Greedy Approximation Algorithms for Active Sequential Hypothesis Testing

Kyra Gan	Su Jia	Andrew Li
Harvard	Cornell	CMU

Example: Liquid Biopsy

- Successful cancer treatment requires early detection
- Early cancer detection remains an open problem

  to other organs.

  - Among the 15% of ovarian cancer cases diagnosed early when cancer is confined to the ovary, over 90% survive five years. Unfortunately, for the nearly two-thirds diagnosed after it has spread, only 28% survive that long. [1]

- Liquid biopsy: a simple blood test to detect early-stage cancer
  - Accurate and minimally invasive

How Liquid Biopsies Work

- Cancer is caused by mutations
  - Tumors contain mutated DNA
- *Cell-free DNA (cfDNA)*: DNA freely circulating in the blood
  - Blood contains tiny amounts of mutated DNA
  - Should be a cancer signal
- Challenge: cost
  - cfDNA Concentration of 10^{-4}
    - Sequencing an address costs $\$10^{-2}$
  - Test must cost at most $\$10^{2}$
  - Thus, can only pre-select a panel ~10^4 addresses

Active Sequential Hypothesis Testing (ASHT)

- H: a set of hypotheses
  - Unknown true hypothesis $h^\star \sim$ known prior π
- A: a set of actions with noisy outcomes & can be adaptively chosen
- Set of outcome distributions for each action-hypothesis pair

0.09	0.01	0.04	0.34	0.05	0.56	0.24	0.10	0.07	0.14	0.01	0.24	0.05	0.25	0.01	0.30	0.02	0.03	0.18	0.12
0.26	0.17	0.68	0.16	0.25	0.01	0.06	0.25	0.05	0.12	0.23	0.49	0.26	0.01	0.23	0.15	0.01	0.05	0.25	0.15
0.01	0.09	0.47	0.69	0.06	0.66	0.41	0.12	0.15	0.02	0.32	0.26	0.15	0.16	0.36	0.34	0.26	0.01	0.12	0.08
0.01	0.28	0.03	0.46	0.01	0.26	0.19	0.44	0.09	0.01	0.31	0.31	0.22	0.04	0.04	0.01	0.15	0.15	0.44	0.21
0.18	0.03	0.11	0.14	0.07	0.3	0.17	0.05	0.02	0.05	0.26	0.1	0.03	0.45	0.26	0.43	0.07	0.06	0.16	0.12
0.29	0.18	0.2	0.08	0.2	0.69	0.04	0.30	0.06	0.2	0.06	0.15	0.19	0.23	0.35	0.36	0.32	0.09	0.14	0.10
0.01	0.05	0.01	0.02	0.09	0.09	0.21	0.09	0.03	0.09	0.3	0.46	0.19	0.07	0.35	0.06	0.15	0.02	0.24	0.06
0.2	0.45	0.09	0.08	0.26	0.15	0.01	0.17	0.02	0.05	0.02	0.02	0.45	0.12	0.4	0.36	0.01	0.06	0.08	0.01

Active Sequential Hypothesis Testing (ASHT)

- H: a set of hypotheses
  - Unknown true hypothesis $h^\star \sim$ known prior π
- A: a set of actions with noisy outcomes & can be adaptively chosen
- Set of outcome distributions for each action-hypothesis pair

0.09	0.01	0.04	0.34	0.05	0.56	0.24	0.10	0.07	0.14	0.01	0.24	0.05	0.25	0.01	0.30	0.02	0.03	0.18	0.12
0.26	0.17	0.68	0.16	0.25	0.01	0.06	0.25	0.05	0.12	0.23	0.49	0.26	0.01	0.23	0.15	0.01	0.05	0.25	0.15
0.01	0.09	0.47	0.69	0.06	0.66	0.41	0.12	0.15	0.02	0.32	0.26	0.15	0.16	0.36	0.34	0.26	0.01	0.12	0.08
0.01	0.28	0.03	0.46	0.01	0.26	0.19	0.44	0.09	0.01	0.31	0.31	0.22	0.04	0.04	0.01	0.15	0.15	0.44	0.21
0.18	0.03	0.11	0.14	0.07	0.3	0.17	0.05	0.02	0.05	0.26	0.1	0.03	0.45	0.26	0.43	0.07	0.06	0.16	0.12
0.29	0.18	0.2	0.08	0.2	0.69	0.04	0.30	0.06	0.2	0.06	0.15	0.19	0.23	0.35	0.36	0.32	0.09	0.14	0.10
0.01	0.05	0.01	0.02	0.09	0.09	0.21	0.09	0.03	0.09	0.3	0.46	0.19	0.07	0.35	0.06	0.15	0.02	0.24	0.06
0.2	0.45	0.09	0.08	0.26	0.15	0.01	0.17	0.02	0.05	0.02	0.02	0.45	0.12	0.4	0.36	0.01	0.06	0.08	0.01

Active Sequential Hypothesis Testing (ASHT)

- H: a set of hypotheses
  - Unknown true hypothesis $h^\star \sim$ known prior π
- A: a set of actions with noisy outcomes & can be adaptively chosen
- Set of outcome distributions for each action-hypothesis pair
- ASHT instance: (π, table)

0.09	0.01	0.04	0.34	0.05	0.56	0.24	0.10	0.07	0.14	0.01	0.24	0.05	0.25	0.01	0.30	0.02	0.03	0.18	0.12
0.26	0.17	0.68	0.16	0.25	0.01	0.06	0.25	0.05	0.12	0.23	0.49	0.26	0.01	0.23	0.15	0.01	0.05	0.25	0.15
0.01	0.09	0.47	0.69	0.06	0.66	0.41	0.12	0.15	0.02	0.32	0.26	0.15	0.16	0.36	0.34	0.26	0.01	0.12	0.08
0.01	0.28	0.03	0.46	0.01	0.26	0.19	0.44	0.09	0.01	0.31	0.31	0.22	0.04	0.04	0.01	0.15	0.15	0.44	0.21
0.18	0.03	0.11	0.14	0.07	0.3	0.17	0.05	0.02	0.05	0.26	0.1	0.03	0.45	0.26	0.43	0.07	0.06	0.16	0.12
0.29	0.18	0.2	0.08	0.2	0.69	0.04	0.30	0.06	0.2	0.06	0.15	0.19	0.23	0.35	0.36	0.32	0.09	0.14	0.10
0.01	0.05	0.01	0.02	0.09	0.09	0.21	0.09	0.03	0.09	0.3	0.46	0.19	0.07	0.35	0.06	0.15	0.02	0.24	0.06
0.2	0.45	0.09	0.08	0.26	0.15	0.01	0.17	0.02	0.05	0.02	0.02	0.45	0.12	0.4	0.36	0.01	0.06	0.08	0.01

Active Sequential Hypothesis Testing (ASHT)

- $\delta \in (0,1)$: probability of identifying a wrong hypothesis
- **Goal**: identify $h^\star$ with probability $1 - \delta$ at minimum expected cost
  - No efficient exact solutions: NP-hard
    - Approximation algorithms
- **Two** types of adaptivity:
  - **Partially adaptive**: fix action sequence, adaptively choosing stopping time
    - Submodular function ranking
  - **Fully adaptive**: choosing action based on previously observed outcomes
    - Optimal decision trees

Summary of Results

Partially Adaptive	Brute Force	LP Heuristic	Our Algorithm
Runtime	$\Omega(\lvert A \rvert !)$	Solving an LP one time with size $\Omega(\lvert A \rvert \lvert H \rvert^2)$	$O(\lvert A \rvert \lvert H \rvert)$ per iteration
Approximation Guarantee	1	—	$O(\log \lvert H \rvert)$

Fully Adaptive	Brute Force	Naghshvar and Javidi (2013)	Our Algorithm
Runtime	$\Omega\left(\lvert A \rvert^{2^{\lvert H \rvert}}\right)$	Solving an LP per iteration with size $\Omega(\lvert A \rvert \lvert H \rvert)$	$O(\lvert A \rvert \lvert H \rvert)$ per iteration
Approximation Guarantee	1	—	$O(\log^2 \lvert H \rvert)$

Algorithmic Intuition: Deterministic Instance

Perform

Observed outcome

Inconsistent outcome

Rule out

	a_1	a_2	a_3	a_4	a_5	…	…	…	…	…	…	…	…	…	…	…	…	…	…	a_n
h_1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0
h_2	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0
h_3	0	0	1	1	0	1	1	0	0	0	1	0	0	0	1	1	1	0	0	1
h_4	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0
h_5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
h_6	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0
h_7	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0
h_8	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0

- Suppose h_1 (Row 1) is the true hypothesis
- h_1 is identified if all other hypotheses are "ruled out"
- Example: perform a_1 (Column 1)
  - Observe outcome 0
  - Rule out h_6

Algorithmic Intuition: Deterministic Instance

	a_1	a_2	a_3	a_4	a_5	…	…	…	…	…	…	…	…	…	…	…	…	…	…	a_n	
Perform →	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	h_1
Rule out →	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	h_2
	0	0	1	1	0	1	1	0	0	0	1	0	0	0	1	1	1	0	0	1	h_3
Rule out →	0	1	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	h_4
Rule out →	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	h_5
Rule out →	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	h_6
Rule out →	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	h_7
Rule out →	0	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	h_8

Algorithmic Intuition: Stochastic Instance

a_1	a_2	a_3	a_4	a_5	…	…	…	…	…	…	…	…	…	…	…	…	…	…	a_n	
0.09	0.01	0.04	0.34	0.05	0.56	0.24	0.10	0.07	0.14	0.01	0.24	0.05	0.25	0.01	0.30	0.02	0.03	0.18	0.12	h_1
0.26	0.17	0.68	0.16	0.25	0.01	0.06	0.25	0.05	0.12	0.23	0.49	0.26	0.01	0.23	0.15	0.01	0.05	0.25	0.15	h_2
0.01	0.09	0.47	0.69	0.06	0.66	0.41	0.12	0.15	0.02	0.32	0.26	0.15	0.16	0.36	0.34	0.26	0.01	0.12	0.08	h_3
0.01	0.28	0.03	0.46	0.01	0.26	0.19	0.44	0.09	0.01	0.31	0.31	0.22	0.04	0.04	0.01	0.15	0.15	0.44	0.21	h_4
0.18	0.03	0.11	0.14	0.07	0.3	0.17	0.05	0.02	0.05	0.26	0.1	0.03	0.45	0.26	0.43	0.07	0.06	0.16	0.12	h_5
0.29	0.18	0.2	0.08	0.2	0.69	0.04	0.30	0.06	0.2	0.06	0.15	0.19	0.23	0.35	0.36	0.32	0.09	0.14	0.10	h_6
0.01	0.01	0.01	0.02	0.09	0.09	0.21	0.09	0.03	0.09	0.3	0.46	0.19	0.07	0.35	0.06	0.15	0.02	0.24	0.06	h_7
0.2	0.45	0.09	0.08	0.26	0.15	0.01	0.17	0.02	0.05	0.02	0.02	0.45	0.12	0.4	0.36	0.01	0.06	0.08	0.01	h_8

- Suppose h_1 (Row 1) is the true hypothesis
- Idea 1: "rule out" h_2 if the log likelihood ratio (LLR) reaches some threshold
  - o_i: outcome at time i when action a_i is performed
  - $\log \Lambda(h_1, h_2) = \sum_i \log \frac{\mathbb{P}(o_i \mid h_1, a_i)}{\mathbb{P}(o_i \mid h_2, a_i)}$ ← Random

    Idea 2: take expectation

Algorithmic Intuition: Stochastic Instance

a_1	a_2	a_3	a_4	a_5	…	…	…	…	…	…	…	…	…	…	…	…	…	…	a_n	
0.09	0.01	0.04	0.34	0.05	0.56	0.24	0.10	0.07	0.14	0.01	0.24	0.05	0.25	0.01	0.30	0.02	0.03	0.18	0.12	h_1
0.26	0.17	0.68	0.16	0.25	0.01	0.06	0.25	0.05	0.12	0.23	0.49	0.26	0.01	0.23	0.15	0.01	0.05	0.25	0.15	h_2
0.01	0.09	0.47	0.69	0.06	0.66	0.41	0.12	0.15	0.02	0.32	0.26	0.15	0.16	0.36	0.34	0.26	0.01	0.12	0.08	h_3
0.01	0.28	0.03	0.46	0.01	0.26	0.19	0.44	0.09	0.01	0.31	0.31	0.22	0.04	0.04	0.01	0.15	0.15	0.44	0.21	h_4
0.18	0.03	0.11	0.14	0.07	0.3	0.17	0.05	0.02	0.05	0.26	0.1	0.03	0.45	0.26	0.43	0.07	0.06	0.16	0.12	h_5
0.29	0.18	0.2	0.08	0.2	0.69	0.04	0.30	0.06	0.2	0.06	0.15	0.19	0.23	0.35	0.36	0.32	0.09	0.14	0.10	h_6
0.01	0.01	0.01	0.02	0.09	0.09	0.21	0.09	0.03	0.09	0.3	0.46	0.19	0.07	0.35	0.06	0.15	0.02	0.24	0.06	h_7
0.2	0.45	0.09	0.08	0.26	0.15	0.01	0.17	0.02	0.05	0.02	0.02	0.45	0.12	0.4	0.36	0.01	0.06	0.08	0.01	h_8

- Suppose h_1 (Row 1) is the true hypothesis
- Idea 2: "rule out" h_2 if the expected LLR reaches some threshold
    - o_i: outcome at time i when action a_i is performed
    - $\mathbb{E}\left[\log \Lambda(h_1, h_2)\right] = \sum_i \mathrm{KL}\left(\mathrm{Ber}(h_1, a_i), \mathrm{Ber}(h_2, a_i)\right) = \sum_i \mathrm{KL}(h_1, h_2; a_i)$

Partially Adaptive Algorithm: Overview

- Input: an ASHT instance (π, table), error tolerance δ
- Algorithm outline:
  - Construct a sequence σ of actions greedily
  - Boost: repeat each action in σ for sufficiently many times, obtain $\tilde{\sigma}$
  - Perform actions in $\tilde{\sigma}$ one by one, and terminate when only one hypothesis is left

Tool: Submodular Function Ranking

Identify/"Covering" a hypothesis

- Consider a true hypothesis h
- S: a set of actions
- B: target threshold for "collecting" KL-divergence ⟵ $B \sim \log \delta^{-1}$
- The progress made towards ruling out g:

$$K_{h,g}(S) = \frac{1}{B} \min\left(B, \sum_{a \in S} \mathrm{KL}(h, g; a)\right)$$

⟵ $0 \leq K_{h,g}(S) \leq 1$

- Total progress towards identifying h:

$$f_h(S) = \frac{1}{|H| - 1} \sum_{g \in H \setminus \{h\}} K_{h,g}(S)$$

⟵ $0 \leq f_h(S) \leq 1$

  - Fact 1: $f_h(S) = 1 \iff K_{h,g}(S) = 1 \, \forall g \neq h \iff h$ is identified
  - Fact 2: submodular

Partially Adaptive Algorithm: Construct Action Sequence

- S: set of actions performed so far
- Choose the action a, with the highest score:

$$\text{Score}(a,S) \leftarrow \sum_{h:f_h(S)<1} \pi(h) \frac{f_h(S \cup \{a\}) - f_h(S)}{1 - f_h(S)}$$

Incremental coverage with one additional action

Give priority to the hypothesis that is about to be covered

Sum over all "uncovered" hypotheses

- Output a sequence σ of actions
- **Boosting**: replace each action in σ by $\sim \log |H|$ copies, and obtain $\tilde{\sigma}$
- Perform actions in $\tilde{\sigma}$
- Stop when the log likelihood ratios between an h and all $g \in H \backslash h$ is above a certain threshold

Partially Adaptive Algorithm: Theoretical Guarantee

- δ**-error** algorithm:
  - Identifies any true hypothesis with probability $1 - \delta$

Theorem 1 [Partially Adaptive]: For any $\delta \in (0,1/2)$, there exists a polynomial-time *partially adaptive* algorithm that achieves δ-error with expected cost at most

$$O((1 + \log_{1/\delta}|H|)\log(|H|\log\delta^{-1}))$$

times the minimum cost among all partially adaptive algorithms that achieve δ-error.

- Special case: $\delta = |H|^{-c}$
  - Approximation ratio: $O(\log|H| + c\log\log|H|)$

Our algorithm is almost tight with respect to $|H|$

Approximation ratio lower bound for ASHT

Fully Adaptive Algorithm

- Input: an ASHT instance (π, table), error tolerance δ
- Algorithm ideas:
  - **Alive set H_{alive}**: hypotheses consistent with previous outcomes
  - **Removing the noise**: repeat each action sufficiently → "meta action"
  - **Greedy**: choose the "meta action" that rules out the most alive hypotheses in the worst case
  - Rule out hypotheses whose outcome is inconsistent with the empirical means
  - Terminate when only one hypothesis remains alive

Fully Adaptive Algorithm: Theoretical Guarantee

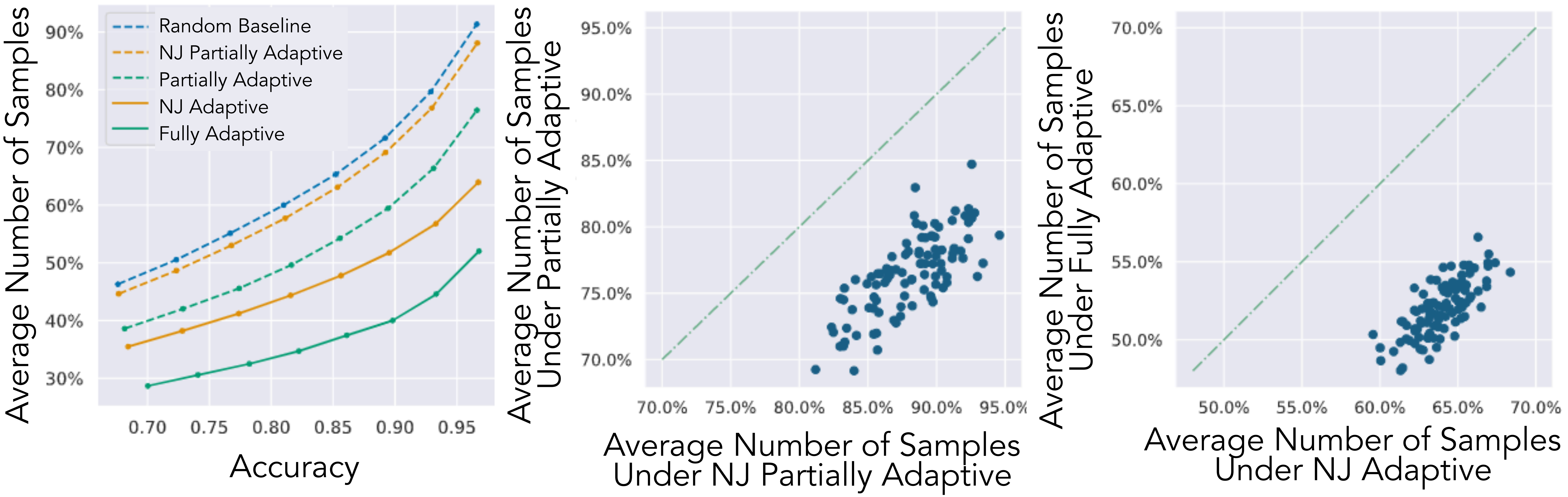
Theorem 2 [Fully Adaptive]: For any $\delta \in (0,1/2)$, there exists a polynomial-time *adaptive* algorithm that identifies any true hypothesis with probability $1-\delta$ with expected cost at most

$$O(\log(|H|\delta^{-1})\log|H|)$$

times the minimum cost among all fully adaptive algorithms that achieve δ-error.

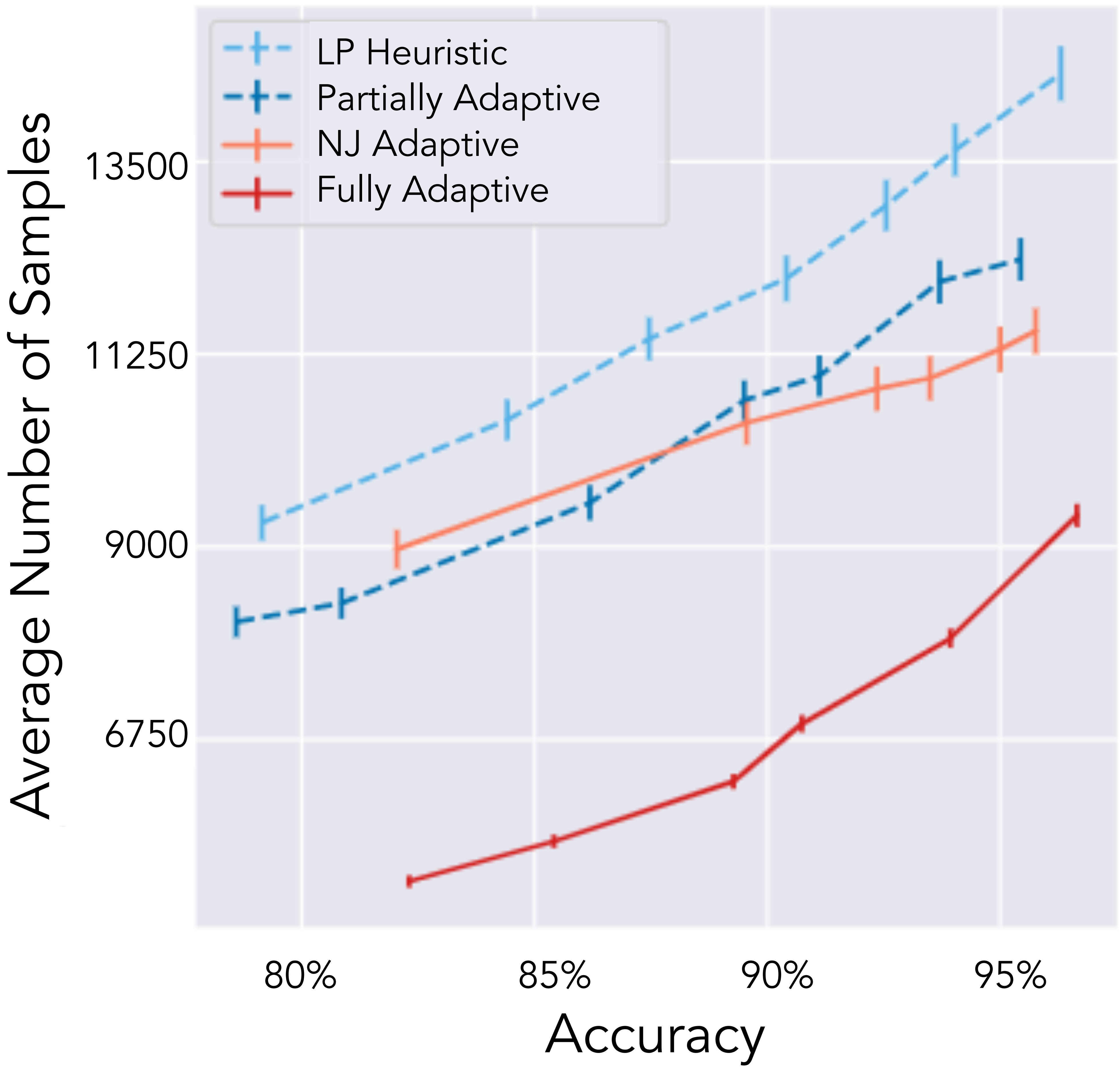
- Special case: $\delta = |H|^{-c}$
  - Approximation ratio: $O(c\log^2|H|)$

Numerical Results on Synthetic Data



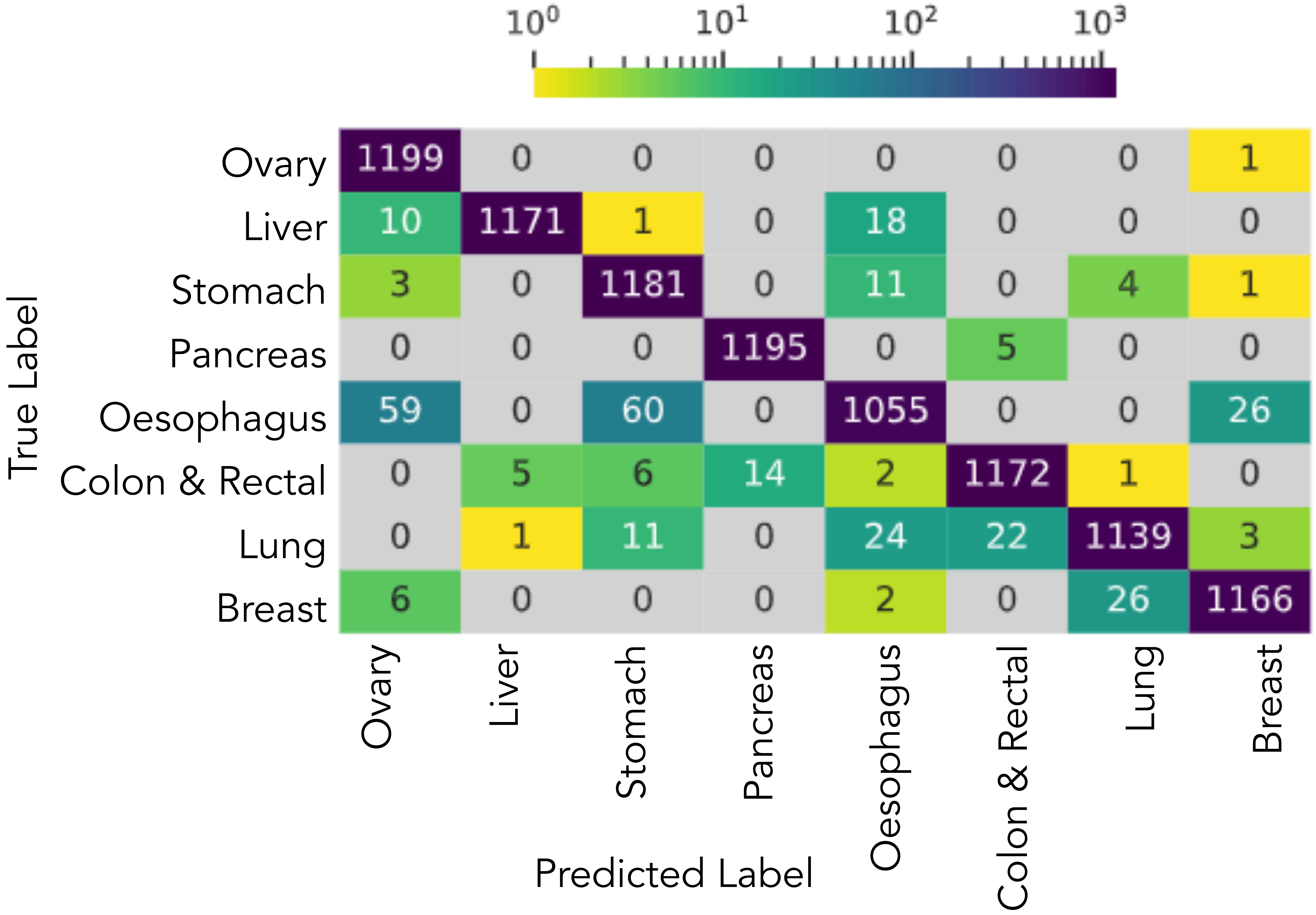
- Number of actions: 40; number of hypotheses: 25
- The average number of samples normalized with respect to Random Baseline
- Number of instances: 100; number of replications: 2,000
- Middle and right: accuracy equals to 0.97

Numerical Results on COSMIC



- Each point: 9,600 replications

Numerical Results on COSMIC



Sensitivity	Specificity
0.9992	0.9907
0.9758	0.9993
0.9842	0.9907
0.9958	0.9983
0.8792	0.9932
0.9767	0.9968
0.9492	0.9963
0.9717	0.9963