Prediction, Belief, and Markets

Jake Abernethy, University of Pennsylvania
Jenn Wortman Vaughan, UCLA

June 26, 2012

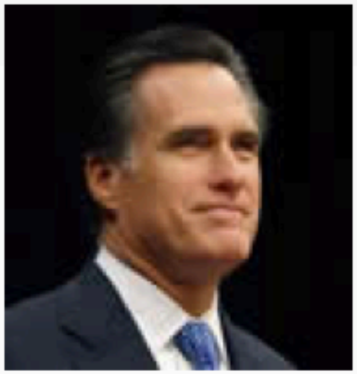
Prediction Markets

"Arrow-Debreu Security": Contract pays $10 if X happens, $0 otherwise. If I think that Pr(X) > p then I should:

- Buy this security at any price less than $10*p*
- Sell this security at any price greater than $10*p*

Current price measures the population's collective beliefs

Example: Intrade



Mitt Romney to be Republican Presidential Nominee in 2012

Last prediction was: **$9.59** / share

Today's Change: ▲ **+$0.02** (+0.2%)

95.9%
CHANCE

NASA to announce discovery of extraterrestrial life before midnight ET 31 Dec 2012

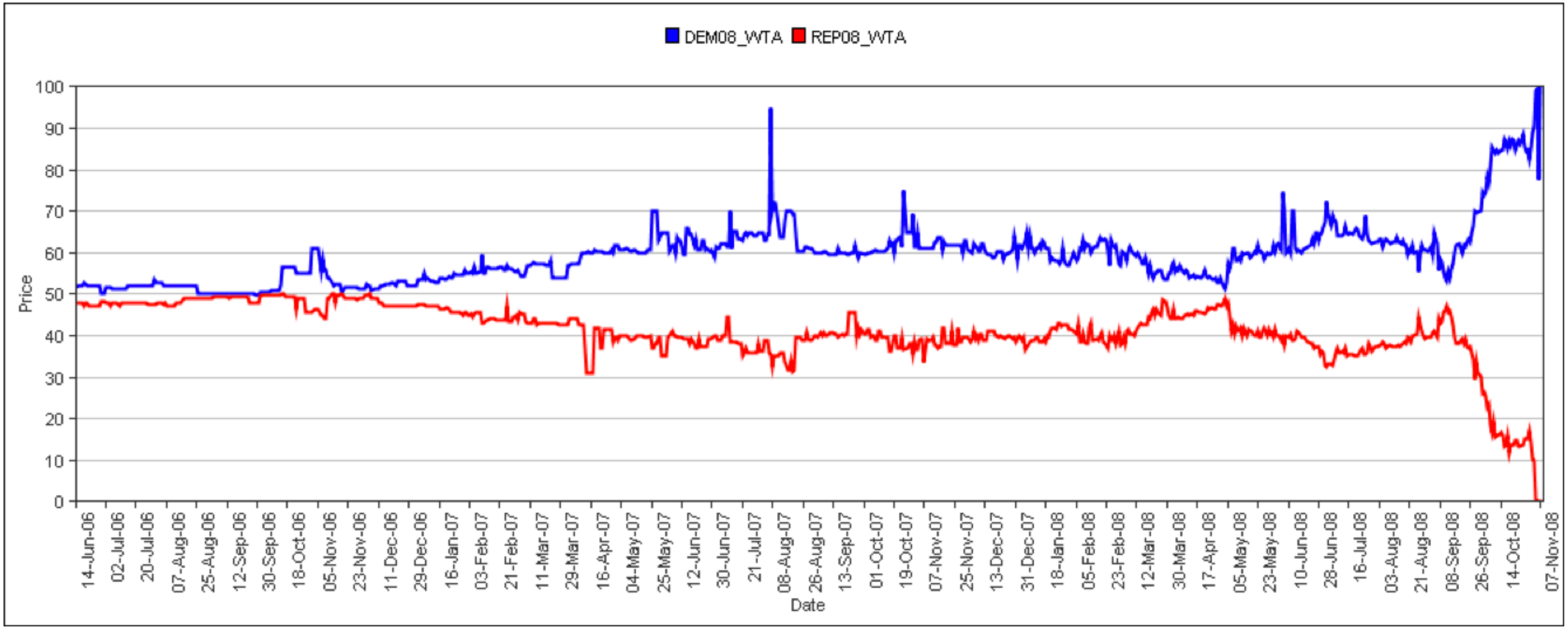
Last prediction was: **$0.21** / share

Today's Change: -

2.1%
CHANCE

Example: Iowa Electronic Market

Popular vote in 2008 presidential election



Example: Inkling Markets

Internal prediction markets used within companies

This Tutorial: Outline

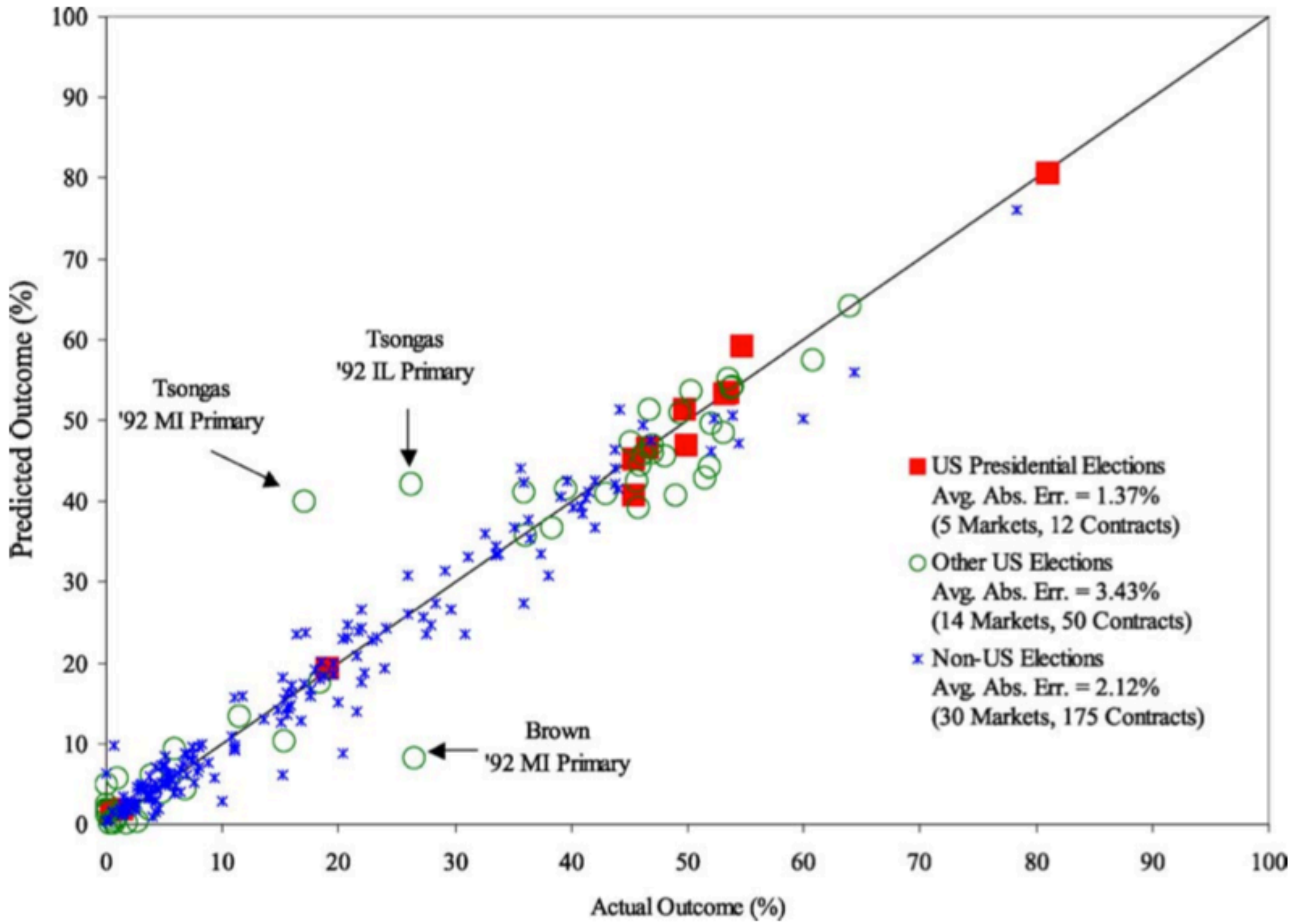
1. Introduction
2. Prediction Markets in Practice: Work Well?
3. Proper Scoring Rules and Eliciting Beliefs
4. Bregman Divergences + Proper Scoring Rules
5. Hanson’s Market Scoring Rule
6. BREAK!
7. Automated Market Makers
8. Beyond Complete Markets
9. Duality & Connection to Online Learning
10. Some Additional Topics

Markets in Practice

Market Prices as a Forecaster

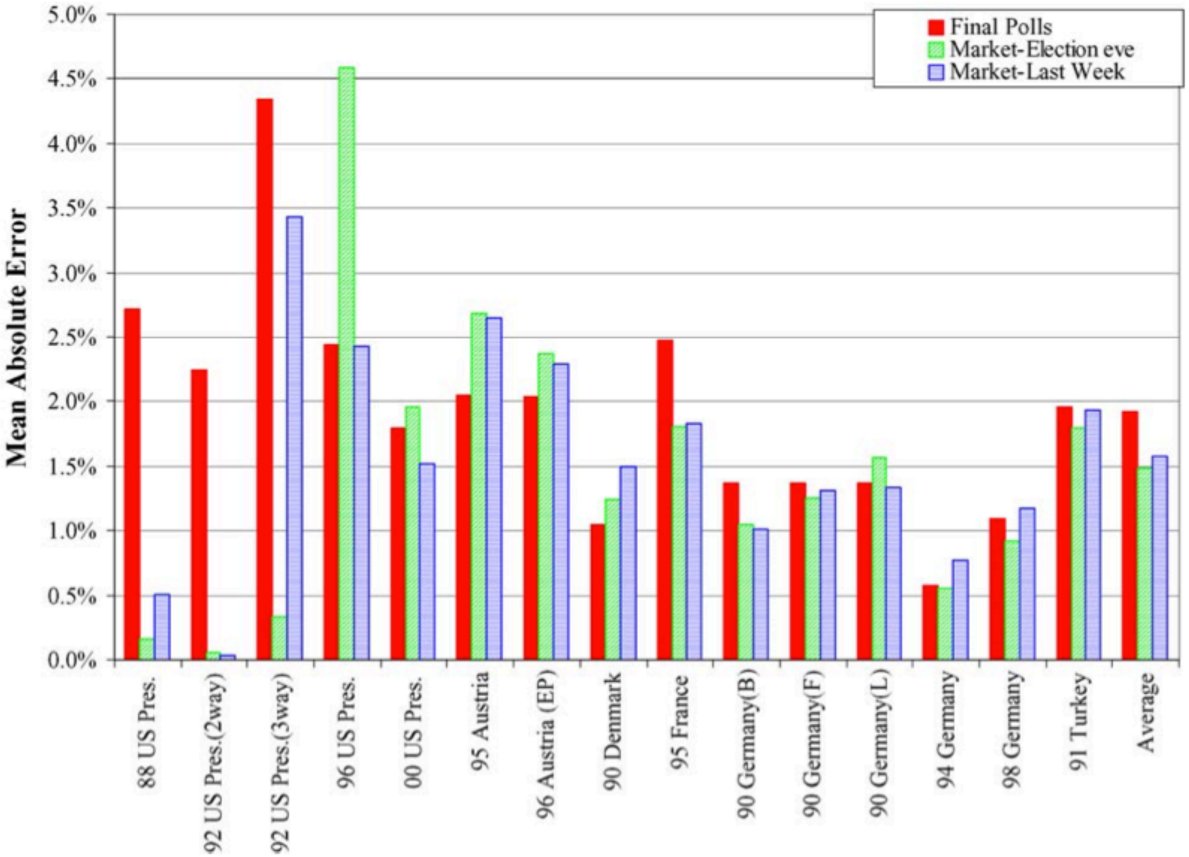
- The market price for Arrow-Debreu security is essentially a "consensus estimate" of the probability of an event
- Are these estimates *accurate*?
- We can check this on historical data!
- Questions:
  - Prices are changing, which do we use?
  - Accurate in which metric?
  - Accurate compared to what?

Market Prediction vs. True Vote Share



Berg et al., 2008: "Results From a Dozen Years of Election Futures Markets Research"

Average Polls vs. Market Prices



Poll Error: (multi poll avg)	1.91%
Market Error: (election eve)	1.49%
Market Error: (week average)	1.58%

Berg et al., 2008: "Results From a Dozen Years of Election Futures Markets Research"

The Basics: Proper Scoring Rules

1950: Brier on Weather Forecasting

MONTHLY WEATHER REVIEW

EDITOR, JAMES E. CASKEY, JR.

Volume 78
Number 1

JANUARY 1950

Closed March 5, 1950
Issued April 15, 1950

VERIFICATION OF FORECASTS EXPRESSED IN TERMS OF PROBABILITY

GLENN W. BRIER

U. S. Weather Bureau, Washington, D. C.

[Manuscript received February 10, 1950]

INTRODUCTION

Verification of weather forecasts has been a controversial subject for more than a half century. There are a number of reasons why this problem has been so perplexing to meteorologists and others but one of the most important difficulties seems to be in reaching an agreement on the specification of a scale of goodness for weather forecasts. Numerous systems have been proposed but one of the greatest arguments raised against forecast verification is that forecasts which may be the "best" according to the accepted system of arbitrary scores may not be the most

numerically have been discussed previously [1, 2, 3, 4] so that the purpose here will not be to emphasize the enhanced usefulness of such forecasts but rather to point out how some aspects of the verification problem are simplified or solved.

VERIFICATION FORMULA

Suppose that on each of n occasions an event can occur in only one of r possible classes or categories and on one such occasion, i, the forecast probabilities are f_{i1}, f_{i2}, . . . f_{ir}, that the event will occur in classes 1, 2, . . . r, respectively. The r classes are chosen to be mutually

INTRODUCTION

Verification of weather forecasts has been a controversial subject for more than a half century. There are a number of reasons why this problem has been so perplexing to meteorologists and others but one of the most important difficulties seems to be in reaching an agreement on the specification of a scale of goodness for weather forecasts. Numerous systems have been proposed but one of the greatest arguments raised against forecast verification is that forecasts which may be the "best" according to the accepted system of arbitrary scores may not be the most useful forecasts. In attempting to resolve this difficulty the forecaster may often find himself in the position of choosing to ignore the verification system or to let it do the forecasting for him by "hedging" or "playing the system." This may lead the forecaster to forecast something other than what he thinks will occur, for it is often easier to analyze the effect of different possible forecasts on the verification score than it is to analyze the weather situation. It is generally agreed that this state of affairs is unsatisfactory, as one essential criterion for satisfactory verification is that the verification scheme should influence the forecaster in no undesirable way. Unfortunately, the criterion is difficult, if not impossible to satisfy, although some schemes will be much worse than others in this respect.

TABLE 2.—*Verification of a series of 85 forecasts expressed in terms of the probability of rain*

Forecast probability of rain	Observed proportion of rain cases	Forecast probability of rain	Observed proportion of rain cases
0.00–0.19	0.07	0.60–0.79	0.40
0.20–0.39	.10	0.80–1.00	.50
0.40–0.59	.29		

Brier's Question: How Should We Pay a Forecaster?

- What is the right "payment scheme" to reward/punish a forecaster who makes a sequence of probability predictions for events that we observe?

- The sequence of outcomes: $y_1, y_2, y_3, \ldots \in \{0,1\}$

- The sequence of forecasts: $p_1, p_2, p_3, \ldots \in [0,1]$

- The forecaster's payment: $\frac{1}{T}\sum_{t=1}^{T} S(y_t, p_t)$

Brier Score ⇔ Quadratic Scoring Rule

$$S(y,p) = -(y-p)^2$$

- For the n-outcome case: $y \in \{1,\ldots,n\}$ $\mathbf{p} \in \Delta_n$

$$S(y,\mathbf{p}) = -\sum_{i=1}^{n} (\mathbf{1}_{y=i} - p_i)^2$$

What's Special About This Function?

$$S(y,p) = -(y-p)^2$$

- Assume y is random: $\Pr(y=1) = q$

$$\operatorname*{argmax}_{p\in[0,1]} \mathrm{E}\left[-(y-p)^2\right]$$

$$= \operatorname*{argmax}_{p\in[0,1]} \left[q(1-p)^2 + (1-q)(0-p)^2\right]$$

$$= \operatorname*{argmax}_{p\in[0,1]} \left[(p-q)^2 + q - q^2\right] = q$$

Proper Scoring Rules

- What we have just introduced is the notion of a **proper scoring rule**, which is any func. S(,) satisfying

$$\mathop{\mathrm{E}}_{y\sim\mathbf{q}}\left[S(y,\mathbf{q})\right] \geq \mathop{\mathrm{E}}_{y\sim\mathbf{q}}\left[S(y,\mathbf{p})\right] \quad \forall \mathbf{p},\mathbf{q} \in \Delta_n$$

- The scoring rule is said to be **strictly proper** if the above inequality is strict unless $\mathbf{p} = \mathbf{q}$

Another Strictly Proper Scoring Rule

$$S(y, \mathbf{p}) = \log \mathbf{p}(y)$$

- This is known as the **logarithmic scoring rule**. For predicting 0/1, it can be written as:

$$S(y, p) = \begin{cases} \log p & y = 1 \\ \log(1 - p) & y = 0 \end{cases}$$

- **EXERCISE:** check that this is proper!

Scoring Rules == -Loss Functions?? Basically, yes.

- Effectively, a scoring rule is just a type of loss function
- Scoring rules measure the *performance* (not loss) of a predicted distribution given a final outcome
- One often sees the use of the term *proper loss function* which is equivalent to *proper scoring rule*
- Research on scoring rules is focused more heavily on the incentives of the associated payment mechanism

Savage 1973

© Journal of the American Statistical Association
December 1971, Volume 66, Number 336
Theory and Methods Section

Elicitation of Personal Probabilities and Expectations

LEONARD J. SAVAGE*

Proper scoring rules, i.e., devices of a certain class for eliciting a person's probabilities and other expectations, are studied, mainly theoretically but with some speculations about application. The relation of proper scoring rules to other economic devices and to the foundations of the personalistic theory of probability is brought out. The implications of various restrictions, especially symmetry restrictions, on scoring rules is explored, usually with a minimum of regularity hypothesis.

1. INTRODUCTION

1.1 Preface

This article is about a class of devices by means of which an idealized *homo economicus*—and therefore, with some approximation, a real person—can be induced to reveal his opinions as expressed by the probabilities that he associates with events or, more generally, his personal expectations of random quantities. My emphasis here is theoretical, though some experimental considerations will be mentioned. The empirical importance of such studies in many areas is now recognized. It was emphasized for the area of economics in an address by Trygve Haavelmo [28, p. 357]:

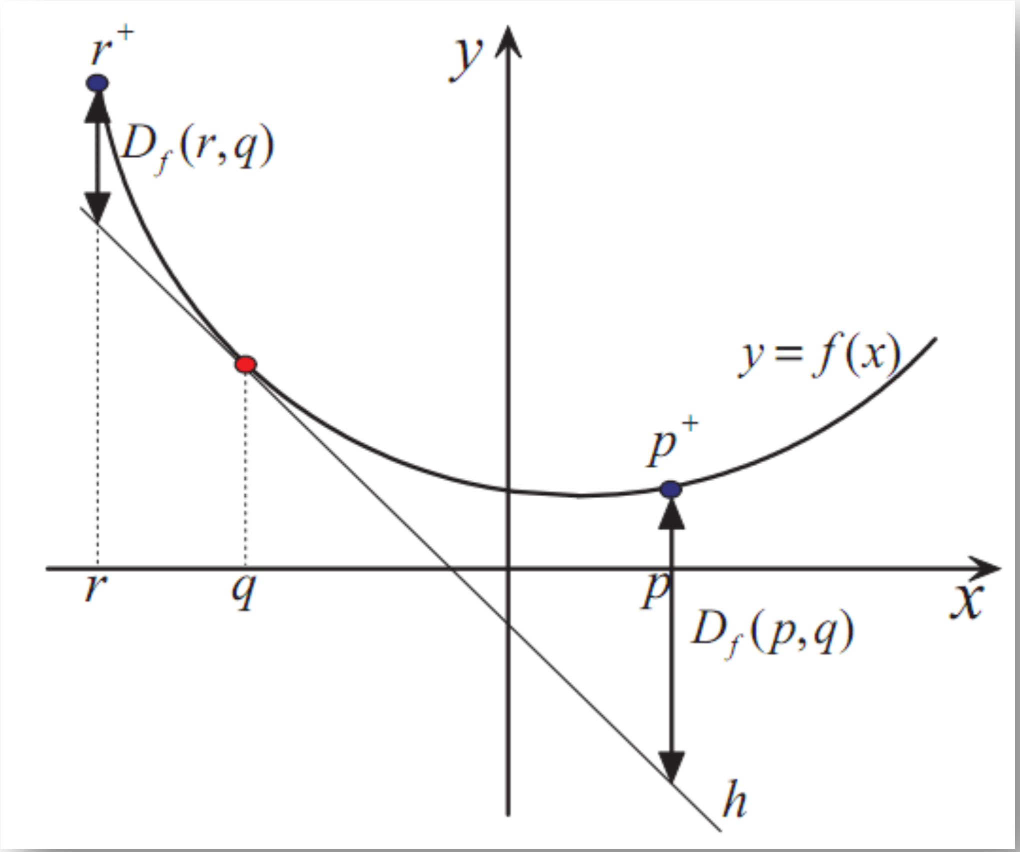
pertaining to it has grown up, some of which will be cited in context and most of which can be found through the references cited, especially the recent and extensive [52] and others that I call "key references."

Bruno de Finetti and I began to write the present article in the spring of 1960, not yet aware of our predecessors and contemporaries. The impetus was de Finetti's, for he had brought us to rediscover McCarthy's [37] insight about convex functions. We expected to make short work of our "little note," but it grew rapidly in many directions and became inordinately delayed. Now we find that the material in the present article is largely mine and that de Finetti has published on diverse aspects of the same subject elsewhere [12, 13, 14, 17]. De Finetti has therefore withdrawn himself from our joint authorship and encouraged me to publish this article alone, though it owes so much to him at every stage, including the final draft.

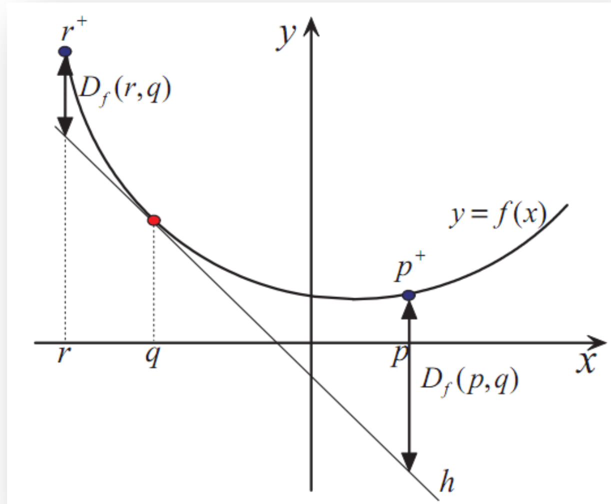
The article is written for a diverse audience. Consequently, some will find parts of it mathematically too

Digression: Bregman Divergences

- *Bregman Divergence*: Measuring dist using convex func's



Bregman Divergences II



- **Def' n**: let f be convex + diff' able, then

$$D_f(x,y) = f(x) - f(y) - \nabla f(y) \cdot (x - y)$$

- **Properties:** $D_f(x,y) \neq D_f(y,x) \quad (not\, nec.)$

$$D_f(x,x) = 0 \quad \forall x \qquad\qquad D_f(x,y) \geq 0 \quad \forall x,y$$

Bregman Divergences III

$$D_f(x,y) := f(x) - f(y) - \nabla f(y) \cdot (x - y)$$

- **Example 1**, quadratic:

$$f(x) = \| x \|^2 \Rightarrow D_f(x,y) = \| x - y \|^2$$

- **Example 2**, entropic:

$$f(\mathbf{p}) = \sum_i p_i \log p_i \Rightarrow D_f(\mathbf{p},\mathbf{q}) = \sum_i p_i \log \frac{p_i}{q_i}$$

Bregman Diverg. ⇔ Scoring Rule??

- Let f be any convex function
- Define: i[th] indicator vector $\mathbf{e}_i := \langle 0,\ldots,0,1,0\ldots\rangle$
- Let **p,q** be any two distributions, then for some g():

$$\underset{i\sim\mathbf{q}}{\mathrm{E}}\left[D_f(\mathbf{e}_i,\mathbf{p})\right] = D_f(\mathbf{q},\mathbf{p}) + g(\mathbf{q})$$

- Then: $\arg\max_{\mathbf{p}} \underset{i\sim\mathbf{q}}{\mathrm{E}}\left[-D_f(\mathbf{e}_i,\mathbf{p})\right] = \mathbf{q}$
- This is the scoring rule property!!

Bregman Diverg. ⇔ Scoring Rule!!

- We now have a recipe for constructing scoring rules!
- Let f be any convex function, h be arbitrary
- We can now let:

$$S(i,\mathbf{p}) = -D_f(\mathbf{e}_i,\mathbf{p}) + h(i)$$

- Quadratic Scoring Rule: $f(\mathbf{p}) = \|\mathbf{p}\|_2^2$
- Log Scoring Rule: $f(\mathbf{p}) = \sum_i p_i \log p_i$

Brief Literature Review

Surrogate regret bounds for proper losses
MD Reid, RC Williamson
Proceedings of the 26th Annual International Conference on Machine Learning ...

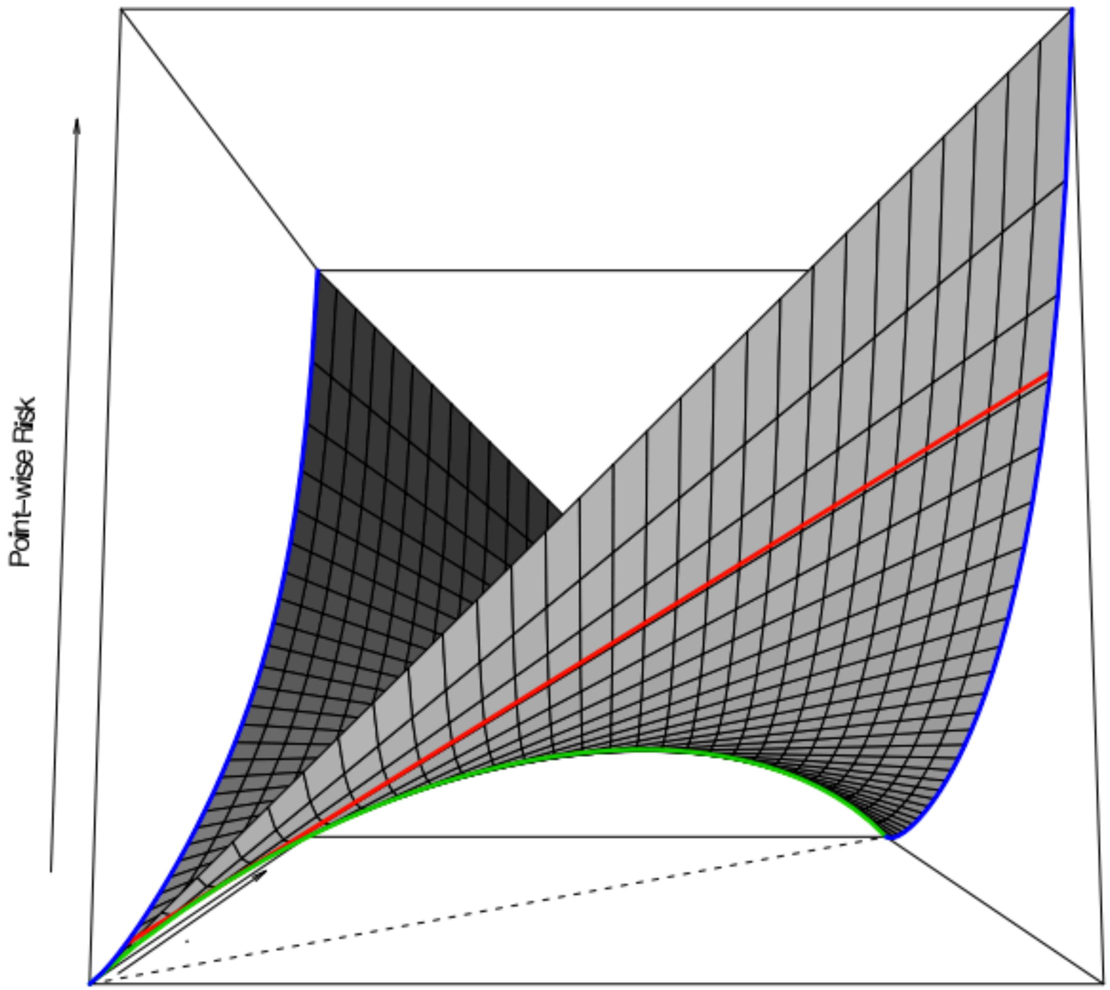
Information, divergence and risk for binary experiments
MD Reid, RC Williamson
Journal of Machine Learning Research 12, 731-817

Composite binary losses
MD Reid, RC Williamson
The Journal of Machine Learning Research 9999, 2387-2422

Composite Multiclass Losses
E Vernet, ENS Cachan, RC Williamson, MD Reid
Neural Information Processing Systems

Mixability is Bayes Risk Curvature Relative to Log Loss
T van Erven, MD Reid, RC Williamson
Proceedings of the 24th Annual Conference on Learning Theory (COLT)

3D Plot of Logarithmic Scoring Rule



From Reid+Williamson, “Composite Binary Losses”

Market Scoring Rules

Learning a Consensus?

- Scoring rules are useful for incentivizing ONE individual to state his beliefs about a probability
- How can we learn "from the crowd"?
- **Proposal**: pay every individual according to a scoring rule!
- **Problems**:
  - This is very expensive!
  - How to combine estimates?
  - How to weed out noise traders?
  - How to weed out copycats?

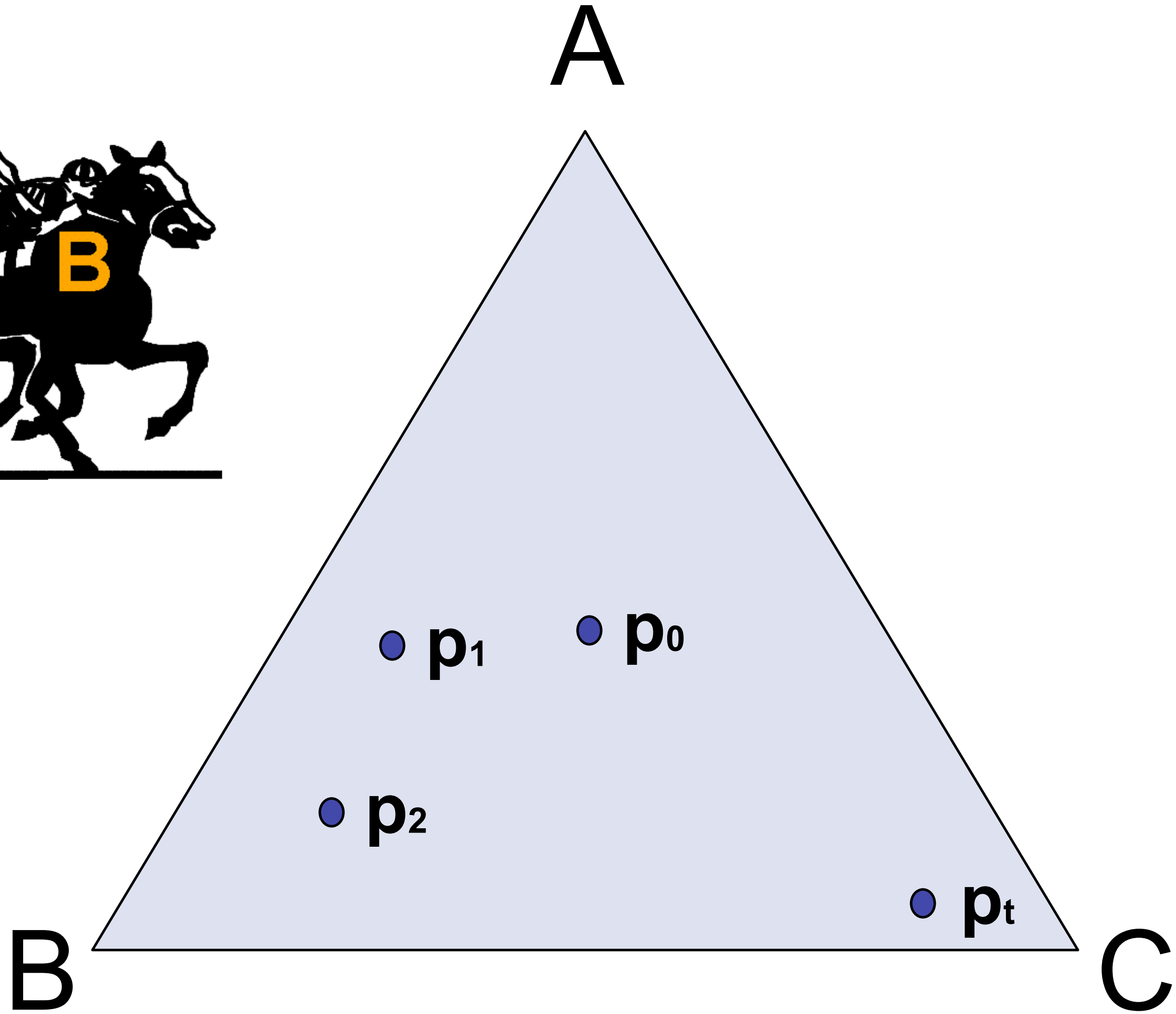
Hanson: *Market* Scoring Rule

- Robin Hanson proposed following idea to create a **prediction market** based on an **automated market maker** (MM) – an agent always willing to bet for/against
- Imagine we have an uncertain event X which will take one of n values {1,2,…,n}
- The MM publishes a *scoring rule* S(,) for everyone to see
- The MM posts an initial distribution (prior) $\mathbf{p}_0$
- Traders arrive, one-by-one, giving *updates* $\mathbf{p}_{t-1} \rightarrow \mathbf{p}_t$
- Eventually, outcome X is revealed, and trader t earns:

$$S(X,\mathbf{p}_t) - S(X,\mathbf{p}_{t-1})$$

- Payment could be negative!

Market Scoring Rule



Incentives and Costs

- Assume trader t has *belief distribution* **p** in outcome X.
- Note: **p** can (should) depend on previous prices!
- He wants to maximize his payment:

$$\operatorname{argmax}_{\mathbf{p}_t} \mathrm{E}_{X \sim \mathbf{p}}[S(X, \mathbf{p}_t) - S(X, \mathbf{p}_{t-1})]$$

- Opt value nonnegative due to scoring rule property!

$$= \operatorname{argmax}_{\mathbf{p}_t} \mathrm{E}_{X \sim \mathbf{p}}[S(X, \mathbf{p}_t)] \quad = \mathbf{p}$$

- The MM must make all payments, which totals

$$\sum_{t=1}^{T} \left[S(X, \mathbf{p}_t) - S(X, \mathbf{p}_{t-1}) \right] = S(X, \mathbf{p}_T) - S(X, \mathbf{p}_0)$$

- This is bounded! This is like MM's subsidy to market.

LMSR: Log Market Scoring Rule

- Initial hypothesis is $\mathbf{p}_0$ uniform distribution
- Trader t posts an update $\mathbf{p}_{t-1} \rightarrow \mathbf{p}_t$
- After X is revealed, trader t earns $\log(\mathbf{p}_t(X)/\mathbf{p}_{t-1}(X))$
- Hanson: the LMSR is an important special case, *only* MSR for which "betting on conditional probabilities does not affect marginal probabilities"
- The market maker's worst case loss is bounded by log n, where n is the number of possible outcomes of X

BREAK TIME!

1. Introduction
2. Prediction Markets in Practice: Work Well?
3. Proper Scoring Rules and Eliciting Beliefs
4. Bregman Divergences + Proper Scoring Rules
5. Hanson's Market Scoring Rule
6. **BREAK!!!!!!!!!!!!**
7. Automated Market Makers
8. Beyond Complete Markets
9. Duality & Connection to Online Learning
10. Some Additional Topics

Automated Market Makers

Back to Arrow–Debreu Securities

Barack Obama to be re-elected President in 2012

Last prediction was **$5.27** / share → **52.7%** CHANCE

Today's Change: ▼ -$0.12 (-2.2%)

Contract Type: **0-100**

How do we arrive at the current price?

- Traditional stock market style pricing (continuous double auction) – low liquidity, huge spreads
- Automated market maker – willing to risk a (bounded) loss in order to encourage trades

Market Makers for Complete Markets

- In a complete market, a security is offered for each of a set of mutually exclusive and exhaustive events

Worth $1 iff

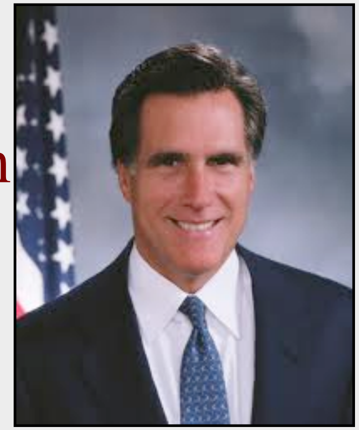
Worth $1 iff

- An automated market maker is always willing to buy and sell these securities at some price

Cost Functions

Cost of purchase:

$$C(\mathbf{q} + \mathbf{r}) - C(\mathbf{q})$$

Worth $1 iff 

Worth $1 iff 

Already purchased:	q_1 shares	q_2 shares
Want to purchase:	r_1 shares	r_2 shares
Instantaneous prices:	$p_1 = \partial C / \partial q_1$	$p_2 = \partial C / \partial q_2$

"predictions"

Back to the LMSR

The logarithmic market scoring rule can be implemented as a cost function based market with cost function

$$C(q_1,\ldots,q_N) = b \log \sum_{i=1}^{N} \exp(q_i/b)$$

and instantaneous prices

$$p_i = \frac{\exp(q_i/b)}{\Sigma_j \exp(q_j/b)}$$

Notice that p_i is increasing in q_i and the prices sum to 1

Equivalence

For all $\mathbf{p}$, $\mathbf{p'}$, $\mathbf{q}$, $\mathbf{q'}$, such that $\nabla C(\mathbf{q}) = \mathbf{p}$ and $\nabla C(\mathbf{q'}) = \mathbf{p'}$, for all outcomes i, a trader who changed the market state from $\mathbf{p}$ to $\mathbf{p'}$ in the MSR would receive the same payoff as a trader who changed the market state from $\mathbf{q}$ to $\mathbf{q'}$ in the cost function based market.

[Hanson 03; Chen & Pennock 07]

A Proof in One Slide

cost function payoff

$$= \left(q'_i - q_i\right) - \left(C(\mathbf{q}') - C(\mathbf{q})\right)$$

$$= \left(\log e^{q'_i} - \log e^{q_i}\right) - \left(\log \sum_j e^{q'_j} - \log \sum_j e^{q_j}\right)$$

$$= \log \frac{e^{q'_i}}{\sum_j e^{q'_j}} - \log \frac{e^{q_i}}{\sum_j e^{q_j}}$$

$$= \log p'_i - \log p_i$$

= scoring rule payoff

More Generally

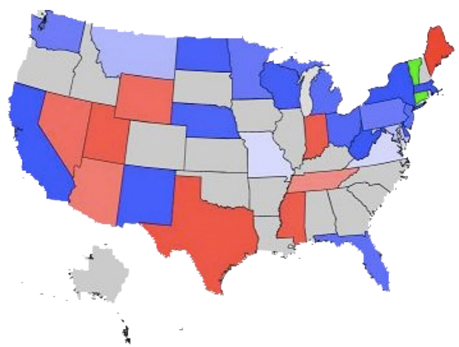
- Any market scoring rule can be implemented as a cost function based market [Chen & Pennock 07; Chen & Vaughan 11; Abernethy & Frongillo 11; …]

- Advantages:
    - Retains the good incentive properties of the MSR
    - Arguably more "natural" for traders
    - Exposure to risk is more transparent

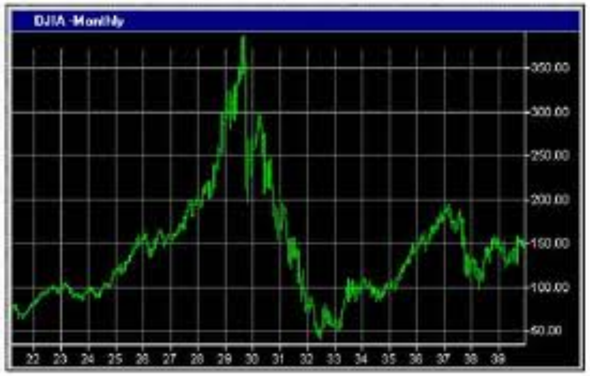
Beyond Complete Markets

Complex Outcome Spaces

$n!$



2^n



Infinite

- Cannot simply run a standard market like LMSR
  - Calculating prices is intractable [Chen et al., 2008]
  - Reasoning about probabilities is too hard for traders
- Can run separate, independent markets (e.g., horses to win, place, or show) but this ignores logical dependences

Complex Outcome Spaces

Given a small set of securities over a very large (or infinite) state space, how can we design a consistent market that can be operated efficiently?

[Abernethy et al., 2011]

Menu of Securities

We would like to offer a menu of securities $\{1, \ldots, K\}$ specified by a payoff function ρ

payoff

securities

outcomes

10	0	5.5	0	17	0
.9	.9	.9	0	.9	.9
0	42	0	10	10	10
0	0	11.5	8	0	0
1	0	0	0	0	1

Example: Pair Betting

$1 if and only if horse *i* finishes ahead of horse *j*

	A<B	B<A	A<C	C<A	B<C	C<B
ABC	1	0	1	0	1	0
ACB	1	0	1	0	0	1
BAC	0	1	1	0	1	0
BCA	0	1	0	1	1	0
CAB	1	0	0	1	0	1
CBA	0	1	0	1	0	1

What are "reasonable" prices?

For complete markets…

$$\sum_i p_i = 1$$

For pair betting…

$$p_{i<j} + p_{j<i} = 1$$

$$1 \le p_{i<j} + p_{j<k} + p_{k<i} \le 2$$

what else?

In general…

???

An Axiomatic Approach

Path independence: The cost of acquiring a bundle **r** of securities must be the same no matter how the trader splits up the purchase. Formally,

$$\text{Cost}(\mathbf{r} + \mathbf{r}' \mid \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t)$$
$$= \text{Cost}(\mathbf{r} \mid \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t) + \text{Cost}(\mathbf{r}' \mid \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t, \mathbf{r})$$

This alone implies the existence of a cost potential function!

$$\text{Cost}(\mathbf{r} \mid \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_t)$$
$$= C(\mathbf{r}_1 + \mathbf{r}_2 + \ldots + \mathbf{r}_t + \mathbf{r}) - C(\mathbf{r}_1 + \mathbf{r}_2 + \ldots + \mathbf{r}_t)$$

An Axiomatic Approach

- **Existence of instantaneous prices:** *C* must be continuous and differentiable
- **Information incorporation:** The purchase of a bundle **r** should never cause the price of **r** to decrease
- **No arbitrage:** It is never possible to purchase a bundle **r** with a guaranteed positive profit regardless of outcome
- **Expressiveness:** A trader must always be able to set the market prices to reflect his beliefs

An Axiomatic Approach

Theorem: Under these five conditions, costs must be determined by a convex cost function C such that

$$\{\nabla C(\mathbf{q}) : \mathbf{q} \in \mathrm{R}^K\} = \mathrm{Hull}(\boldsymbol{\rho})$$

reachable price vectors

securities

outcomes

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

Cost Functions Via Duality &
The Connection to Online Learning

How do we find these cost functions?

- **Fact:** A closed, differentiable function C is convex if and only if it can be written in the form

$$C(\mathbf{q}) = \sup_{\mathbf{x} \in \mathrm{dom}(R)} \mathbf{x} \cdot \mathbf{q} - R(\mathbf{x})$$

for a strictly convex function R called the conjugate.

Furthermore, $\nabla C(\mathbf{q}) = \arg\max_{\mathbf{x} \in \mathrm{dom}(R)} \mathbf{x} \cdot \mathbf{q} - R(\mathbf{x})$

To generate a convex cost function C, we just have to choose an appropriate conjugate function and domain!

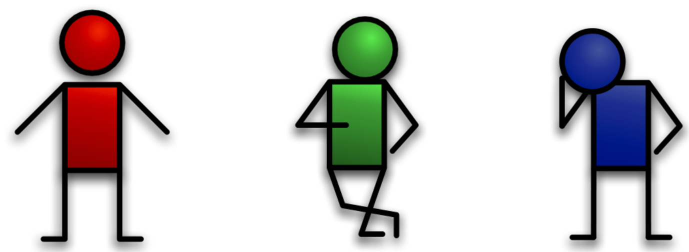
But how do we choose R?

We can borrow ideas from online linear optimization (or the simpler expert advice setting) and in particular, Follow the Regularized Leader algorithms

- Market's conjugate function ≈ regularizer

Learning from Expert Advice

Suppose we would like to choose actions based on the advice of n "experts" (people, algorithms, features…)



At each round t,

- Algorithm selects weights $w_{t,1}$ $w_{t,2}$ $w_{t,3}$
- Experts suffer a loss $l_{t,1}$ $l_{t,2}$ $l_{t,3}$
- Algorithm suffers a loss $\mathbf{w}_t \cdot \mathbf{l}_t$

What is the goal?

- Ideally, we'd like to bound the cumulative loss $\sum_{t=1}^{T} \mathbf{w}_t \cdot \mathbf{l}_t$

- Instead, we look at the algorithm's regret

$$\sum_{t=1}^{T} \mathbf{w}_t \cdot \mathbf{l}_t - \min_{\mathbf{w} \in K} \mathbf{w} \cdot \mathbf{L}_T$$

cumulative loss vector

algorithm's loss

loss of the best fixed weight vector in hindsight

What is the goal?

- Ideally, we'd like to bound the cumulative loss $\sum_{t=1}^{T} \mathbf{w}_t \cdot \mathbf{l}_t$
- Instead, we look at the algorithm's regret

$$\sum_{t=1}^{T} \mathbf{w}_t \cdot \mathbf{l}_t - \min_{\mathbf{w} \in K} \mathbf{w} \cdot \mathbf{L}_T$$

- Can achieve optimal ($O(T^{1/2})$) regret with Follow the Regularized Leader

$$\mathbf{w}_{t+1} = \arg\min_{\mathbf{w} \in K} \mathbf{w} \cdot \mathbf{L}_t + \frac{1}{\eta} R(\mathbf{w})$$

cumulative loss regularizer

Online Linear Opt.	Market Making
• Learner maintains weights $\mathbf{w}_t \in K$ over n items/experts	• Market maker maintains prices $\mathbf{p}_t \in \Pi$ over n contracts
• Items have loss vector $\mathbf{l}_t$, cumulatively $\mathbf{L}_{t+1} = \mathbf{L}_t + \mathbf{l}_{t+1}$	• Contracts are purchased in bundles $\mathbf{r}_t$, and $\mathbf{q}_{t+1} = \mathbf{q}_t + \mathbf{r}_{t+1}$
• FTRL selects weights $\mathbf{w}_{t+1} = \arg\min_{\mathbf{w}\in K} \mathbf{w} \cdot \mathbf{L}_t + \frac{1}{\eta} R(\mathbf{w})$	• Market maker selects prices $\mathbf{p}_{t+1} = \arg\max_{\mathbf{p}\in\Pi} \mathbf{p} \cdot \mathbf{q}_t - R(\mathbf{p})$
• Learner suffers regret $\sum_{t=1}^{T} \mathbf{w}_t \cdot \mathbf{l}_t - \min_{\mathbf{w}\in K} \mathbf{w} \cdot \mathbf{L}_T$	• MM has worst-case loss $\max_{\mathbf{p}\in\Pi} \mathbf{p} \cdot \mathbf{q}_T - \sum_{t=1}^{T} \left(C(\mathbf{q}_t) - C(\mathbf{q}_{t-1}) \right)$

More on Choosing R

- Interesting market properties can be described in terms of the conjugate…
  - Worst-case market maker loss can be bounded by

$$\sup_{\mathbf{x} \in \mathrm{Hull}(\boldsymbol{\rho})} R(\mathbf{x}) - \inf_{\mathbf{x} \in \mathrm{Hull}(\boldsymbol{\rho})} R(\mathbf{x})$$

  - Information loss (or the bid-ask spread, or the speed at which prices change) can be bounded too

Gives us a way to optimize trade-offs in market design!

Example: Permutations

- Suppose our state space is all permutations of n items (e.g., candidates in an election, or horses in a race)
  - Pair bets: Bets on events of the form "horse i finishes ahead of horse j" for any i, j
  - Subset bets: Bets on events of the form "horse i finishes in position j" for any i, j
- Both known to be #P-hard to price using LMSR [Chen et al., 2008]

- The complex market framework handles both

Example: Permutations

Subset bets ("horse *i* finishes in position *j*")

- Hull($\boldsymbol{\rho}$) can be described by a small number of constraints:

$$\sum_j \text{price}(i \text{ in slot } j) = 1 \qquad \sum_i \text{price}(i \text{ in slot } j) = 1$$

- Easily handled

Example: Permutations

Pair bets ("horse *i* finishes ahead of horse *j*")

- Hull($\boldsymbol{\rho}$) is a bit uglier…
- Solution: Relax the no-arbitrage axiom
  - Allows us to to work with a larger, efficiently specified price space
  - But does it increase worst case loss? No!

Some Additional Topics

Markets & Variational Inference

- The math behind these markets also parallels the math behind variational inference

mean parameter	$\leftrightarrow$	prices
natural parameter	$\leftrightarrow$	quantity vector
sufficient statistics	$\leftrightarrow$	payoff function

- This connection can be used to design new scoring rules [Lahaie et al., working paper, 2012]

Making a Profit

- We have assumed that the market maker is willing to take a potential (bounded) loss in order to obtain information

- The ideas presented here can be modified to yield market makers guaranteed to earn a profit if the volume of trades is sufficiently high [e.g., Othman & Sandholm, 2011]

- In the complete market setting, this requires that prices sum to something more than one – adds some ambiguity when backing out probability estimates

Convergence of Prices

- Under what circumstances do security prices converge and reflect the private information of the traders?

- A bad example for convergence:
  - We flip two fair coins
  - Trader 1 learns the result of the first flip, and trader 2 learns the result of the second
  - Suppose that our security pays off $1 if the flips are the same (HH, TT) and $0 otherwise (HT, TH)
  - If the price starts at $0.5, neither trader will trade

Convergence of Prices

- Ostrovsky [2012] characterized separable securities for which prices can never get "stuck"
- Separability is a property of both the security and the information structure of the traders
- If a security is separable with respect to the traders' information structure, then at equilibrium, the final price of the security will reflect all traders' private information

Also at ICML/COLT

- COLT talk: *A Characterization of Scoring Rules for Linear Properties* by Rafael Frongillo and Jake in the Appleton Tower at 3:05pm (RIGHT NOW!)

- ICML workshop: Markets, Mechanisms, and Multi-Agent Models, all day Sunday

Questions?

http://icmlmarketstutorial.pbworks.com

Reminder: Your coffee break is in the Informatics Forum across the street, starting now