MS&E 233
Game Theory, Data Science and AI
Lecture 14

Vasilis Syrgkanis

Assistant Professor

Management Science and Engineering

(by courtesy) Computer Science and Electrical Engineering

Institute for Computational and Mathematical Engineering

Computational Game Theory for Complex Games

1
- Basics of game theory and zero-sum games (T)
- Basics of online learning theory (T)
- Solving zero-sum games via online learning (T)
- *HW1: implement simple algorithms to solve zero-sum games*
- Applications to ML and AI (T+A)
- *HW2: implement boosting as solving a zero-sum game*

2
- Basics of extensive-form games
- Solving extensive-form games via online learning (T)
- *HW3: implement agents to solve very simple variants of poker*

3
- General games, equilibria and online learning (T)
- Online learning in general games
- *HW4: implement no-regret algorithms that converge to correlated equilibria in general games*

Data Science for Auctions and Mechanisms

4
- Basics and applications of auction theory (T+A)
- Basic Auctions and Learning to bid in auctions (T)
- *HW5: implement bandit algorithms to bid in ad auctions*

5
- Optimal auctions and mechanisms (T)
- Simple vs optimal mechanisms (T)
- *HW6: implement simple and optimal auctions, analyze revenue empirically*

6
- Basics of Statistical Learning Theory (T)
- **Optimizing Mechanisms from Samples (T)**
- *HW7: implement procedures to learn approximately optimal auctions from historical samples*

Further Topics

7
- Econometrics in games and auctions (T+A)
- A/B testing in markets (T+A)
- *HW8: implement procedure to estimate values from bids in an auction, empirically analyze inaccuracy of A/B tests in markets*

Guest Lectures

- Mechanism Design for LLMs, Renato Paes Leme, Google Research
- Auto-bidding in Sponsored Search Auctions, Kshipra Bhawalkar, Google Research

Statistical Learning Theory

General Framework

- Given samples $S = \{v_1, \dots, v_m\}$ that are i.i.d. from distribution F
- Given a hypothesis/function space H
- Given a reward function $r(v; h)$

- Goal is to maximize the expected reward over distribution F

$$R(h) = E_{v \sim F}[r(v; h)]$$

Desiderata

- Without knowledge of distribution F, we want to produce a hypothesis h_S, that achieves good reward on this distribution
- For some $\epsilon(m) \to 0$ as the number of samples grows:

$$R(h_S) \stackrel{\text{def}}{=} E_{v\sim F}[r(v;h)] \geq \max_{h\in H} R(h) - \epsilon(m)$$

- Either in expectation over the draw of the samples, i.e.

$$E_S[R(h_S)] \geq \max_{h\in H} R(h) - \epsilon(m)$$

- Or with high-probability over the draw of the samples, i.e.

$$\text{w.p. } 1-\delta: \quad R(h_S) \geq \max_{h\in H} R(h) - \epsilon_\delta(m)$$

Desiderata (Mechanism Design from Samples)

- Without knowledge of Distribution of value profiles F, we want to produce a hypothesis h_S, that achieves good Revenue on this distribution
- For some $\epsilon(m) \to 0$ as the number of samples grows:

$$R(h_S) \overset{\text{def}}{=} E_{v\sim F}\left[\sum_i p_i(v)\right] \geq \max_{h\in H} R(h) - \epsilon(m)$$

- Either in expectation over the draw of the samples, i.e.

$$E_S[R(h_S)] \geq \max_{h\in H} R(h) - \epsilon(m)$$

- Or with high-probability over the draw of the samples, i.e.

$$\text{w.p. } 1-\delta: \quad R(h_S) \geq \max_{h\in H} R(h) - \epsilon_\delta(m)$$

The Obvious Algorithm

- We want to choose r that maximizes

$$\max_{h \in H} R(h) \stackrel{\text{def}}{=} E_{v \sim F}[r(v; h)], \qquad \text{(population objective)}$$

- With m samples, we can optimize average reward on samples!

$$\max_{h \in H} R_S(h) \stackrel{\text{def}}{=} \frac{1}{m} \sum_{j=1}^{m} r(v_j; h), \qquad \text{(empirical objective)}$$

- This approach is called Empirical Reward Maximization (ERM)
- **Intuition.** Since each value is drawn from distribution F the empirical average over i.i.d. draws from F, by law of large numbers, should be very close to expected value

Bounding Error via Representativeness

- We will try to argue the expected performance

$$E_S[R(h_S)] \geq \max_{h \in H} R(h) - \epsilon(m)$$

- Expected Sample Average *Representativeness*: suppose that

$$\text{Rep} = \boxed{E_S\left[\sup_{h \in H} R_S(h) - R(h)\right]} \leq \epsilon(m)$$

How good is the sample average in terms of **representing** the population expectation, in the worst case over H

- Then we can prove expected error of $\epsilon(m)$

$$E_S[R(h_S)] = E[R_S(h_S)] - E[R_S(h_S) - R(h_S)] \geq E[R_S(h_S)] - \epsilon(m)$$

- Since h_S optimizes $R_S(h)$ and $h_* = \text{argmax}_{h \in H} R(h)$ is feasible

$$E[R_S(h_S)] \geq \boxed{E[R_S(h_*)] = R(h_*)}$$

h_* does not depend on the samples

$$E[R_S(h_*)] \overset{\text{def}}{=} \frac{1}{m} \sum_j E\left[h\left(v^j; h_*\right)\right] = E[h(v; h_*)] = R(h_*)$$

If we can bound representativeness

$$\mathrm{Rep} = E_S\left[\sup_h R_S(h) - R(h)\right] \leq \epsilon(m)$$

Then we can bound expected performance

$$E[R(h_S)] \geq E[R(h_*)] - \epsilon(m)$$

Bounding Error via Representativeness*

- We will try to argue the expected performance

$$E_S[R(h_S)] \geq \max_{h \in H} R(h) - \epsilon(m)$$

- Expected Sample Average *Representativeness*: suppose that

$$\mathrm{Rep}_* = \boxed{E_S[R_S(h_S) - R(h_S)]} \leq \epsilon(m)$$

How good is the sample average in terms of **representing** the population expectation, **of the ERM hypothesis**

- Then we can prove expected error of $\epsilon(m)$

$$E_S[R(h_S)] = E[R_S(h_S)] - E[R_S(h_S) - R(h_S)] \geq E[R_S(h_S)] - \epsilon(m)$$

- Since h_S optimizes $R_S(h)$ and $h_* = \mathrm{argmax}_{h \in H} R(h)$ is feasible

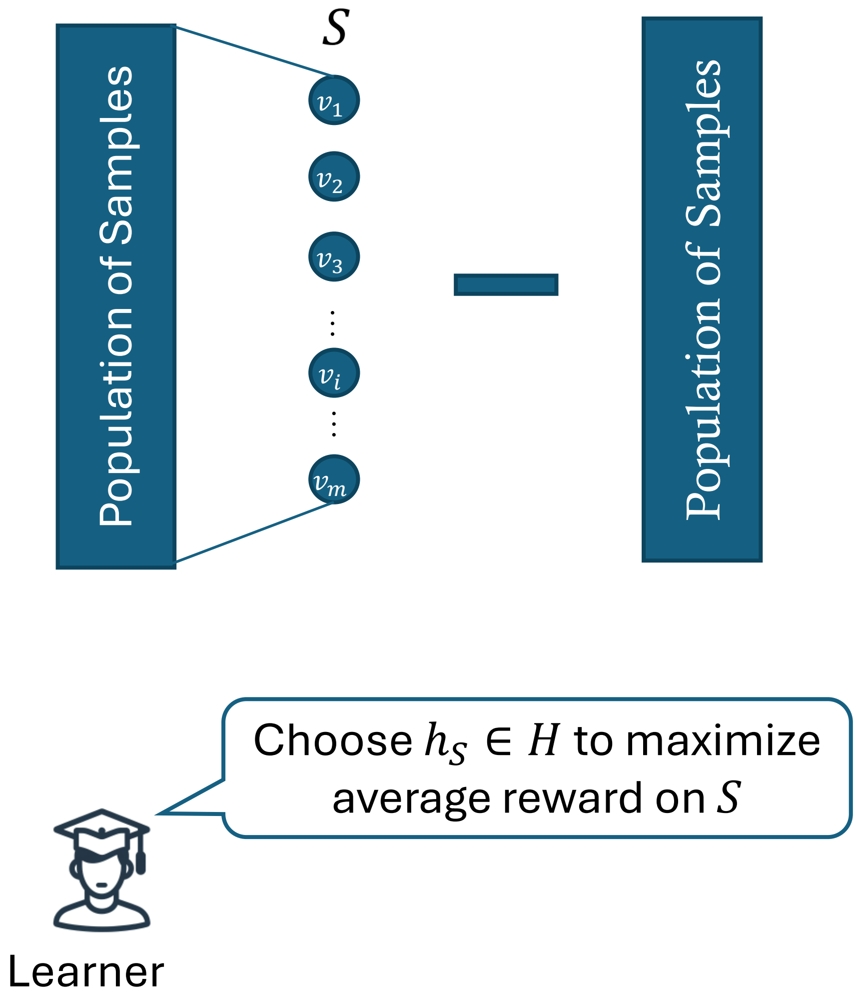
$$E[R_S(h_S)] \geq \boxed{E[R_S(h_*)] = R(h_*)}$$

h_* does not depend on the samples

$$E[R_S(h_*)] \stackrel{\text{def}}{=} \frac{1}{m} \sum_j E\left[h\left(v^j; h_*\right)\right] = E[h(v; h_*)] = R(h_*)$$

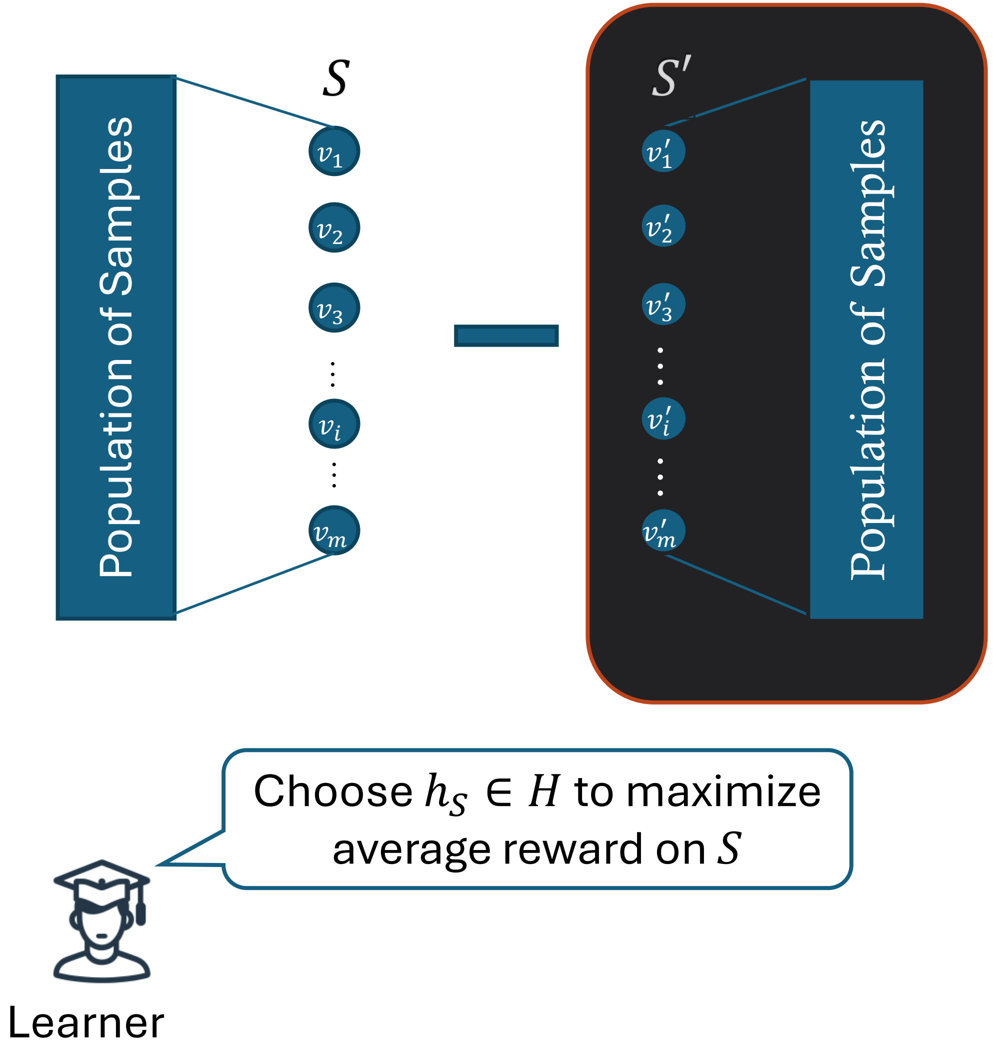
$$\text{Rep}_* = \boxed{E[R_S(h_S) - R_D(h_S)]}$$

How good is the sample average in terms of **representing** the population expectation, **of the ERM hypothesis**



$$\text{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$
$$\text{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

How good is the sample average in terms of **representing** the population expectation, **of the ERM hypothesis**



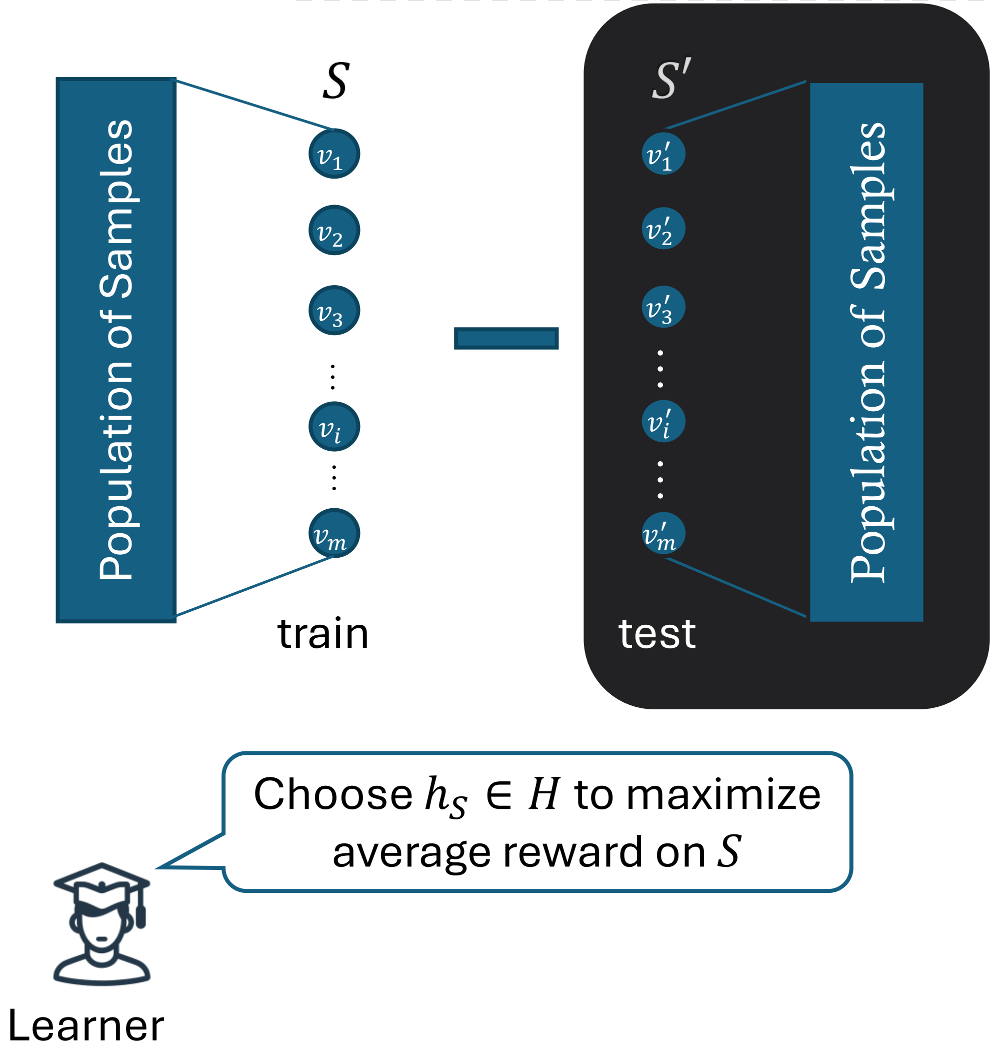
$$\text{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

$$\text{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

$$\text{Rep}_* = \boxed{E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]}$$

How good is the sample average in terms of **representing** the population expectation, **of the ERM hypothesis**

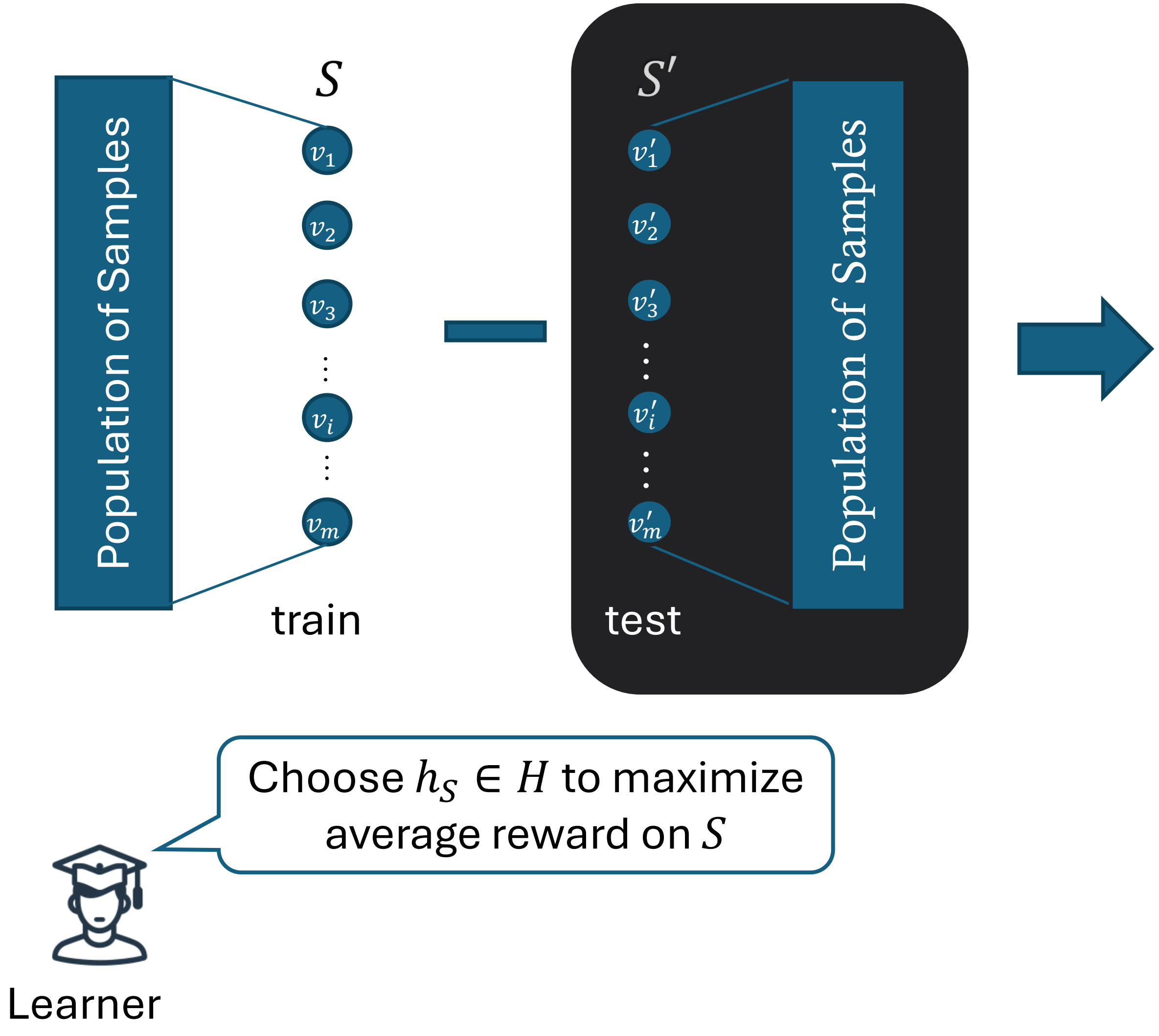
If we randomly split $2m$ samples into a **train and test split**, and choose a hypothesis based on train, **how different is the training reward from the test reward**

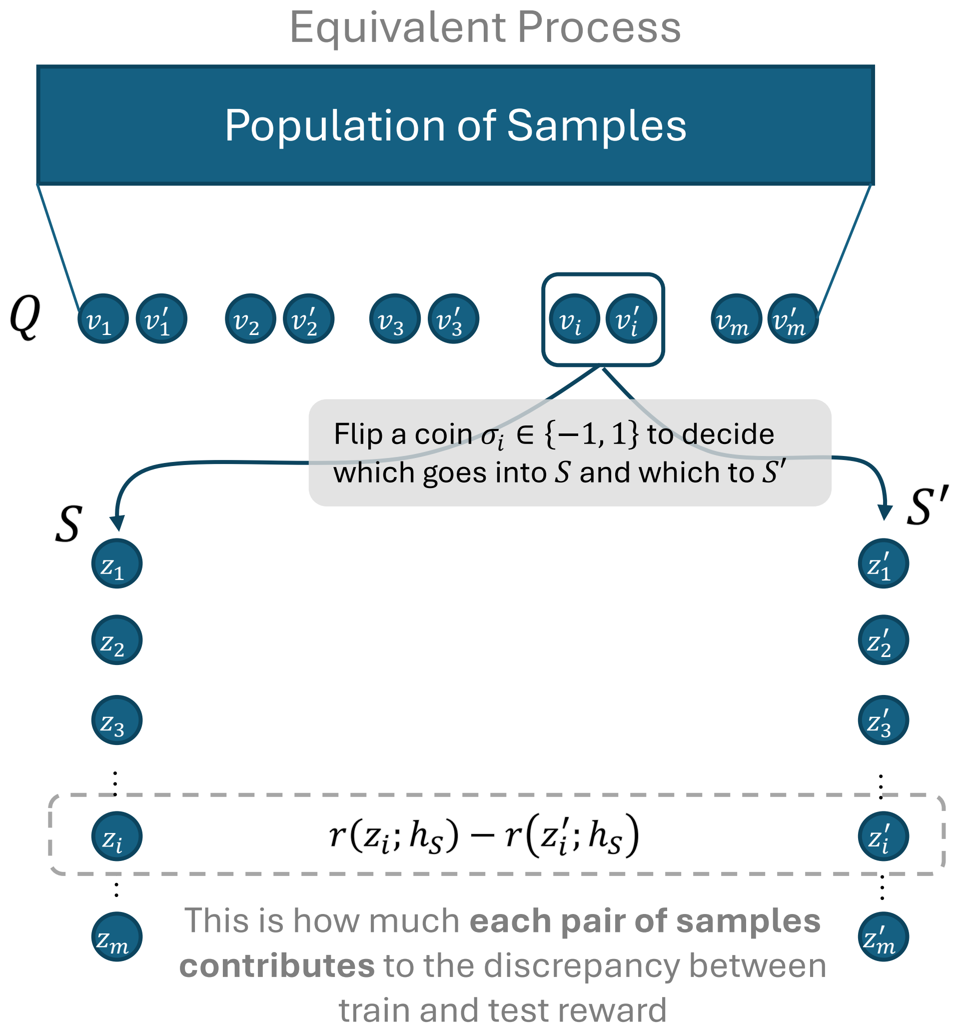


$$\mathrm{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

$$\mathrm{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

$$\mathrm{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$



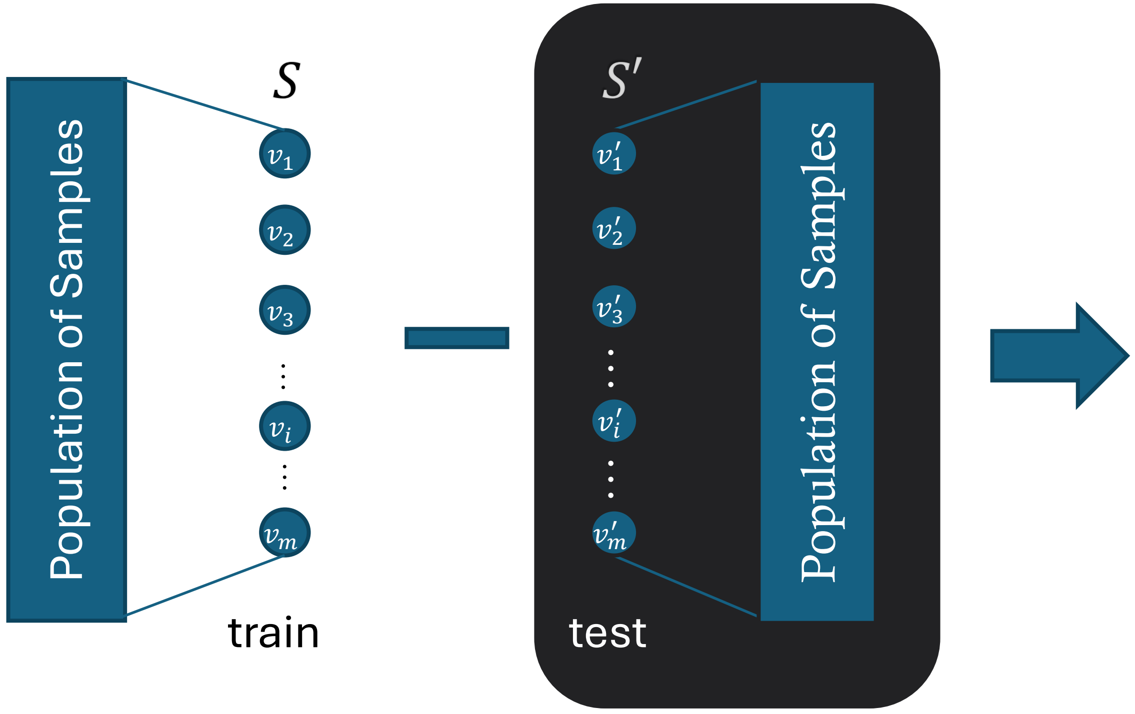


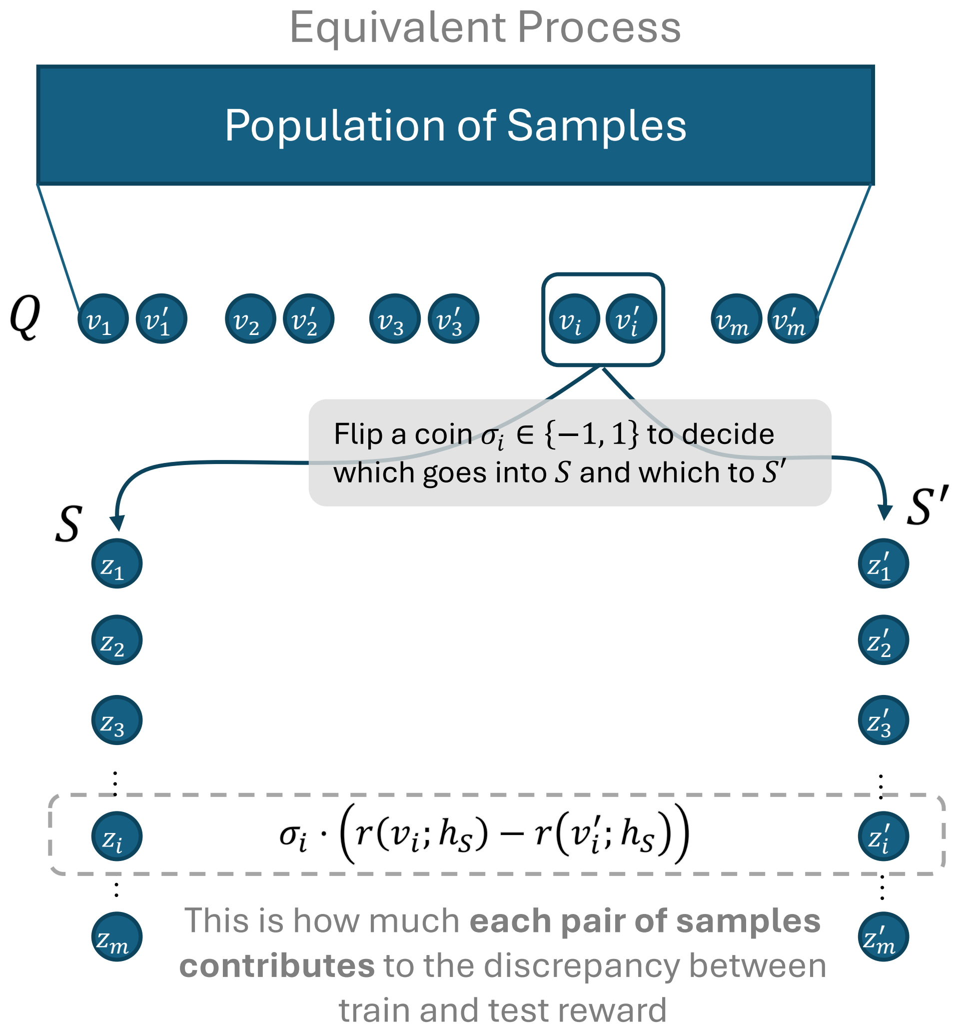
If we randomly split $2m$ samples into a **train and test split**, and choose a hypothesis based on train, **how different is the training reward from the test reward**

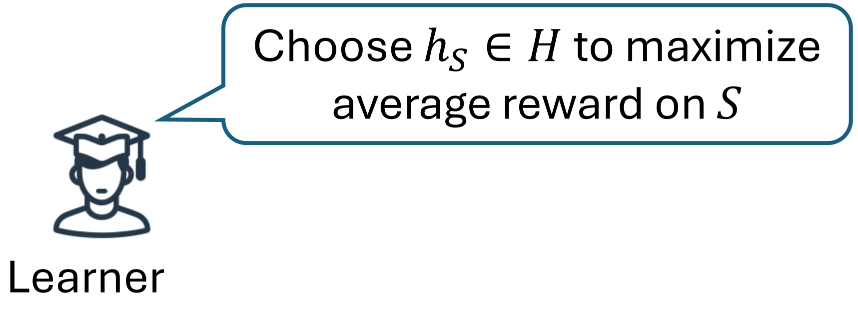
$$\mathrm{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

$$\mathrm{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

$$\mathrm{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$





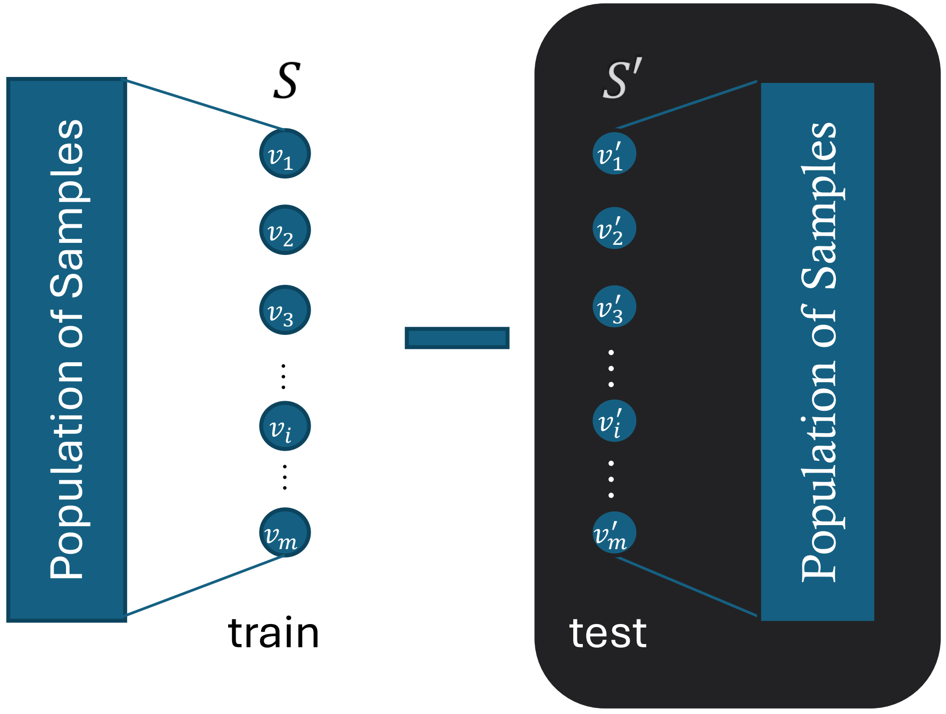


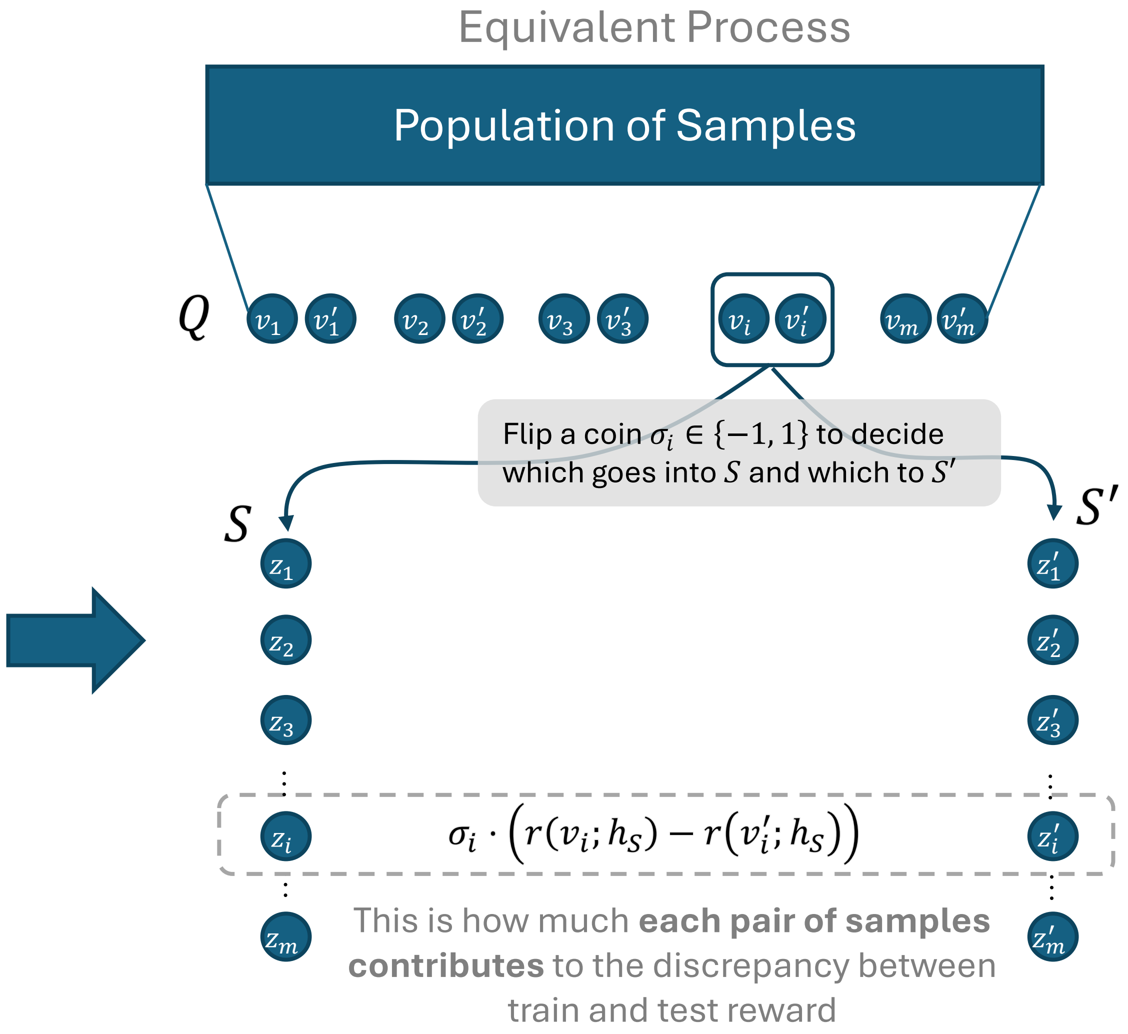
If we randomly split $2m$ samples into a **train and test split**, and choose a hypothesis based on train, **how different is the training reward from the test reward**

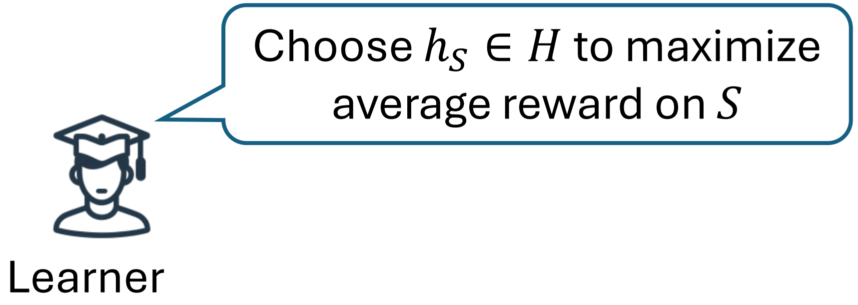
$$\text{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

$$\text{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

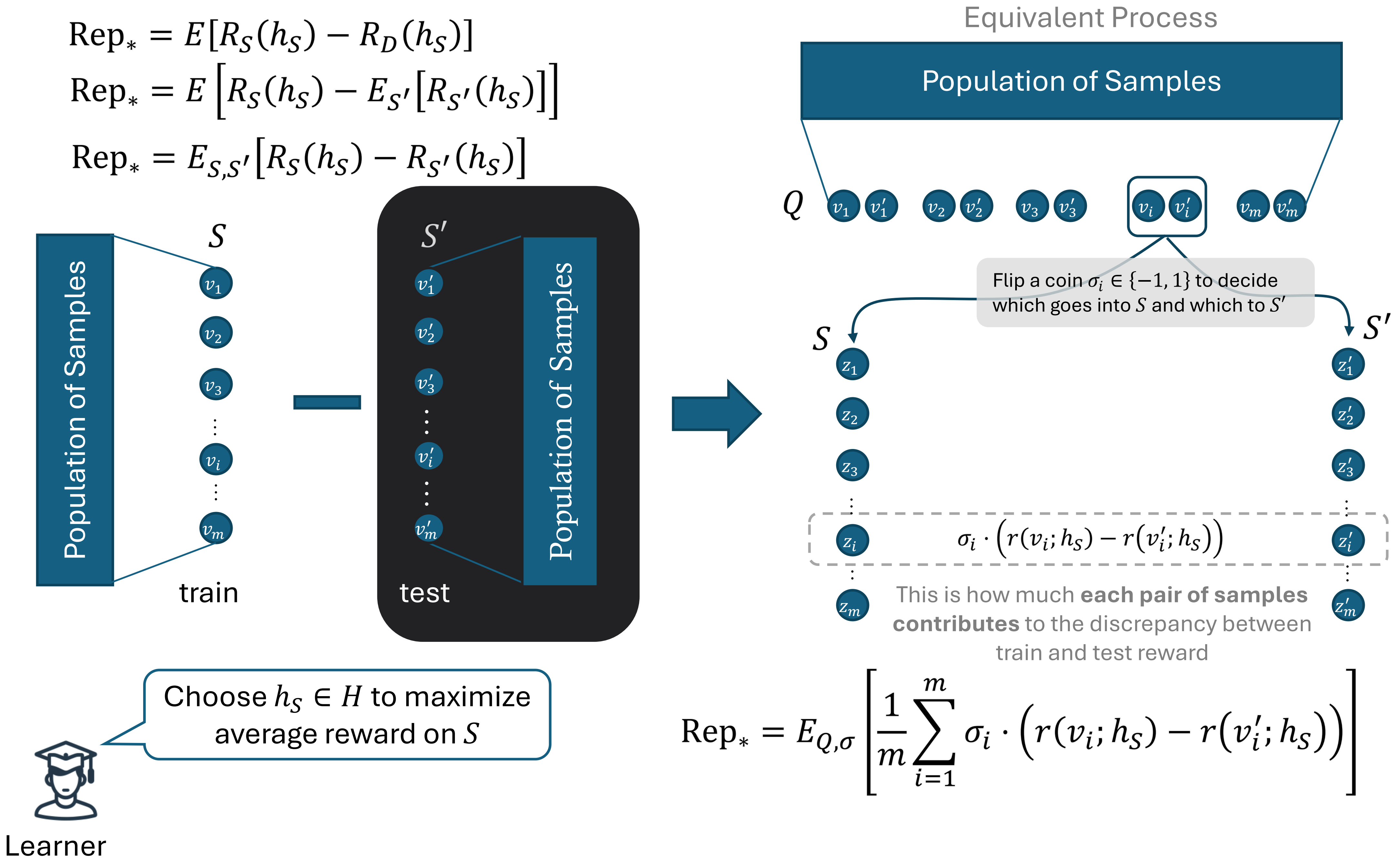
$$\text{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$







Important. If h was not chosen based on S but was some fixed $h \in H$. Then this contribution is mean zero in expectation over the random split σ_i, even when we condition on the sample values Q

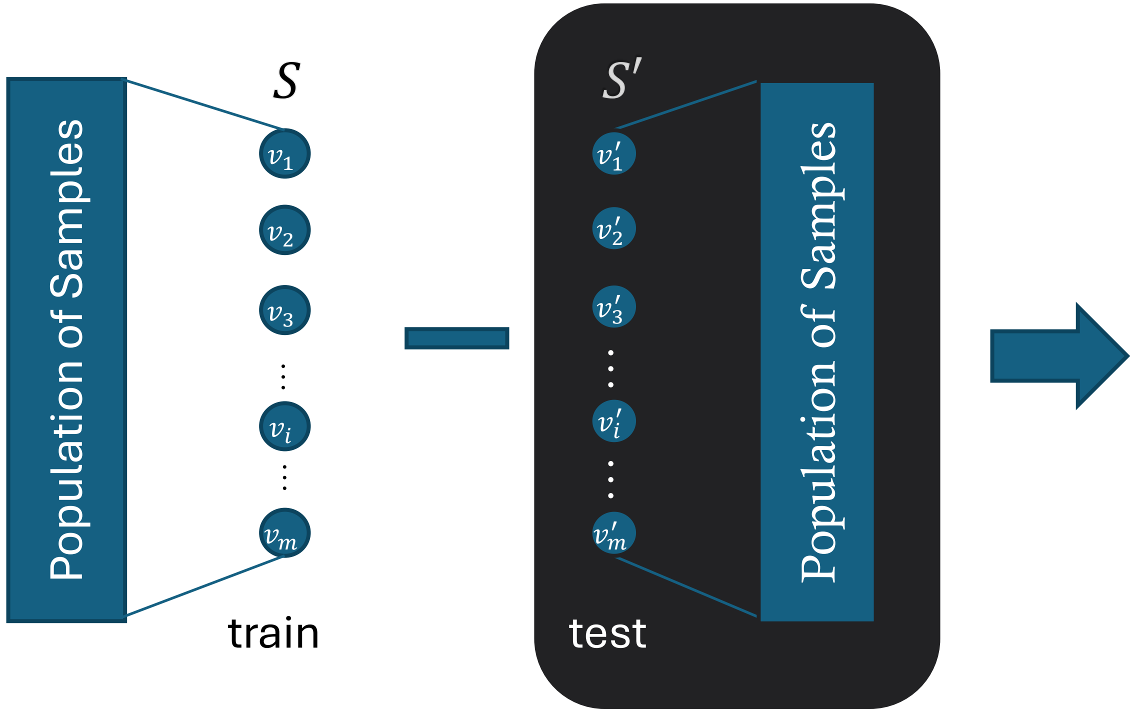


$\text{Rep}_* = E[R_S(h_S) - R_D(h_S)]$
$\text{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$
$\text{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$
Population of Samples
S
v_1 v_2 v_3 v_i v_m
train
S'
v_1' v_2' v_3' v_i' v_m'
Population of Samples
test
Choose $h_S \in H$ to maximize average reward on S
Learner
Equivalent Process
Population of Samples
Q
v_1 v_1' v_2 v_2' v_3 v_3' v_i v_i' v_m v_m'
Flip a coin $\sigma_i \in \{-1, 1\}$ to decide which goes into S and which to S'
S
z_1 z_2 z_3 z_i z_m
S'
z_1' z_2' z_3' z_i' z_m'
$\sigma_i \cdot \left(r(v_i; h_S) - r(v_i'; h_S)\right)$
This is how much **each pair of samples contributes** to the discrepancy between train and test reward
$\text{Rep}_* = E_{Q,\sigma}\left[\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot \left(r(v_i; h_S) - r(v_i'; h_S)\right)\right]$

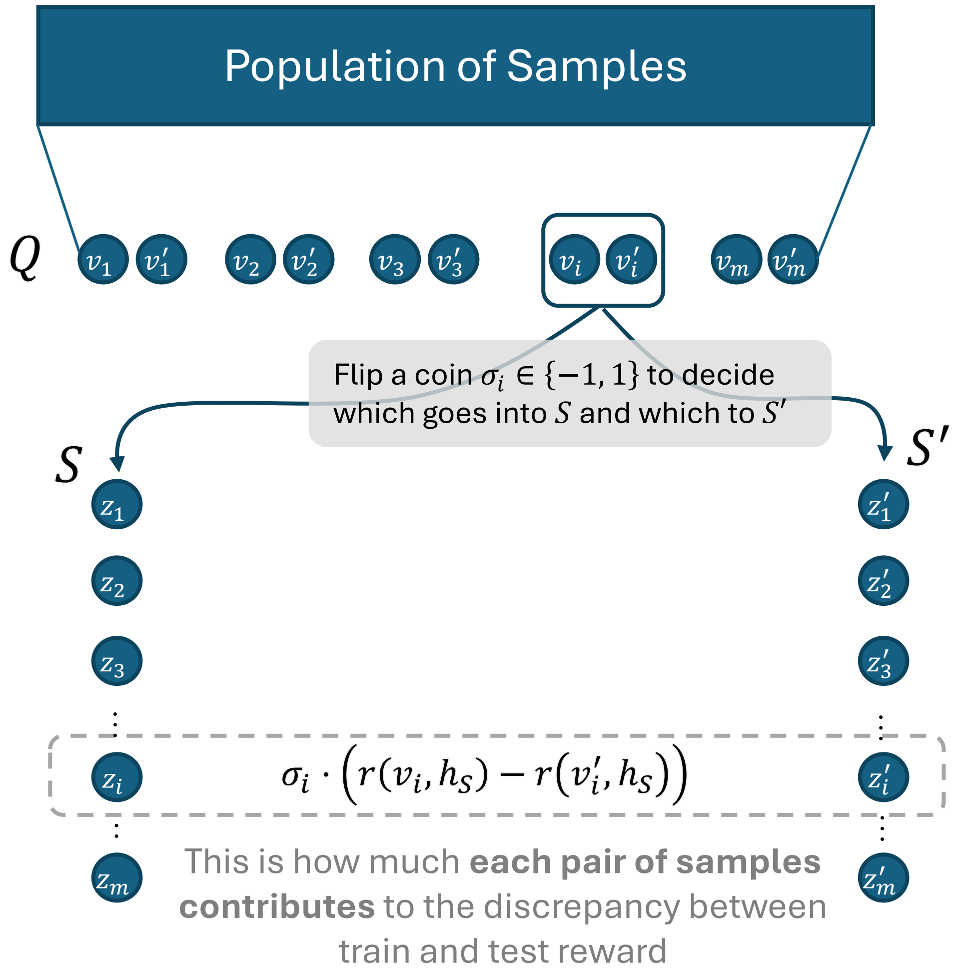
$$\text{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

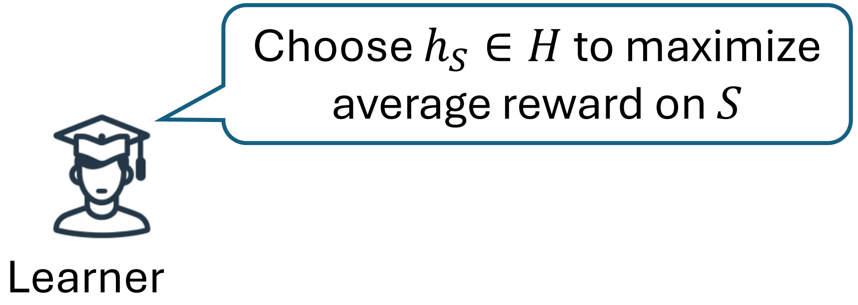
$$\text{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

$$\text{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$



Equivalent Process





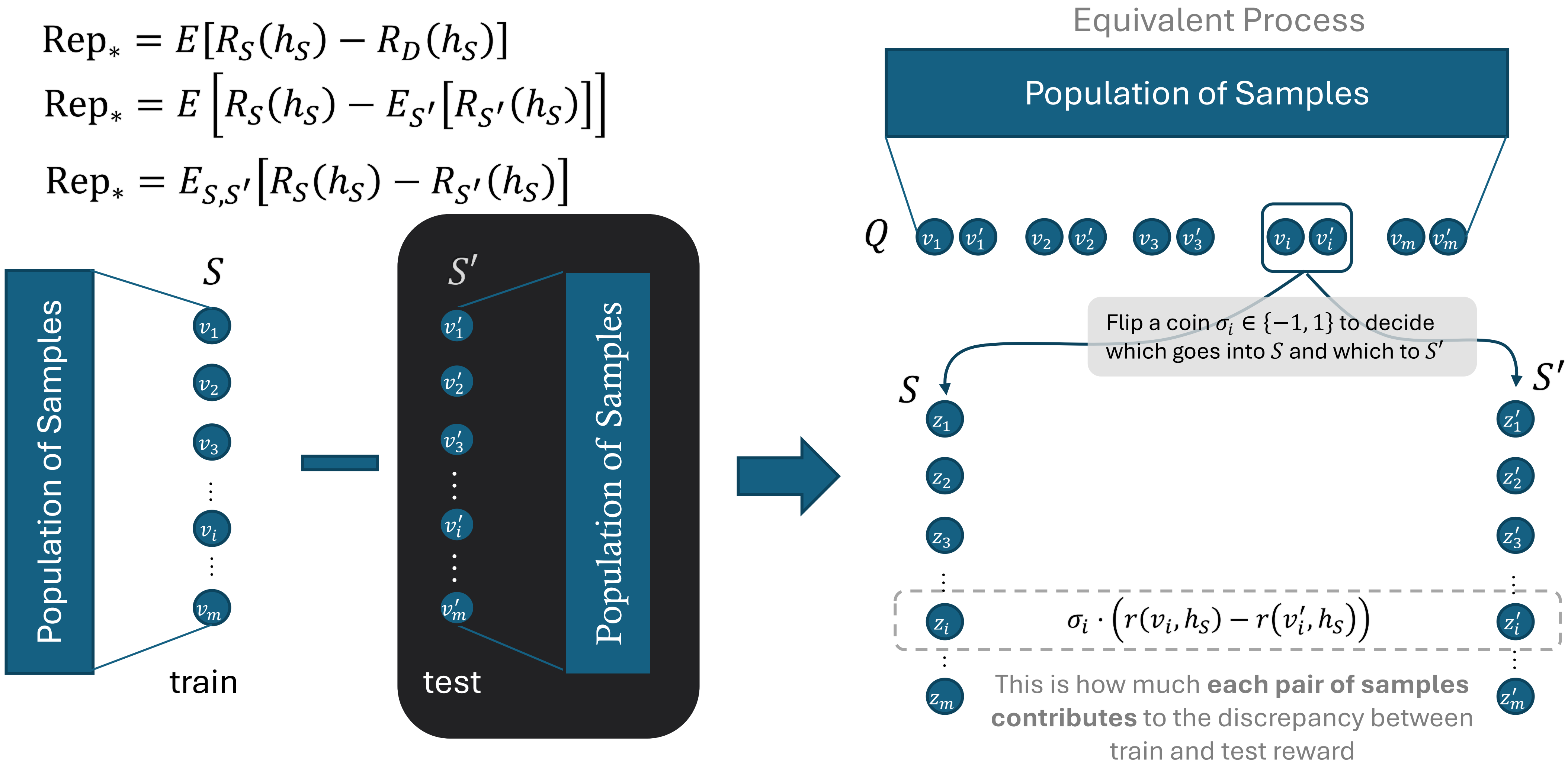
$$\text{Rep}_* = E_{Q,\sigma}\left[\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot \left(r(v_i; h_S) - r(v_i'; h_S)\right)\right]$$

Since it is hard to argue about the "inner-workings" of ERM, let's be pessimistic here and replace h_S **with an adversarial choice**

$$\mathrm{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

$$\mathrm{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

$$\mathrm{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$



$$\mathrm{Rep}_* \leq E_{Q,\sigma}\left[\max_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \cdot \left(r(v_i; h) - r(v_i'; h)\right)\right]$$

Choose $h_S \in H$ to maximize average reward on S

Learner

Adversary

For any Q, σ, choose $h \in H$ to maximize the average discrepancy

Symmetrization and Rademacher Complexity

- We can upper bound representativeness by

$$\mathrm{Rep}_* \leq E_{S,S',\sigma}\left[\max_{h\in H}\frac{1}{m}\sum_{j=1}^{m}\sigma_j\left(r(v_j;h)-r\left(v_j';h\right)\right)\right]$$

- We can upper bound by splitting the max into the two separate

$$\mathrm{Rep}_* \leq E_{S,\sigma}\left[\max_{h\in H}\frac{1}{m}\sum_{j=1}^{m}\sigma_j r(v_j;h)\right]+E_{S',\sigma}\left[\max_{h\in H}-\frac{1}{m}\sum_{j=1}^{m}\sigma_j r\left(v_j';h\right)\right]$$

- But these two quantities are the same

$$\mathrm{Rep} \leq \boxed{2\,E_{S,\sigma}\left[\max_{h\in H}\frac{1}{m}\sum_{j=1}^{m}\sigma_j r(v_j;h)\right]}$$

Rademacher Complexity of Hypothesis Space H

Empirical Rademacher Complexity

Empirical Rademacher Complexity of hypothesis space H on samples S:

$$\mathrm{Rad}(S, H) \coloneqq 2E_\sigma \left[\max_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \cdot r(v_i; h) \right]$$

Theorem. We have thus proven that:

$$E[R(h_S)] \geq R(h_*) - E_S[\mathrm{Rad}(S, H)]$$

Bounding Empirical Rademacher Complexity

- We have now conditioned on the value of the samples $S = \{v_1, \ldots, v_m\}$
- For a fixed h, contribution of each sample v_i, in expectation over the random split σ_i, is zero

$$E_\sigma\left[\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot r(v_i; h)\right] = 0$$

- We have reduced to arguing how large this "almost mean zero" quantity can be for a fixed set of samples $\{v_1, \ldots, v_m\}$
- Requires arguing properties of the hypothesis space H on the samples S and not on the whole unknown support of the unknown distribution F

Simple Case

- Suppose that H was finite, i.e. $H = \{h_1, \dots, h_K\}$ and reward bounded in $[0,1]$
- By Hoeffding concentration inequality and the mean-zero property, for each h_t, w.p. $1-\delta$

$$\left|\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot r(v_i; h_t)\right| \leq \sqrt{\frac{\log(2/\delta)}{2m}}$$

- By the union bound, w.p. $1-\delta$

$$\max_{t=1}^{K} \left|\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot r(v_i; h_t)\right| \leq \sqrt{\frac{\log(2K/\delta)}{2m}}$$

- This implies that the expected value of this quantity is of the same order

$$\mathrm{Rad}(S, H) \coloneqq 2E_\sigma\left[\max_{h\in H}\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot r(v_i; h)\right] \approx \sqrt{\frac{\log(2K)}{2m}}$$

Massart's lemma. For any finite hypothesis space H:

$$\mathrm{Rad}(S, H) \leq 2\sqrt{\frac{2\log(|H|)}{m}}$$

Beyond Simple Case

- Suppose we can find a finite subspace $\widetilde{H}_S \subseteq H$ such that every $h \in H$ has a **representative** $\tilde{h} \in \widetilde{H}_S$ that has the exact same behavior on the samples S

$$\forall v_i \in S: r(v_i; h) = r\left(v_i; \tilde{h}\right)$$

- Then Empirical Rademacher Complexity of H is upper bounded by that of $\widetilde{H}_S$

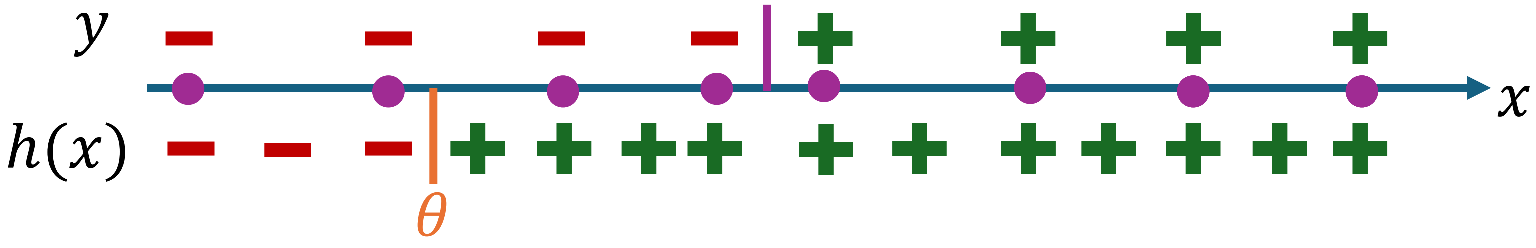
$$\mathrm{Rad}(S, H) \coloneqq 2E_\sigma\left[\max_{h\in H} \frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot r(v_i; h)\right]$$

$$\leq 2E_\sigma\left[\max_{h\in \widetilde{H}_S} \frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot r(v_i; h)\right] \leq 2\sqrt{\frac{2\log\left(\left|\widetilde{H}_S\right|\right)}{m}}$$

Beyond Simple Case

- Suppose we can find a finite subspace $\widetilde{H}_S \subseteq H$ such that every $h \in H$ has a **representative** $\tilde{h} \in \widetilde{H}_S$ that has the exact same behavior on the samples S
- Classification Example with Threshold Functions: Data $v = (x, y)$, with $x \in [-1, 1]$ and $y \in \{-1,1\}$. Label every $x \geq \theta$ with $+1$ and every $x < \theta$ with -1.

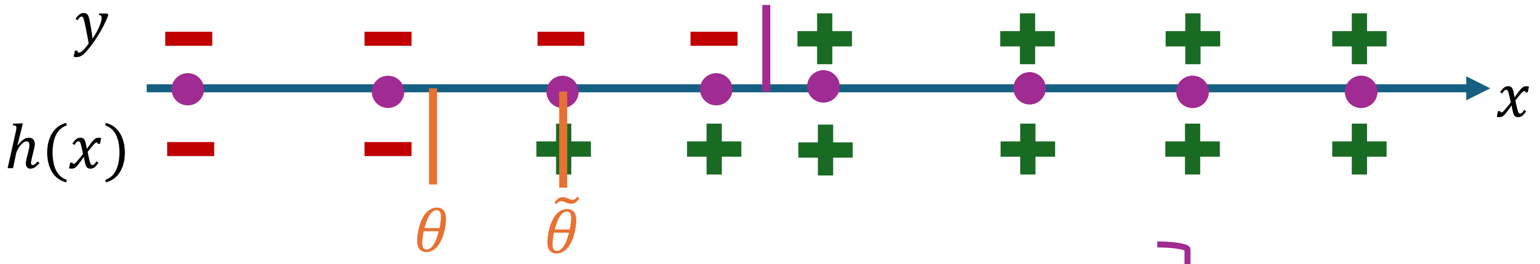
$$H = \{x \to 1(x \geq \theta) - 1(x < \theta) : \theta \in \Theta\}$$



Beyond Simple Case

- Suppose we can find a finite subspace $\widetilde{H}_S \subseteq H$ such that every $h \in H$ has a **representative** $\tilde{h} \in \widetilde{H}_S$ that has the exact same behavior on the samples S
- Classification Example with Threshold Functions: Data $v = (x, y)$, with $x \in [-1, 1]$ and $y \in \{-1,1\}$. Label every $x \geq \theta$ with $+1$ and every $x < \theta$ with -1

$$H = \{x \to 1(x \geq \theta) - 1(x < \theta) : \theta \in \Theta\}$$



- We only look at the behavior on the samples
- Suffices to look at thresholds equal to a sample or 1

$$\mathrm{Rad}(S, H) \leq 2\sqrt{\frac{2\log(m+1)}{m}}$$

Growth Rate of Function Space

- Suppose we can find a finite subspace $\widetilde{H}_S \subseteq H$ such that every $h \in H$ has a **representative** $\tilde{h} \in \widetilde{H}_S$ that has the exact same behavior on the samples S
$$\forall v_i \in S: r(v_i; h) = r\left(v_i; \tilde{h}\right)$$
- Empirical Rademacher Complexity of H is upper bounded by that of $\widetilde{H}_S$
- Growth Rate $\tau(m, H)$: the size of the smallest $\widetilde{H}_S$ that satisfies the above property, in the worst case over sample dataset of size m
- Example. For threshold classifiers $\tau(m, H) = m + 1$

Theorem. For any hypothesis H

$$\mathrm{Rad}(S, H) \leq \boxed{2\sqrt{\frac{2\log(\tau(m, H))}{m}}}$$

SideNote For classification, a seminal notion is the Vapnik-Chervonenkis **(VC) dimension**: size d of largest dataset that the hypothesis can assign labels in all possible manners

Cannot be assigned by threshold classifiers $\Rightarrow d = 2$

Sauer's Lemma. If has VC-dim $\leq d$ then $\tau(m, H) \lesssim 2^d \Rightarrow \mathrm{Rad}(S, H) \lesssim \sqrt{d/m}$

Discretization on Samples

- Suppose we can find a finite subspace $\widetilde{H}_{S,\epsilon} \subseteq H$ such that every $h \in H$ has a **representative** $\tilde{h} \in \widetilde{H}_{S,\epsilon}$ that has approximately the same behavior on the samples S

$$\forall v_i \in S: \left| r(v_i; h) - r\left(v_i; \tilde{h}\right) \right| \leq \epsilon$$

- Empirical Rademacher Complexity of H upper bounded approximately by $\widetilde{H}_{S,\epsilon}$

$$\mathrm{Rad}(S, H) \coloneqq 2E_\sigma \left[\max_{h \in H} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \cdot r(v_i; h) \right]$$

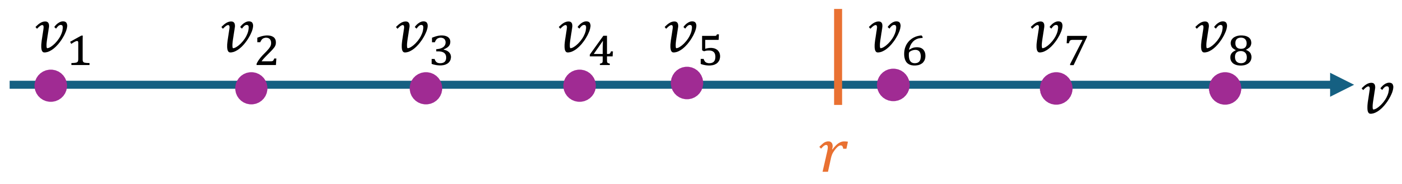
$$\leq 2E_\sigma \left[\max_{h \in \widetilde{H}_{S,\epsilon}} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \cdot r(v_i; h) \right] + 2\epsilon \leq 2\sqrt{\frac{2\log\left(\left|\widetilde{H}_{S,\epsilon}\right|\right)}{m}} + 2\epsilon$$

Back to Mechanism Design from Samples

Example: Pricing from Samples

- Suppose we are given a set of samples S of a bidder's value
- We optimize over the space of posted prices
- For every price r we want to find a price $\tilde{r}$ that achieves almost the same revenue as r for every value in the samples

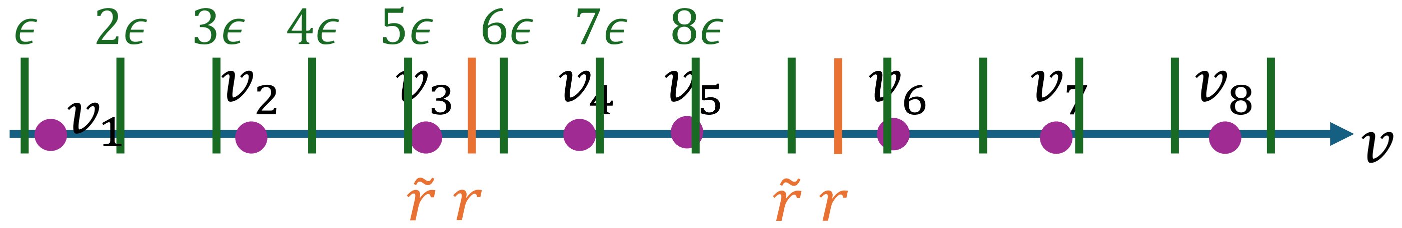
$$\forall v_i \in S\colon\ |r \cdot 1\{v_i \geq r\} - \tilde{r} \cdot 1\{v_i \geq \tilde{r}\}| \leq \epsilon$$



Example: Pricing from Samples

- Suppose we are given a set of samples S of a bidder's value
- We optimize over the space of posted prices
- For every price r we want to find a price $\tilde{r}$ that achieves almost the same revenue as r for every value in the samples

$$\forall v_i \in S\colon\ |r \cdot 1\{v_i \geq r\} - \tilde{r} \cdot 1\{v_i \geq \tilde{r}\}| \leq \epsilon$$



- For every r, pick maximum of {largest multiple of ϵ below r, largest sampled value below r}. At most $m + 1/\epsilon$ prices.

Example: Pricing from Samples

- Suppose we are given a set of samples S of a bidder's value
- We optimize over the space of posted prices
- For every price r we want to find a price $\tilde{r}$ that achieves almost the same revenue as r for every value in the samples

$$\forall v_i \in S\colon\ |r \cdot 1\{v_i \geq r\} - \tilde{r} \cdot 1\{v_i \geq \tilde{r}\}| \leq \epsilon$$

- For every r, pick maximum of {largest multiple of ϵ below r, largest sampled value below r}. At most $m + 1/\epsilon$ prices.

$$\mathrm{Rad}(S, H) \leq 2\sqrt{\frac{2\log\left(m + \frac{1}{\epsilon}\right)}{m}} + 2\epsilon \underset{\epsilon = 1/m}{\leq} 4\sqrt{\frac{2\log(2m)}{m}}$$

Second Price with a Reserve

- Suppose we are given a set of samples S of n bidder value profiles
- Optimize over the space of Second-Price Auctions with a Reserve
- For every price r we want to find a price $\tilde{r}$ that achieves almost the same revenue as r for every value in the samples
$$\forall v_i = (v_{i1}, \dots, v_{in}) \in S\colon\ |\mathrm{rev}(v_i; r) - \mathrm{rev}(v_i; \tilde{r})| \leq \epsilon$$
- For every r, pick maximum of {largest multiple of ϵ below r, largest sampled value below r}. At most $m \cdot n + 1/\epsilon$ prices.

$$\mathrm{Rad}(S, H) \leq 2\sqrt{\frac{2\log(m \cdot n + 1/\epsilon)}{m}} + 2\epsilon \leq 4\sqrt{\frac{2\log(2m \cdot n)}{m}}$$

$\epsilon = 1/m$

Second Price with Player-Specific Reserves

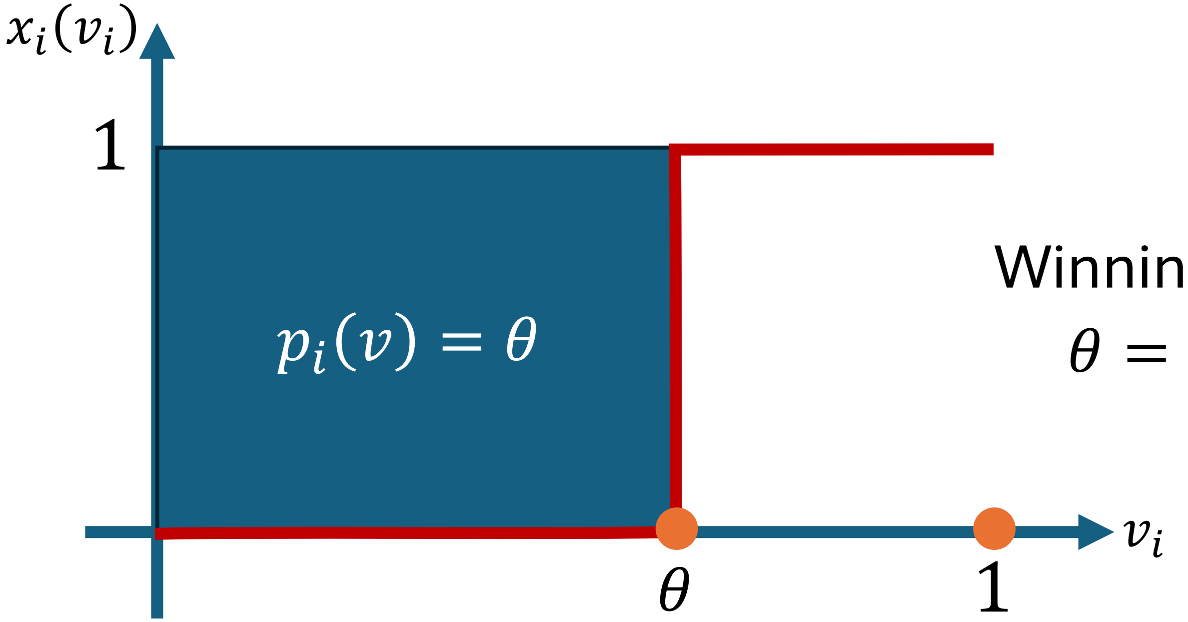
- Suppose we are given a set of samples S of n bidder value profiles
- Optimize over the space of Second-Price with Player-Specific Reserves
- For every price vector $r = (r_1, \ldots, r_n)$ we want to find a vector $\tilde{r}$ that achieves almost the same revenue as r for every value in the samples
$$\forall v_i = (v_{i1}, \ldots, v_{in}) \in S\text{:}\ |\mathrm{rev}(v_i; r) - \mathrm{rev}(v_i; \tilde{r})| \leq \epsilon$$
- For every r_j, pick maximum of {largest multiple of ϵ below r, largest sampled value for bidder j below r}. At most $(m + 1/\epsilon)^n$ prices.

$$\mathrm{Rad}(S, H) \leq 2\sqrt{\frac{2n\log(m + 1/\epsilon)}{m}} + 2\epsilon \leq 4\sqrt{\frac{2n\log(2m)}{m}}$$

$\epsilon = 1/m$

Competing with the Myerson Auction

- Want to optimize over virtual welfare maximizing mechanisms
- For each bidder i, we assign a monotone virtual value function ϕ_i
- Allocate to the bidder with highest positive virtual value $\phi_i(v_i)$
- Charge dominant strategy truthful payments

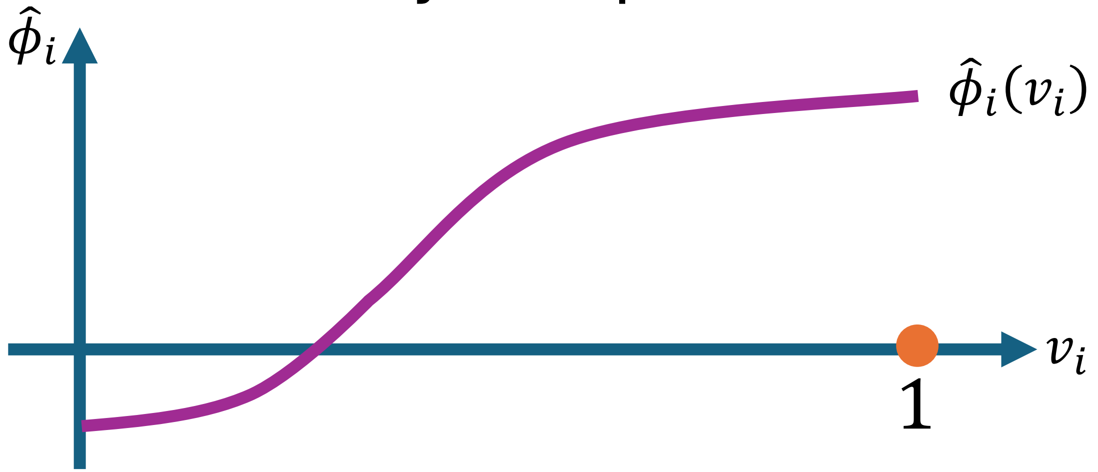


Winning threshold:

$$\theta = \phi_i^{-1}\left(\max_{j \neq i} \phi_j(v_j), \eta_i\right)$$

Optimizing over Virtual Value Functions

- ERM optimizes over all monotone functions for each bidder
- This space is infinite and a bit harder to discretize
- We will see that monotonicity is important!



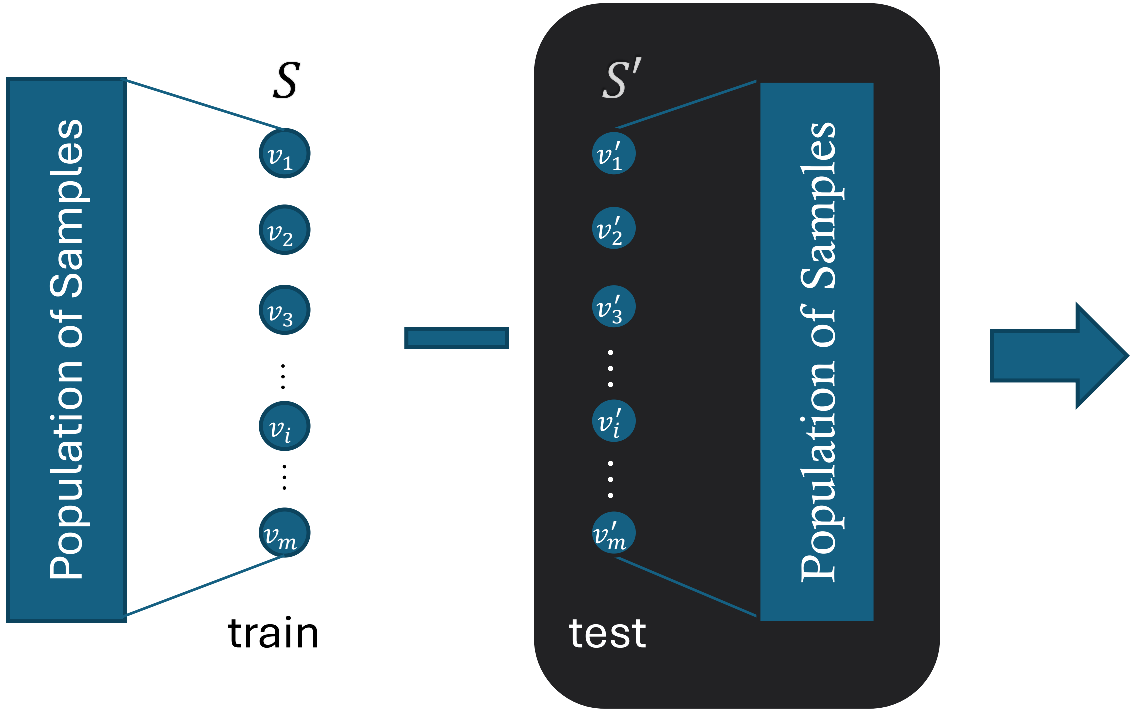
- We introduce a variant of Rademacher complexity analysis that will help us in the analysis of ERM over virtual welfare maximizers

Back to Statistical Learning Theory

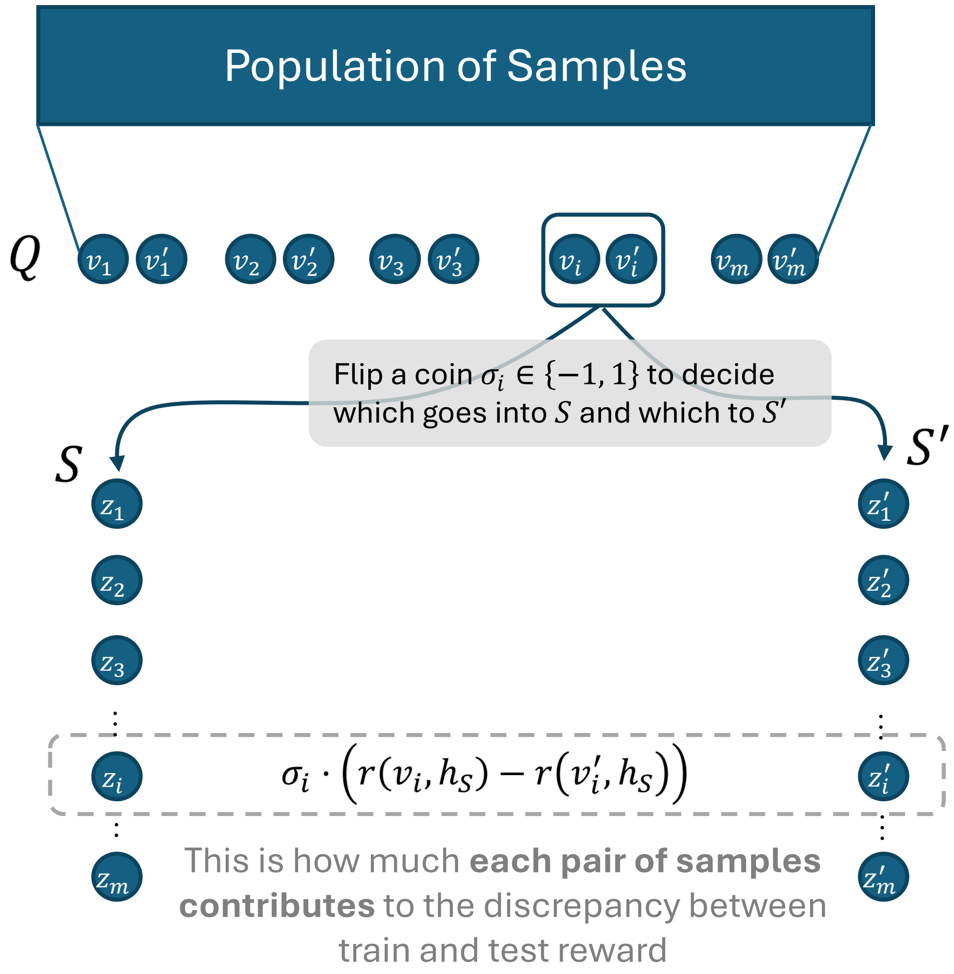
$$\text{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

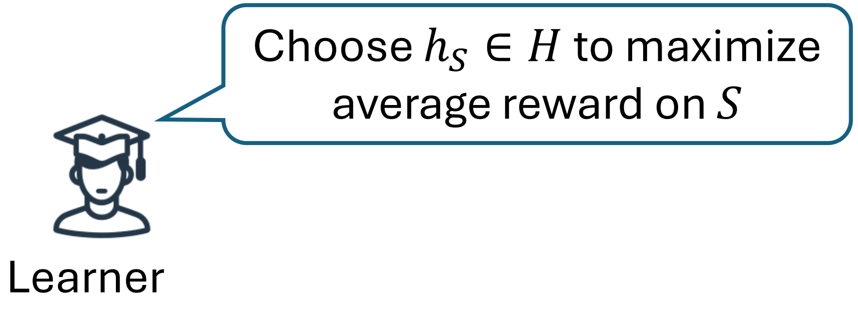
$$\text{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

$$\text{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$



Equivalent Process





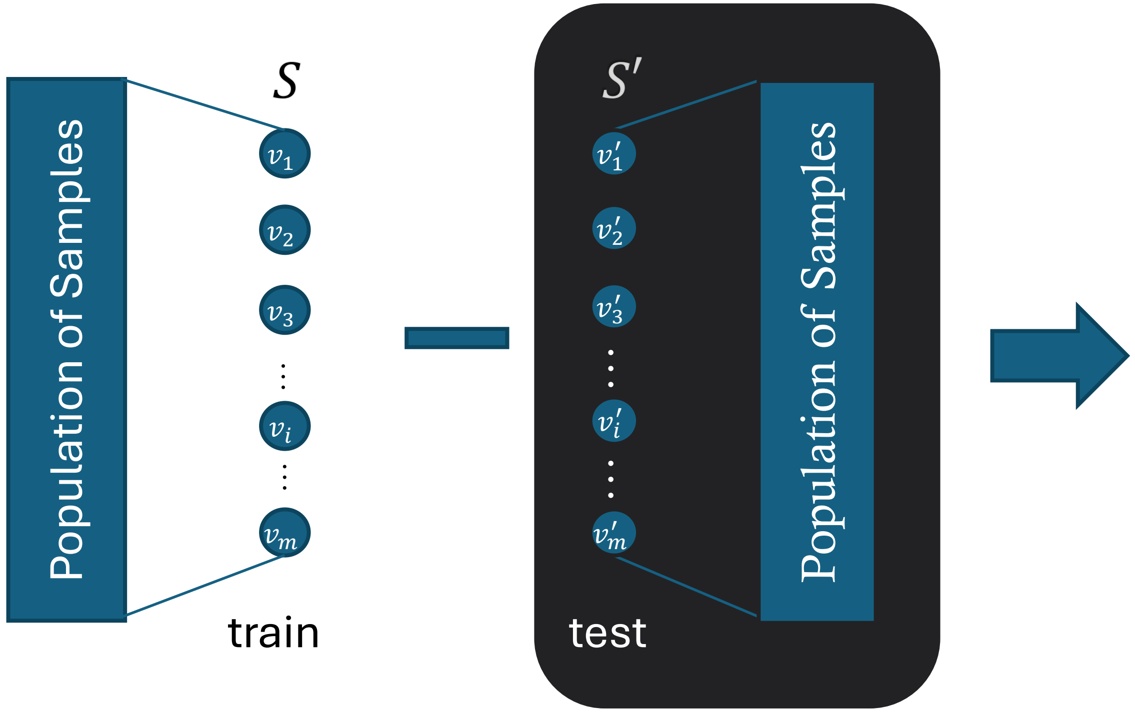
$$\text{Rep}_* = E_{Q,\sigma}\left[\frac{1}{m}\sum_{i=1}^{m} \sigma_i \cdot \left(r(v_i; h_S) - r(v'_i; h_S)\right)\right]$$

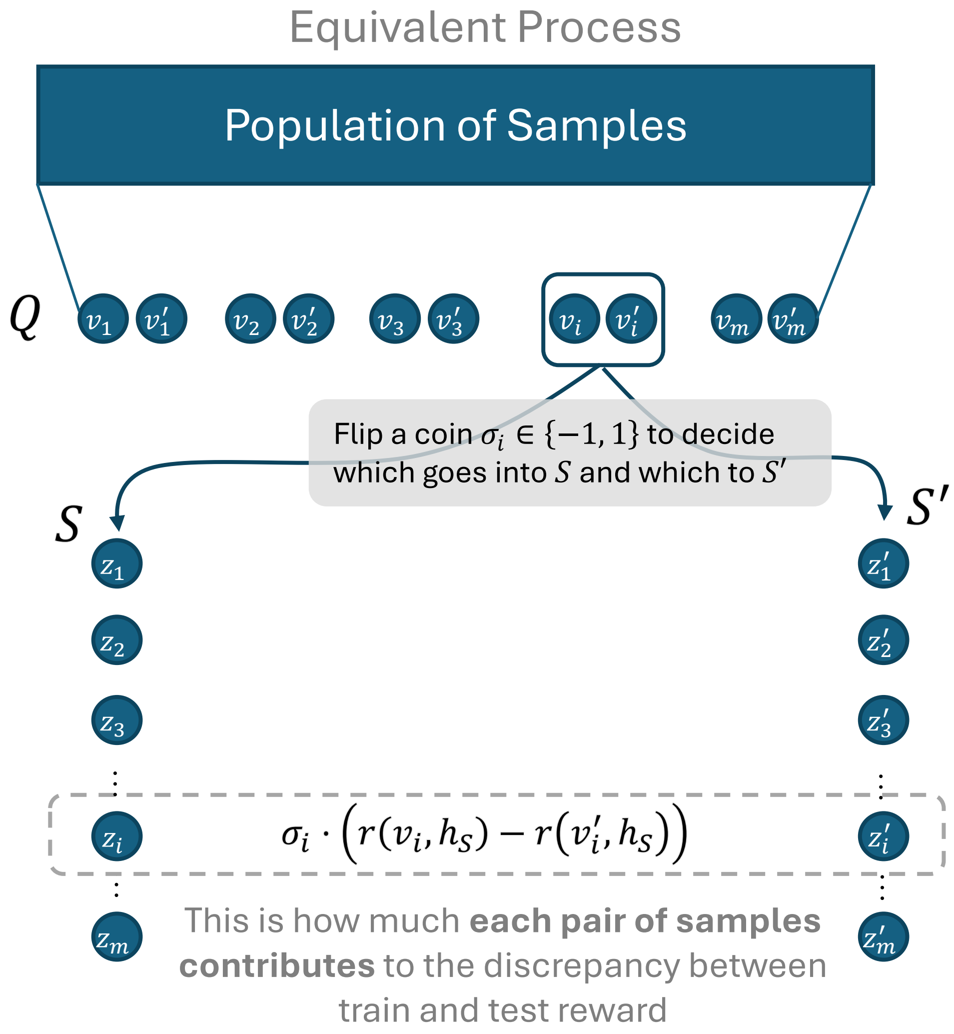
Since it is hard to argue about the "inner-workings" of ERM, let's be pessimistic here and replace h_S **with an adversarial choice over all possible outputs of ERM on a half-sample of Q**

$$\mathrm{Rep}_* = E[R_S(h_S) - R_D(h_S)]$$

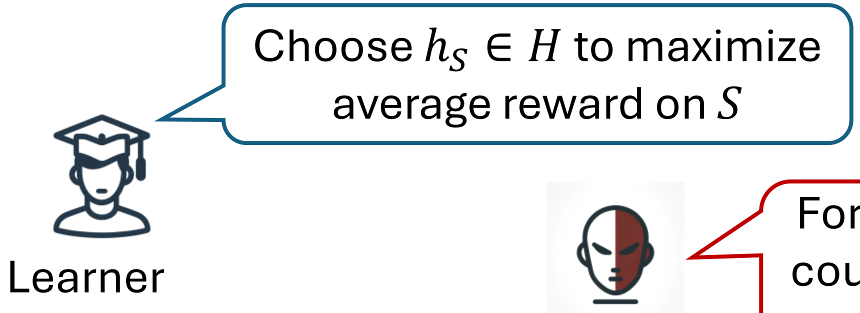
$$\mathrm{Rep}_* = E\left[R_S(h_S) - E_{S'}[R_{S'}(h_S)]\right]$$

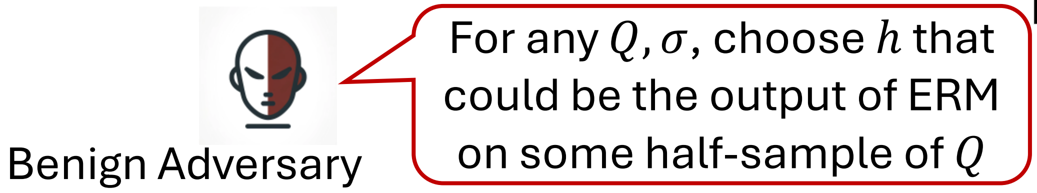
$$\mathrm{Rep}_* = E_{S,S'}[R_S(h_S) - R_{S'}(h_S)]$$





$$\mathrm{Rep}_* \le E_{Q,\sigma}\left[\max_{h \in H_Q} \frac{1}{m} \sum_{i=1}^{m} \sigma_i \cdot \left(r(v_i; h) - r(v_i'; h)\right)\right]$$





Refined Rademacher Complexity

- We can upper bound representativeness by

$$\mathrm{Rep}_* \leq E_{S,S',\sigma}\left[\max_{h\in H_{S\cup S'}} \frac{1}{m}\sum_{j=1}^{m} \sigma_j \left(r(v_j;h) - r\left(v_j';h\right)\right)\right]$$

- We can upper bound by splitting the max into the two separate

$$\mathrm{Rep}_* \leq E_{S,S',\sigma}\left[\max_{h\in H_{S\cup S'}} \frac{1}{m}\sum_{j=1}^{m} \sigma_j r(v_j;h)\right] + E_{S,\,S',\sigma}\left[\max_{h\in H_{S\cup S'}} -\frac{1}{m}\sum_{j=1}^{m} \sigma_j r\left(v_j';h\right)\right]$$

- But these two quantities are the same

$$\mathrm{Rep} \leq 2\, E_{S,S',\sigma}\left[\max_{h\in H_{S\cup S'}} \frac{1}{m}\sum_{j=1}^{m} \sigma_j r(v_j;h)\right]$$

Empirical Rademacher Complexity

Empirical Rademacher Complexity of hypothesis space H on samples S:

$$\mathrm{Rad}(S,H) \coloneqq 2E_\sigma\left[\max_{h\in H}\frac{1}{m}\sum_{i=1}^{m}\sigma_i\cdot r(v_i;h)\right]$$

Theorem. We have thus proven that:

$$E[R(h_S)] \geq R(h_*) - E_{S,S'}\left[\mathrm{Rad}\left(S, H_{S\cup S'}\right)\right]$$

$H_{S\cup S'} \stackrel{\text{def}}{=}$ all possible outputs of ERM on some subset of $S \cup S'$ of size m

Back to Mechanism Design from Samples

Example: Pricing from Samples

- Suppose we are given a set of samples S of a bidder's value
- We optimize over the space of posted prices
- On any subset of size m of a set of samples Q of size $2m$, ERM will return a reserve price that is equal to one of the $2m$ values!

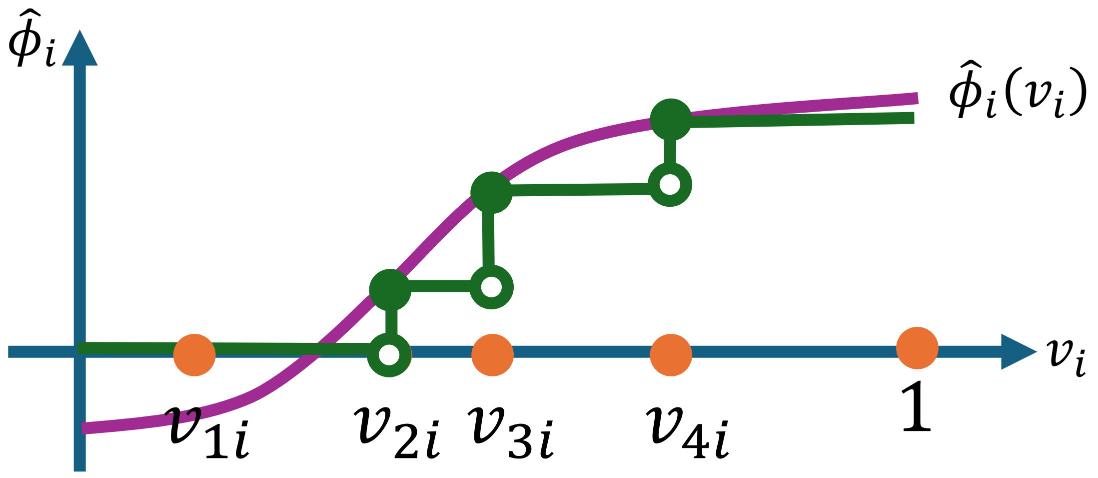
$$H_Q \subseteq \{r : r = v_i \text{ for some } v_i \in Q\}$$

- By Masart's Lemma

$$\mathrm{Rad}\left(S, H_{S \cup S'}\right) \leq 2\sqrt{\frac{2\log(2m)}{m}}$$

Optimizing over Virtual Value Functions

- ERM optimizes over all monotone functions for each bidder
- For any monotone function, we receive strictly larger payment had we used step-function on the samples (threshold to win is higher)!



Samples of bidder i values

- H_Q contains only monotone step functions that change on one of the $2m$ samples for each bidder

Equivalent Representation

- These mechanisms can be thought as follows
- Construct a set of $n \cdot m$ ranked positions
- For each sampled value of a bidder, assign a rank, in a manner that it is monotone across the values of the bidder
- Assign the item to the bidder with the highest rank

How Many are these Mechanisms?

- By monotonicity, each assignment of values to ranks, can be described by:

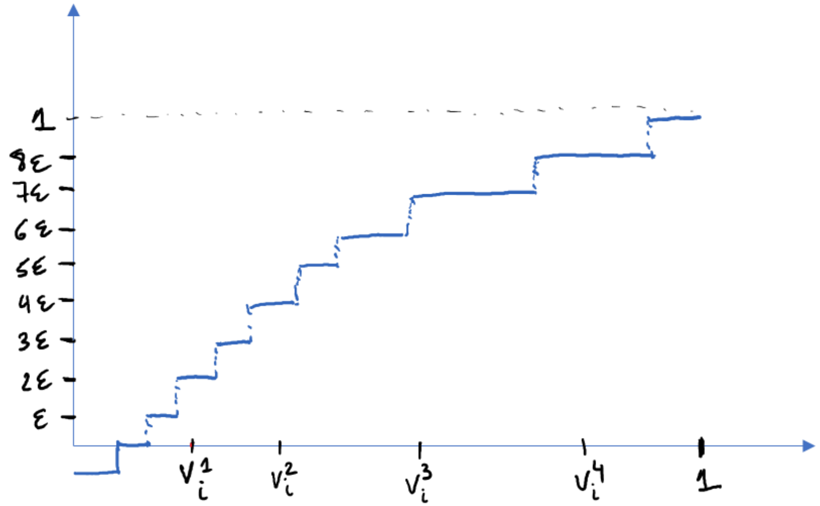
  "for each rank r, specify the smallest of the $2m$ sampled values for which the rank of the bidder goes above r"

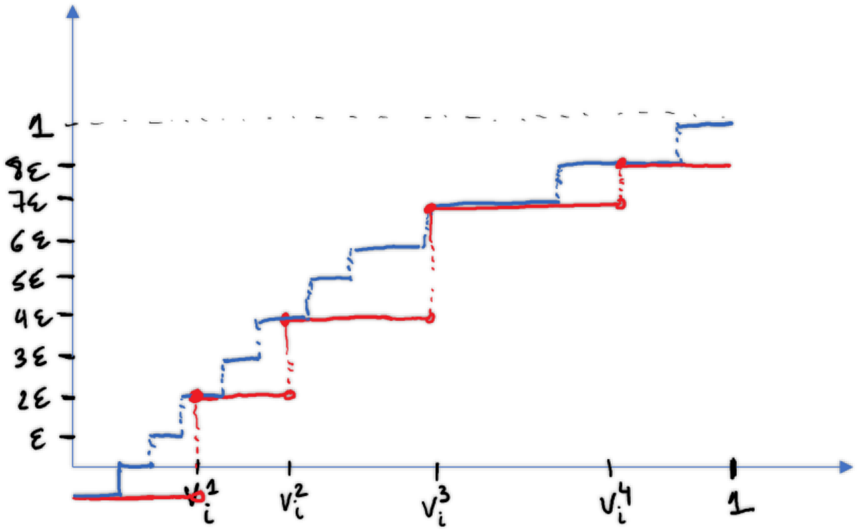
- Roughly $m^{n \cdot m}$ such combinations

$$\mathrm{Rad}\big(S, H_{S \cup S'}\big) \lesssim \sqrt{\frac{n \cdot m \cdot \log(m \cdot n)}{m}} = \sqrt{n \cdot m \cdot \log(m \cdot n)} \to \infty$$

Coarsen Space of Mechanisms we Optimize

- Consider only virtual value functions that take values on an ϵ-grid
$$\phi_i(v_i) \in \{-\epsilon, 0, \epsilon, \dots, 1\}$$





- These step functions in H_Q can be described by
"for each value r on the grid, specify the smallest of the $2m$ sampled values for which the rank of the bidder goes above r"
- These are $\approx (2m)^{\frac{1}{\epsilon}}$ combinations for each player

Coarsen Space of Mechanisms we Optimize

- Optimize over virtual value functions that takes values on the grid
- On any subset of size m of a set of samples Q of size $2m$, ERM will return a monotone step function that takes these values and changes only on the $2m$ sampled points for each bidder
- These are $\approx (2m)^{\frac{1}{\epsilon}}$ functions for each bidder
- By Masart’s Lemma

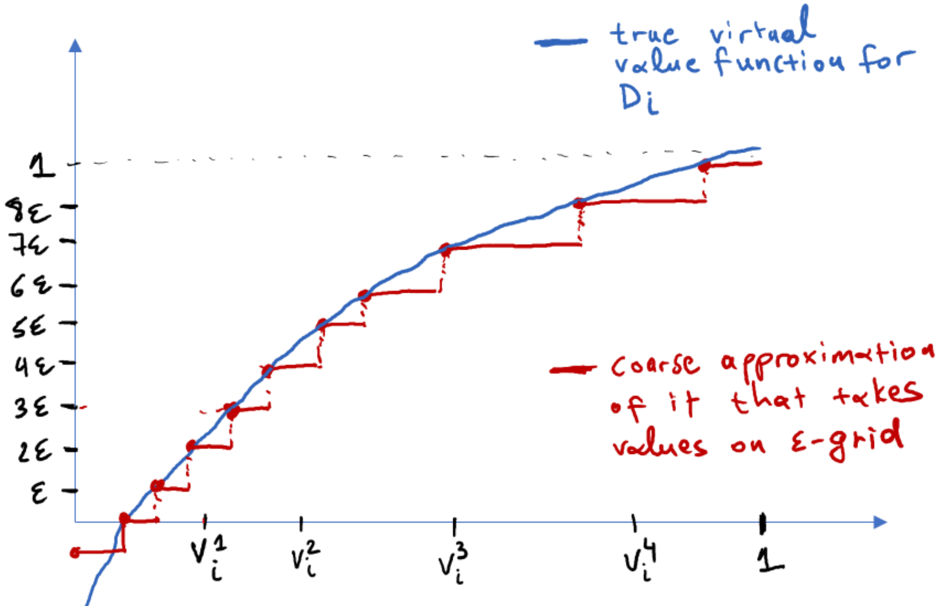
$$\mathrm{Rad}\left(S, H_{S\cup S',\epsilon}\right) \lesssim 2\sqrt{\frac{2n\log(4m)}{\epsilon \cdot m}}$$

Approximation Error

- By optimizing over these ϵ-grid virtual values we don't lose more than $\approx \epsilon$
- If h_* is optimal mechanism and h_ϵ the mechanism where each player's virtual value is rounded down to the closest multiple of ϵ

$$\mathrm{Rev}(h_\epsilon) = E\left[\left[\phi_{i_\epsilon}(v_i)\right]_+\right] \geq E\left[\left[\phi^\epsilon_{i_\epsilon}(v_i)\right]_+\right] = E\left[\max_i\left[\phi^\epsilon_i(v_i)\right]_+\right]$$

$$\geq E\left[\max_i[\phi_i(v_i)]_+ - \epsilon\right] = \mathrm{Rev}(h_*) - \epsilon$$



Putting it all together

- If we output the mechanism h_S that optimizes the empirical revenue among all monotone virtual welfare maximizers, with virtual value functions taking values in an ϵ-grid

$$E_S[\mathrm{Rev}(h_\epsilon)] \geq \mathrm{Rev}(h_*) - \sqrt{\frac{2\ \mathrm{nlog}(2m)}{\epsilon \cdot m}} - \epsilon$$

- For $\epsilon = \left(\frac{2n\log(2m)}{m}\right)^{\frac{1}{3}}$

$$E_S[\mathrm{Rev}(h_\epsilon)] \geq \mathrm{Rev}(h_*) - 2\left(\frac{2n\log(2m)}{m}\right)^{\frac{1}{3}}$$

Zooming out

Setting	Lower Bound	Upper Bound
Regular	$\Omega(n\epsilon^{-3})$	$\tilde{O}(n\epsilon^{-3})$
MHR	$\tilde{\Omega}(n\epsilon^{-2})$	$\tilde{O}(n\epsilon^{-2})$
$[1, H]$	$\Omega(nH\epsilon^{-2})$	$\tilde{O}(nH\epsilon^{-2})$
$[0, 1]$-additive	$\Omega(n\epsilon^{-2})$	$\tilde{O}(n\epsilon^{-2})$

Table II. Sample complexity bounds in [Guo et al. 2019]

- Study initiated by [Cole, Roughgarden, 2014]
- Approaches either
  - Discretize the value space and use revenue monotonicity arguments
  - Discretize the virtual value space and use statistical learning theory arguments
  - Try to learn the virtual value functions and the CDF functions and use bounds on learning CDFs (DKW inequality and more elaborate inequalities that better control errors in the extreme quantiles)

- Statistical learning theory approaches side-step the computational question
  - Several papers on computionally efficient algorithms [Devanur et al, 2016], [Gonczarowski and Nisan 2017], [Roughgarden and Schrijvers 2016], [Guo et al, 19]

Multi-Item Auctions

- n bidders and m items and additive valuations
- A very active research area
- One well-explored direction: Sample complexity of simple mechanisms
  - Characterize simple mechanisms with constant-factor approximate revenue [Chawla et al. (2007, 2010, 2015); Hart and Nisan (2012); Babaioff et al. (2014); Rubinstein and Weinberg (2015); Yao (2015); Cai et al. (2016); Chawla and Miller (2016); Cai and Zhao (2017)]
  - Show that learning such simple mechanisms can be done with polynomial sample complexity [Morgenstern and Roughgarden (2016); Balcan et al. (2016, 2018); Cai and Daskalakis (2017); Syrgkanis (2017)]
- Optimal auctions do not have a simple characterization as virtual welfare maximizers
- One key property (revenue monotonicity) used in single-dimensional does not hold
- [Gonczarowski, Weinberg,'18] Prove a sample complexity result in this general setting with independent valuations, even without the need to use some characterization of how the optimal auction looks like