \in S, a_i \in \triangle^n \text{ for } n \in \mathbb{Z}^+ \right\}$$

For any $x \in \operatorname{conv}(S)$,

$$f(x) = f\left(\sum_i a_i x_i\right) \leq \sum_i a_i f(x_i) \leq f(x^*)$$

where $x^* = \arg\max_{x \in S} f(x)$. In particular, $\max_{x \in S} f(x) \geq \max_{x \in \operatorname{conv}(S)} f(x)$ □

Proposition 4.2. *The variance of a distribution is* 0 *exactly when the outcome is the mean of the distribution with probability* 1.

Proof.

$$\begin{aligned} \operatorname{var}(X) &\stackrel{\text{def}}{=} \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \int_X \Pr(X)(X - \mathbb{E}[X])^2 dX \end{aligned}$$

Since $(X - \mathbb{E}[X])^2 \geq 0$, $\int_X \Pr(X) dX = 1$ and $\Pr(X) \geq 0$, the variance is 0 exactly when $X = \mathbb{E}[X]$. □

Proposition 4.3. *An arbitrage opportunity exists if and only if there also exists a trade that is guaranteed to* lose *money.*

Proof. An arbitrage opportunity may be written as

$$\exists q, r \forall x : C(q + r) - C(q) < r \cdot \phi(x)$$

But this means that a sale of r from a starting point of $q + r$ is guaranteed to lose money since this can be equivalently written as

$$\exists q, r \forall x : C(q) - C(q + r) > -r \cdot \phi(x)$$

□

4.1 Exponential Smoothing

One way to model the exponential family prediction market might be through exponential smoothing.

- A *time series* is a set of data points sampled at periodic instances.

- *Smoothing* is the process of creating an approximation function based on time series that while capturing patterns in the data is relatively insensitive to noise.
- Natural first step is to use moving average. But average on how many samples (say k)? What to do until the first k samples have been received?
- Next attempt might be to use weighted moving average. Advantage is that you can give higher weight to more recent terms. But still same disadvantage as simple moving average technique.
- *Exponential Smoothing* attempts to remove this disadvantage.

Exponential smoothing is defined iteratively for $\alpha \in (0,1)$ as

$$\begin{aligned} s_1 &= x_0 \\ s_t &= \alpha x_{t-1} + (1-\alpha)s_{t-1} \\ &= \alpha\left[x_{t-i} + \sum_{i=2}^{t}(1-\alpha)^{i-1}x_{t-i}\right] + (1-\alpha)^{t-1}x_0 \end{aligned}$$

Smaller values of α means greater smoothing e.g., $\alpha = 0$ gives a constant function $(= x_0)$. Note that the selection of x_0 is important – and increases in importance as α. Some advantages over moving average: all data points matter (decreasing importance with time) and computationally only last data point need be stored.

References

[1] Jacob Abernethy, Yiling Chen and Jennifer Wortman Vaughan, "Efficient Market Making via Convex Optimization, and a Connection to Online Learning". In *ACM Trans. Econ. Comput.* Vol. 1. ACM, New York, NY, USA