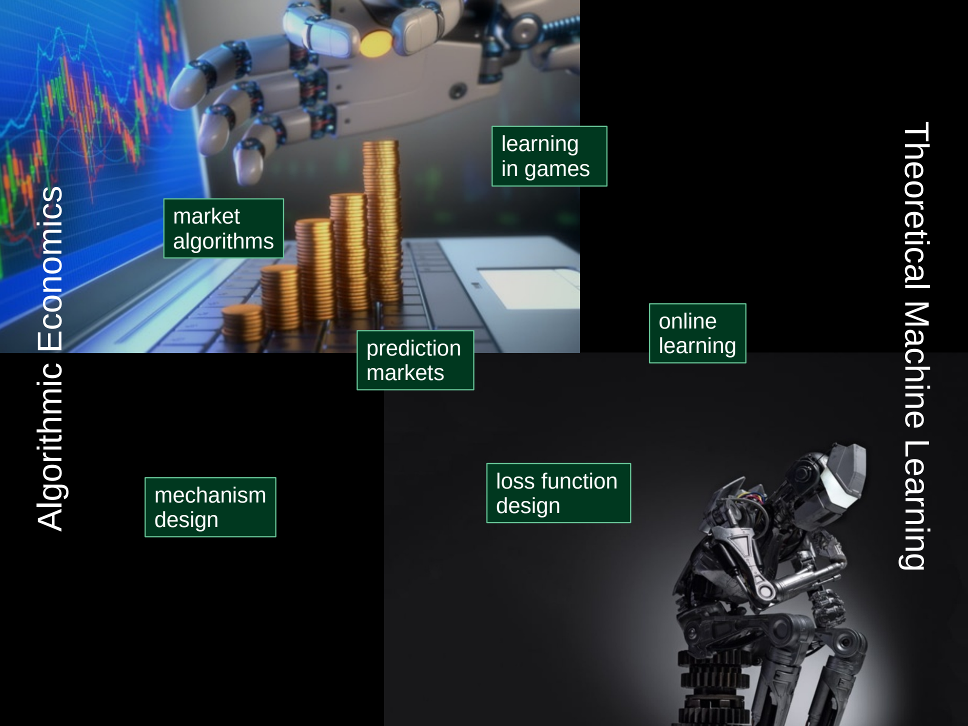
How (Not) to Run a Forecasting Competition: Incentives and Efficiency

Rafael Frongillo, Robert Gomez, Anish Thilagar, Bo Waggoner
University of Colorado Boulder

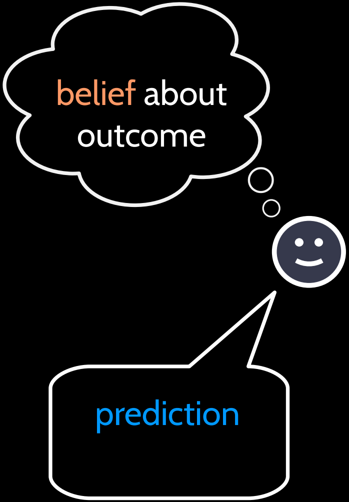
March 9, 2022

WU Vienna Institute for Statistics and Mathematics

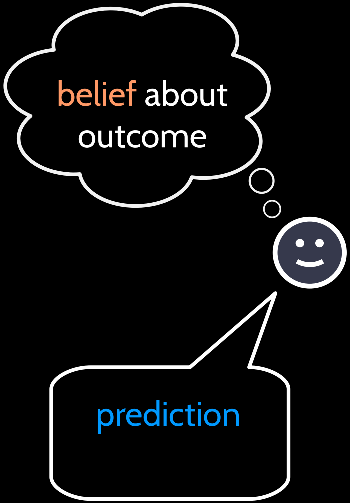

Algorithmic Economics
Theoretical Machine Learning
market algorithms
learning in games
prediction markets
online learning
mechanism design
loss function design

Information Elicitation

Information Elicitation

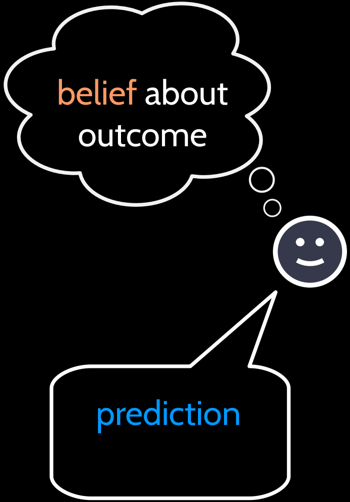


Information Elicitation



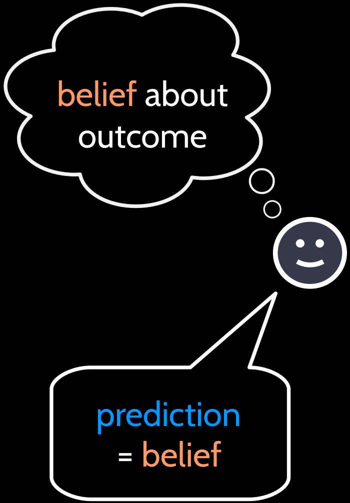
Score (prediction , outcome)

Information Elicitation



Score (prediction , outcome)

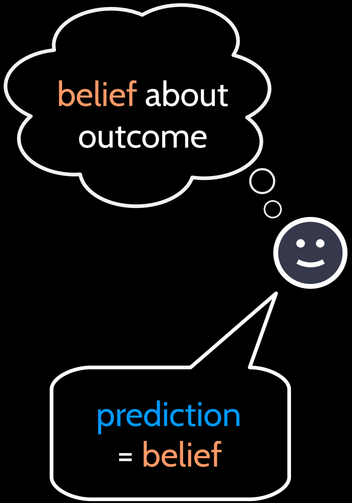
$$\max \; \mathbb{E}_{\text{outcome} \sim \text{belief}} [\text{Score}]$$



Score (prediction , outcome)

$$\max \ \mathbb{E}_{\text{outcome} \sim \text{belief}} [\ \text{Score} \ ]$$

Information Elicitation



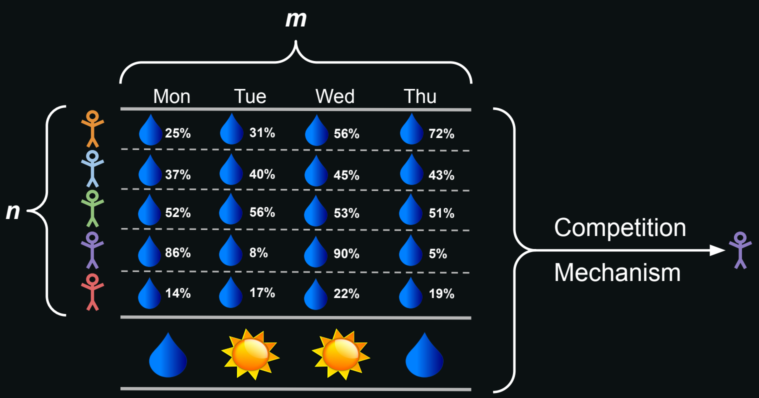
$$\text{Score}(\text{prediction}, \text{outcome})$$

$$\max \; \mathbb{E}_{\text{outcome} \sim \text{belief}} [\, \text{Score} \,]$$

$$\min \sum_{y \text{ in data}} \text{Loss}(\text{pred}, y)$$

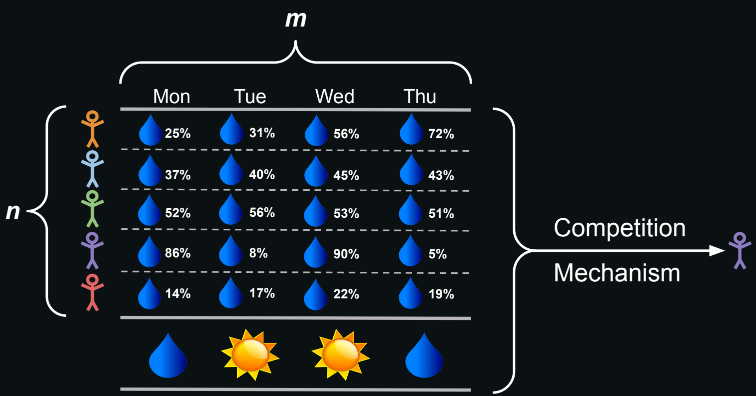
Forecasting Competitions

Kaggle, Good Judgement Project, Hybrid Forecasting Competition, ...



Forecasting Competitions

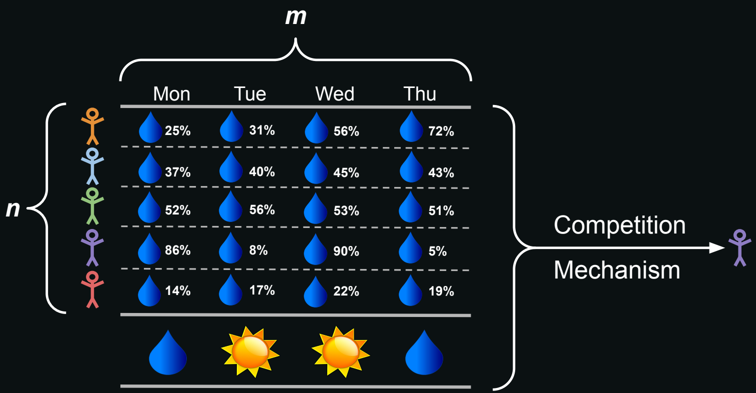
Kaggle, Good Judgement Project, Hybrid Forecasting Competition, ...



How to pick the winner?

Forecasting Competitions

Kaggle, Good Judgement Project, Hybrid Forecasting Competition, ...



How to pick the winner? Usually with Simple Max.

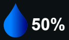
Strategic Forecasters

Toy example: $n = 3$, $m = 1$, truth is 50%. Who wins?

Strategic Forecasters

Toy example: $n = 3$, $m = 1$, truth is 50%. Who wins?

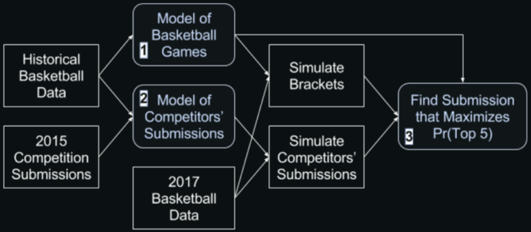
Optimal strategy: deviate to 0 or 1.

Toy example: $n = 3$, $m = 1$, truth is 

. Who wins?

Optimal strategy: deviate to 0 or 1.

Real life: Kaggle March Mania 2017



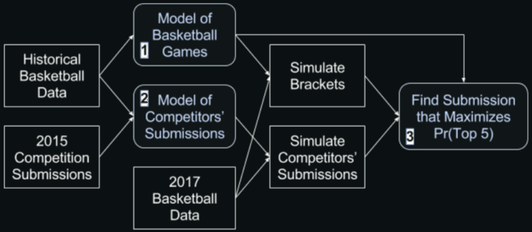
Andrew Landgraf, Kaggle [2017]

Strategic Forecasters

Toy example: $n = 3$, $m = 1$, truth is 50%. Who wins?

Optimal strategy: deviate to 0 or 1.

Real life: Kaggle March Mania 2017



Andrew Landgraf, Kaggle [2017]

What's the problem?

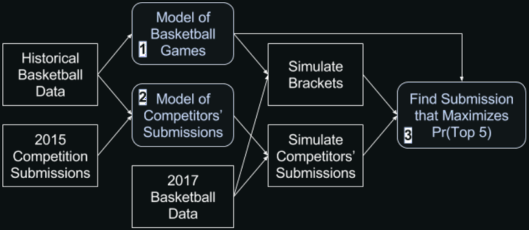
1. Accuracy: Picking the *best forecaster*? beliefs ≈ true probabilities

Strategic Forecasters

Toy example: $n = 3$, $m = 1$, truth is 50%. Who wins?

Optimal strategy: deviate to 0 or 1.

Real life: Kaggle March Mania 2017



Andrew Landgraf, Kaggle [2017]

What's the problem?

1. Accuracy: Picking the *best forecaster*? — beliefs ≈ true probabilities
2. Wasted effort: Forecasting vs. strategizing

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

Truthful: impossible to game.

Accurate: picks best forecaster w.h.p. when $m = O(n^2 \log n)$.

Recall: n forecasters, m events

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

Truthful: impossible to game.

Accurate: picks best forecaster w.h.p. when $m = O(n^2 \log n)$.

Recall: n forecasters, m events

Is $O(n^2 \log n)$ events practical?

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

Truthful: impossible to game.

Accurate: picks best forecaster w.h.p. when $m = O(n^2 \log n)$.

Recall: n forecasters, m events

Is $O(n^2 \log n)$ events practical?

Kaggle 2017 March Mania: $n = 442$, but only $m = 2278$ events!

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

Truthful: impossible to game.

Accurate: picks best forecaster w.h.p. when $m = O(n^2 \log n)$.

Recall: n forecasters, m events

Is $O(n^2 \log n)$ events practical?

Kaggle 2017 March Mania: $n = 442$, but only $m = 2278$ events!

Central Question

Can we pick the best forecaster using fewer events?

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

Truthful: impossible to game.

Accurate: picks best forecaster w.h.p. when $m = O(n^2 \log n)$.

Recall: n forecasters, m events

Is $O(n^2 \log n)$ events practical?

Kaggle 2017 March Mania: $n = 442$, but only $m = 2278$ events!

Central Question

Can we pick the best forecaster using fewer events?

Best possible: $\Omega(\log n)$ PAC learning lower bounds

Overview

Central Question

Can we pick the best forecaster using fewer than $O(n^2 \log n)$ events?

Overview

Central Question

Can we pick the best forecaster using fewer than $O(n^2 \log n)$ events?

1. A tight analysis of ELF
   ELF only needs $\Theta(n \log n)$ events

Overview

Central Question

Can we pick the best forecaster using fewer than $O(n^2 \log n)$ events?

1. A tight analysis of ELF
   ELF only needs $\Theta(n \log n)$ events
2. A new competition mechanism
   Even fewer events...

Overview

Central Question

Can we pick the best forecaster using fewer than $O(n^2 \log n)$ events?

1. A tight analysis of ELF
   ELF only needs $\Theta(n \log n)$ events
2. A new competition mechanism
   Even fewer events...
3. Application to online learning
   No-regret even with strategic experts

1. A tight analysis of ELF

Notation, Accuracy

For each event $t = 1, \ldots, m$:

$$
\begin{aligned}
p_{it} &\in [0,1] && \text{forecaster } i\text{'s belief} \\
r_{it} &\in [0,1] && \text{forecaster } i\text{'s report} \\
\theta_t &\in [0,1] && \text{ground truth probability} \\
y_t &\in \{0,1\} && \text{actual outcome} \\
S(r,y) &\in [0,1] && \text{scoring rule, e.g. } S(r,y) = 1 - (r-y)^2
\end{aligned}
$$

Notation, Accuracy

For each event $t = 1, \ldots, m$:

$$\begin{aligned} p_{it} &\in [0,1] && \text{forecaster } i\text{'s belief} \\ r_{it} &\in [0,1] && \text{forecaster } i\text{'s report} \\ \theta_t &\in [0,1] && \text{ground truth probability} \\ y_t &\in \{0,1\} && \text{actual outcome} \\ S(r,y) &\in [0,1] && \text{scoring rule, e.g. } S(r,y) = 1 - (r-y)^2 \end{aligned}$$

A forecaster's accuracy: the true expected score of their beliefs,

$$a_i = \frac{1}{m} \sum_{t=1}^{m} \mathop{\mathbb{E}}_{y_t \sim \theta_t} S(p_{it}, y_t) \ .$$

Goal: pick a forecaster whose accuracy is within ϵ of the best (w.h.p).

BTW: dependence on ϵ is always $1/\epsilon^2$

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = O(n^2 \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Hoeffding

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = O(n^2 \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Hoeffding

Observation: $\Pr[i \text{ wins point } t] \leq \frac{2}{n}$

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = O(n^2 \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Hoeffding

Observation: $\Pr[i \text{ wins point } t] \leq \frac{2}{n} \implies$ Low variance

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1} \right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = O(n \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Bernstein

Observation: $\Pr[i \text{ wins point } t] \leq \frac{2}{n} \implies$ Low variance $\implies$ Upper bound!

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = O(n \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Bernstein

Observation: $\Pr[i \text{ wins point } t] \leq \frac{2}{n} \implies$ Almost uniform

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = O(n \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Bernstein

Observation: $\Pr[i \text{ wins point } t] \leq \frac{2}{n} \implies$ Almost uniform $\implies$ Lower bound!

Best case scenario for ELF:

Forecaster i is perfect. $\Pr[i \text{ wins point } t] = \frac{2}{n}$
Forecaster $j \neq i$ is terrible. $\Pr[j \text{ wins point } t] \approx \frac{1}{n}$

A Tighter ELF Analysis: $\Theta(n \log n)$

Event Lotteries Forecaster (ELF) [Witkowski et al., 2018]

For each event t, assign a point with a lottery:

$$\Pr[i \text{ wins point } t] = \frac{1}{n} + \frac{1}{n}\left(S(r_{it}, y_t) - \frac{\sum_{j \neq i} S(r_{jt}, y_t)}{n-1}\right)$$

Forecaster with highest point total wins.

Truthful: Optimal strategy is $r_{it} = p_{it}$. $\max \Pr[\text{point}] \equiv \max \mathbb{E}[\text{score}]$

Accurate: Picks best forecaster with $m = \Theta(n \log n)$ events.

Proof: m large enough so point totals $\approx$ true accuracies via Bernstein

Observation: $\Pr[i \text{ wins point } t] \leq \frac{2}{n} \implies$ Almost uniform $\implies$ Lower bound!

Best case scenario for ELF:

Forecaster i is perfect. $\Pr[i \text{ wins point } t] = \frac{2}{n}$
Forecaster $j \neq i$ is terrible. $\Pr[j \text{ wins point } t] \approx \frac{1}{n}$

Balls and bins: if $m < \frac{1}{8} n \log n$, a bad forecaster wins w.h.p.

So, ELF needs $\Theta(n \log n)$ events.

So, ELF needs $\Theta(n \log n)$ events.

Can we do better?

2. A New (Old) Mechanism

Follow The Regularized Leader

Why didn't Simple Max work?

Follow The Regularized Leader

Why didn't Simple Max work? One answer: sensitivity.

Small change in input (report) $\implies$ big change in output (winner)

Follow The Regularized Leader

Why didn't Simple Max work? One answer: sensitivity.

Small change in input (report) $\implies$ big change in output (winner)

In machine learning, we address sensitivity with regularization.

Follow The Regularized Leader

Why didn't Simple Max work? One answer: sensitivity.

Small change in input (report) $\implies$ big change in output (winner)

In machine learning, we address sensitivity with regularization.

FTRL: Choose a distribution π over the forecasters which maximizes the expected forecaster score (under π) minus a regularization term $\mathcal{R}(\pi)$.

$$\arg\max_{\pi} \sum_i \pi_i \sum_{t=1}^{m} S(r_{it}, y_t) - \mathcal{R}(\pi)$$

Follow The Regularized Leader

Why didn't Simple Max work? One answer: sensitivity.

Small change in input (report) $\implies$ big change in output (winner)

In machine learning, we address sensitivity with regularization.

FTRL: Choose a distribution π over the forecasters which maximizes the expected forecaster score (under π) minus a regularization term $\mathcal{R}(\pi)$.

$$\arg\max_{\pi} \sum_i \pi_i \sum_{t=1}^{m} S(r_{it}, y_t) - \mathcal{R}(\pi)$$

Simple Max: $\mathcal{R} = 0$, and π puts all weight on one forecaster.

Follow The Regularized Leader

Why didn't Simple Max work? One answer: sensitivity.

Small change in input (report) $\implies$ big change in output (winner)

In machine learning, we address sensitivity with regularization.

FTRL: Choose a distribution π over the forecasters which maximizes the expected forecaster score (under π) minus a regularization term $\mathcal{R}(\pi)$.

$$\arg\max_{\pi} \sum_i \pi_i \sum_{t=1}^{m} S(r_{it}, y_t) - \mathcal{R}(\pi)$$

Simple Max: $\mathcal{R} = 0$, and π puts all weight on one forecaster.
Multiplicative Weights: $\mathcal{R}$ is negative entropy $\mathcal{R}(\pi) = \sum_i \pi_i \log \pi_i$.

Follow The Regularized Leader

Why didn't Simple Max work? One answer: sensitivity.

Small change in input (report) $\implies$ big change in output (winner)

In machine learning, we address sensitivity with regularization.

FTRL: Choose a distribution π over the forecasters which maximizes the expected forecaster score (under π) minus a regularization term $\mathcal{R}(\pi)$.

$$\arg\max_{\pi} \sum_i \pi_i \sum_{t=1}^{m} S(r_{it}, y_t) - \mathcal{R}(\pi)$$

Simple Max: $\mathcal{R} = 0$, and π puts all weight on one forecaster.
Multiplicative Weights: $\mathcal{R}$ is negative entropy $\mathcal{R}(\pi) = \sum_i \pi_i \log \pi_i$.

Can we use FTRL as a *batch* algorithm for forecasting competitions?

Multiplicative Weights

Choose forecasters using the distribution:

$$\pi_i = \frac{\exp\left(\eta \sum_{t=1}^{m} S(r_{it}, y_t)\right)}{\sum_{j=1}^{n} \exp\left(\eta \sum_{t=1}^{m} S(r_{jt}, y_t)\right)}$$

Multiplicative Weights

Choose forecasters using the distribution:

$$\pi_i = \frac{\exp\left(\eta \sum_{t=1}^{m} S(r_{it}, y_t)\right)}{\sum_{j=1}^{n} \exp\left(\eta \sum_{t=1}^{m} S(r_{jt}, y_t)\right)}$$

Is Multiplicative Weights truthful? Not quite...

Multiplicative Weights

Choose forecasters using the distribution:

$$\pi_i = \frac{\exp\left(\eta \sum_{t=1}^{m} S(r_{it}, y_t)\right)}{\sum_{j=1}^{n} \exp\left(\eta \sum_{t=1}^{m} S(r_{jt}, y_t)\right)}$$

Is Multiplicative Weights truthful? Not quite...

Game theory: dominated strategies

Let A, B be strategies for player i. We say A dominates B if, no matter what other players do, A is better for player i than B.

Multiplicative Weights

Choose forecasters using the distribution:

$$\pi_i = \frac{\exp\left(\eta \sum_{t=1}^{m} S(r_{it}, y_t)\right)}{\sum_{j=1}^{n} \exp\left(\eta \sum_{t=1}^{m} S(r_{jt}, y_t)\right)}$$

Is Multiplicative Weights truthful? Not quite...

Game theory: dominated strategies

Let A, B be strategies for player i. We say A dominates B if, no matter what other players do, A is better for player i than B.

Truthful: reporting $r_{it} = p_{it}$ for all t is a dominant strategy.

Multiplicative Weights

Choose forecasters using the distribution:

$$\pi_i = \frac{\exp\left(\eta \sum_{t=1}^{m} S(r_{it}, y_t)\right)}{\sum_{j=1}^{n} \exp\left(\eta \sum_{t=1}^{m} S(r_{jt}, y_t)\right)}$$

Is Multiplicative Weights truthful? Not quite...

Game theory: dominated strategies

Let A, B be strategies for player i. We say A dominates B if, no matter what other players do, A is better for player i than B.

Truthful: reporting $r_{it} = p_{it}$ for all t is a dominant strategy.
γ-Approximately Truthful: reports with $|r_{it} - p_{it}| > \gamma$ are dominated.

Multiplicative Weights

Choose forecasters using the distribution:

$$\pi_i = \frac{\exp\left(\eta \sum_{t=1}^{m} S(r_{it}, y_t)\right)}{\sum_{j=1}^{n} \exp\left(\eta \sum_{t=1}^{m} S(r_{jt}, y_t)\right)}$$

Is Multiplicative Weights truthful? Not quite...

Game theory: dominated strategies

Let A, B be strategies for player i. We say A dominates B if, no matter what other players do, A is better for player i than B.

Truthful: reporting $r_{it} = p_{it}$ for all t is a dominant strategy.

γ-Approximately Truthful: reports with $|r_{it} - p_{it}| > \gamma$ are dominated.

Theorem

Multiplicative Weights is 4η-approximately truthful.

For small η, reports $\approx$ beliefs!

Multiplicative Weights is Accurate

Theorem

Multiplicative Weights chooses an ϵ-accurate forecaster with high probability if $m = O(\log n/\epsilon^2)$.

Matches the best possible bound!

Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$.

Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$. Define:

$$\hat{S}_i = \frac{1}{m}\sum_t S(r_{it}, y_t) \qquad i\text{'s empirical score}$$

$$S_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(r_{it}, y_t)\right] \qquad i\text{'s expected score}$$

$$a_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(p_{it}, y_t)\right] \qquad i\text{'s accuracy}$$

Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$. Define:

$$\hat{S}_i = \quad \frac{1}{m}\sum_t S(r_{it}, y_t) \qquad \text{i's empirical score}$$

$$S_i = \mathbb{E}_{\vec{\theta}}[\, \frac{1}{m}\sum_t S(r_{it}, y_t)\,] \qquad \text{i's expected score}$$

$$a_i = \mathbb{E}_{\vec{\theta}}[\, \frac{1}{m}\sum_t S(p_{it}, y_t)\,] \qquad \text{i's accuracy}$$

Let i be the best forecaster, and j be any ϵ-bad one. So $a_i - a_j > \epsilon$.

Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$. Define:

$$\hat{S}_i = \quad \frac{1}{m}\sum_t S(r_{it}, y_t) \qquad \text{i's empirical score}$$

$$S_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(r_{it}, y_t)\right] \qquad \text{i's expected score}$$

$$a_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(p_{it}, y_t)\right] \qquad \text{i's accuracy}$$

Let i be the best forecaster, and j be any ϵ-bad one. So $a_i - a_j > \epsilon$.

Accuracy gap

Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$. Define:

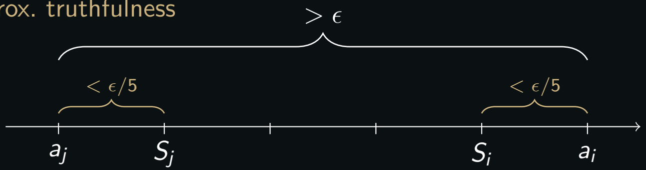
$$\hat{S}_i = \frac{1}{m}\sum_t S(r_{it}, y_t) \qquad i\text{'s empirical score}$$

$$S_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(r_{it}, y_t)\right] \qquad i\text{'s expected score}$$

$$a_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(p_{it}, y_t)\right] \qquad i\text{'s accuracy}$$

Let i be the best forecaster, and j be any ϵ-bad one. So $a_i - a_j > \epsilon$.

Approx. truthfulness



Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$. Define:

$$\hat{S}_i = \frac{1}{m}\sum_t S(r_{it}, y_t) \qquad i\text{'s empirical score}$$

$$S_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(r_{it}, y_t)\right] \qquad i\text{'s expected score}$$

$$a_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(p_{it}, y_t)\right] \qquad i\text{'s accuracy}$$

Let i be the best forecaster, and j be any ϵ-bad one. So $a_i - a_j > \epsilon$.

Hoeffding

Proof

Take $\eta = O(\epsilon)$, $m = O\left(\frac{\log n}{\epsilon^2}\right)$. Define:

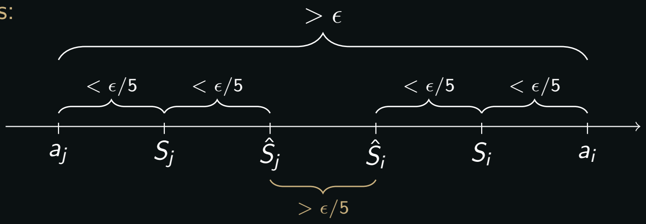
$$\hat{S}_i = \frac{1}{m}\sum_t S(r_{it}, y_t) \qquad \text{i's empirical score}$$

$$S_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(r_{it}, y_t)\right] \qquad \text{i's expected score}$$

$$a_i = \mathbb{E}_{\vec{\theta}}\left[\frac{1}{m}\sum_t S(p_{it}, y_t)\right] \qquad \text{i's accuracy}$$

Let i be the best forecaster, and j be any ϵ-bad one. So $a_i - a_j > \epsilon$.

Thus:



Proof

So w.h.p $\hat{S}_i - \hat{S}_j > \epsilon/5$.

Proof

So w.h.p $\hat{S}_i - \hat{S}_j > \epsilon/5$.

Multiplicative Weights chooses the winner according to

$$\pi_i = \frac{\exp\left(\eta m \hat{S}_i\right)}{\sum_{k=1}^{n} \exp\left(\eta m \hat{S}_k\right)}$$

Proof

So w.h.p $\hat{S}_i - \hat{S}_j > \epsilon/5$.

Multiplicative Weights chooses the winner according to

$$\pi_i = \frac{\exp\left(\eta m \hat{S}_i\right)}{\sum_{k=1}^{n} \exp\left(\eta m \hat{S}_k\right)}$$

How much more likely is i to win than j?

Proof

So w.h.p $\hat{S}_i - \hat{S}_j > \epsilon/5$.

Multiplicative Weights chooses the winner according to

$$\pi_i = \frac{\exp\left(\eta m \hat{S}_i\right)}{\sum_{k=1}^{n} \exp\left(\eta m \hat{S}_k\right)}$$

How much more likely is i to win than j?

$$\pi_i/\pi_j = \exp\left(\eta m(\hat{S}_i - \hat{S}_j)\right)$$

Proof

So w.h.p $\hat{S}_i - \hat{S}_j > \epsilon/5$.

Multiplicative Weights chooses the winner according to

$$\pi_i = \frac{\exp\left(\eta m \hat{S}_i\right)}{\sum_{k=1}^{n} \exp\left(\eta m \hat{S}_k\right)}$$

How much more likely is i to win than j?

$$\pi_i/\pi_j = \exp\left(\eta m(\hat{S}_i - \hat{S}_j)\right) > \exp(\eta m \epsilon/5) = \exp(\log(n/\delta)) = n/\delta \ .$$

Recall $\eta = O(\epsilon)$. Take $m = O(\frac{\log(n/\delta)}{\epsilon^2})$ for small constant δ.

Proof

So w.h.p $\hat{S}_i - \hat{S}_j > \epsilon/5$.

Multiplicative Weights chooses the winner according to

$$\pi_i = \frac{\exp\left(\eta m \hat{S}_i\right)}{\sum_{k=1}^n \exp\left(\eta m \hat{S}_k\right)}$$

How much more likely is i to win than j?

$$\pi_i/\pi_j = \exp\left(\eta m(\hat{S}_i - \hat{S}_j)\right) > \exp(\eta m \epsilon/5) = \exp(\log(n/\delta)) = n/\delta \ .$$

Recall $\eta = O(\epsilon)$. Take $m = O(\frac{\log(n/\delta)}{\epsilon^2})$ for small constant δ.

Finally, since $\pi_i \leq 1$, we have

$$\pi_j \leq \pi_j/\pi_i < \delta/n \ ,$$

so

$$\Pr[\epsilon\text{-bad forecaster wins}] \leq \sum_{j \text{ is } \epsilon\text{-bad}} \pi_j < n(\delta/n) = \delta \ . \qquad \square$$

3. Online Learning from Strategic Experts

When you can't trust your experts...

Classic online learning from expert advice:

- Experts give advice p_{it} on each round $t = 1, \ldots, T$
- Algorithm chooses some expert i and predicts their p_{it}
- Algorithm's goal: perform not much worse than the best expert

When you can't trust your experts...

Classic online learning from expert advice:

- Experts give advice p_{it} on each round $t = 1, \ldots, T$
- Algorithm chooses some expert i and predicts their p_{it}
- Algorithm's goal: perform not much worse than the best expert

Strategic experts:

- Now experts report some r_{it}, potentially $\neq p_{it}$
- Experts' goal: maximize chance of being chosen!
- Algorithm's goal: perform not much worse than the best expert

  Best: still according to their p_{it}'s

When you can't trust your experts...

Classic online learning from expert advice:

- Experts give advice p_{it} on each round $t = 1, \dots, T$
- Algorithm chooses some expert i and predicts their p_{it}
- Algorithm's goal: perform not much worse than the best expert

Strategic experts:

- Now experts report some r_{it}, potentially $\neq p_{it}$
- Experts' goal: maximize chance of being chosen!
- Algorithm's goal: perform not much worse than the best expert

    Best: still according to their p_{it}'s

Open: find such an algorithm for strategic experts

Myopic case: [Roughgarden and Schrijvers, 2017, Freeman et al., 2020]

When you can't trust your experts...

Classic online learning from expert advice:

- Experts give advice p_{it} on each round $t = 1, \dots, T$
- Algorithm chooses some expert i and predicts their p_{it}
- Algorithm's goal: perform not much worse than the best expert

Strategic experts:

- Now experts report some r_{it}, potentially $\neq p_{it}$
- Experts' goal: maximize chance of being chosen!
- Algorithm's goal: perform not much worse than the best expert
  Best: still according to their p_{it}'s

Open: find such an algorithm for strategic experts

Myopic case: [Roughgarden and Schrijvers, 2017, Freeman et al., 2020]

Theorem

Mult. Weights achieves $O(\sqrt{T})$ regret even with strategic experts.

Proof Sketch

FTRL with $\eta = O(1/\sqrt{T})$ has $O(\sqrt{T})$ regret w.r.t. expert reports.

Approx. truthfulness: reports are within $O(\eta) = O(1/\sqrt{T})$ of beliefs.

So we pick up at most $T \cdot O(1/\sqrt{T}) = O(\sqrt{T})$ extra regret.

Recap and Future Directions

What We Showed

1. Competitions: better event complexity

   ELF needs $\Theta(n \log n)$ events, FTRL achieves *optimal* $\Theta(\log n)$
2. Online learning: $O(\sqrt{T})$ regret with strategic experts

Recap and Future Directions

What We Showed

1. Competitions: better event complexity

   ELF needs $\Theta(n \log n)$ events, FTRL achieves *optimal* $\Theta(\log n)$
2. Online learning: $O(\sqrt{T})$ regret with strategic experts

Takeaways

1. Forecasting competitions: Kaggle should add regularization!

Recap and Future Directions

What We Showed

1. Competitions: better event complexity

   ELF needs $\Theta(n \log n)$ events, FTRL achieves *optimal* $\Theta(\log n)$

2. Online learning: $O(\sqrt{T})$ regret with strategic experts

Takeaways

1. Forecasting competitions: Kaggle should add regularization!
2. Algorithmic economics: *what is the cost of strategic behavior?*

   Price of anarchy: social welfare of Nash vs optimal; sometimes large

Recap and Future Directions

What We Showed

1. Competitions: better event complexity
   ELF needs $\Theta(n \log n)$ events, FTRL achieves *optimal* $\Theta(\log n)$
2. Online learning: $O(\sqrt{T})$ regret with strategic experts

Takeaways

1. Forecasting competitions: Kaggle should add regularization!
2. Algorithmic economics: *what is the cost of strategic behavior?*
   Price of anarchy: social welfare of Nash vs optimal; sometimes large
   Forecasting competitions and online learning: no cost!

Recap and Future Directions

What We Showed

1. Competitions: better event complexity
   ELF needs $\Theta(n \log n)$ events, FTRL achieves *optimal* $\Theta(\log n)$
2. Online learning: $O(\sqrt{T})$ regret with strategic experts

Takeaways

1. Forecasting competitions: Kaggle should add regularization!
2. Algorithmic economics: *what is the cost of strategic behavior?*
   Price of anarchy: social welfare of Nash vs optimal; sometimes large
   Forecasting competitions and online learning: no cost!

Extensions and Open Questions

- ✓ Beyond binary outcomes, other scoring rules & regularizers
- ✓ More realistic beliefs — Can correlate with truth
- ?? Exactly truthful mechanism
- ?? Correlated events

Recap and Future Directions

Thanks for listening!

What We Showed

1. Competitions: better event complexity
   ELF needs $\Theta(n \log n)$ events, FTRL achieves *optimal* $\Theta(\log n)$
2. Online learning: $O(\sqrt{T})$ regret with strategic experts

Takeaways

1. Forecasting competitions: Kaggle should add regularization!
2. Algorithmic economics: *what is the cost of strategic behavior?*
   Price of anarchy: social welfare of Nash vs optimal; sometimes large
   Forecasting competitions and online learning: no cost!

Extensions and Open Questions

✓ Beyond binary outcomes, other scoring rules & regularizers
✓ More realistic beliefs — Can correlate with truth
?? Exactly truthful mechanism
?? Correlated events

References

Rupert Freeman, David Pennock, Chara Podimata, and Jennifer Wortman Vaughan. No-regret and incentive-compatible online learning. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 3270–3279. PMLR, 13–18 Jul 2020. URL http://proceedings.mlr.press/v119/freeman20a.html.

Kaggle. March machine learning mania, 1st place winner's interview: Andrew landgraf. http://blog.kaggle.com/2017/05/19/march-machine-learning-mania-1st-place-winners-interview-andrew-landgraf/, May 2017. Accessed: 6/29/2019.

Tim Roughgarden and Okke Schrijvers. Online prediction with selfish experts. *arXiv preprint arXiv:1702.03615*, 2017.

Jens Witkowski, Rupert Freeman, Jennifer Wortman Vaughan, David M. Pennock, and Andreas Krause. Incentive-compatible forecasting competitions. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, AAAI 2018, 2018.