

# Summary about Prediction Market

Zhihao Ruan, Jiayu Yi, Naihao Deng  
{ruanzh, yijiaiyu, dnaihao}@umich.edu

August 4, 2019

## Contents

<b>1</b>	<b>Exponential Family</b>	<b>2</b>
1.1	Sufficient Statistics . . . . .	2
1.2	Conjugate Priors . . . . .	3
1.2.1	Bernoulli Distribution . . . . .	3
1.2.2	Normal Distribution . . . . .	3
<b>2</b>	<b>Scoring Rules</b>	<b>5</b>
2.1	Motivations . . . . .	5
2.2	Incentive Compatibility . . . . .	5
2.3	Logarithmic Scoring Rule . . . . .	5
2.4	Maximum Entropy Optimization . . . . .	6
<b>3</b>	<b>Cost Function based Prediction Market with Bayesian Traders</b>	<b>7</b>

# 1 Exponential Family

Distributions over  $\mathbf{x}$  with *natural parameters*  $\boldsymbol{\eta}$  that are of the form

$$p(\mathbf{x} | \boldsymbol{\eta}) = h(\mathbf{x})g(\boldsymbol{\eta}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\} \quad (1)$$

We need to make sure that  $\int p(\mathbf{x} | \boldsymbol{\eta}) d\mathbf{x} = 1$ , therefore

$$g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\} d\mathbf{x} = 1 \quad (2)$$

$g(\boldsymbol{\eta})$  is a coefficient that is needed to ensure that the distribution is normalized.

## 1.1 Sufficient Statistics

If we take the partial derivative with respect to  $\boldsymbol{\eta}$  from both sides of equation 2, we have

$$\nabla g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\} d\mathbf{x} + g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\} \mathbf{u}(\mathbf{x}) d\mathbf{x} = 0 \quad (3)$$

If we plug in 2 to 3, we have

$$-\frac{1}{g(\boldsymbol{\eta})} \nabla g(\boldsymbol{\eta}) = g(\boldsymbol{\eta}) \int h(\mathbf{x}) \exp\{\boldsymbol{\eta}^T \mathbf{u}(\mathbf{x})\} \mathbf{u}(\mathbf{x}) d\mathbf{x} = \mathbb{E}[\mathbf{u}(\mathbf{x})] \quad (4)$$

Since

$$\nabla \ln g(\boldsymbol{\eta}) = \frac{1}{g(\boldsymbol{\eta})} \nabla g(\boldsymbol{\eta})$$

It can be deduced that

$$-\nabla \ln g(\boldsymbol{\eta}) = \mathbb{E}[\mathbf{u}(\mathbf{x})] \quad (5)$$

If we consider a set of independent identically distributed data denoted by  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ , the likelihood function is given by

$$p(\mathbf{X} | \boldsymbol{\eta}) = \left( \prod_{n=1}^N h(\mathbf{x}_n) \right) g(\boldsymbol{\eta})^N \exp \left\{ \boldsymbol{\eta}^T \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right\} \quad (6)$$

To produce the maximum likelihood estimator  $\boldsymbol{\eta}_{ML}$ , we need to take the partial derivative of 6 and set its value to 0.

$$\nabla \left( \left( \prod_{n=1}^N h(\mathbf{x}_n) \right) g(\boldsymbol{\eta})^N \exp \left\{ \boldsymbol{\eta}^T \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right\} \right) = 0 \quad (7)$$

By expanding 7 it can be seen that

$$\left( \prod_{n=1}^N h(\mathbf{x}_n) \right) \left( N g(\boldsymbol{\eta})^{N-1} \nabla g(\boldsymbol{\eta}) \exp \left\{ \boldsymbol{\eta}^T \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right\} + g(\boldsymbol{\eta})^N \exp \left\{ \boldsymbol{\eta}^T \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right\} \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right) = 0$$

Therefore,

$$-\nabla \ln g(\boldsymbol{\eta}_{ML}) = \frac{1}{N} \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \quad (8)$$

From equation 8 it can be seen that  $\boldsymbol{\eta}_{ML}$  can be determined if  $\mathbf{u}(\mathbf{x}_n)$  is known,  $\mathbf{u}(\mathbf{x}_n)$  is therefore called the *sufficient statistics*. If  $N \rightarrow \infty$ , the right hand side of 8 would become  $\mathbb{E}[\mathbf{u}(\mathbf{x})]$  and  $\boldsymbol{\eta}_{ML}$  would be the true  $\boldsymbol{\eta}$ .

## 1.2 Conjugate Priors

Conjugate priors are priors that lead to their corresponding posterior distributions to have the same form as they do. For exponential family distributions, their conjugate priors are of the form

$$p(\boldsymbol{\eta}|\boldsymbol{\chi}, \nu) = f(\boldsymbol{\chi}, \nu)g(\boldsymbol{\eta})^\nu \exp \{ \nu \boldsymbol{\eta}^T \boldsymbol{\chi} \} \quad (9)$$

$\nu$  can be interpreted as the number of samples that are observed.  $\boldsymbol{\chi}$  is the pseudo observation of the sufficient statistics  $\mathbf{u}(\mathbf{x})$ . Each one of the  $\nu$  observations take the value of  $\mathbf{u}(\mathbf{x})$ . The updated posterior can be shown as

$$\begin{aligned} p(\boldsymbol{\eta}|\mathbf{X}, \boldsymbol{\chi}, \nu) &= \left( \prod_{n=1}^N h(\mathbf{x}_n) \right) g(\boldsymbol{\eta})^N \exp \left\{ \boldsymbol{\eta}^T \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right\} \times f(\boldsymbol{\chi}, \nu) g(\boldsymbol{\eta})^\nu \exp \{ \nu \boldsymbol{\eta}^T \boldsymbol{\chi} \} \\ &\propto g(\boldsymbol{\eta})^{\nu+N} \exp \left\{ \boldsymbol{\eta}^T \left( \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) + \nu \boldsymbol{\chi} \right) \right\} \end{aligned} \quad (10)$$

It can be seen from here that we can update the posterior with  $\nu \leftarrow \nu + N$  and  $\nu \boldsymbol{\chi} \leftarrow \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) + \nu \boldsymbol{\chi}$

### 1.2.1 Bernoulli Distribution

The conjugate prior for Bernoulli distribution is the Beta distribution. The probability density function for Beta distribution is

$$p(x | a, b) = \frac{x^{a-1}(1-x)^{b-1}}{\mathbf{B}(a, b)} \quad (11)$$

When it is written in the form of the conjugate prior

$$\frac{\theta^{a-1}(1-\theta)^{b-1}}{\mathbf{B}(a, b)} = \frac{1}{\mathbf{B}(a, b)} (1-\theta)^{a+b-2} \exp \left[ (a+b-2) \ln \left( \frac{\theta}{1-\theta} \right) \frac{a-1}{a+b-2} \right] \quad (12)$$

From equation 12 we can see that  $\nu$  corresponds to  $a+b-2$  here and  $\boldsymbol{\chi}$  corresponds to  $\frac{a-1}{a+b-2}$ .

### 1.2.2 Normal Distribution

Given that the conjugate priors for exponential family distributions are of the form  $p(\boldsymbol{\eta}|\boldsymbol{\chi}, \nu) = f(\boldsymbol{\chi}, \nu)g(\boldsymbol{\eta})^\nu \exp \{ \nu \boldsymbol{\eta}^T \boldsymbol{\chi} \}$ , we were trying to find the  $\nu$  and  $\boldsymbol{\chi}$  for univariate Gaussian distribution given that the variance  $\sigma^2$  is fixed.

The probability density function of univariate Gaussian distribution is given by

$$p(x | \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (13)$$

First we can start by proving that the conjugate prior for Gaussian distribution is also a Gaussian distribution (when the variance is fixed).

$$\begin{aligned} p(x|\mu) &\sim N(\mu, \sigma^2) \\ p(\mu) &\sim N(\mu_0, \sigma_0^2) \end{aligned}$$

Suppose  $D$  is the set of data points that we have observed. We have

$$\begin{aligned}
p(\mu|D) &\propto \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left( \frac{x^{(i)} - \mu}{\sigma} \right)^2 \right\} \times \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - \mu_0}{\sigma_0} \right)^2 \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \sum_{i=1}^N \left( \frac{x^{(i)} - \mu}{\sigma} \right)^2 + \left( \frac{\mu - \mu_0}{\sigma_0} \right)^2 \right] \right\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[ \left( \frac{N}{\sigma^2} + \frac{1}{\sigma_0^2} \right) \mu^2 - 2 \left( \frac{\sum_{i=1}^N x^{(i)}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right) \mu \right] \right\}
\end{aligned} \tag{14}$$

We can see that equation 14 can be easily written in the form of a Gaussian distribution if we extract multiple from the coefficient in the front.

***Can all this be still written in the form of exponential family distribution and exponential family distribution's conjugate priors?***

We want to write equation

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

in the form of  $p(\mathbf{x} | \boldsymbol{\eta}) = \left( \prod_{n=1}^N h(\mathbf{x}_n) \right) g(\boldsymbol{\eta})^N \exp \left\{ \boldsymbol{\eta}^T \sum_{n=1}^N \mathbf{u}(\mathbf{x}_n) \right\}$ , and the prior

$$\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - \mu_0}{\sigma_0} \right)^2 \right\}$$

in the form of  $p(\boldsymbol{\eta} | \boldsymbol{\chi}, \nu) = f(\boldsymbol{\chi}, \nu) g(\boldsymbol{\eta})^\nu \exp \left\{ \nu \boldsymbol{\eta}^T \boldsymbol{\chi} \right\}$ .

We can first start with priors.

$$\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{1}{2} \left( \frac{\mu - \mu_0}{\sigma_0} \right)^2 \right\} = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{\mu^2 - 2\mu_0\mu + \mu_0^2}{2\sigma_0^2} \right\} \tag{15}$$

From the equation above we can deduce that the  $\boldsymbol{\eta}$  should be  $\begin{bmatrix} \mu \\ \mu^2 \end{bmatrix}$ . Is this OK? Continue with equation 15,

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ -\frac{\mu^2 - 2\mu_0\mu + \mu_0^2}{2\sigma_0^2} \right\} &= \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left\{ \boldsymbol{\eta}^T \begin{bmatrix} \frac{\mu_0}{\sigma_0^2} \\ -\frac{1}{2\sigma_0^2} \end{bmatrix} - \frac{\mu_0^2}{2\sigma_0^2} \right\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left( \frac{-\mu_0^2}{2\sigma_0^2} \right) \exp \left[ \boldsymbol{\eta}^T \begin{bmatrix} \frac{\mu_0}{\sigma_0^2} \\ -\frac{1}{2\sigma_0^2} \end{bmatrix} \right]
\end{aligned}$$

We know from equation 1.2.2 that we need  $\mu$  outside exp to construct  $g(\boldsymbol{\eta})$ . Therefore we have

$$\frac{1}{\sqrt{2\pi}\sigma_0} \exp \left( \frac{-\mu_0^2}{2\sigma_0^2} \right) \exp \left[ \boldsymbol{\eta}^T \begin{bmatrix} \frac{\mu_0}{\sigma_0^2} \\ -\frac{1}{2\sigma_0^2} \end{bmatrix} \right] = \frac{1}{\sqrt{2\pi}\sigma_0} \exp \left[ \frac{\alpha}{\sigma_0^2} \mu^2 + \frac{\beta\mu_0}{\sigma_0^2} \mu \right] \left\{ -\frac{(1 + 2\alpha)\mu^2 - 2\mu_0(1 - \beta)\mu + \mu_0^2}{2\sigma_0^2} \right\} \tag{16}$$

We can assume that  $\nu$  here is the common denominator of  $\frac{\alpha}{\sigma_0^2}$  and  $\frac{\beta\mu_0}{\sigma_0^2}$ , which is  $\frac{1}{\sigma_0^2}$ . Under this assumption

$$g(\boldsymbol{\eta}) = \exp(\alpha\boldsymbol{\eta}_2 + \beta\mu_0\boldsymbol{\eta}_1)$$

There's so far no way for us to determine what values  $\alpha$  and  $\beta$ .

## 2 Scoring Rules

Scoring rules is the simplest form of prediction mechanism. For every agent with some information, a scoring rule evaluates how close the information is from the actual outcome, and pays the agent in return for his/her information.

### 2.1 Motivations

Taking an event  $X$  with outcome space  $\mathcal{X}$ , we want to know something about each agent's belief  $p(x)$  on the actual outcome  $x \in \mathcal{X}$ , compare it with the actual outcome, and see how accurate the agent's prediction is. However, it is impossible to ask agent for his/her entire  $p(x)$  probability distribution. What should we do?

We've already known that sufficient statistic is a very suitable parameter to characterize a probability distribution. Therefore, we can use it to represent agent's belief. We just ask each agent for the sufficient statistic  $u(x)$  as a representation of his/her belief  $p(x)$ . Then we are able to measure how accurate the agent can predict with a scoring rule based on the actual outcome.

Assume  $\mu = \mathbb{E}_p[u(x)]$  is the *expected* sufficient statistic over some  $p(x)$  that takes  $u(x)$  as its sufficient statistic and is taken as an estimate of the agent's belief. Assume  $\hat{\mu}$  is the agent's report as an estimate of  $\mu$ . Then, with the actual outcome denoted  $x$ , the scoring rule takes the form:

$$S(\hat{\mu}, x).$$

We can see that the scoring rule is a measure of how close the agent's belief is from the actual outcome.

### 2.2 Incentive Compatibility

**Incentive compatibility** is a property of prediction mechanism with which the best strategy for an agent to earn the most profit is to *honestly* report all the information as soon as he/she has it. As a prediction mechanism, a proper scoring rule should leverage **incentive compatibility** in order to get real information from agents.

Assume that we already set the sufficient statistic to be  $u(x)$ . Then for each  $p \in \mathcal{P}$  that takes this  $u(x)$  as its sufficient statistic, we can calculate its *expected* sufficient statistic  $\mu = \mathbb{E}_p[u(x)]$ . A scoring rule is thus called **proper** if it satisfies the following, for all such  $p$ , for all  $\hat{\mu} \neq \mu$ :

$$\mathbb{E}_p[S(\mu, x)] \geq \mathbb{E}_p[S(\hat{\mu}, x)].$$

Any scoring rule with this property actually encourages agents to report a probability distribution as close to the actual probability distribution of  $X$  as possible, which is in accordance with the essence of incentive compatibility.

### 2.3 Logarithmic Scoring Rule

Suppose that we have set a form of  $u(x)$ . For some  $p$  that takes  $u(x)$  as its sufficient statistic, a classic logarithmic scoring rule takes the form:

$$S(\mu, x) = \ln p(x; \mu),$$

where  $\mu = \mathbb{E}_p[u(x)]$  is the expected sufficient statistic over  $p(x; \mu)$ . **For the following sections, we will be only talking about logarithmic scoring rules.**

## 2.4 Maximum Entropy Optimization

Now that we have the general expression of logarithmic scoring rules, how can we determine which  $p(\mathbf{x}; \boldsymbol{\mu})$  to choose as an estimate of agents' beliefs, given a specific  $\mathbf{u}(\mathbf{x})$ ?

For some particular form of  $\mathbf{u}(\mathbf{x})$ , denote the set  $\mathcal{P}$  which contains all  $p$  that takes  $\mathbf{u}(\mathbf{x})$  as its sufficient statistic. Then, we formulate the following optimization problem:

$$\begin{aligned} \min_{p \in \mathcal{P}} \quad & F(p) = \int_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \ln p(\mathbf{x}) \, dh(\mathbf{x}) \\ \text{s.t.} \quad & \mathbb{E}_p[\mathbf{u}(\mathbf{x})] = \boldsymbol{\mu} \end{aligned}$$

where  $\boldsymbol{\mu}$  is the expected sufficient statistic over  $p$ , and  $h(\mathbf{x})$  is the base measure.

It is clear to see that  $F(p)$  is actually the negative of entropy of  $p(\mathbf{x})$ . Therefore, minimizing  $F(p)$  is thus maximizing the entropy of  $p$ . This actually mean that of all  $p \in \mathcal{P}$ , we tend to choose a  $p$  that shows the most uncertainty of  $\mathbf{X}$  as an estimate of agents' beliefs. This estimate is not necessarily exactly the same as agents' beliefs, as the general goal is just to find something that we can use to construct a metric in order to measure how close the agents' beliefs are from the actual outcome.

It has been proved that  $S(\boldsymbol{\mu}, \mathbf{x}) = \ln p(\mathbf{x}; \boldsymbol{\mu})$  is **proper** if and only if  $p$  is the exponential family. It has also been proved that the solutions to the optimization problem are exponential family distributions which takes the form

$$\begin{aligned} p(\mathbf{x}; \boldsymbol{\eta}) &= h(\mathbf{x})g(\boldsymbol{\eta}) \exp \{ \boldsymbol{\eta} \cdot \mathbf{u}(\mathbf{x}) \} \\ &= h(\mathbf{x}) \exp \{ \boldsymbol{\eta} \cdot \mathbf{u}(\mathbf{x}) - T(\boldsymbol{\eta}) \} \end{aligned}$$

where  $\boldsymbol{\eta}$  is the natural parameter of  $p$ ,  $T(\boldsymbol{\eta}) = -\ln g(\boldsymbol{\eta})$  is the log-partition function of  $p$ .

Either *expected* sufficient statistic  $\boldsymbol{\mu}$  or natural parameter  $\boldsymbol{\eta}$  can parametrize a probability distribution  $p$ , and actually they are closely related. If we denote the convex conjugate of  $T(\boldsymbol{\eta})$  as  $G(\boldsymbol{\mu})$ , then we have  $\boldsymbol{\mu} = \nabla T(\boldsymbol{\eta})$ ,  $\boldsymbol{\eta} = \nabla G(\boldsymbol{\mu})$ .

### **3 Cost Function based Prediction Market with Bayesian Traders**