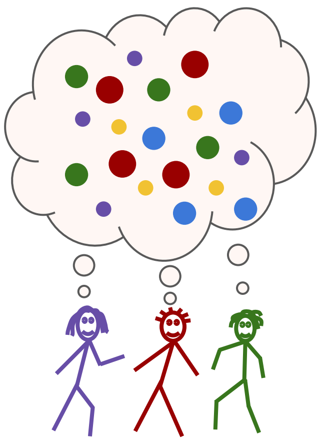
Buying and Learning from User Data, Privately

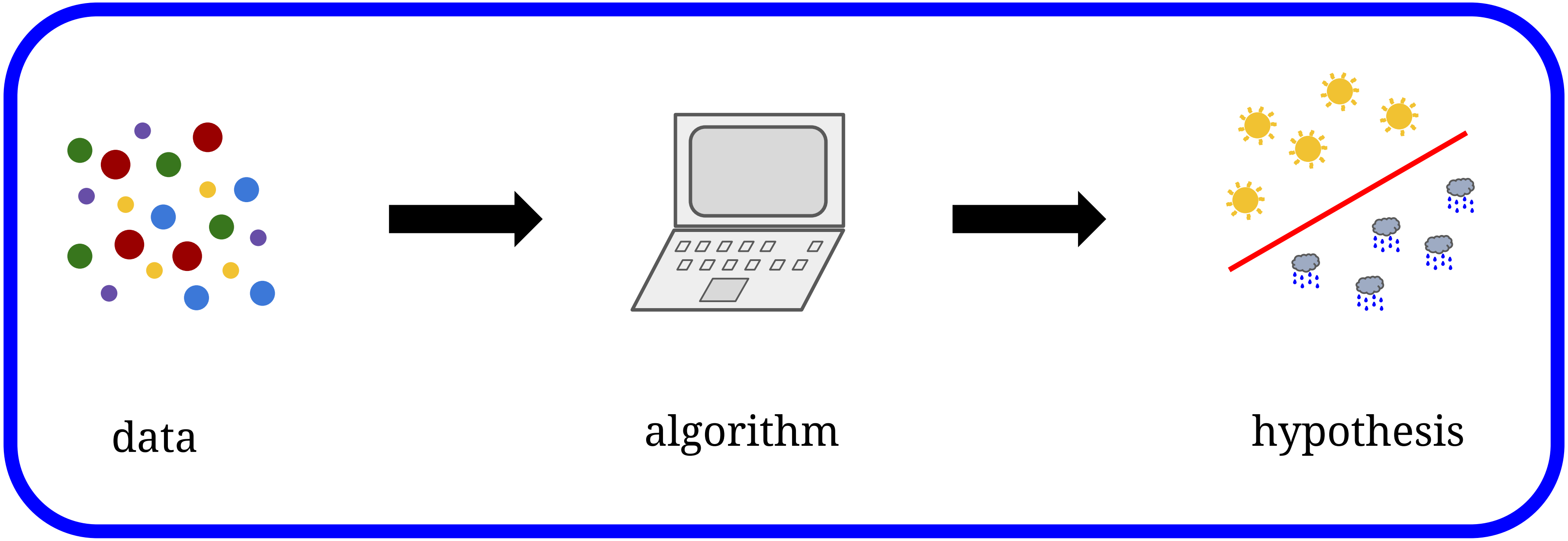


Bo Waggoner
University of Pennsylvania

December 13, 2017

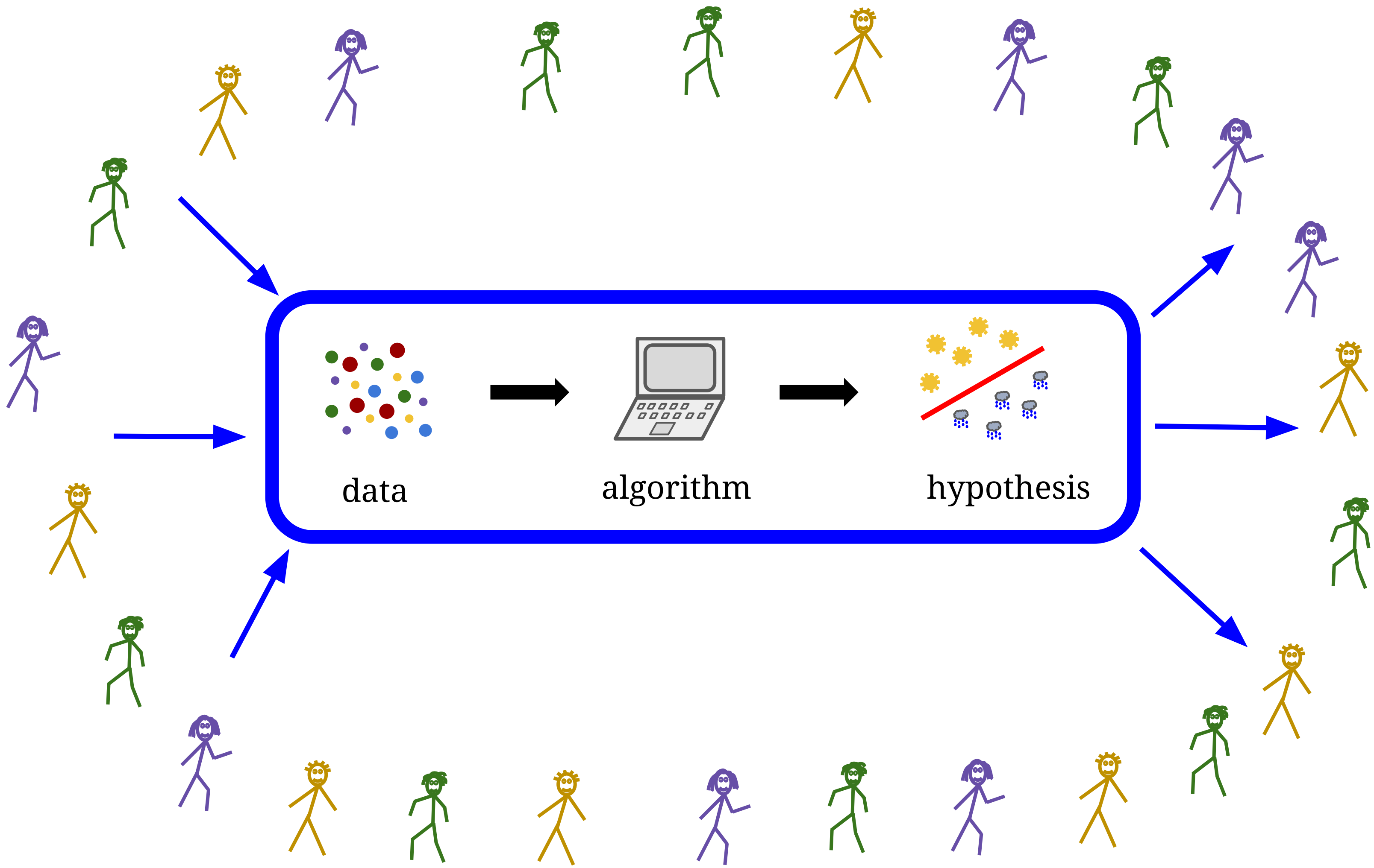
Georgetown

The machine learning paradigm



*drawing not to scale

In context



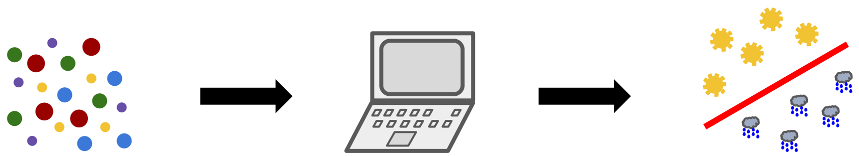
*drawing not to scale

In context

Technical and *social* challenges:

- privacy for data
- fairness for outcomes
- “strategic” behavior

  → conflicts between simple, ethical, and optimal (strategic) behavior



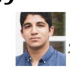
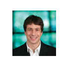
In context

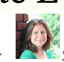
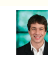
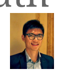
Technical and *social* challenges:

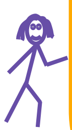
- privacy for data
- fairness for outcomes
- “strategic” behavior

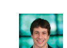
  → conflicts between simple, ethical, and optimal (strategic) behavior

(Hot topics at Penn!)

Accuracy First: Selecting a Differential Privacy Level for Accuracy-Constrained ERM. Katrina Ligett , Seth Neel , Aaron Roth , Bo Waggoner, and Steven Wu , NIPS 2017.

A Smoothed Analysis of the Greedy Algorithm for the Linear Contextual Bandit Problem. Sampath Kannan , Jamie Morgenstern , Aaron Roth , Bo Waggoner, and Steven Wu . (draft) 2017.



Strategic Classification from Revealed Preferences. Jinshuo Dong , Aaron Roth , Zachary Schutzman , Bo Waggoner, and Steven Wu . (draft) 2017.

Outline

I. “Take It Or Leave It”

Interlude: information, privacy, and tech

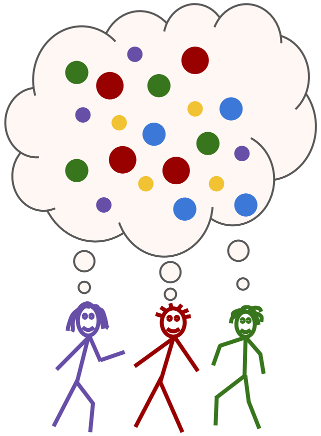
II. “Markets”

III. Going Forward

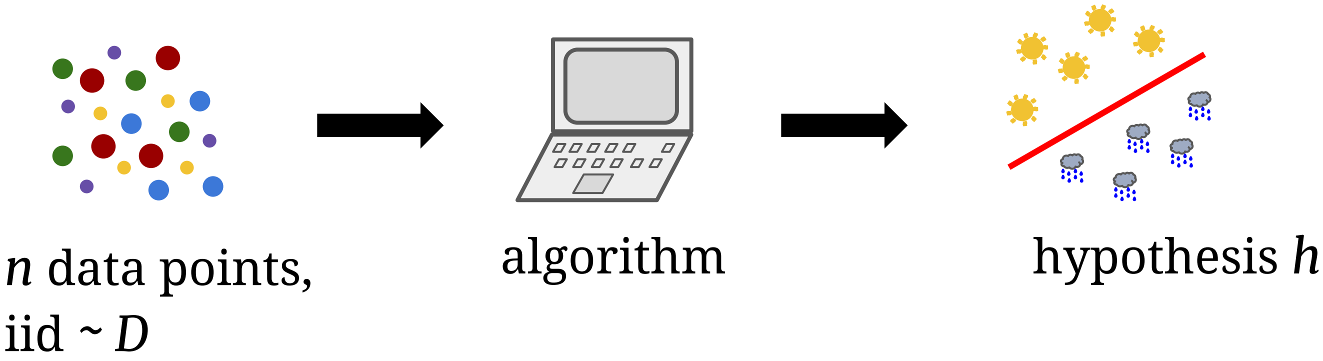
“Take it or leave it”

Low-Cost Learning via Active Data Procurement. Jacob Abernethy , Yiling Chen , Chien-Ju Ho , and Bo Waggoner, EC 2015.

How to obtain **theoretical guarantees** for machine learning when data must be **purchased** from strategic agents?



Classic supervised learning problem



Goal: for a given loss function *loss*(h, z), predict well on new data.

Classic supervised learning problem (cont.)

Example theorem form

Given n data points iid, an algorithm can produce h with roughly

$$\mathbb{E}\text{loss}(h, z) \;\leq\; \mathbb{E}\text{loss}(h^*, z) \;+\; \sqrt{\frac{\text{VC-dim}}{n}}$$

where h^* is the optimal hypothesis.

Classic supervised learning problem (cont.)

Example theorem form

Given n data points iid, an algorithm can produce h with roughly

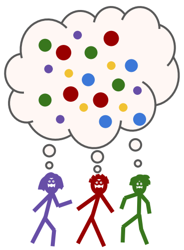
$$\mathbb{E}\mathrm{loss}(h, z) \;\leq\; \mathbb{E}\mathrm{loss}(h^*, z) \;+\; \sqrt{\frac{\boxed{\text{VC-dim}}}{\boxed{n}}}$$

where h^* is the optimal hypothesis.

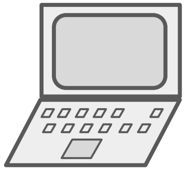
Error depends on:

- problem difficulty
- quantity of resources

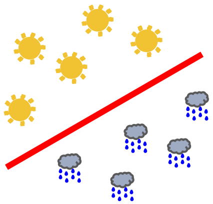
Proposed model of strategic data-holders



agents with secret data iid ~ *D* and **costs** in [0,1]

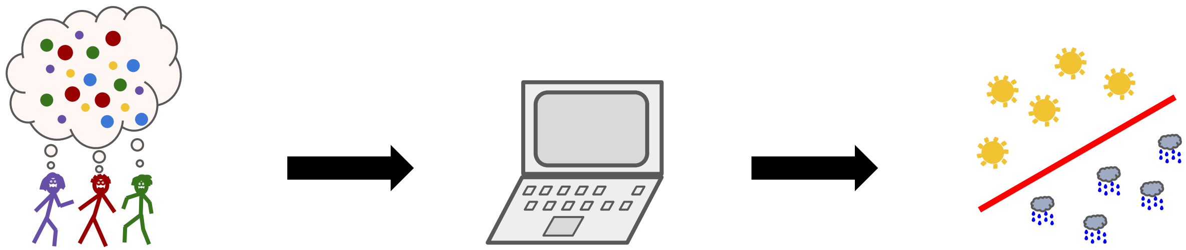


algorithm with budget *B*



hypothesis *h*

Proposed model of strategic data-holders



agents with secret data iid ~ D and **costs** in [0,1]

algorithm with budget B

hypothesis h

Challenge:

- want to only purchase **valuable** and **cheap** data points…
- …but this biases the data!

Approach:

- offer random prices skewed toward “value”
- “de-bias” (importance weighting)

Main result

Theorem* (ACHW'15)

Given budget B and iid data, our algorithm has roughly

$$\mathbb{E}\mathrm{loss}(h, z) \;\leq\; \mathbb{E}\mathrm{loss}(h^*, z) \;+\; \sqrt{\frac{\gamma}{B}}$$

where h^* is the optimal hypothesis.

*Low-order terms and Lipschitz conditions apply.

Main result

Theorem* (ACHW'15)

Given budget B and iid data, our algorithm has roughly

$$\mathbb{E}\mathrm{loss}(h, z) \leq \mathbb{E}\mathrm{loss}(h^*, z) + \sqrt{\frac{\boxed{\gamma}}{\boxed{B}}}$$

where h^* is the optimal hypothesis.

Error depends on:

- problem difficulty
- quantity of resources

*Low-order terms and Lipschitz conditions apply.

Takeaways

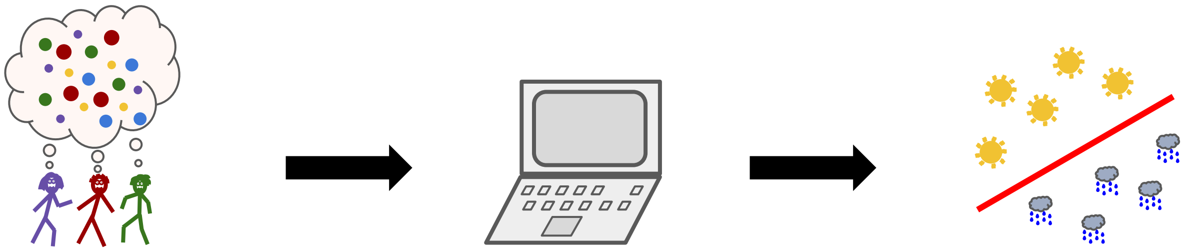
Model:

- people control their data
- will reveal it for at least (unknown) cost

Results:

- theoretical guarantees
- analogues of classical results in this new setting

Lots of future work!



Outline

I. “Take It Or Leave It”

Interlude: information, privacy, and tech

II. “Markets”

III. Going Forward

Outline

I. “Take It Or Leave It”

Interlude: information, privacy, and tech

II. “Markets”

a. Non-private construction
b. Making it private
c. Properties and extensions

III. Going Forward

“Markets”

A Market Framework for Eliciting Private Data. Bo Waggoner, Rafael Frongillo , and Jacob Abernethy . NIPS 2015.

prediction market

[Abernethy, Frongillo 2011]

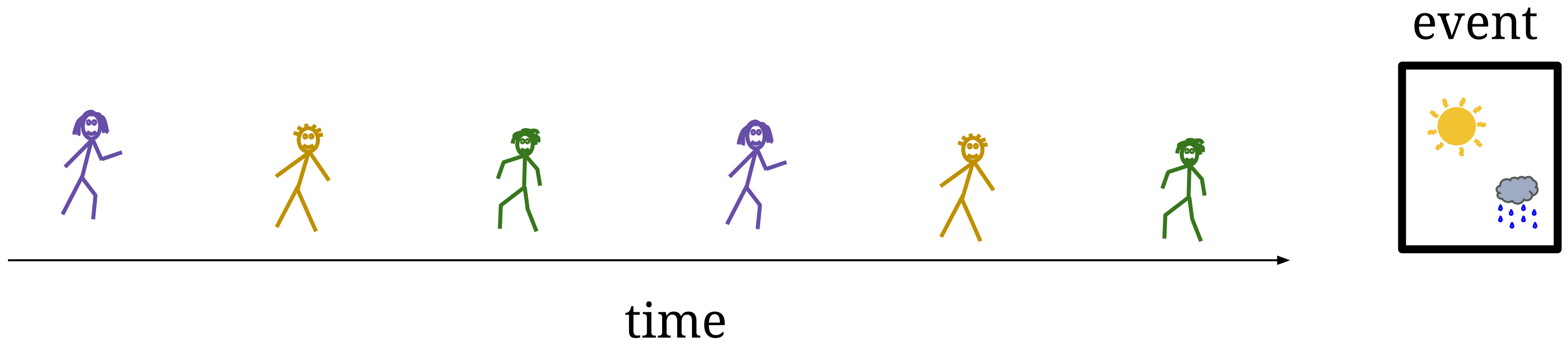
machine-learning contest

market for purchasing data

private market for purchasing data

"Scoring Rule" Prediction Market

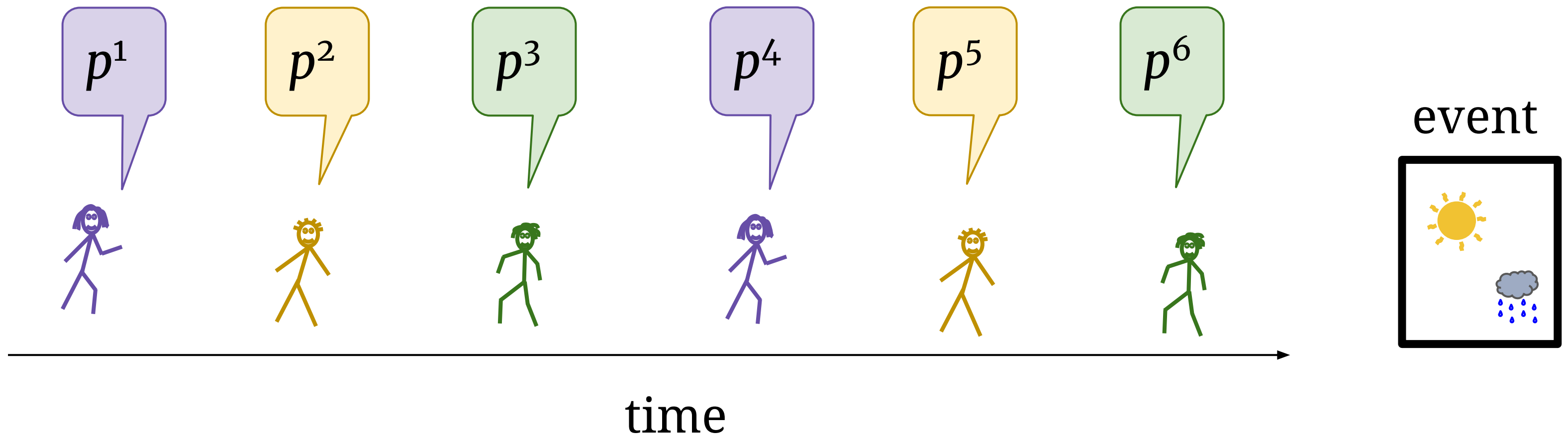
1. Designer chooses initial **public** prediction p^0



[Hanson 2003; Lambert, Pennock, Shoham 2008]

“Scoring Rule” Prediction Market

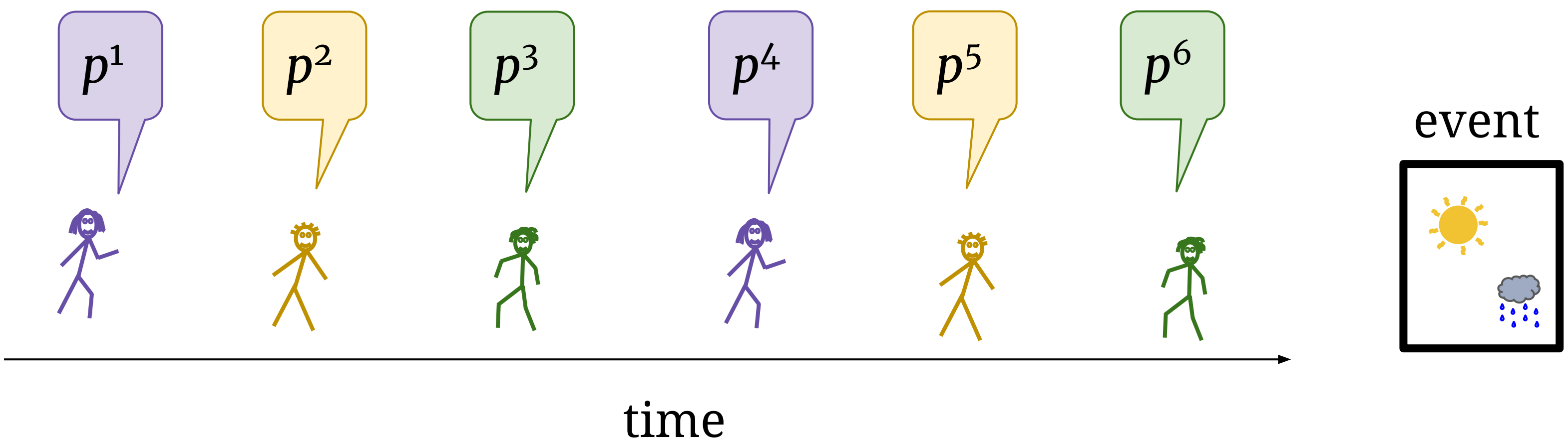
1. Designer chooses initial **public** prediction p^0
2. Participant t=1,..., proposes **public update** $p^{t-1} \rightarrow p^t$



[Hanson 2003; Lambert, Pennock, Shoham 2008]

“Scoring Rule” Prediction Market

1. Designer chooses initial **public** prediction p^0
2. Participant t=1,..., proposes **public update** $p^{t-1} \to p^t$
3. Outcome is observed
4. Reward for t is $S(p^t, \cdot) - S(p^{t-1}, \cdot)$.

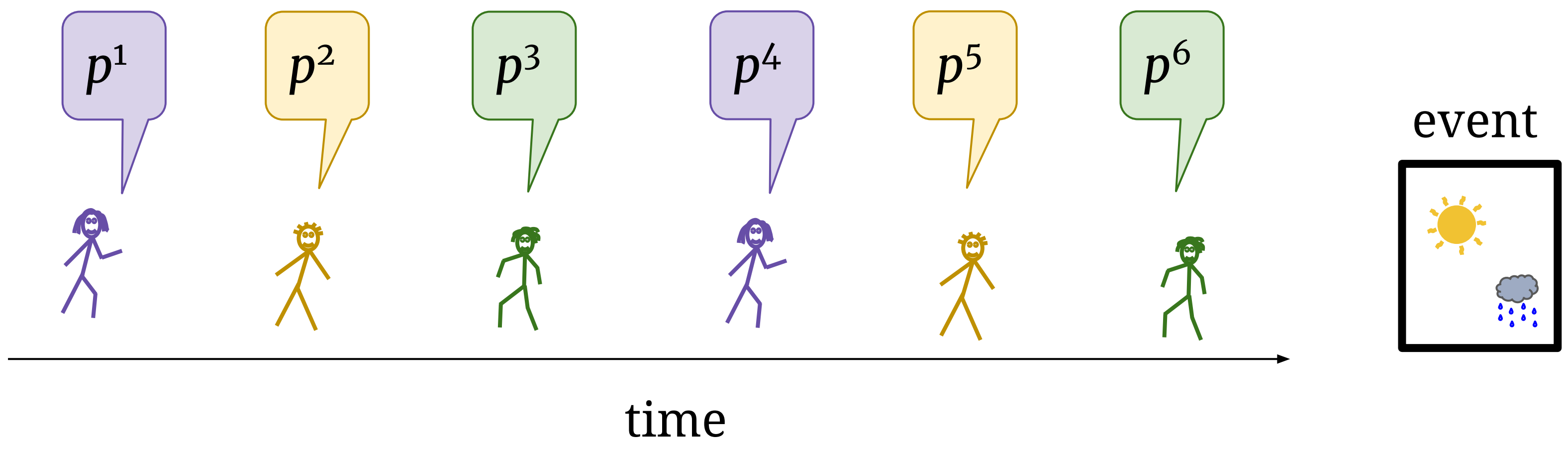


[Hanson 2003; Lambert, Pennock, Shoham 2008]

“Scoring Rule” Prediction Market

1. Designer chooses initial **public** prediction p^0
2. Participant t=1,…, proposes **public update** $p^{t-1} \to p^t$
3. Outcome ☀ is observed
4. Reward for t is $S(p^t,☀) - S(p^{t-1},☀)$.

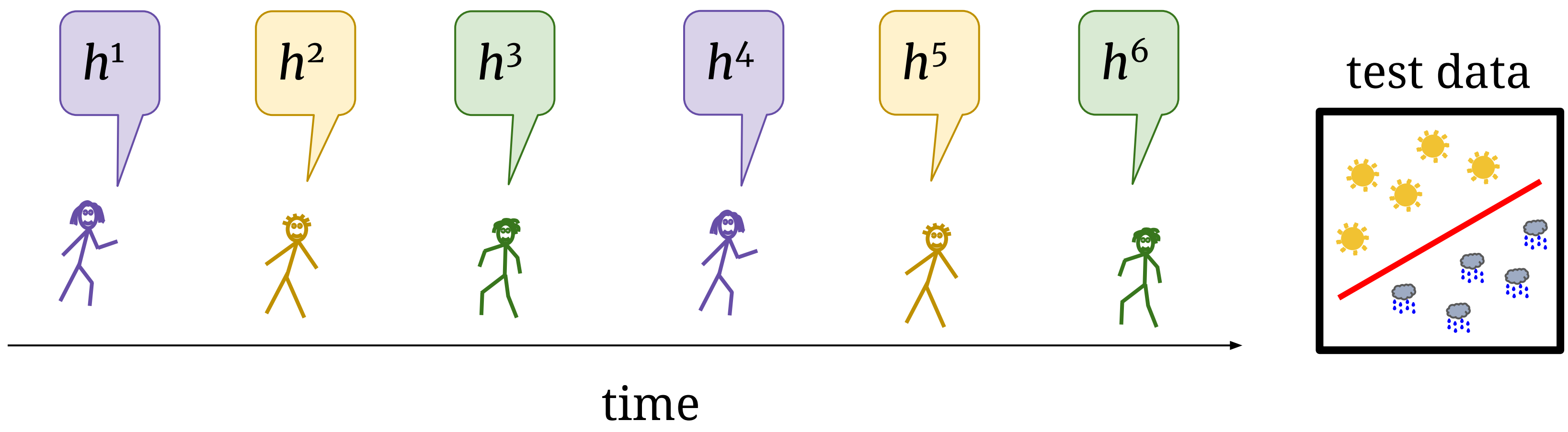
Example **proper scoring rule**: S(p, ☀) = log p(☀).



[Hanson 2003; Lambert, Pennock, Shoham 2008]

SRM for machine learning

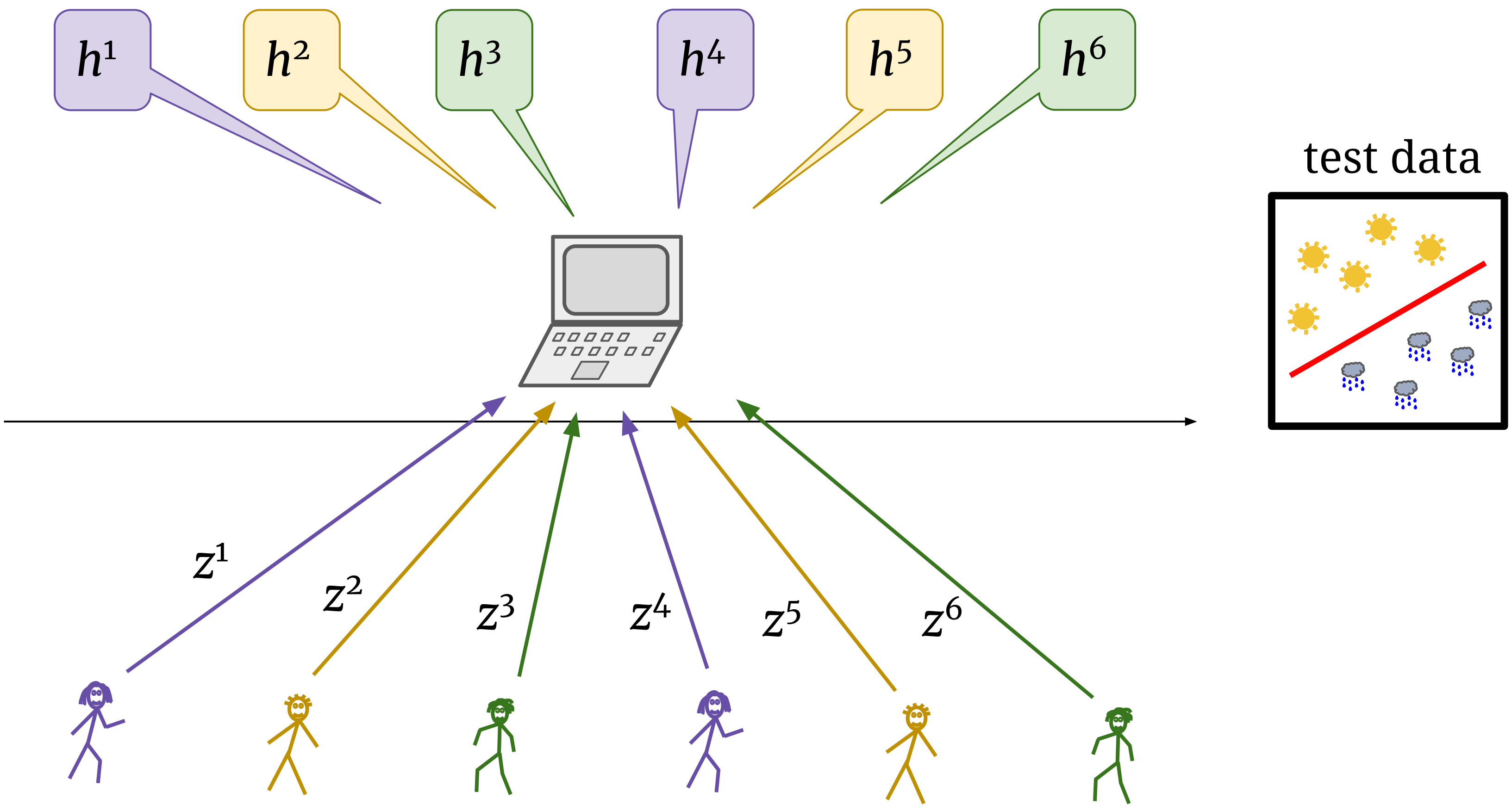
1. Designer chooses initial **public** hypothesis h^0
2. Participant t=1,…, proposes **public update** $h^{t-1} \rightarrow h^t$
3. Data point z is observed
4. Reward for t is loss(h^{t-1}, z) - loss(h^t, z).



[Abernethy, Frongillo 2011]

Buying data, idea #1 (WFA '15)

Use an **online learning algorithm** on agents' behalfs.

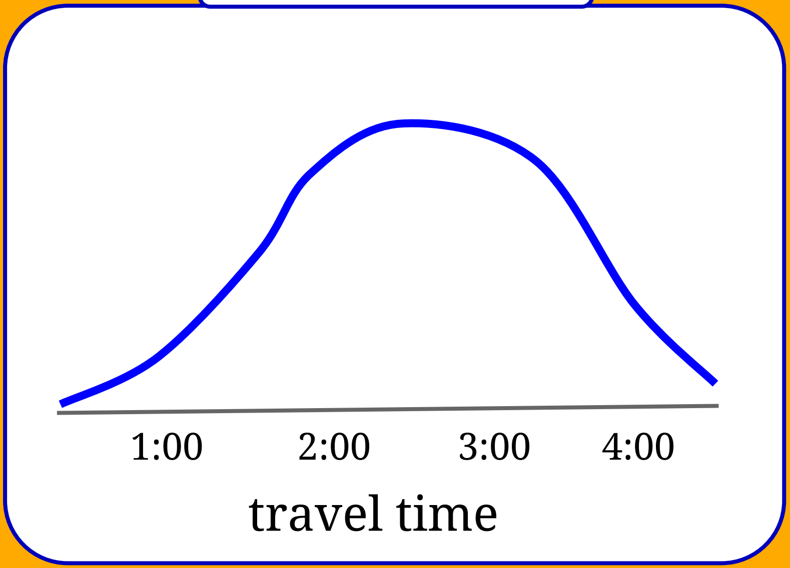


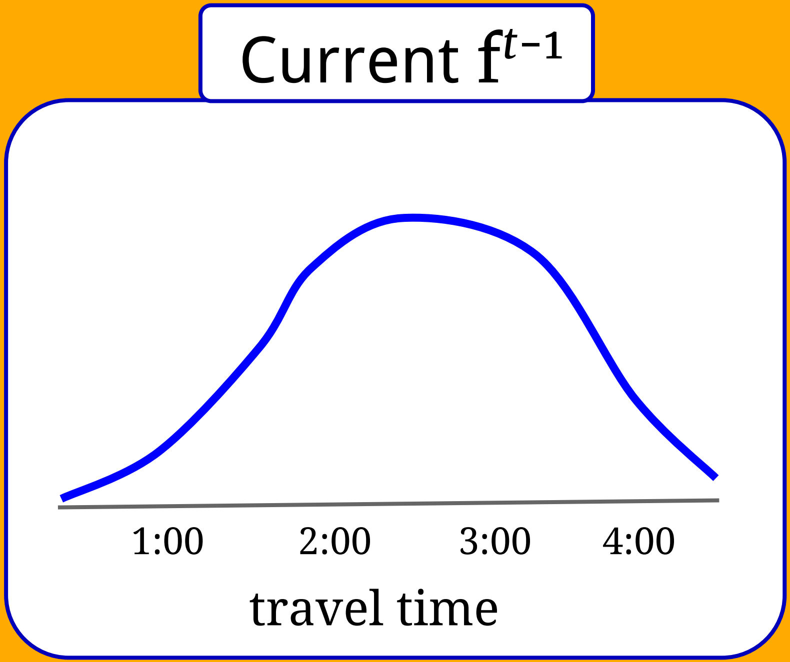
Buying data, idea #2 (WFA ‘15)

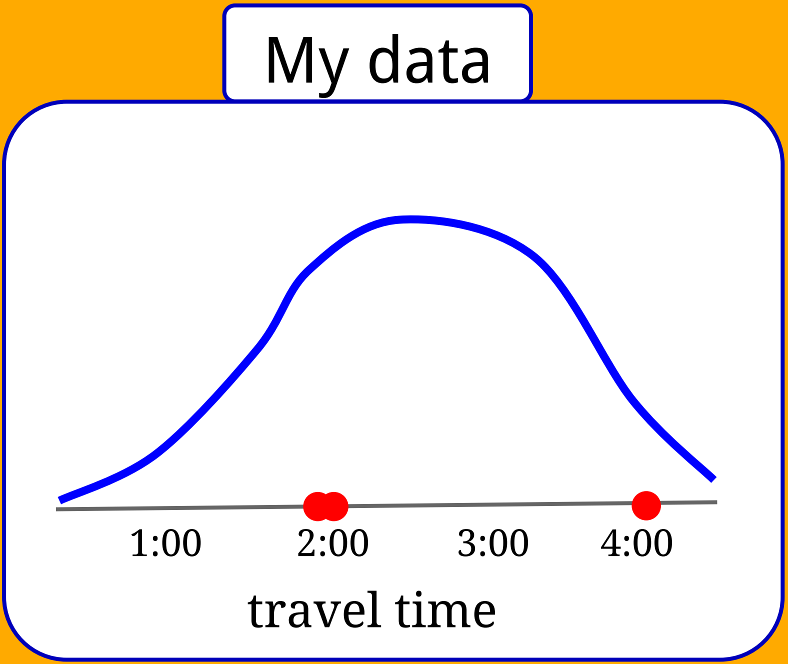
Use **kernels** and a **market interface**.

By example: predict the **travel time** from Philadelphia to D.C.

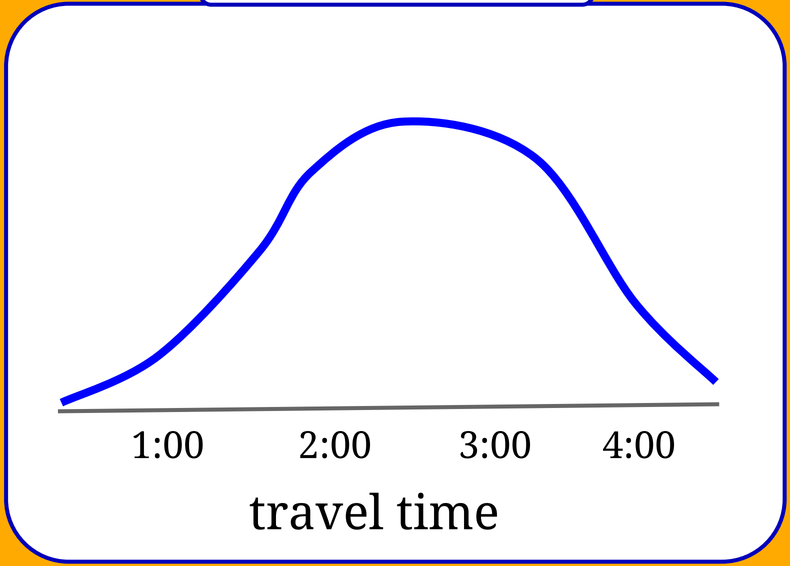
Current f^{t-1}



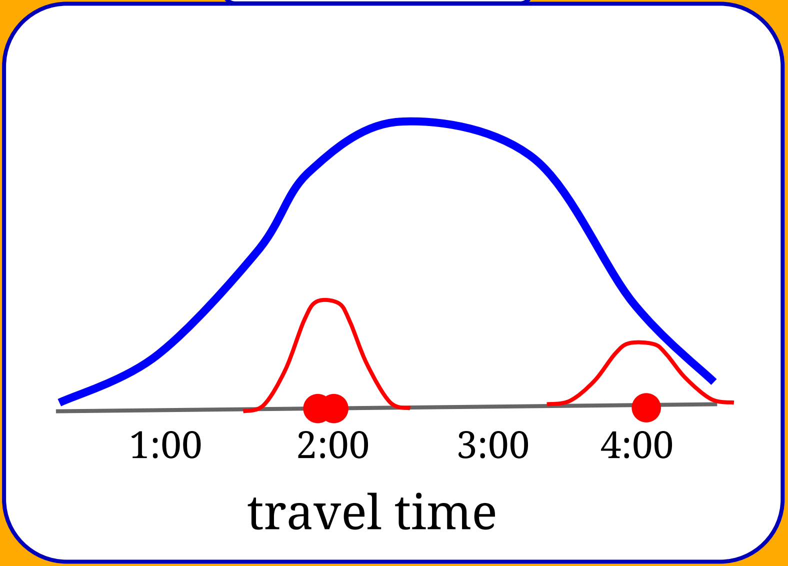

Current f^{t-1}
1:00
2:00
3:00
4:00
travel time

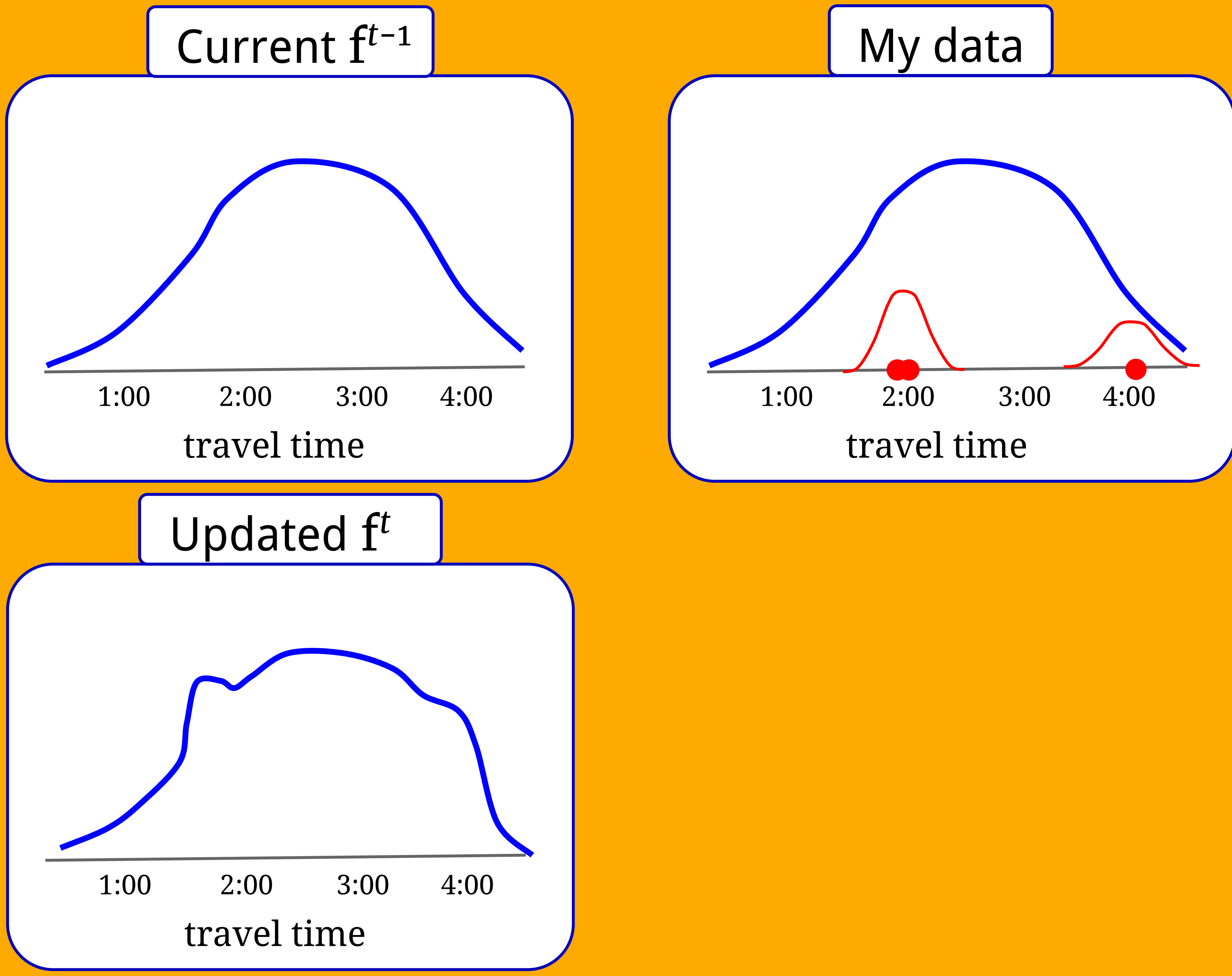

My data
1:00
2:00
3:00
4:00
travel time

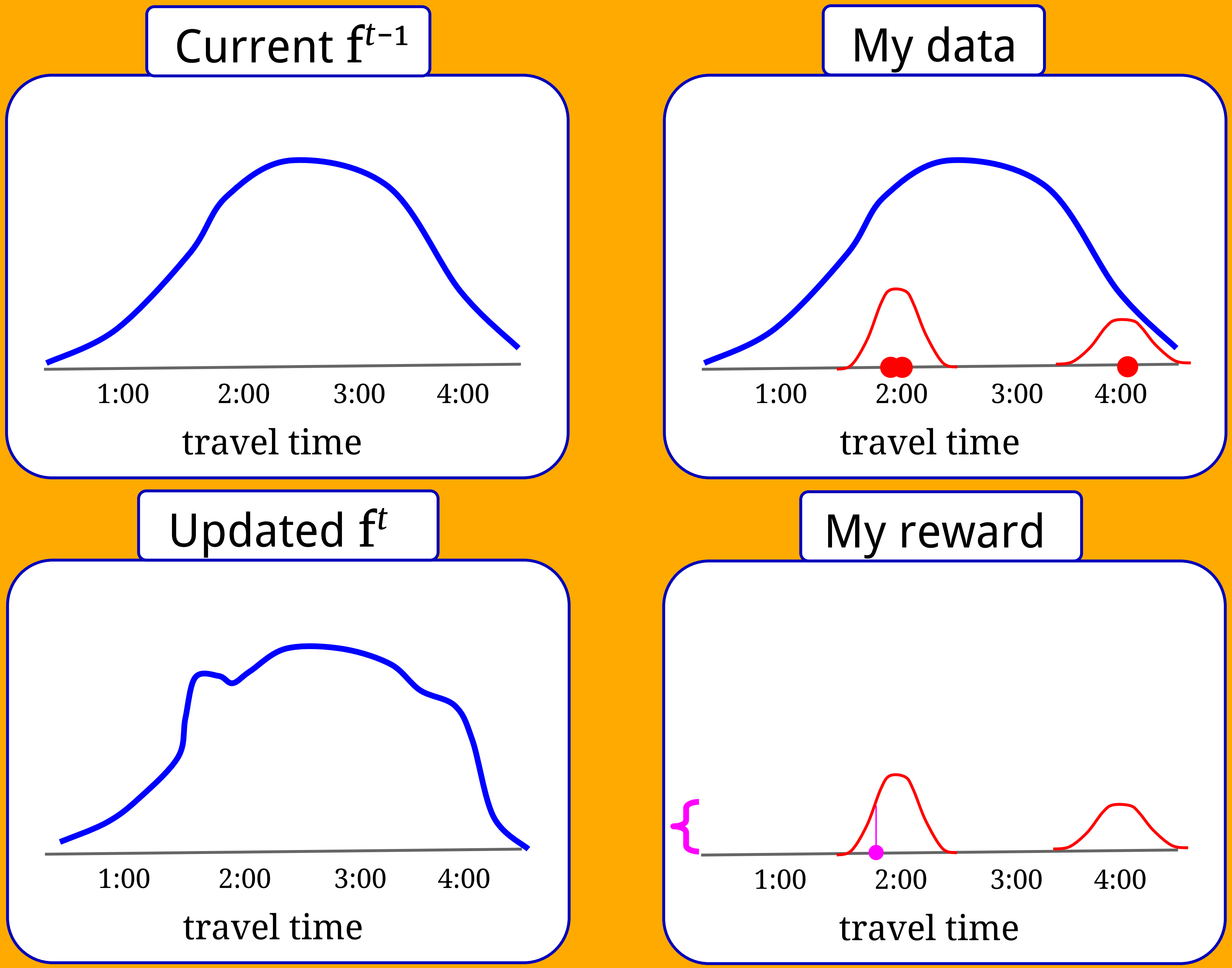
Current f^{t-1}



My data




Current f^{t-1}
1:00
2:00
3:00
4:00
travel time
My data
1:00
2:00
3:00
4:00
travel time
Updated f^t
1:00
2:00
3:00
4:00
travel time


Current f^{t-1}
1:00
2:00
3:00
4:00
travel time
My data
1:00
2:00
3:00
4:00
travel time
Updated f^{t}
1:00
2:00
3:00
4:00
travel time
My reward
1:00
2:00
3:00
4:00
travel time

Buying data, idea #2 (WFA ‘15)

Use **kernels** and a **market interface**.

By example: predict the **travel time** from Philadelphia to D.C.

1. Designer chooses initial public “feature function” f^0
2. Participant t=1,..., purchases “bundle” d^t; updates $\mathrm{f}^{t-1} + d^t \rightarrow \mathrm{f}^t$
3. Data point z is observed
4. Reward for t is $d^t(z)$.

see also: [Abernethy, Chen, Wortman-Vaughan 2013]

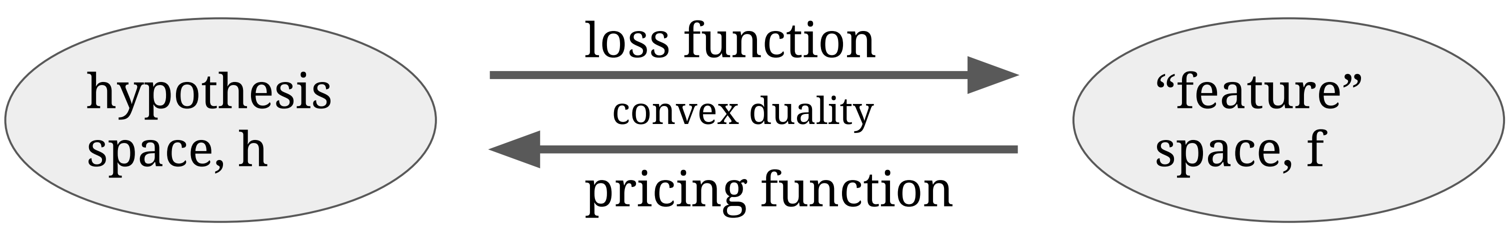
Buying data, idea #2 (WFA '15)

Use **kernels** and a **market interface**.

By example: predict the **travel time** from Philadelphia to D.C.

1. Designer chooses initial public "feature function" f^0
2. Participant t=1,..., purchases "bundle" d^t; updates $f^{t-1} + d^t \rightarrow f^t$
3. Data point z is observed
4. Reward for t is $d^t(z)$.

Equivalent to prior model!



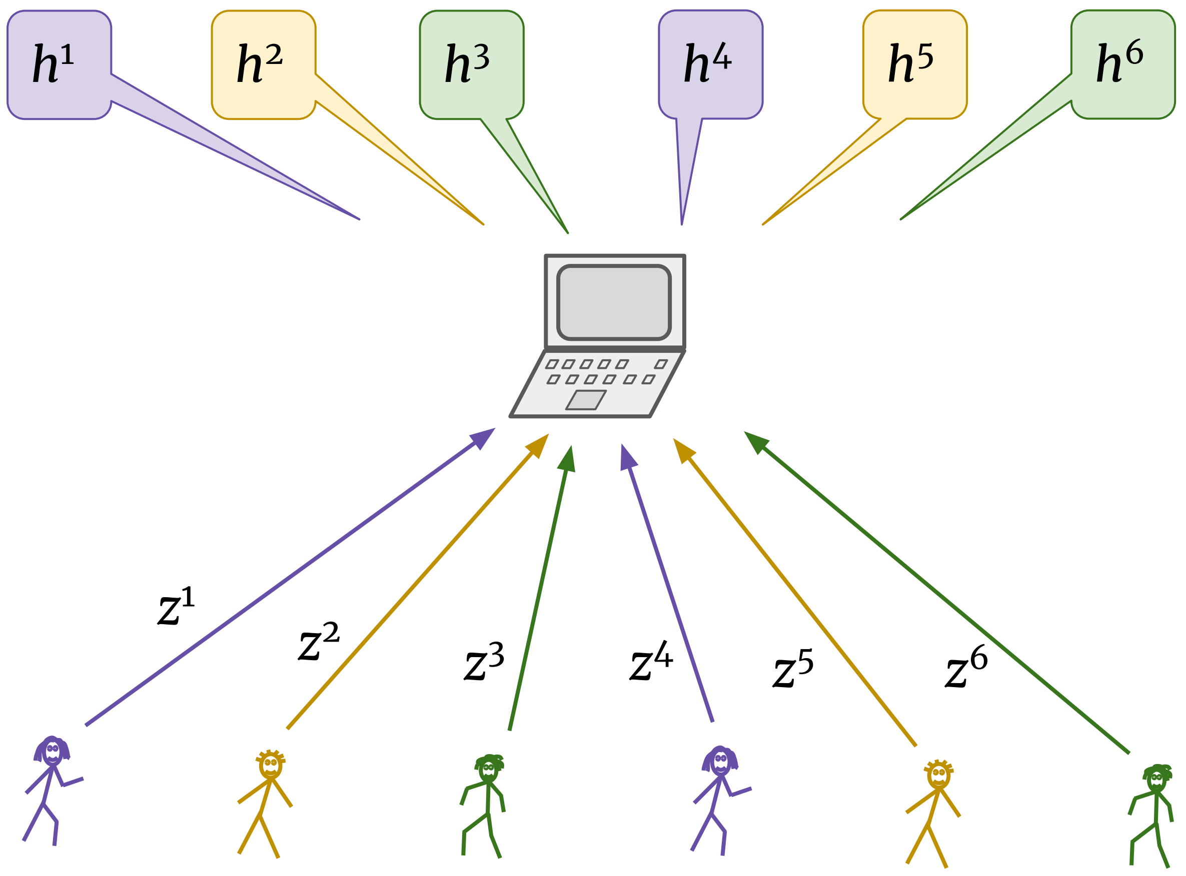
see also: [Abernethy, Chen, Wortman-Vaughan 2013]

Differential privacy

A randomized algorithm A: data $\rightarrow$ information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.

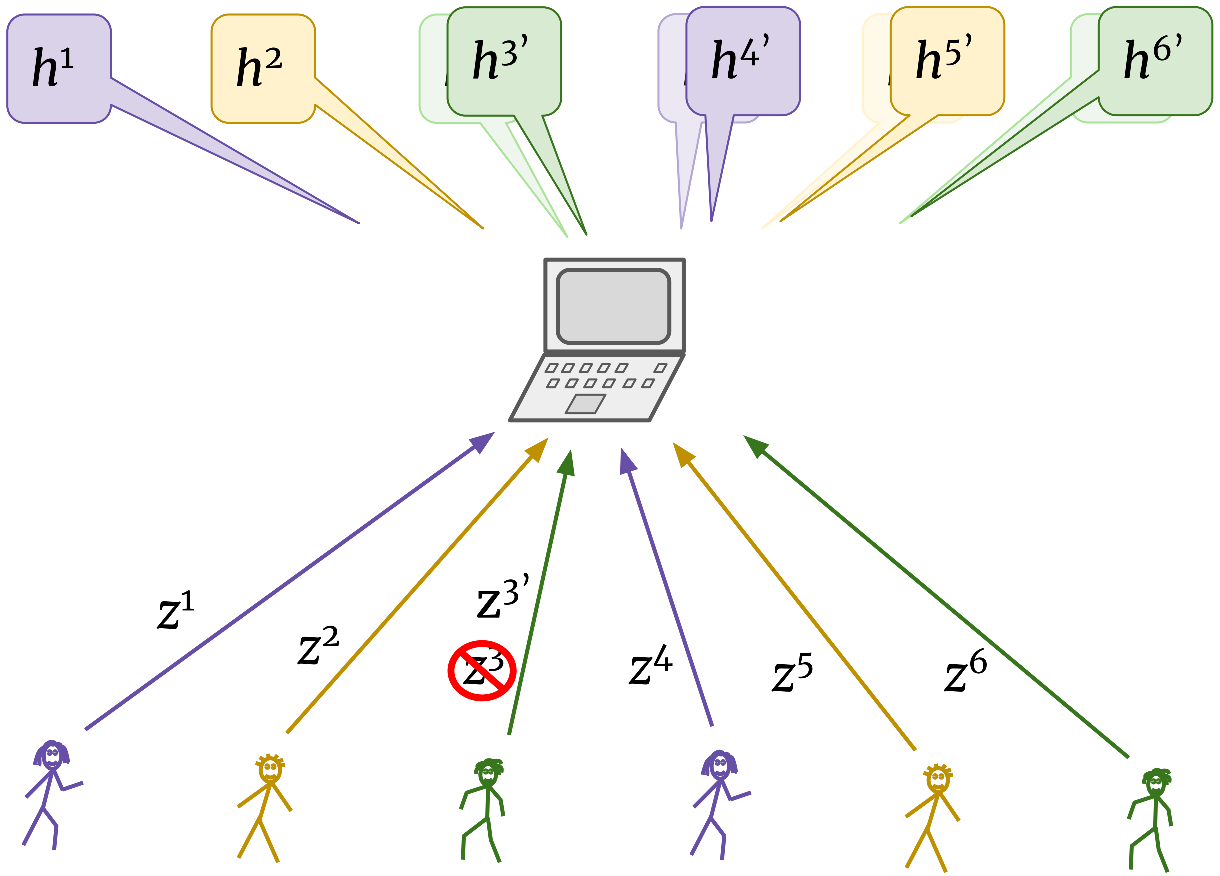
Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.



Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.



Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.

$$\Pr\left[A(\mathrm{data}) = \mathrm{output}\right] \approx \Pr\left[A(\mathrm{data}') = \mathrm{output}\right]$$

Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.

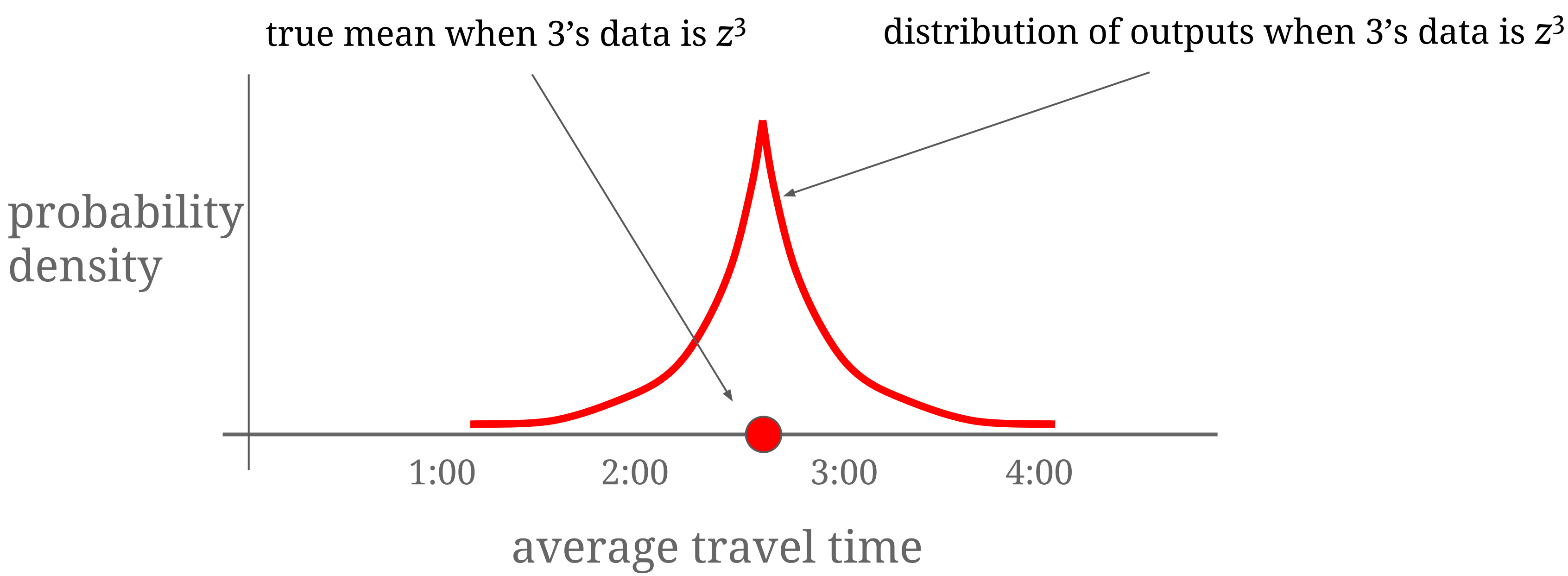
$$\Pr\left[A(\mathrm{data}) = \mathrm{output}\right] \leq e^{\varepsilon} \Pr\left[A(\mathrm{data}') = \mathrm{output}\right]$$

Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.

$$\Pr\left[A(\text{data}) = \text{output}\right] \leq e^{\varepsilon} \Pr\left[A(\text{data}') = \text{output}\right]$$

Example (average travel time): $A(x) = x + \text{Laplace}(1/\varepsilon)$

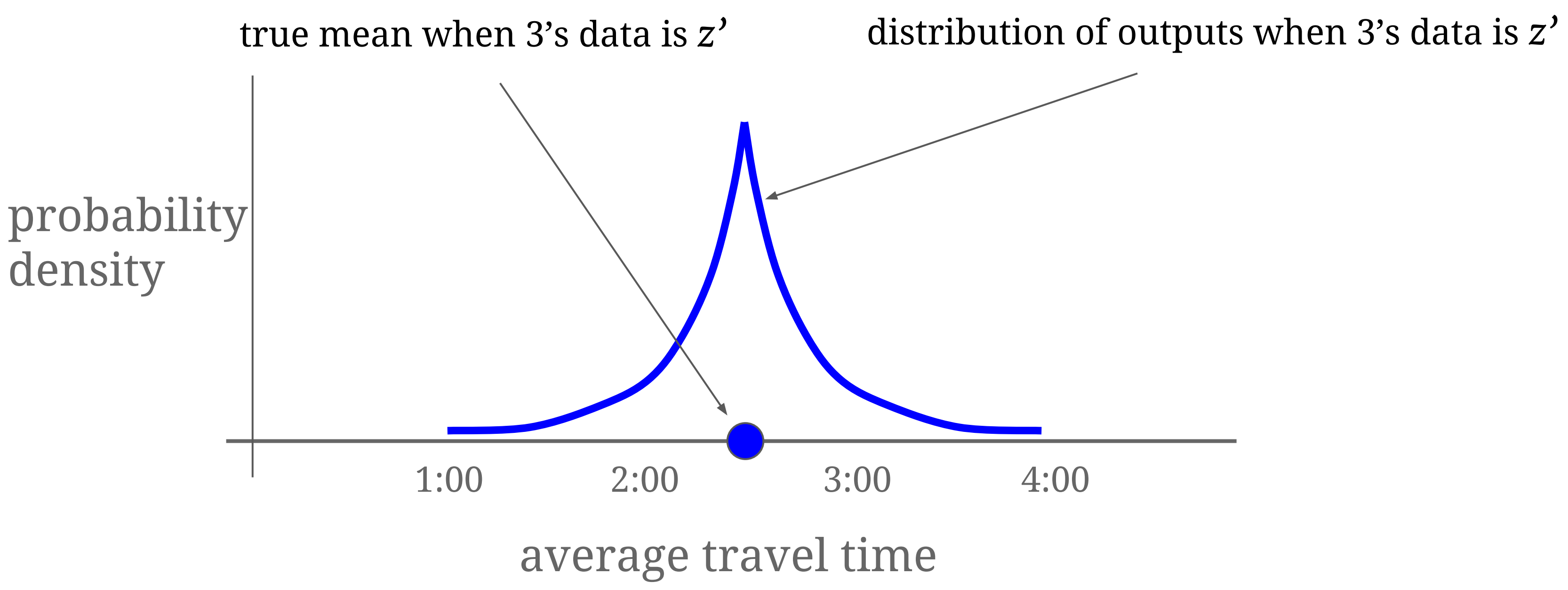


Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.

$$\Pr\left[A(\text{data}) = \text{output}\right] \leq e^{\varepsilon} \Pr\left[A(\text{data}') = \text{output}\right]$$

Example (average travel time): $A(z) = z + \text{Laplace}(1/\varepsilon)$

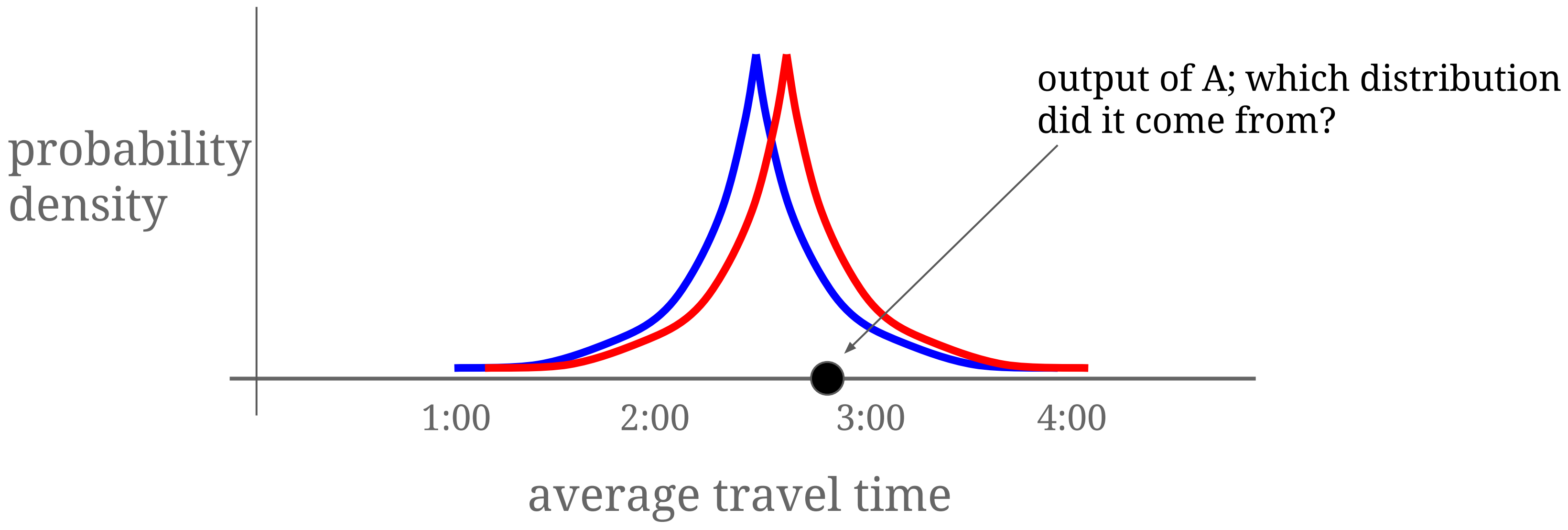


Differential privacy

A randomized algorithm A: data → information is **ε-differentially private** if when one piece of data changes, the output distribution is about the same.

$$\Pr\left[A(\text{data}) = \text{output}\right] \leq e^{\varepsilon} \Pr\left[A(\text{data}') = \text{output}\right]$$

Example (average travel time): $A(z) = z + \text{Laplace}(1/\varepsilon)$



The privacy-preserving prediction market

1. Designer chooses initial public f^0
2. For t=1,...:
   a. participant purchases "bundle" d^t
   b. designer "purchases" **noisy bundle** e^t
   c. updates $f^{t-1} + d^t + e^t \rightarrow f^t$
3. Test data point z
4. Reward $d^t(z)$, not observable by others.

The privacy-preserving prediction market

1. Designer chooses initial public f^0
2. For t=1,...:
   a. participant purchases "bundle" d^t
   b. designer "purchases" **noisy bundle** e^t
   c. updates $f^{t-1} + d^t + e^t \rightarrow f^t$
3. Test data point z
4. Reward $d^t(z)$, not observable by others.

Good: preserves privacy.

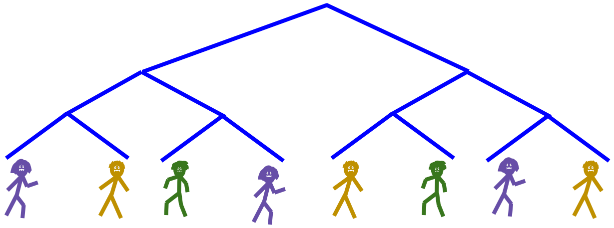
Bad: doesn't work (well).

Why: the noise overwhelms the useful information!

The privacy-preserving prediction market

1. Designer chooses initial public f^0
2. For t=1,...:
   a. participant purchases "bundle" d^t
   b. designer "purchases" **noisy bundle** e^t
   c. updates $\mathrm{f}^{t-1} + d^t + e^t \rightarrow \mathrm{f}^t$
3. Test data point z
4. Reward $d^t(z)$, not observable by others.

Fix: "continual observation" technique: add correlated noise over time.
→ designer sometimes "sells back" noisy bundles to herself



[Dwork, Naor, Pitassi, Rothblum 2010; Chan, Shi, Song 2011]

Results for private prediction markets

Theorem* (WFA'15)

With T participants, the market is ε-differentially private and guarantees accuracy α with high probability when **scaling loss function** by

$$O\left(\frac{(\log T)^{5/2}}{\alpha\ \varepsilon}\right)$$

Privacy for kernel functions: [Hall, Rinaldo, Wasserman 2013]

Results for private prediction markets

Theorem* (WFA'15)

With T participants, the market is ε-differentially private and guarantees accuracy α with high probability when **scaling loss function** by

$$O\left(\frac{(\log T)^{5/2}}{\alpha\ \varepsilon}\right)$$

Implications:

- **budget** "should" be bounded by this quantity (but it's not)
- after **relatively few** participants, predictions converge

Privacy for kernel functions: [Hall, Rinaldo, Wasserman 2013]

Results for private prediction markets

Theorem* (WFA'15)

With T participants, the market is ε-differentially private and guarantees accuracy α with high probability when **scaling loss function** by

$$O\left(\frac{(\log T)^{5/2}}{\alpha\ \varepsilon}\right)$$

Theorem (Cummings, Pennock, Wortman Vaughan 2016)

The private prediction market **cannot** have bounded budget!

$\longrightarrow$ Noisy bundles + smart participants = bad news.

Results for private prediction markets

Theorem* (WFA'15)

With T participants, the market is ε-differentially private and guarantees accuracy α with high probability when **scaling loss function** by

$$O\left(\frac{(\log T)^{5/2}}{\alpha\ \varepsilon}\right)$$

Theorem* (WFA 2017)

By introducing a small transaction fee:

- Budget is bounded independent of T
- Accuracy guarantee α is maintained
- Privacy guarantee ε is maintained
- If prices are wrong by 2α, participants have incentive to update.

Related work on elicition and markets

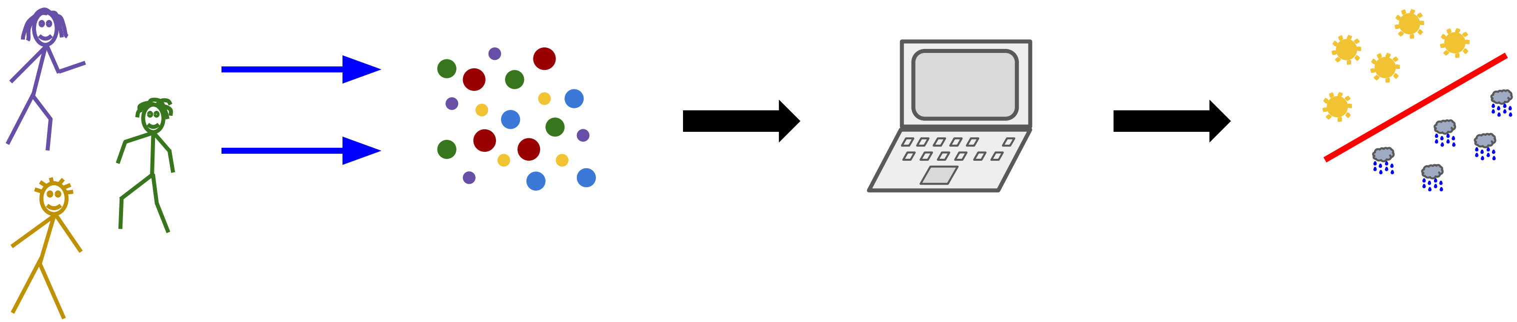
- Strategic participation; timing.
  Informational Substitutes. Yiling Chen  and Bo Waggoner, FOCS 2016.

- Predicting higher-order relationships in data.
  Multi-Observation Elicitation. Sebastian Casalaina-Martin , Rafael Frongillo , Tom Morgan , and Bo Waggoner. COLT 2017.

- Usability and “market-like” properties.
  An Axiomatic Study of Scoring Rule Markets. Rafael Frongillo  and Bo Waggoner. ITCS 2018.

Recap: properties of these mechanisms

- Incentives aligned
- Privacy-preserving
- End to end



Practical challenges remaining: many!

Outline

I. "Take It Or Leave It"

Interlude: information, privacy, and tech

II. "Markets"

III. Going Forward

What makes information **valuable**?

What makes information **valuable**?

Information creates value by
changing (improving) our decisions

See also: [Howard 1966],
Informational Substitutes. Yiling Chen  and Bo Waggoner, FOCS 2016.

"The best way to control someone's actions is to **control the information** upon which he makes his decisions."

Steven Brust, author

"The best way to control someone's actions is to **control the information** upon which he makes his decisions."

Steven Brust, author

World's most "*valuable*" companies

(one probably-wrong ranking I saw online)

1. Apple
2. Alphabet Google
3. Microsoft
4. Amazon
5. Berkshire Hathaway
6. Facebook

World's most "*valuable*" companies

(one probably-wrong ranking I saw online)

1. Apple
2. **Alphabet**
3. **Microsoft**
4. **Amazon**
5. Berkshire Hathaway
6. **Facebook**

Value is **entirely** (**partially**) from:

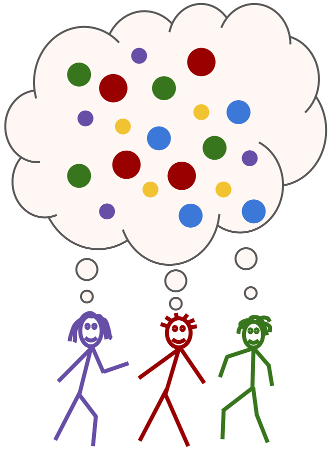
- Our *data*
- Our *attention*

...via non-monetary transactions!

Future directions

To understand these systems, and engineer better ones:

- Value of information to people and algorithms
- Microfounding the costs of privacy loss
- Exposure to (mis)information; **persuasion**
- More end-to-end systems for buying + learning from data!



Future directions

To understand these systems, and engineer better ones:

- Value of information to people and algorithms
- Microfounding the costs of privacy loss
- Exposure to (mis)information; **persuasion**
- More end-to-end systems for buying + learning from data!

Thanks!