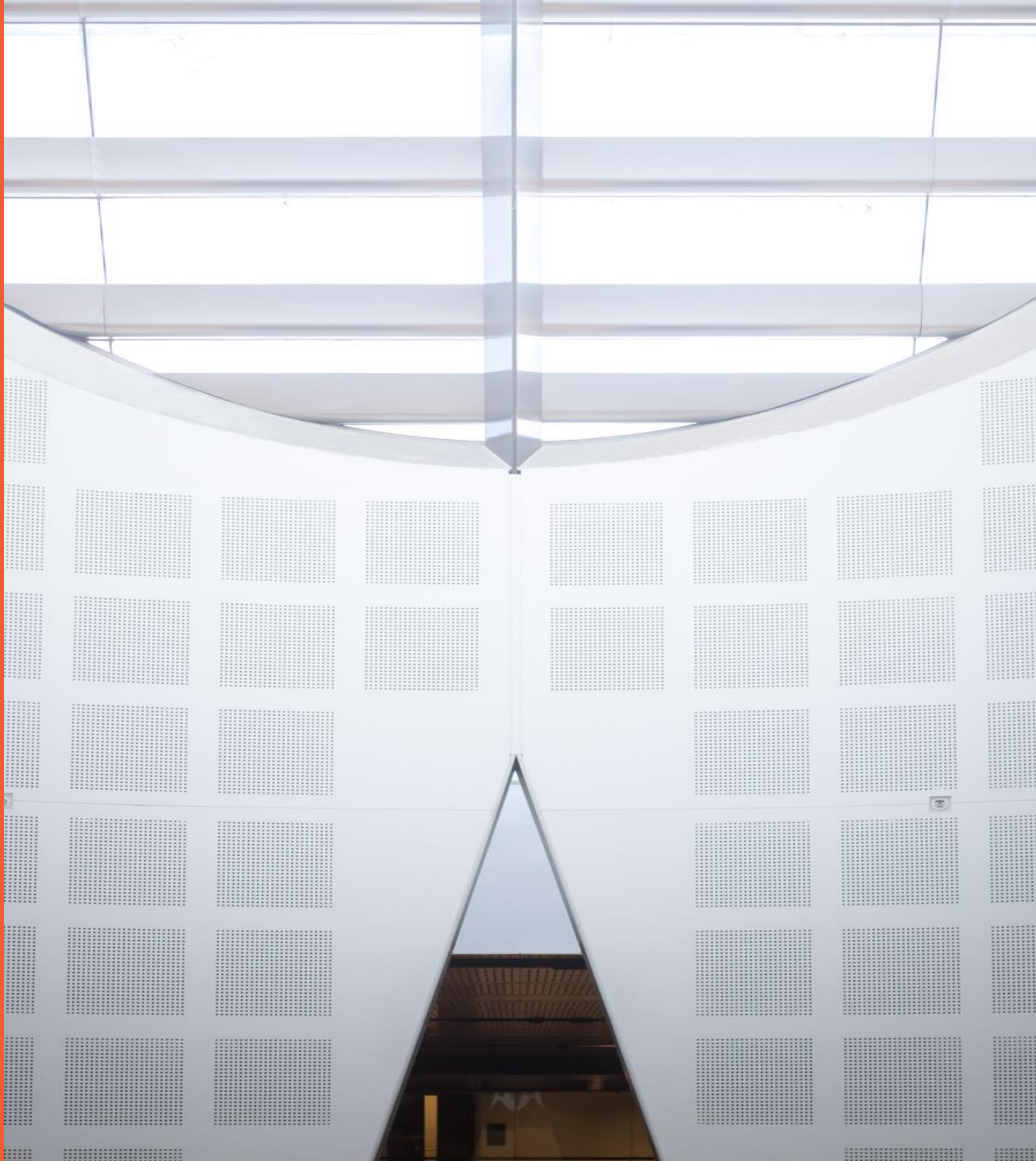
COMPx270: Randomised and Advanced Algorithms

Lecture 11: Learning and testing probability distributions

Clément Canonne

School of Computer Science

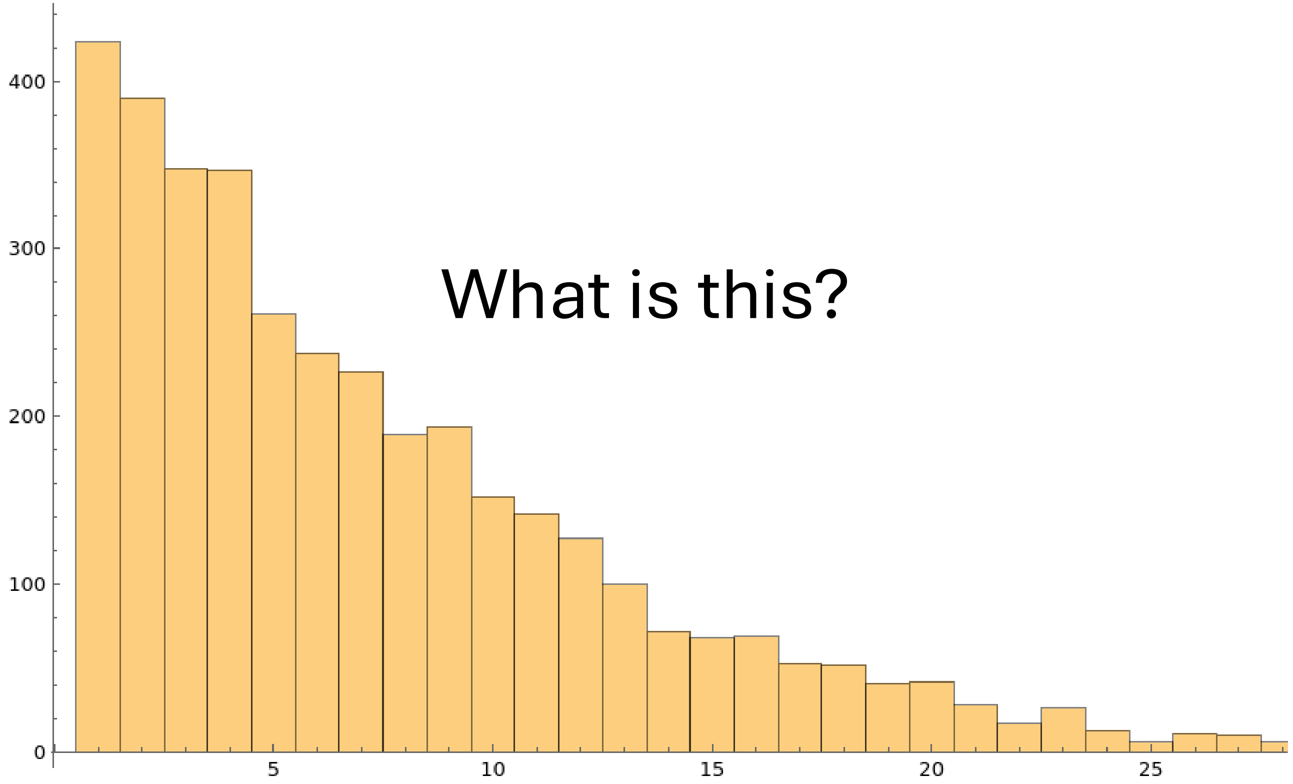
THE UNIVERSITY OF SYDNEY



Some housekeeping

- A2 still being marked: deepest apologies (my fault)
- A3 (after Simple Extension) due tomorrow
- Don't forget the "participation" assignment (Oct 18)
- Sample exam is out, will be the topic of Week 13
- Feedback welcome: https://forms.office.com/r/DymMcfn47n
- Final exam on Tues, Nov 12 (9am) → what is allowed?

A question 📊



What is this?

Learning and testing (discrete) probability distributions *

p over $\mathcal{X}$ (discrete, finite domain) of size k

- Learn p ⇝ get $\hat{p}$ (distr over $\mathcal{X}$) st $d(p,\hat{p}) \leq \varepsilon$
  (k numbers)
  $\uparrow$ TBD
- learn a parameter of p (estimation) : a number
- learn a bit (test if p satisfies some property)

"TBD" : $d(p,q)$ = total variation distance
(quantifies how far p, q are)

Access $x_1, \dots, x_n \overset{\text{iid}}{\sim} p$ (algo only sees $x_1, \dots, x_n$)

* $p : \mathcal{X} \to [0,1]$

$\sum_{i \in \mathcal{X}} p(i) = 1$

Preliminaries on probability distributions

TV distance p, q over $\mathcal{X}$

A <u>metric</u> would be nice

Bounded (in $[0,1]$?) also?

Meaning?

$$TV(p,q) = \sup_{S \subseteq \mathcal{X}} \left(p(S) - q(S) \right) = \sup_{S \subseteq \mathcal{X}} \left| p(S) - q(S) \right|$$

$$p(S) = \sum_{i \in S} p(i)$$

$$= \sup_{\substack{\underline{\text{any}}\ \text{algo} \\ A : \mathcal{X} \to \{0,1\}}} \left(\Pr_{x \sim p}[A(x) = 1] - \Pr_{x \sim q}[A(x) = 1] \right)$$

Preliminaries on probability distributions

Fact

$$TV(p,q) = \frac{1}{2}\sum_{i\in\mathcal{X}} |p(i)-q(i)| = \frac{1}{2}\|p-q\|_1$$

Pf

Take $S^* = \{x : p(x) > q(x)\}$

$$TV(p,q) \geq p(S^*) - q(S^*) = \sum_{i\in S^*}(p(i)-q(i)) = \sum_{i\in S^*}|p(i)-q(i)|$$

But
$$\sum_{i\in\mathcal{X}}|p(i)-q(i)| = \sum_{i\in S^*}(p(i)-q(i)) + \underbrace{\sum_{i\notin S^*}(q(i)-p(i))}_{\sum_{i\notin S^*} q(i) - \sum_{i\notin S^*} p(i) = 1-q(S^*) - (1-p(S^*))}$$

$$= 2\sum_{i\in S^*}(p(i)-q(i))$$

$$= 2(p(S^*)-q(S^*))$$

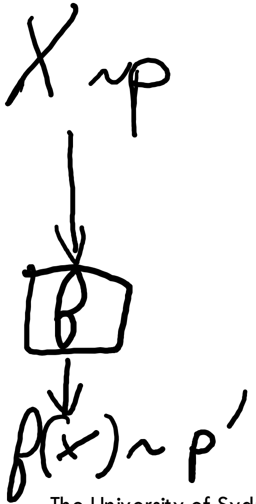
Preliminaries on probability distributions

DPI (Data Processing Inequality)

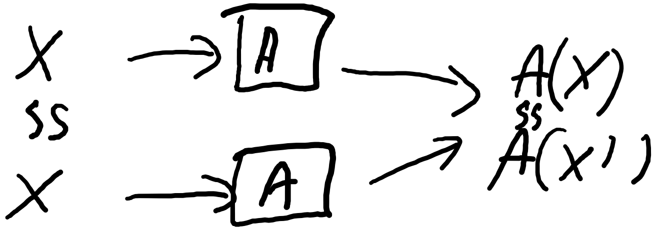
Take any $f: \mathcal{X} \to \mathcal{Y}$

If $X \sim p$ let p' be the distr of $f(X)$

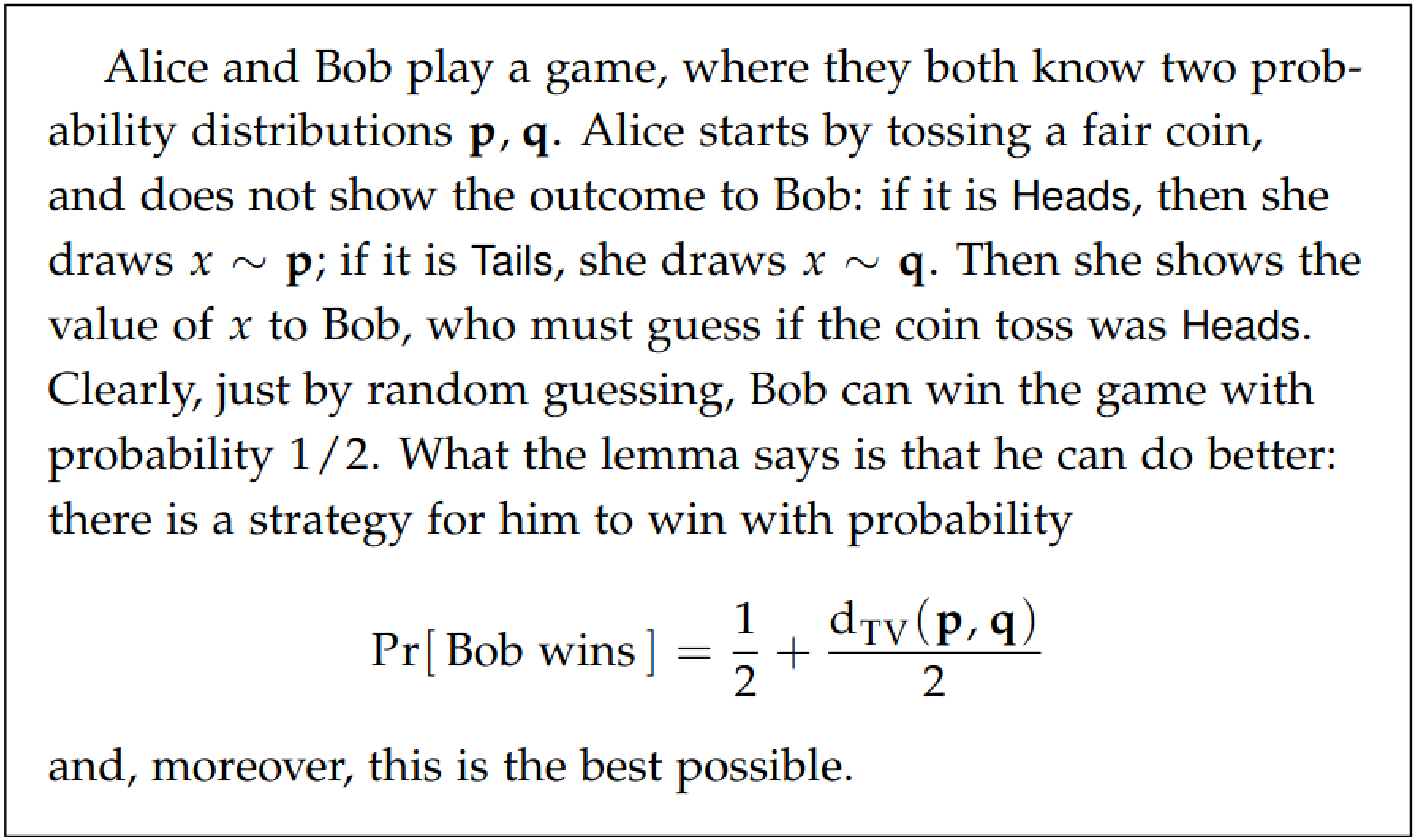
$X' \sim q$ let q' ——— $f(X')$



$$TV(p', q') \leq TV(p, q)$$



A view of TV distance

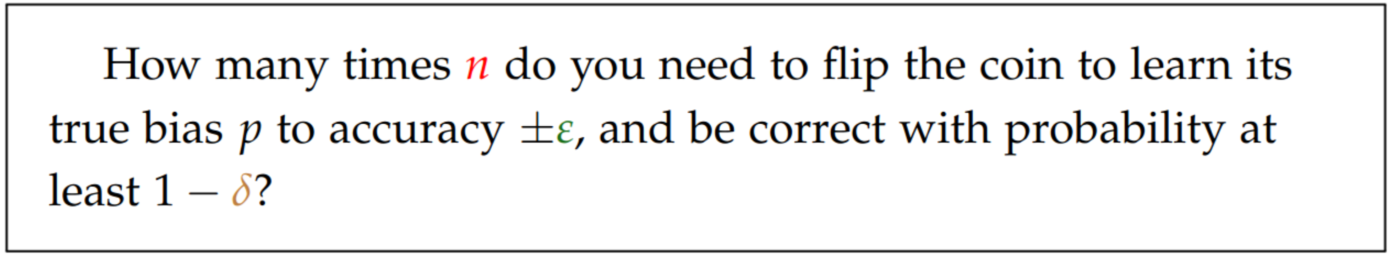


Alice and Bob play a game, where they both know two probability distributions $\mathbf{p}, \mathbf{q}$. Alice starts by tossing a fair coin, and does not show the outcome to Bob: if it is Heads, then she draws $x \sim \mathbf{p}$; if it is Tails, she draws $x \sim \mathbf{q}$. Then she shows the value of x to Bob, who must guess if the coin toss was Heads. Clearly, just by random guessing, Bob can win the game with probability $1/2$. What the lemma says is that he can do better: there is a strategy for him to win with probability

$$\Pr[\,\text{Bob wins}\,] = \frac{1}{2} + \frac{d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})}{2}$$

and, moreover, this is the best possible.

The case of a coin ☐



How many times n do you need to flip the coin to learn its true bias p to accuracy $\pm\varepsilon$, and be correct with probability at least $1 - \delta$?

The case of a coin ☐

Theorem 50. *Suppose we are promised that the true bias p of the coin satisfies* $0 \leq p < q \leq \frac{1}{2}$*, for some known value q. Then estimating the bias of the coin to an additive* ε*, with probability at least* $1 - \delta$*, can be done with* $n = O\left(\frac{q}{\varepsilon^2} \log \frac{1}{\delta}\right)$ *i.i.d. samples. (Moreover, this is optimal.)*

The case of a coin ☐

Theorem 50. *Suppose we are promised that the true bias* p *of the coin satisfies* $0 \leq p < q \leq \frac{1}{2}$*, for some known value* q*. Then estimating the bias of the coin to an additive* ε*, with probability at least* $1 - \delta$*, can be done with* $n = O\left(\frac{q}{\varepsilon^2} \log \frac{1}{\delta}\right)$ *i.i.d. samples. (Moreover, this is optimal.)*

Corollary 50.1. *Estimating the bias of a coin to an additive* ε*, with probability at least* $1 - \delta$*, can be done with* $n = O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta}\right)$ *i.i.d. samples. (Moreover, this is optimal.)*

The case of a coin ☐

Theorem 50. *Suppose we are promised that the true bias p of the coin satisfies $0 \leq p < q \leq \frac{1}{2}$, for some known value q. Then estimating the bias of the coin to an additive ε, with probability at least $1 - \delta$, can be done with $n = O\left(\frac{q}{\varepsilon^2} \log \frac{1}{\delta}\right)$ i.i.d. samples. (Moreover, this is optimal.)*

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Hoeffding ↓ want ↓

$$\Pr[|\hat{p} - p| > \varepsilon] \leq 2e^{-2\varepsilon^2 n} \leq \delta$$

take $n = \left\lceil \frac{1}{2\varepsilon^2} \ln \frac{2}{\delta} \right\rceil$

$$\leq 2e^{-\frac{\varepsilon^2 q^{-1} n}{3}} \leq \delta$$

↑ Chernoff (if $q > \varepsilon$ (if not, easy...)) ↑ want

$n = \left\lceil \frac{3q}{\varepsilon^2} \right\rceil \ln \frac{2}{\delta}$

The case of a coin: what about testing? ☐

Is my coin a fair coin?

The case of a coin: what about testing? ☐

Theorem 51. *Testing whether the bias of a coin is* $1/2$ *or at least* $1/2 + \varepsilon$*, with probability at least* $1 - \delta$*, can be done with* $n = O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\delta}\right)$ *i.i.d. samples. (Moreover, this is optimal.)*

Not better than learning

p.

The case of a coin: what about testing? ☐

Theorem 52. *For any* $0 < \alpha \leq 1/2$ *and* $\varepsilon \in (0,1]$, *testing whether the bias of a coin is at most* α *or at least* $\alpha(1+\varepsilon)$, *with probability at least* $1-\delta$, *can be done with* $n = O\left(\frac{1}{\alpha\varepsilon^2}\log\frac{1}{\delta}\right)$ *i.i.d. samples.*

$\alpha = \frac{1}{2}$ → previous theorem

α small, $\varepsilon = 1$

$\leq \alpha$ vs $\geq 2\alpha$: $O\left(\frac{1}{\alpha}\log\frac{1}{\delta}\right)$

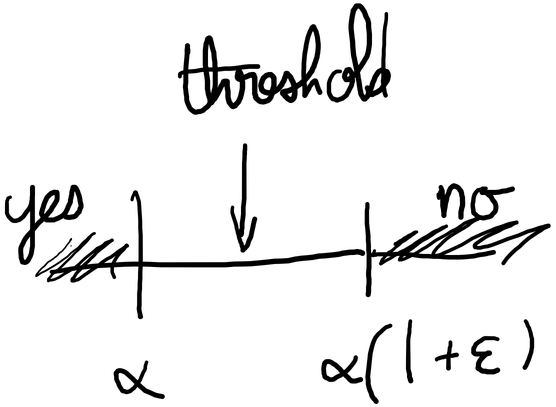
The case of a coin: what about testing? ☐

Theorem 52. *For any* $0 < \alpha \leq 1/2$ *and* $\varepsilon \in (0,1]$*, testing whether the bias of a coin is at most* α *or at least* $\alpha(1+\varepsilon)$*, with probability at least* $1-\delta$*, can be done with* $n = O\left(\frac{1}{\alpha\varepsilon^2}\log\frac{1}{\delta}\right)$ *i.i.d. samples.*

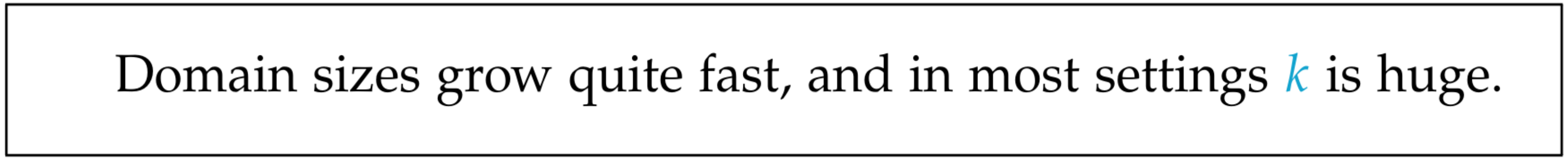
Can prove via Chernoff:

Compute $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Output "yes" if $p \leq \alpha\left(1+\frac{\varepsilon}{2}\right)$



Beyond coins: k is large



Domain sizes grow quite fast, and in most settings k is huge.

Learning in TV distance

Unknown p over $\mathcal{X}$ $(|\mathcal{X}| = k)$

Parameter ε, δ

Get $x_1, \dots, x_n \overset{iid}{\sim} p$ for n to be chosen

Goal: output $\hat{p}$ (distrib. over $\mathcal{X}$) such that

$$\Pr[\,\mathrm{TV}(p, \hat{p}) > \varepsilon\,] \leq \delta$$

Minimise $\underline{n}$ (sample complexity)

$\hookrightarrow n(k, \varepsilon, \delta)$

Learning in TV distance: first attempt

$$p = (p_1, p_2, \dots, p_k)$$

Want: $\hat{p} \in [0,1]^k$ st

★ $\|\hat{p} - p\|_1 \le 2\varepsilon$ (w/p $\ge 1-\delta$)

$$\| \quad \sum_{i=1}^{k} |\hat{p}_i - p_i|$$

(If $\|\hat{p}\|_1 \neq 1$, can normalise)

★ implies

$$\big| \|\hat{p}\|_1 - \underbrace{\|p\|_1}_{=1} \big| \le 2\varepsilon$$

Enough to estimate each p_i by $\hat{p}_i$ to $\pm\frac{2\varepsilon}{k}$ w/p $1 - \frac{\delta}{k}$ each

k instances of "estimate bias p_i of a coin"

Gives us $n = O\left(\frac{1}{(\varepsilon/k)^2} \log \frac{1}{(\delta/k)}\right) = O\left(\frac{k^2}{\varepsilon^2} \log \frac{k}{\delta}\right)$

Learning in TV distance: second attempt

What if instead of $\hat{p}_i \simeq p_i \pm \frac{2\varepsilon}{k}$, we try $\hat{p}_i \simeq (1 \pm 2\varepsilon) p_i$

Then $\|\hat{p} - p\|_1 = \sum_{i=1}^{k} |p_i - \hat{p}_i| \leq \sum_{i=1}^{k} 2\varepsilon p_i = 2\varepsilon$

☆ If $\min_{1 \leq i \leq k} p_i \geq \frac{\varepsilon}{k}$, we can (if no lower bound on p_i, can't)

Chernoff bound + union bound $(\delta' = \frac{\delta}{k})$

$\left(\leq e^{-\varepsilon^2 \cdot \boxed{\frac{\varepsilon}{k}} n} \right)$
when using Chernoff

$$n = O\left(\frac{k}{\varepsilon^3} \log \frac{k}{\delta} \right)$$

☆ replace p by
$p' = (1 - \frac{\varepsilon}{2}) p + \frac{\varepsilon}{2} u_k$
Learn p' to TV $\frac{\varepsilon}{2}$.

Learning in TV distance: third attempt

Theorem 53. *Learning an unknown distribution* $\mathbf{p} \in \Delta(k)$ *to total variation distance* ε *(with success probability* $1-\delta$*) can be done with*

$$n = O\left(\frac{k + \log\frac{1}{\delta}}{\varepsilon^2}\right)$$

i.i.d. samples. (Moreover, this is optimal.)

↓

The empirical estimator works

$$\hat{p}_i = \frac{N_i}{n} \leftarrow \text{\# of times } i \text{ is seen in the } n \text{ samples}$$

Theorem 53. *Learning an unknown distribution* $\mathbf{p} \in \Delta(k)$ *to total variation distance* ε *(with success probability* $1-\delta$*) can be done with*

$$n = O\left(\frac{k + \log\frac{1}{\delta}}{\varepsilon^2}\right)$$

i.i.d. samples. (Moreover, this is optimal.)

Learning in TV distance: third attempt

Want $TV(p,\hat{p}) \le \varepsilon$

$$TV(p,\hat{p}) > \varepsilon \iff \exists S \subseteq \mathcal{X}, \quad |p(S) - \hat{p}(S)| > \varepsilon$$

Fix any $S \subseteq \mathcal{X}$.

$$\Pr[|p(S) - \hat{p}(S)| > \varepsilon] \le 2e^{-2\varepsilon^2 n} \overset{\text{want}}{\le} \frac{\delta}{2^k}$$

this is Hoeffding. Bias of a coin!

$$p(S) = \sum_{i \in S} p_i$$

$$\hat{p}(S) = \sum_{i \in S} \hat{p}_i = \frac{\#\text{samples falling in } S}{n}$$

$$n = \left\lceil \frac{\ln\left(\frac{2^{k+1}}{\delta}\right)}{2\varepsilon^2} \right\rceil = O\left(\frac{k + \log(1/\delta)}{\varepsilon^2}\right)$$

+ union bound over 2^k subsets.

Learning in TV distance: second third attempt

Theorem 53. *Learning an unknown distribution* $\mathbf{p} \in \Delta(k)$ *to total variation distance* ε *(with success probability* $1-\delta$*) can be done with*

$$n = O\left(\frac{k+\log\frac{1}{\delta}}{\varepsilon^2}\right)$$

i.i.d. samples. (Moreover, this is optimal.)

$$\mathbb{E}[TV(p,\hat{p})] = \frac{1}{2}\mathbb{E}[\|p-\hat{p}\|_1] = \frac{1}{2}\sum_{i=1}^{k}\mathbb{E}[|p_i-\hat{p}_i|]$$

$$\underset{\substack{\text{Jensen}\\ f(x)=x^2}}{\le} \frac{1}{2}\sum_{i=1}^{k}\sqrt{\underbrace{\mathbb{E}[(p_i-\hat{p}_i)^2]}_{\mathrm{Var}\,\hat{p}_i}} = \frac{1}{2n}\sum_{i=1}^{k}\sqrt{\underbrace{\mathrm{Var}(n\hat{p}_i)}_{np_i(1-p_i)\le np_i}}$$

$$\le \frac{1}{2\sqrt{n}}\sum_{i=1}^{k}\sqrt{p_i} \underset{\substack{\text{Cor.}\\ \text{(Jensen)}}}{\le} \frac{\sqrt{k}}{2\sqrt{n}}$$

$$\underset{\substack{\uparrow\\ \text{want}}}{\le} \frac{\varepsilon}{10}$$

$\mathbb{E}\hat{p}_i = p_i$

$n\hat{p}_i \sim \mathrm{Bin}(n,p_i)$

True for $n \ge \frac{25k}{\varepsilon^2}$ + Markov.

$$k\cdot\frac{1}{k}\sum_{i=1}^{k}\sqrt{p_i} = k\,\mathbb{E}_{i\sim u}\sqrt{p_i} \le k\sqrt{\mathbb{E}_{i\sim u}p_i} = k\sqrt{\sum_{i=1}^{k}\frac{p_i}{k}} = \sqrt{k}$$

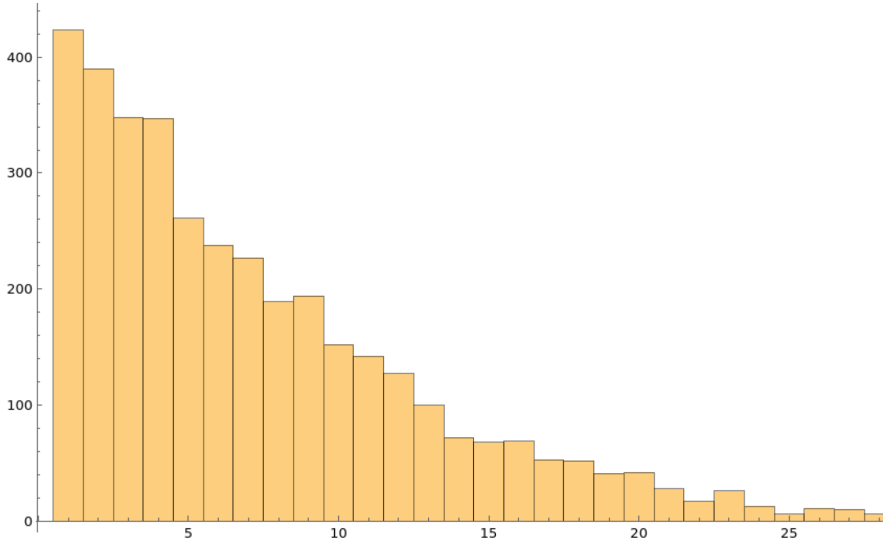
Learning in TV distance: second third attempt

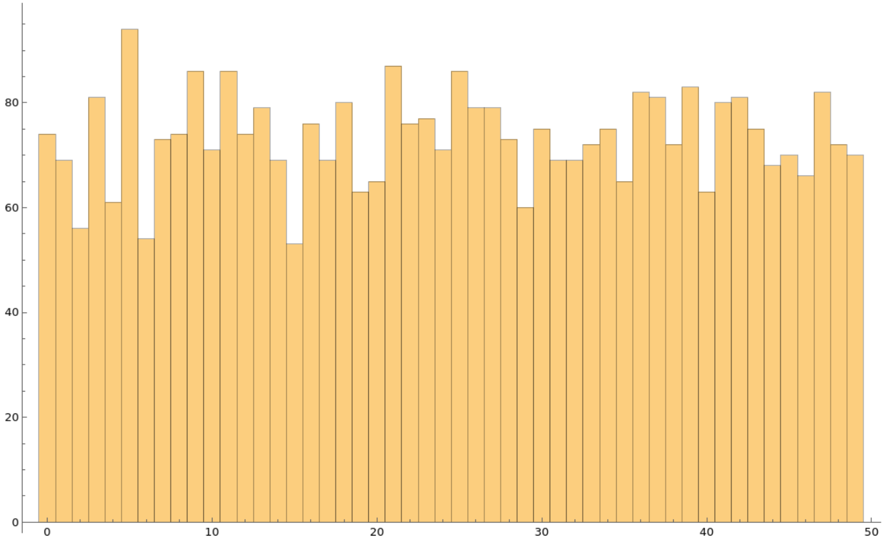
Theorem 53. *Learning an unknown distribution* $\mathbf{p} \in \Delta(k)$ *to total variation distance* ε *(with success probability* $1-\delta$*) can be done with*

$$n = O\left(\frac{k + \log\frac{1}{\delta}}{\varepsilon^2}\right)$$

i.i.d. samples. (Moreover, this is optimal.)

Testing in TV distance





Testing in TV distance: identity testing

Give an algorithm A which takes parameters $\varepsilon, \delta \in (0, 1]$ and n samples from $\mathbf{p}$, and:

- If $\mathbf{p} = \mathbf{q}$, then $\Pr[\, A \text{ outputs yes}\,] \geq 1 - \delta$;
- If $d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) > \varepsilon$, then $\Pr[\, A \text{ outputs no}\,] \geq 1 - \delta$

(if $0 < d_{\mathrm{TV}}(\mathbf{p}, \mathbf{q}) \leq \varepsilon$, then A is off the hook and can output whatever).

Testing in TV distance: identity testing via learning

$n = O\left(\frac{k + \log(1/\delta)}{\varepsilon^2}\right)$ is an upper bound

$p \longrightarrow \hat{p}$ st $TV(p, \hat{p}) \leq \frac{\varepsilon}{2}$ $\longrightarrow$ check if $TV(\hat{p}, q) \leq \frac{\varepsilon}{2}$ $\rightsquigarrow$ if not, say "no"

Key: $TV(p, q) \leq TV(p, \hat{p}) + TV(\hat{p}, q)$

Testing in TV distance: uniformity is all you need

Theorem 54 (Identity to uniformity reduction). *Suppose there is an algorithm A for uniformity testing, which takes $n = n(k, \varepsilon, \delta)$ i.i.d. samples from the unknown distribution. Then there is an algorithm A' for identity testing over a domain of size k to any fixed $\mathbf{q} \in \Delta(k)$, which takes $n = n(4k, \varepsilon/4, \delta)$ i.i.d. samples from the unknown distribution. Moreover, A' is efficient if A is.*

Upshot: "wlog, test $p = u_k$ vs $\mathrm{TV}(p, u_k) > \varepsilon$"

(no need to worry about other q)

Testing in TV distance: uniformity testing

Uniform is simpler

(Focus on $\delta = \frac{1}{3}$)

Birthday paradox:

If p is uniform on $\frac{k}{2}$ elem'ts

vs p is <u>the</u> uniform distr (over k elem'ts)

$\Big\} \; n = \Omega(\sqrt{k})$

Testing in TV distance: uniformity testing

Theorem 55. *Testing uniformity of an unknown distribution* $\mathbf{p} \in \Delta(k)$ *to total variation distance* ε *(with success probability* $2/3$*) can be done with*

$$n = O\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$$

i.i.d. samples, using Algorithm 21. (Moreover, this is optimal for constant success probability.)

Testing in TV distance: uniformity testing, key ideas

$\mathrm{TV} \longrightarrow \ell_2$

$$\|p - u_k\|_1 \leq \sqrt{k}\,\|p - u_k\|_2$$

so if $p = u_k$, $\|p - u_k\|_2 = 0$

if $\mathrm{TV}(p, u_k) > \varepsilon$, $\|p - u_k\|_2 > \frac{2\varepsilon}{\sqrt{k}}$

Testing in TV distance: uniformity testing, key ideas

②

$$\|p - u_k\|_2^2 = \|p\|_2^2 - \frac{1}{k}$$

$$\|$$

$$\sum_i \left(p_i - \frac{1}{k}\right)^2 = \sum_i p_i^2 - \frac{2}{k}\underbrace{\boxed{\sum_i p_i}}_{=1} + \frac{1}{k} = \|p\|_2^2 - \frac{1}{k}$$

- $\|p\|_2^2 = \frac{1}{k}$ if $p = u_k$
- $\|p\|_2^2 > \frac{1+4\varepsilon^2}{k}$ if $TV(p, u_k) > \varepsilon$

$$\|p\|_2^2 = \Pr_{x,y\sim p}[x = y]$$

$$\|$$

$$\sum_i \underbrace{\Pr[x = y = i]}_{p_i \cdot p_i}$$

Testing in TV distance: uniformity testing, algorithm

```
Input: Multiset of $n$ i.i.d. samples $x_1, \dots, x_n \in \mathcal{X}$, parameters $\varepsilon \in (0,1]$ and $k = |\mathcal{X}|$
Set $\tau \leftarrow \frac{1+2\varepsilon^2}{k}$
Compute                                         ▷ $O(n)$ time if $\mathcal{X}$ is known
    $Z = \frac{1}{\binom{n}{2}} \sum_{1 \leq s < t \leq n} \mathbb{1}_{\{x_s = x_t\}} = \frac{1}{\binom{n}{2}} \sum_{j \in \mathcal{X}} \binom{N_j}{2}$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}_{\{x_t = j\}}$.
if $Z \geq \tau$ then return no                 ▷ Not uniform
else return yes                                 ▷ Uniform
```

Testing in TV distance: uniformity testing

```
Input: Multiset of $n$ i.i.d. samples $x_1, \dots, x_n \in \mathcal{X}$, parameters $\varepsilon \in (0,1]$ and $k = |\mathcal{X}|$
Set $\tau \leftarrow \frac{1+2\varepsilon^2}{k}$
Compute                                          ▷ $O(n)$ time if $\mathcal{X}$ is known
    $Z = \frac{1}{\binom{n}{2}} \sum_{1 \le s < t \le n} \mathbb{1}_{\{x_s = x_t\}} = \frac{1}{\binom{n}{2}} \sum_{j \in \mathcal{X}} \binom{N_j}{2}$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}_{\{x_t = j\}}$.
if $Z \ge \tau$ then return no                  ▷ Not uniform
else return yes                                  ▷ Uniform
```

① $\mathbb{E} Z = \|p\|_2^2$ (easy)

② $\mathrm{Var}\, Z <$ small (really hard) } + Chebyshev

"Easy": can prove a bound on Var Z

giving

$$n = O\left(\frac{k}{\varepsilon^4}\right)$$

Getting $O\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$ is hard

Testing in TV distance: uniformity testing

```
Input: Multiset of $n$ i.i.d. samples $x_1, \dots, x_n \in \mathcal{X}$, parameters $\varepsilon \in (0,1]$ and $k = |\mathcal{X}|$
Set $\tau \leftarrow \frac{1+2\varepsilon^2}{k}$
Compute                                                   ▷ $O(n)$ time if $\mathcal{X}$ is known
    $$Z = \frac{1}{\binom{n}{2}} \sum_{1 \le s < t \le n} \mathbb{1}_{\{x_s = x_t\}} = \frac{1}{\binom{n}{2}} \sum_{j \in \mathcal{X}} \binom{N_j}{2}$$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}_{\{x_t = j\}}$.
if $Z \ge \tau$ then return no                            ▷ Not uniform
else return yes                                           ▷ Uniform
```

Testing in TV distance: uniformity testing

```
Input: Multiset of $n$ i.i.d. samples $x_1, \dots, x_n \in \mathcal{X}$, parameters $\varepsilon \in (0,1]$ and $k = |\mathcal{X}|$
Set $\tau \leftarrow \frac{1+2\varepsilon^2}{k}$
Compute                                          ▷ $O(n)$ time if $\mathcal{X}$ is known
    $Z = \frac{1}{\binom{n}{2}} \sum_{1 \leq s < t \leq n} \mathbb{1}_{\{x_s = x_t\}} = \frac{1}{\binom{n}{2}} \sum_{j \in \mathcal{X}} \binom{N_j}{2}$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}_{\{x_t = j\}}$.
if $Z \geq \tau$ then return no                  ▷ Not uniform
else return yes                                  ▷ Uniform
```

Testing in TV distance: uniformity testing

```
Input: Multiset of $n$ i.i.d. samples $x_1, \dots, x_n \in \mathcal{X}$, parameters $\varepsilon \in (0,1]$ and $k = |\mathcal{X}|$
Set $\tau \leftarrow \frac{1+2\varepsilon^2}{k}$
Compute                                                  ▷ $O(n)$ time if $\mathcal{X}$ is known
    $Z = \frac{1}{\binom{n}{2}} \sum_{1 \leq s < t \leq n} \mathbb{1}_{\{x_s = x_t\}} = \frac{1}{\binom{n}{2}} \sum_{j \in \mathcal{X}} \binom{N_j}{2}$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}_{\{x_t = j\}}$.
if $Z \geq \tau$ then return no                          ▷ Not uniform
else return yes                                          ▷ Uniform
```

Testing in TV distance: uniformity testing

```
Input: Multiset of $n$ i.i.d. samples $x_1, \dots, x_n \in \mathcal{X}$, parameters $\varepsilon \in (0,1]$ and $k = |\mathcal{X}|$
Set $\tau \leftarrow \frac{1+2\varepsilon^2}{k}$
Compute                                          ▷ $O(n)$ time if $\mathcal{X}$ is known
    $$Z = \frac{1}{\binom{n}{2}} \sum_{1 \le s < t \le n} \mathbb{1}_{\{x_s = x_t\}} = \frac{1}{\binom{n}{2}} \sum_{j \in \mathcal{X}} \binom{N_j}{2}$$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}_{\{x_t = j\}}$.
if $Z \ge \tau$ then return no                    ▷ Not uniform
else return yes                                   ▷ Uniform
```

Testing in TV distance: uniformity testing, summary

Theorem 55. *Testing uniformity of an unknown distribution* $\mathbf{p} \in \Delta(k)$ *to total variation distance* ε *(with success probability* $2/3$*) can be done with*

$$n = O\left(\frac{\sqrt{k}}{\varepsilon^2}\right)$$

i.i.d. samples, using Algorithm 21. (Moreover, this is optimal for constant success probability.)

Tight bound (other algo)

$$n = O\left(\frac{\sqrt{k \log 1/\delta} + \log 1/\delta}{\varepsilon^2}\right)$$

Summary

- Learning p $\boxed{k/\varepsilon^2}$

  VI

- Testing $p = u_k$ vs $TV(p, u_k) > \varepsilon$ $\boxed{\sqrt{k}/\varepsilon^2}$

- Estimating TV distance $TV(p, u_k)$ $\boxed{\frac{k}{\varepsilon^2 \log k}}$