

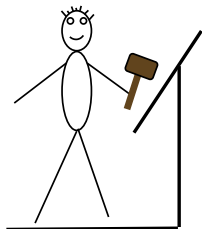
Information Elicitation: Past, Present, and Future



Bo Waggoner
CU Boulder / Gensyn

CU Denver
October 19, 2026

$$\operatorname{argmax}_{r \in \mathcal{R}} \mathbb{E}_{y \sim p} S(r, y)$$



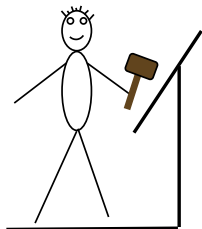
Designer



time



event



Designer



time

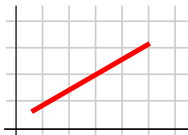
Score
 $S(p, \text{rain})$



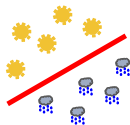
event



algorithm



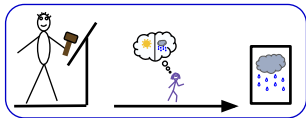
hypothesis



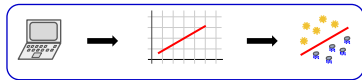
performance

$$\operatorname{argmax}_{r \in \mathcal{R}} \mathbb{E}_{y \sim p} S(r, y)$$

game theory, prediction



machine learning, hypotheses



Information Elicitation

Outline

1 Scoring rules and prediction markets

Eliciting and aggregating information

2 Statistics and machine learning

Structure of prediction spaces

3 Market frameworks for ML

Bringing it together

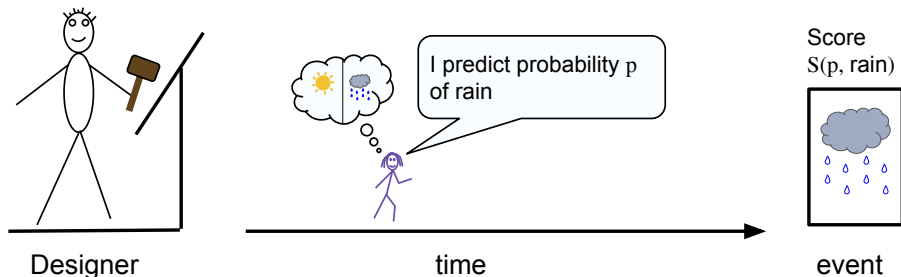
Predicting a two-outcome event

- 1 agent has belief q
- 2 agent predicts p
- 3 event y is observed
- 4 agent gets **score** $S(p, y)$.

e.g. $q = \Pr[\text{rain}]$

hopefully $p = q$

e.g. $y = \text{rain}$



The designer's question

Agent acts to maximize **expected score**.

$$\operatorname{argmax}_p \mathbb{E}_{y \sim q} S(p, y).$$

The designer's question

Agent acts to maximize **expected score**.

$$\operatorname{argmax}_p \mathbb{E}_{y \sim q} S(p, y).$$

Question: *How should we design S to be **proper**: the agent reports her true belief q ?*

The designer's question

Agent acts to maximize **expected score**.

$$\operatorname{argmax}_p \mathbb{E}_{y \sim q} S(p, y).$$

Question: *How should we design S to be **proper**: the agent reports her true belief q ?*

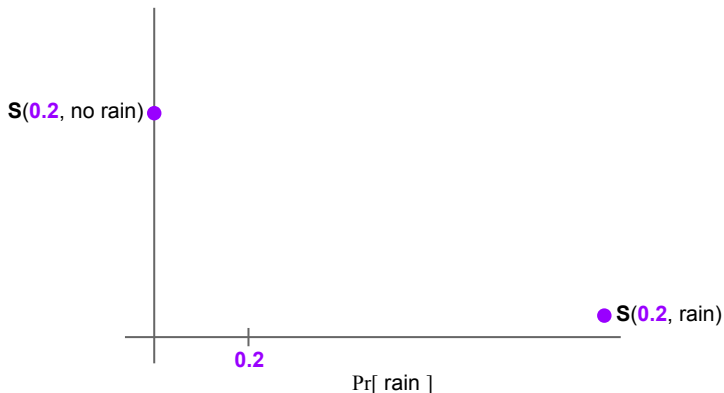
Answer (McCarthy 1956; Savage 1971): *Use convex functions!*

Proper scoring rules

Example (Good, 1951): $S(p, y) = \begin{cases} \log(p) & \text{if } y = \text{rain} \\ \log(1 - p) & \text{otherwise.} \end{cases}$

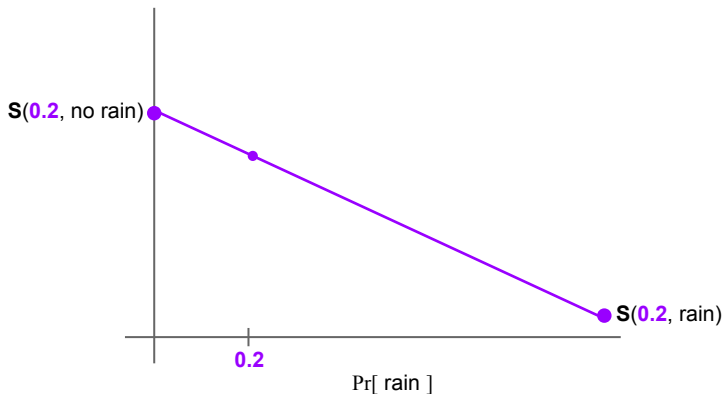
Proper scoring rules

Example (Good, 1951): $S(p, y) = \begin{cases} \log(p) & \text{if } y = \text{rain} \\ \log(1 - p) & \text{otherwise.} \end{cases}$



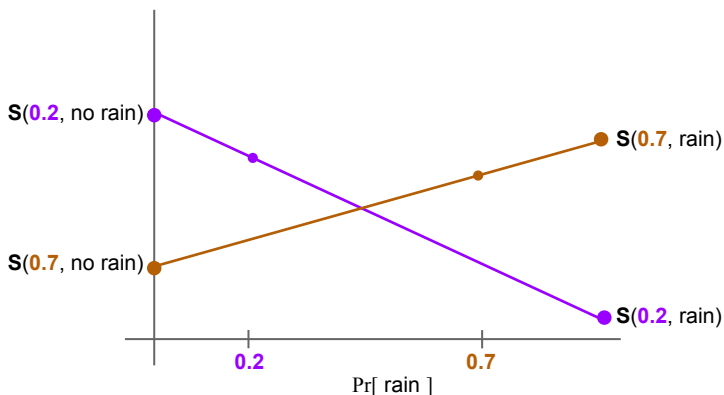
Proper scoring rules

Example (Good, 1951): $S(p, y) = \begin{cases} \log(p) & \text{if } y = \text{rain} \\ \log(1 - p) & \text{otherwise.} \end{cases}$



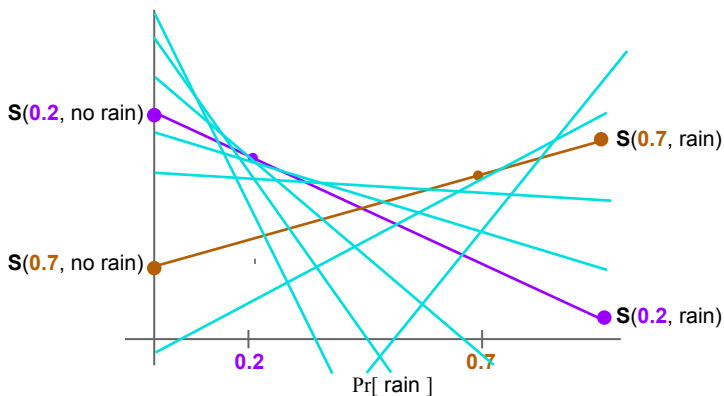
Proper scoring rules

Example (Good, 1951): $S(p, y) = \begin{cases} \log(p) & \text{if } y = \text{rain} \\ \log(1 - p) & \text{otherwise.} \end{cases}$



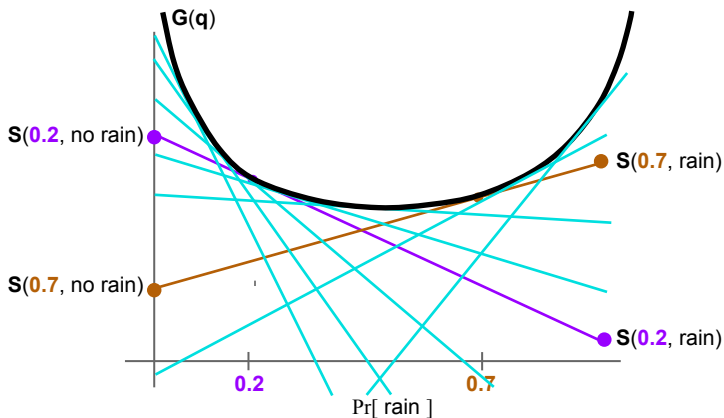
Proper scoring rules

Example (Good, 1951): $S(p, y) = \begin{cases} \log(p) & \text{if } y = \text{rain} \\ \log(1 - p) & \text{otherwise.} \end{cases}$



Proper scoring rules

Example (Good, 1951): $S(p, y) = \begin{cases} \log(p) & \text{if } y = \text{rain} \\ \log(1 - p) & \text{otherwise.} \end{cases}$



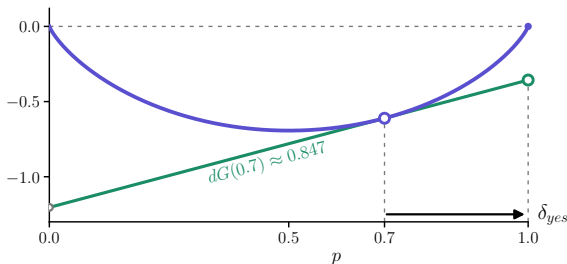
Characterization of scoring rules

Theorem (McCarthy 1956; Savage 1971; ...)

A scoring rule $S : \Delta_{\mathcal{Y}} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ is (strictly) proper if and only if there exists a (strictly) convex $G : \Delta_{\mathcal{Y}} \rightarrow \mathbb{R}$ such that

$$S(p, y) = G(p) + \langle dG(p), \delta_y - p \rangle$$

where $dG(p)$ is a subgradient and δ_y is the indicator on y .



Characterization of scoring rules

Theorem (McCarthy 1956; Savage 1971; ...)

A scoring rule $S : \Delta_{\mathcal{Y}} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ is (strictly) proper if and only if there exists a (strictly) convex $G : \Delta_{\mathcal{Y}} \rightarrow \mathbb{R}$ such that

$$S(p, y) = G(p) + \langle dG(p), \delta_y - p \rangle$$

where $dG(p)$ is a subgradient and δ_y is the indicator on y .

Example (Good, 1951):

$$S(p, y) = \log p(y)$$

$$G(p) = \sum_{y \in \mathcal{Y}} p(y) \log p(y)$$

Characterization of scoring rules

Theorem (McCarthy 1956; Savage 1971; ...)

A scoring rule $S : \Delta_{\mathcal{Y}} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ is (strictly) proper if and only if there exists a (strictly) convex $G : \Delta_{\mathcal{Y}} \rightarrow \mathbb{R}$ such that

$$S(p, y) = G(p) + \langle dG(p), \delta_y - p \rangle$$

where $dG(p)$ is a subgradient and δ_y is the indicator on y .

Example (Brier, 1950):

$$S(p, y) = -\|\delta_y - p\|_2^2$$

$$G(p) = \|p\|_2^2$$

Characterization of scoring rules

Theorem (McCarthy 1956; Savage 1971; ...)

A scoring rule $S : \Delta_{\mathcal{Y}} \times \mathcal{Y} \rightarrow \mathbb{R} \cup \{-\infty\}$ is (strictly) proper if and only if there exists a (strictly) convex $G : \Delta_{\mathcal{Y}} \rightarrow \mathbb{R}$ such that

$$S(p, y) = G(p) + \langle dG(p), \delta_y - p \rangle$$

where $dG(p)$ is a subgradient and δ_y is the indicator on y .

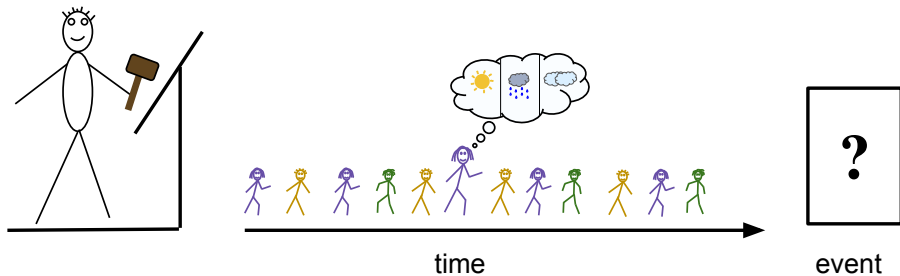
Corollary

If an agent believes p ,

- *The expected score for truthfulness is $G(p)$.*
- *The expected penalty for misreporting p' is $D_G(p, p')$
the Bregman divergence*

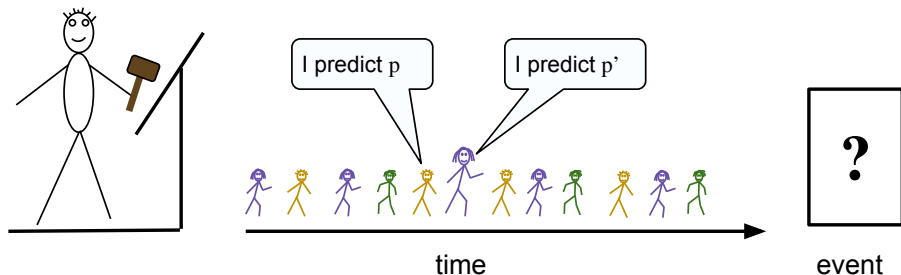
Prediction markets [Hanson 2003]

- agents arrive sequentially and update the current prediction
- when event occurs, payoff is **difference in score**



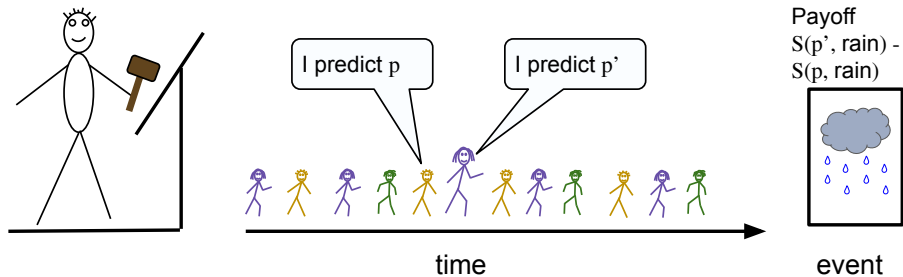
Prediction markets [Hanson 2003]

- agents arrive sequentially and update the current prediction
- when event occurs, payoff is **difference in score**



Prediction markets [Hanson 2003]

- agents arrive sequentially and update the current prediction
- when event occurs, payoff is **difference in score**



Markets: convex optimization (ACW 2013)

Equivalent process:

- Initial prediction $p^{(0)}$

Markets: convex optimization (ACW 2013)

Equivalent process:

- Initial prediction $p^{(0)}$
- Agent arrives, demands “bundle” $\theta^{(1)} \in \mathbb{R}^{\mathcal{Y}}$

Markets: convex optimization (ACW 2013)

Equivalent process:

- Initial prediction $p^{(0)}$
- Agent arrives, demands “bundle” $\theta^{(1)} \in \mathbb{R}^{\mathcal{Y}}$
- Agent pays $\approx p^{(0)} \cdot \theta^{(1)}$

Markets: convex optimization (ACW 2013)

Equivalent process:

- Initial prediction $p^{(0)}$
- Agent arrives, demands “bundle” $\theta^{(1)} \in \mathbb{R}^{\mathcal{Y}}$
- Agent pays $\approx p^{(0)} \cdot \theta^{(1)}$
 - If y happens, receives payoff $\theta^{(1)}(y)$
it's a market

Markets: convex optimization (ACW 2013)

Equivalent process:

- Initial prediction $p^{(0)}$
- Agent arrives, demands “bundle” $\theta^{(1)} \in \mathbb{R}^{\mathcal{Y}}$
- Agent pays $\approx p^{(0)} \cdot \theta^{(1)}$
 - If y happens, receives payoff $\theta^{(1)}(y)$
it's a market
- New prediction $p^{(1)} = \arg \max_p [p \cdot \theta^{(1)} - G(p)]$
it's an optimization algorithm

Markets: convex optimization (ACW 2013)

Equivalent process:

- Previous prediction $p^{(t)}$
- Agent arrives, demands “bundle” $\theta^{(t+1)} \in \mathbb{R}^{\mathcal{Y}}$
- Agent pays $\approx p^{(t)} \cdot \theta^{(t+1)}$
 - If y happens, receives payoff $\theta^{(t+1)}(y)$
it's a market
- New prediction $p^{(t+1)} = \arg \max_p \left[\sum_{s=1}^{t+1} p \cdot \theta^{(s)} - G(p) \right]$
it's an optimization algorithm

Specifically, **Follow the Regularized Leader (FTRL)**

- Log scoring rule: **Multiplicative Weights**
- Quadratic (Brier) score: **Online Gradient Descent**

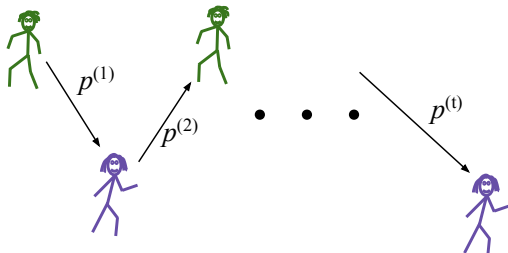
Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?

Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?

Aumann (1976), Geanakoplos and Polemarchakis (1981):
Bayesians, exchanging current beliefs p , **eventually agree**.

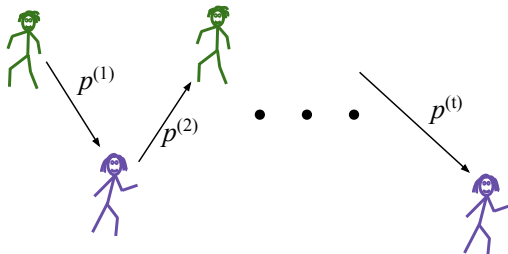


Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?

Aumann (1976), Geanakoplos and Polemarchakis (1981):
Bayesians, exchanging current beliefs p , **eventually agree**.

Aaronson (2003): they agree to within ϵ in $O(1/\epsilon^2)$ rounds!



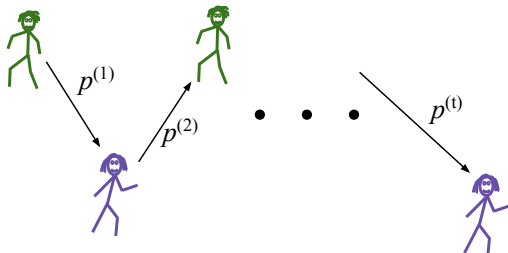
Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?

Aumann (1976), Geanakoplos and Polemarchakis (1981): Bayesians, exchanging current beliefs p , **eventually agree**.

Aaronson (2003): they agree to within ϵ in $O(1/\epsilon^2)$ rounds!

Frongillo, Neyman, W. (2023): if *informational substitutes*, they are furthermore **accurate**. *see also: Kong and Schoenenbeck (2024)*



Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?

Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?

Hanson (2003): if each agent participates once, **truthful**.

Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?

Hanson (2003): if each agent participates once, **truthful**.

Ostrovsky (2012): if “separable” information, *eventually* aggregated in equilibrium.

Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?

Hanson (2003): if each agent participates once, **truthful**.

Ostrovsky (2012): if “separable” information, *eventually* aggregated in equilibrium.

Chen and W. (2016): iff *informational substitutes*, **immediately** aggregated in equilibrium.

Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?

Hanson (2003): if each agent participates once, **truthful**.

Ostrovsky (2012): if “separable” information, *eventually* aggregated in equilibrium.

Chen and W. (2016): iff *informational substitutes*, **immediately** aggregated in equilibrium.

Open problem (Kong and Schoenebeck 2018; Anunrojwong, Chen, W., Xu 2020): is there a polynomial-time algorithm for computing an equilibrium?

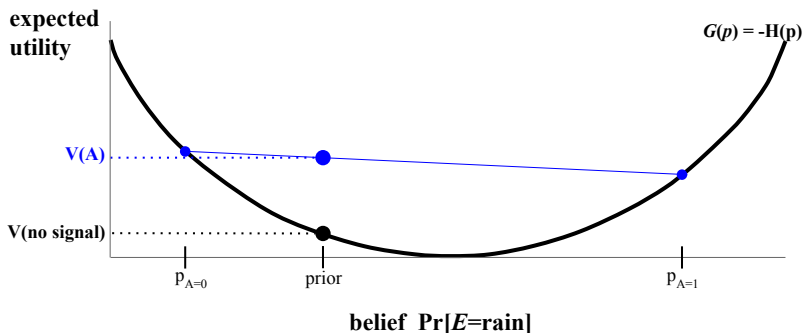
Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?
- How do we *design* markets to encourage aggregation?

Some big questions (1976 - ?)

- How is information structured, communicated, and aggregated?
- How do agents act strategically in (prediction) markets?
- How do we *design* markets to encourage aggregation?

Incentives for gathering information? e.g. *Papireddygari and W. (2022)*



Outline

1 Scoring rules and prediction markets

Eliciting and aggregating information

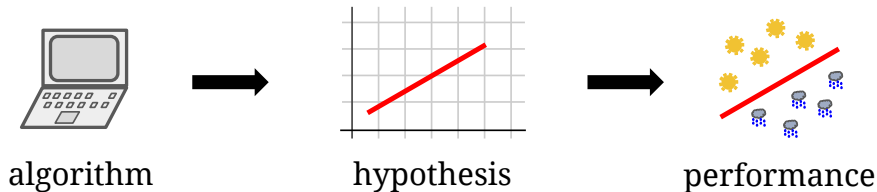
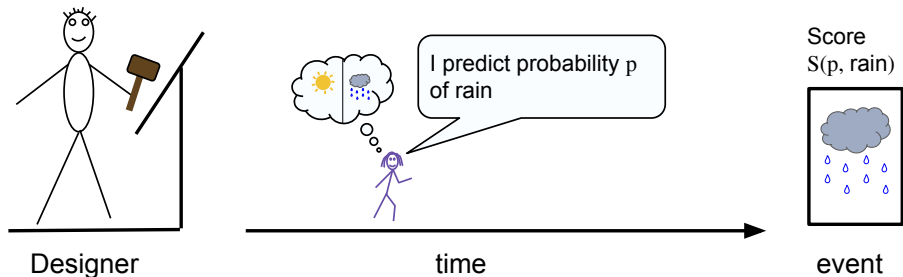
2 Statistics and machine learning

Structure of prediction spaces

3 Market frameworks for ML

Bringing it together

People $\rightarrow$ predictions $\leftarrow$ algorithms



Properties of distributions

What if we want to predict the **median** or **mean**?
More generally, some **property** of the distribution?

$$\operatorname{argmax}_{r \in \mathcal{R}} \mathbb{E}_{y \sim p} S(r, y).$$

$\mathcal{R} =$ “report space” or “hypothesis space”

Properties of distributions

Examples:

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

mean

Properties of distributions

Examples:

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

mean

- $S(r, y) = -|r - y|$

median

Properties of distributions

Examples:

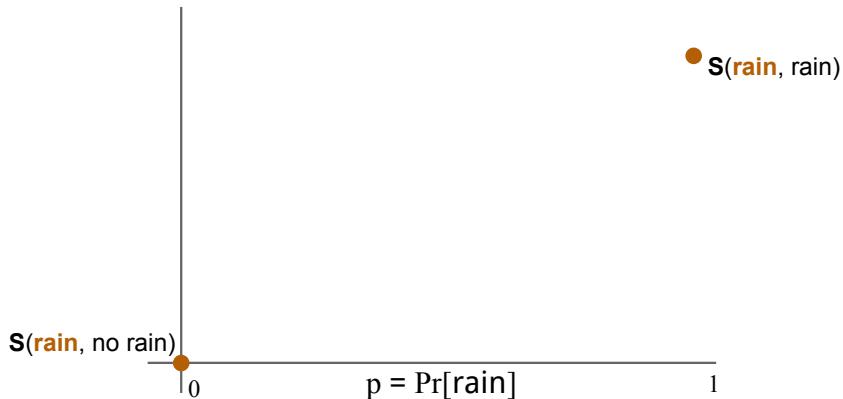
- $S(r, y) = -(r - y)^2$ *mean*
- $S(r, y) = -|r - y|$ *median*
- $S(r, y) = \begin{cases} 1 & r = y \\ 0 & \text{otherwise} \end{cases}$ *mode*

Properties of distributions

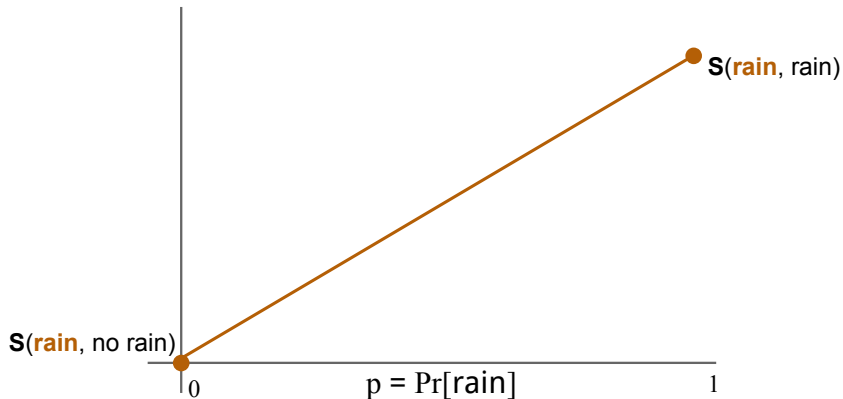
Examples:

- $S(r, y) = -(r - y)^2$ *mean*
- $S(r, y) = -|r - y|$ *median*
- $S(r, y) = \begin{cases} 1 & r = y \\ 0 & \text{otherwise} \end{cases}$ *mode*
- $S(r, y) = ??$ *variance*

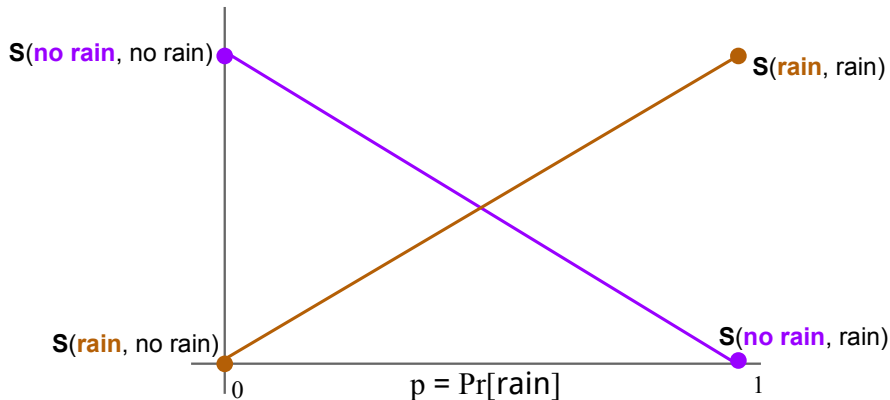
Mode: the convex function view



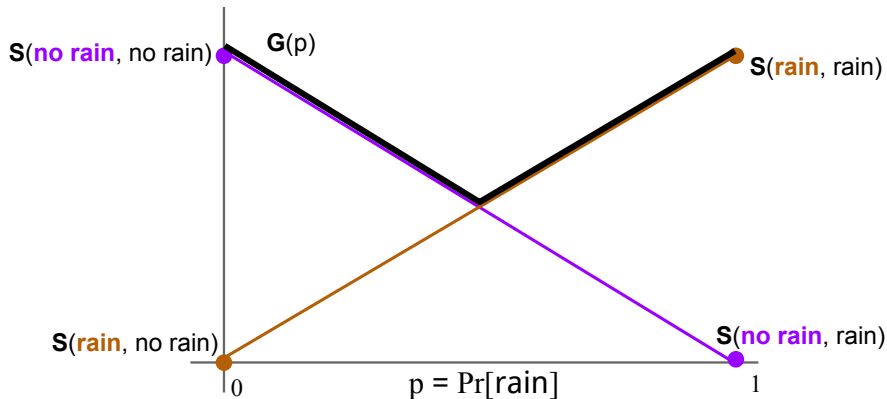
Mode: the convex function view



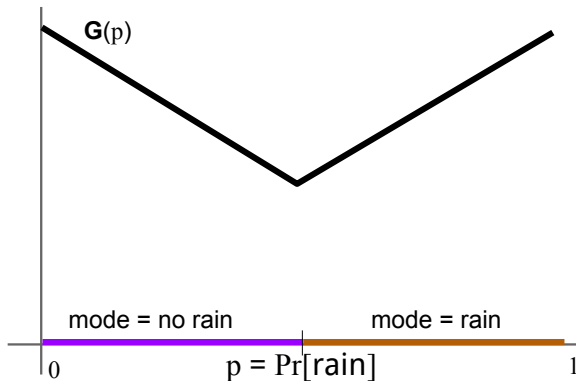
Mode: the convex function view



Mode: the convex function view

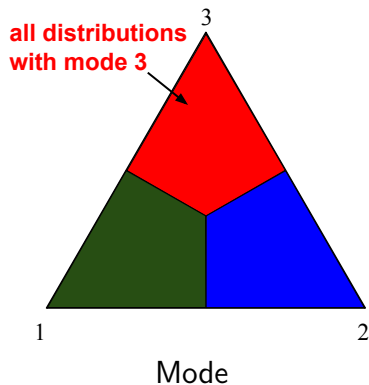


Mode: the convex function view



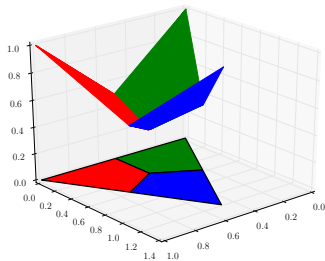
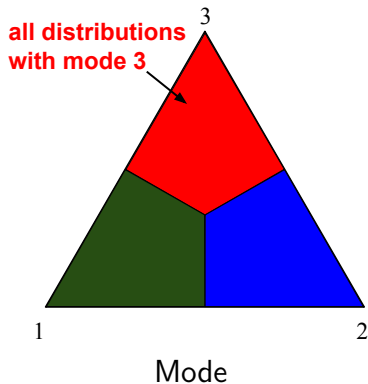
Mode: the geometric view

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



Mode: the geometric view

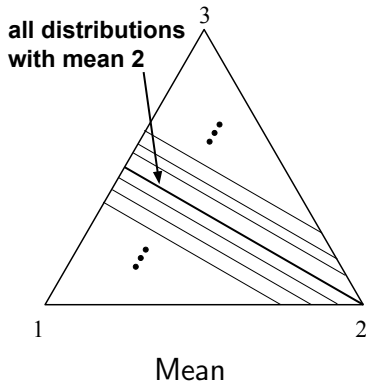
Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



Mode - convex function

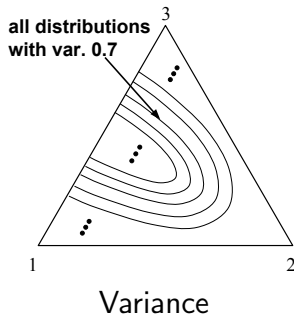
The geometric view - mean

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



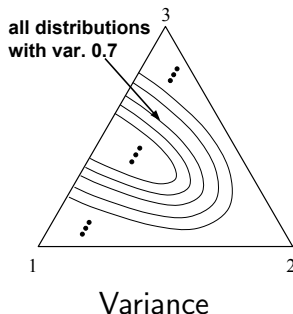
Variance

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



Variance

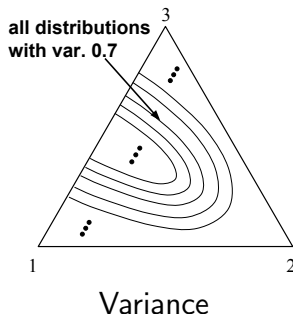
Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



Fact: Let $S : \mathcal{R} \times \mathcal{Y} \rightarrow \mathbb{R}$. Given r , the set $\left\{ p : r = \operatorname{argmax}_{r \in \mathcal{R}} \mathbb{E}_{y \sim p} S(r, y) \right\}$ is convex.

Variance

Simplex on $\mathcal{Y} = \{1, 2, 3\}$:



Fact: Let $S : \mathcal{R} \times \mathcal{Y} \rightarrow \mathbb{R}$. Given r , the set $\left\{ p : r = \operatorname{argmax}_{r \in \mathcal{R}} \mathbb{E}_{y \sim p} S(r, y) \right\}$ is convex.

Cor: Variance is not elicitable: $\nexists S$ s.t.
 $\operatorname{Var}(p) = \operatorname{argmax}_r \mathbb{E}_{y \sim p} S(r, y)$.

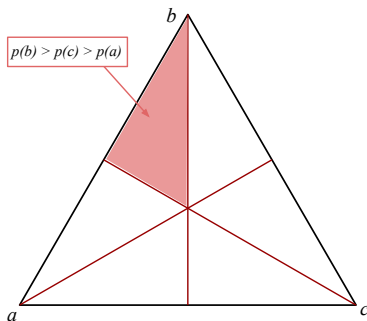
Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

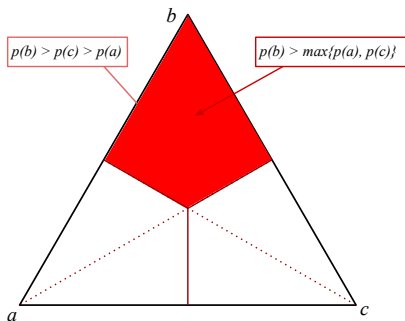
Surrogate loss L for target ℓ (informally): minimizer of L determines a minimizer of ℓ .



Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

Surrogate loss L for target ℓ (informally): minimizer of L determines a minimizer of ℓ .



Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

Surrogate loss L for target ℓ (informally): minimizer of L determines a minimizer of ℓ .

Fisher (1922): *consistent estimators*

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

Surrogate loss L for target ℓ (informally): minimizer of L determines a minimizer of ℓ .

Fisher (1922): *consistent estimators*

Bartlett et al. (2003), e.g.: *classification-calibration* - minimizing hinge loss on the distribution $\implies$ minimizing 0-1 loss

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

Surrogate loss L for target ℓ (informally): minimizer of L determines a minimizer of ℓ .

Fisher (1922): *consistent estimators*

Bartlett et al. (2003), e.g.: *classification-calibration* - minimizing hinge loss on the distribution $\implies$ minimizing 0-1 loss

Agarwal, Agarwal (2015): surrogate loss $\approx$ **indirect elicitation**

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?

Surrogate loss L for target ℓ (informally): minimizer of L determines a minimizer of ℓ .

Fisher (1922): *consistent estimators*

Bartlett et al. (2003), e.g.: *classification-calibration* - minimizing hinge loss on the distribution $\implies$ minimizing 0-1 loss

Agarwal, Agarwal (2015): surrogate loss $\approx$ **indirect elicitation**

Finocchiaro, Frongillo, W. (2024): characterization of elicitability for *polyhedral* surrogate loss functions

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?
- Which is the “best”?

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?
- Which is the “best”?

Questions:

- dimensionality of prediction space (FFW 2021, NWKF 2024)
- convergence rates to optimality (FW 2021)

Some big questions (1922 - ?)

- What are the possible loss functions for a learning task?
- Which is the “best”?

Questions:

- dimensionality of prediction space (FFW 2021, NWKF 2024)
- convergence rates to optimality (FW 2021)

Open problem: given a target loss $\ell : \mathcal{R} \times \mathcal{Y} \rightarrow \mathbb{R}$,

- what is the least d such that a (nice, convex) surrogate $L : \mathbb{R}^d \times \mathcal{Y} \rightarrow \mathbb{R}$ exists?
- ditto, but with a “fast” convergence rate to optimality?

Outline

1 Scoring rules and prediction markets

Eliciting and aggregating information

2 Statistics and machine learning

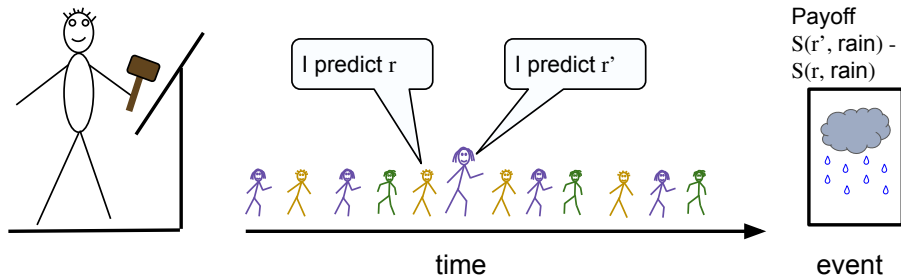
Structure of prediction spaces

3 Market frameworks for ML

Bringing it together

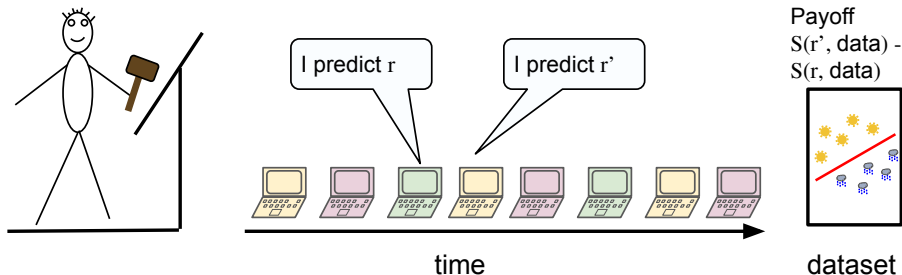
Prediction markets for properties

- agents arrive sequentially and update the current report
- when event occurs, payoff is difference in score



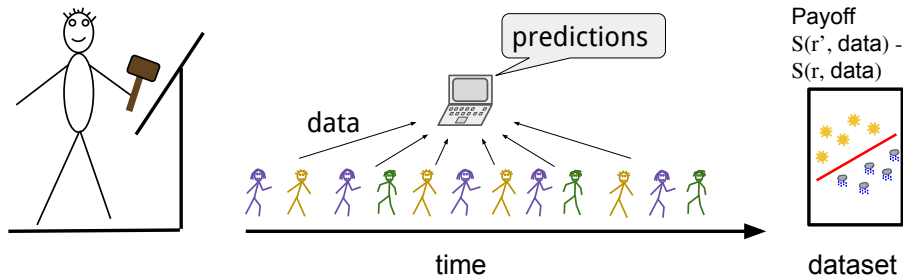
Markets for machine-learning contests

- agents arrive sequentially and update the **current hypothesis**
- when event occurs, payoff is difference in score



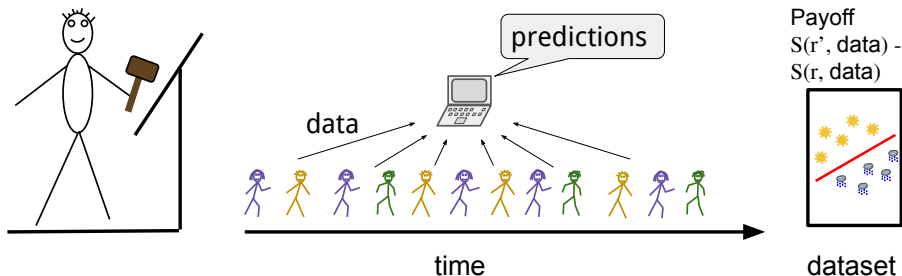
Markets for data

- agents arrive sequentially and **provide data**, updating the model
- when event occurs, payoff is difference in score



Some properties of data markets

- reward = **improvement in accuracy** due to one's data
- can bound total amount “spent on purchasing data”
- can be made **differentially private**



Big Questions

How can we better acquire and aggregate information?



Big Questions

How can we better acquire and aggregate information?

- Understand information elicitation more deeply
- Design interesting mechanisms to aggregate info.
- Use our powers for good! *privacy implications*



Big Questions

How can we better acquire and aggregate information?

- Understand information elicitation more deeply
- Design interesting mechanisms to aggregate info.
- Use our powers for good! *privacy implications*



Thanks!