Sample Complexity of Robust Hypothesis Testing

Shankar Vallinayagam Alston Xu

Summer Research Festival

Supervised by Professor Varun Jog

Simple hypothesis testing

Simple hypothesis testing

- Let p and q be known distributions over $\{1, 2, \ldots, k\}$

Simple hypothesis testing

- Let p and q be known distributions over $\{1, 2, \ldots, k\}$
- Observe n i.i.d. samples from one of them, equally likely

Simple hypothesis testing

- Let p and q be known distributions over $\{1, 2, \ldots, k\}$
- Observe n i.i.d. samples from one of them, equally likely
- Goal: Decide whether p or q using samples

Simple hypothesis testing

- Let p and q be known distributions over $\{1, 2, \ldots, k\}$
- Observe n i.i.d. samples from one of them, equally likely
- Goal: Decide whether p or q using samples
- Sample complexity n^*: Smallest n such that $\mathbb{P}(\text{error}) \leq 0.1$ for optimal test

Characterising n^* in terms of TV and Hellinger distance

Characterising n^* in terms of TV and Hellinger distance

- TV-distance defined as

$$d_{TV}(p,q) = \sup_S p(S) - q(S) = \frac{1}{2}\sum |p(i) - q(i)|$$

Characterising n^* in terms of TV and Hellinger distance

- TV-distance defined as

$$d_{TV}(p, q) = \sup_S p(S) - q(S) = \frac{1}{2} \sum |p(i) - q(i)|$$

Sample complexity using d_{TV}

$$\frac{1}{d_{TV}(p, q)} \lesssim n^* \lesssim \frac{1}{d_{TV}(p, q)^2}$$

Characterising n^* in terms of TV and Hellinger distance

- TV-distance defined as

$$d_{TV}(p,q) = \sup_S p(S) - q(S) = \frac{1}{2}\sum |p(i) - q(i)|$$

Sample complexity using d_{TV}

$$\frac{1}{d_{TV}(p,q)} \lesssim n^* \lesssim \frac{1}{d_{TV}(p,q)^2}$$

- Hellinger distance defined as

$$d_h^2(p,q) = \sum_x (\sqrt{p}(x) - \sqrt{q}(x))^2 = 2\left(1 - \langle \sqrt{p}, \sqrt{q} \rangle\right)$$

Characterising n^* in terms of TV and Hellinger distance

- TV-distance defined as

$$d_{TV}(p,q) = \sup_S p(S) - q(S) = \frac{1}{2}\sum |p(i) - q(i)|$$

Sample complexity using d_{TV}

$$\frac{1}{d_{TV}(p,q)} \lesssim n^* \lesssim \frac{1}{d_{TV}(p,q)^2}$$

- Hellinger distance defined as

$$d_h^2(p,q) = \sum_x (\sqrt{p}(x) - \sqrt{q}(x))^2 = 2\left(1 - \langle \sqrt{p}, \sqrt{q} \rangle\right)$$

Sample complexity is characterized by Hellinger

$$n^* = \Theta\left(\frac{1}{d_h^2(p,q)}\right)$$

Robust statistics

Robust statistics

- Robust statistical procedures are those that demonstrate insensitivity to small deviations from the assumptions made.

Robust statistics

- Robust statistical procedures are those that demonstrate insensitivity to small deviations from the assumptions made.
- The optimal solution to this hypothesis testing problem in the non-robust setting is the likelihood ratio test:

$$\prod_{i=1}^{k} \frac{p(i)}{q(i)}$$

Robust statistics

- Robust statistical procedures are those that demonstrate insensitivity to small deviations from the assumptions made.
- The optimal solution to this hypothesis testing problem in the non-robust setting is the likelihood ratio test:

$$\prod_{i=1}^{k} \frac{p(i)}{q(i)}$$

- Our adversaries are
  - (Oblivious) Huber contamination
  - TV-contamination
  - Adaptive Huber contamination

Adversaries

Adversaries

Huber Adversary

$$B_{\text{Huber}}(p, \epsilon) = \{p^* : p^* = (1 - \epsilon)p + \epsilon h\}$$

Given the underlying distribution is p, adversary selects distribution in this ball for data to be drawn from

Adversaries

Huber Adversary

$$B_{\text{Huber}}(p,\epsilon) = \{p^* : p^* = (1-\epsilon)p + \epsilon h\}$$

Given the underlying distribution is p, adversary selects distribution in this ball for data to be drawn from

TV Adversary

$$B_{\text{TV}}(p,\epsilon) = \{p^* : d_{TV}(p,p^*) \le \epsilon\}$$

Given the underlying distribution is p, adversary selects distribution in this ball for data to be drawn from

Adversaries

Huber Adversary

$$B_{\text{Huber}}(p, \epsilon) = \{p^* : p^* = (1 - \epsilon)p + \epsilon h\}$$

Given the underlying distribution is p, adversary selects distribution in this ball for data to be drawn from

TV Adversary

$$B_{\text{TV}}(p, \epsilon) = \{p^* : d_{TV}(p, p^*) \leq \epsilon\}$$

Given the underlying distribution is p, adversary selects distribution in this ball for data to be drawn from

Adaptive Huber adversary

Take a sample of size n, then adversary selects ϵn of these samples and changes them

Optimal test for oblivious adversaries

Optimal test for oblivious adversaries

- We want to find a test that minimises

$$\max_{p^* \in B(p,\epsilon), q^* \in B(q,\epsilon)} n^*(p^*, q^*)$$

Optimal test for oblivious adversaries

- We want to find a test that minimises

$$\max_{p^* \in B(p,\epsilon), q^* \in B(q,\epsilon)} n^*(p^*, q^*)$$

- The way Huber, 1964 does this is by finding Least Favourable Distributions that maximise error for all LRTs

Optimal test for oblivious adversaries

- We want to find a test that minimises

$$\max_{p^* \in B(p,\epsilon), q^* \in B(q,\epsilon)} n^*(p^*, q^*)$$

- The way Huber, 1964 does this is by finding Least Favourable Distributions that maximise error for all LRTs
- In both the oblivious cases this leads to a Censored Likelihood Ratio Test; we set thresholds c', c'' such that for any observation if the likelihood ratio is above c'' or below c' we censor down to one of these values

Optimal test for oblivious adversaries

- We want to find a test that minimises

$$\max_{p^* \in B(p,\epsilon), q^* \in B(q,\epsilon)} n^*(p^*, q^*)$$

- The way Huber, 1964 does this is by finding Least Favourable Distributions that maximise error for all LRTs
- In both the oblivious cases this leads to a Censored Likelihood Ratio Test; we set thresholds c', c'' such that for any observation if the likelihood ratio is above c'' or below c' we censor down to one of these values
- The problem reduces to understanding the Hellinger distance between the LFDs

Small ϵ Regime

Small ϵ Regime

- For small epsilon $\epsilon \le \frac{d_H(p,q)^2}{9}$

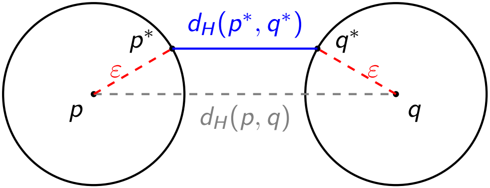
$$n^*_{TV}(\epsilon) \asymp n^*_{Huber}(\epsilon) \asymp n^*(p, q)$$

Small ϵ Regime

- For small epsilon $\epsilon \leq \frac{d_H(p,q)^2}{9}$

$$n^*_{TV}(\epsilon) \asymp n^*_{Huber}(\epsilon) \asymp n^*(p,q)$$

- This follows trivially from the triangle inequality



Conjectures

Conjectures

- We conjectured the sample complexity for Huber and TV stays within constant for more general ϵ: $n^*_{TV}(\epsilon) \asymp n^*_{Huber}(\epsilon) \qquad \forall \epsilon$

Conjectures

- We conjectured the sample complexity for Huber and TV stays within constant for more general ϵ: $n^*_{TV}(\epsilon) \asymp n^*_{Huber}(\epsilon) \qquad \forall \epsilon$
- We also conjectured the sample complexity stays within constants if we scale ϵ to $\epsilon/2$:

$$n^*_{TV}(\epsilon) \asymp n^*_{TV}(\epsilon/2)$$
$$n^*_{Huber}(\epsilon) \asymp n^*_{Huber}(\epsilon/2)$$

Counterexample

Counterexample

- We disproved the above statements by coming up with the following pair of 7-point distribution counterexample.

i	**p**	**q**	$\frac{p_i}{q_i}$
3	$2kt^{1+\delta}$	$kt^{1+\delta}$	2
2	$k(1+t^{1-\delta})t^{2\delta}$	$kt^{2\delta}$	$1+t^{1-\delta}$
1	$14k(1+t^{1+\delta})$	$14k$	$1+t^{1+\delta}$
0	$1-(\cdots)$	$1-(\cdots)$	1
-1	$14k$	$14k(1+t^{1+\delta})$	$\frac{1}{1+t^{1+\delta}}$
-2	$kt^{2\delta}$	$k(1+t^{1-\delta})t^{2\delta}$	$\frac{1}{1+t^{1-\delta}}$
-3	$kt^{1+\delta}$	$2kt^{1+\delta}$	$\frac{1}{2}$

Counterexample

Counterexample

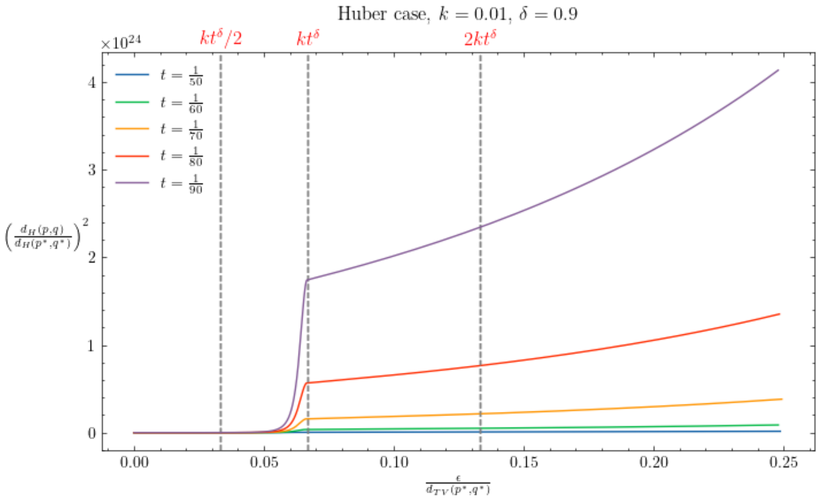
- Key idea: Clipping at $1+t$ and $1+t^{1-\delta}$ and take $t \to 0$

Counterexample

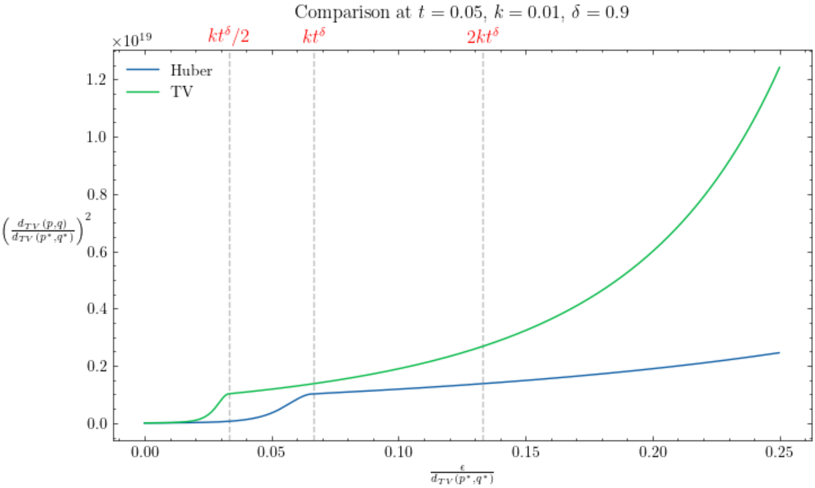
- Key idea: Clipping at $1 + t$ and $1 + t^{1-\delta}$ and take $t \to 0$
- Here, ϵ is up to a negligible error term kt^{δ}

$$
\begin{tikzcd}
 & t^2 & t^{2+2\delta} \\
\text{Huber} & \epsilon \arrow[r] \arrow[dr] & 2\epsilon \\
\text{TV} & \epsilon/2 \arrow[r] & \epsilon
\end{tikzcd}
$$

Huber contamination

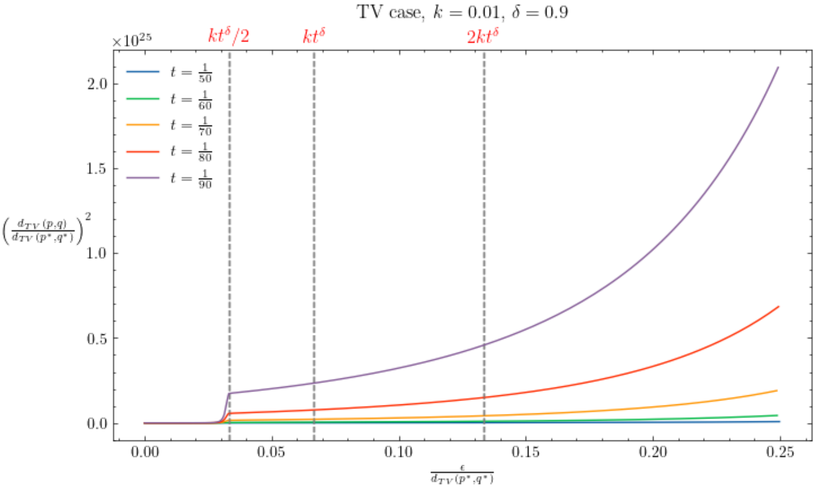


Huber vs TV



Comparison at $t = 0.05$, $k = 0.01$, $\delta = 0.9$
$\times 10^{19}$
$kt^\delta/2$
kt^δ
$2kt^\delta$
Huber
TV
1.2
1.0
0.8
0.6
0.4
0.2
0.0
$\left(\frac{d_{TV}(p,q)}{d_{TV}(p^*,q^*)}\right)^2$
0.00
0.05
0.10
0.15
0.20
0.25
$\frac{\epsilon}{d_{TV}(p^*,q^*)}$

TV contamination



Additional/Ongoing results

Additional/Ongoing results

- Motivated by our counter example, we are working on proving:

$$n^*_{TV}(\epsilon/2) \lesssim n^*_{Huber}(\epsilon)$$

Additional/Ongoing results

- Motivated by our counter example, we are working on proving:

$$n^*_{TV}(\epsilon/2) \lesssim n^*_{Huber}(\epsilon)$$

- We are also working on the following result:

$$n^*_{TV}(\epsilon/2) \lesssim n^*_{Adv}(\epsilon) \lesssim n^*_{TV}(\epsilon)$$

Thank you!