A useful tool: Hypothesis Testing

Given collection of distributions, $\mathcal{H}$, at least one (via complete description) has high accuracy for describing p $\Leftarrow$ given via samples

output one of collection that is close to p.

How many samples in terms of $|\mathcal{H}|$ + domain size?

Why is this different than testing closeness, uniformity? Do we have the same lower bounds?

NO

Since p is guaranteed to be close to some $q \in \mathcal{H}$, all bets are off!!

A "subtool": allows comparing two hypothesis

Thm Given sample access to p

Given h_1, h_2 hypothesis distributions (fully known to algorithm)

Given accuracy parameter ε', confidence δ'

Algorithm "Choose" takes $O(\log(1/\delta')/(\varepsilon')^2)$ samples + outputs $h \in \{h_1, h_2\}$. If one of h_1, h_2 has $\|h_i - p\|_1 < \varepsilon'$

Then with prob $\geq 1-\delta'$, output h_j has $\|h_j - p\|_1 \leq 12\varepsilon'$

Actually, will prove something stronger:

Thm p given via samples

h_1, h_2 fully known

ε', δ' given

Algorithm "Choose" takes $O(\log(1/\delta')(1/\varepsilon')^2)$ samples

+ outputs $h \in \{h_1, h_2\}$ satisfying:

(1) if h_i more than $\underbrace{12\varepsilon'\text{-far}}_{\text{very bad}}$ from p, $\underbrace{\text{unlikely}}_{2e^{-m\varepsilon'^2/2}}$ to ouput it as winner <u>or</u> tie

(2) if h_i more than $\underbrace{1[illegible]\varepsilon'\text{-far}}_{\text{not that bad}}$, unlikely to output as winner

might tie but won't win

Proof of "Subtool":

idea: if h_1 is ε'-close, show will output h_1 whp

else if h_1 is $12\varepsilon'$-far, show will not output h_1 + will output h_2

else (h_1 is not $12\varepsilon'$-far, but not ε'-close, so h_2 must be ε'-close)

we don't know what will happen,

but either way we are golden (neither h_1 or h_2 are that bad)

Algorithm Choose:

Input p, h_1, h_2 (samples: p; explicit description: h_1, h_2)

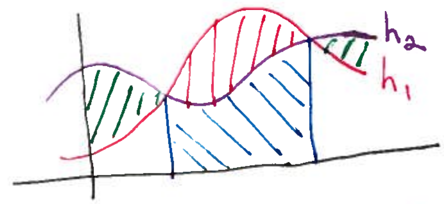
First some definitions:

$$A = \{x \mid h_1(x) > h_2(x)\}$$

$$a_1 = h_1(A)$$

$$a_2 = h_2(A)$$

note $\|h_1 - h_2\|_1 = 2(a_1 - a_2)$



green area = red area

L_1 dist = green + red

red area = $a_1 - a_2$

blue area = a_2

blue + red area = a_1

this is important! ⟹ 1. if $a_1 - a_2 \le 5\varepsilon'$ declare "tie" + return h_1 (no samples needed)

2. draw $m = 2 \cdot \dfrac{\log \frac{1}{\delta'}}{(\varepsilon')^2}$ samples $s_1 \cdots s_m$ from p

3. $\alpha \leftarrow \frac{1}{m} \left| \{ i \mid s_i \in A \} \right|$

4. if $\alpha > a_1 - \frac{3}{2}\varepsilon'$ return h_1

else if $\alpha < a_2 + \frac{3}{2}\varepsilon'$ return h_2

else declare "tie" + return h_1

Why does it work?

$$E[\alpha] = p(A)$$

- if reach step 2, whp (via Chernoff) $|\alpha - E[\alpha]| \le \frac{\varepsilon'}{2}$

  if $\|p - h_1\|_1 > 12\varepsilon'$ then since other is $\le \varepsilon'$ distance,
  or $\|p - h_2\|_1 > 12\varepsilon'$ $\qquad \|h_1 - h_2\|_1 > 11\varepsilon'$

  so will reach step 2

  if p ε'-close to h_1, whp $\alpha > a_1 - \varepsilon' - \frac{\varepsilon'}{2}$

  from closeness to h_1 $\qquad$ sampling error

  so output h_1

  else, p is $12\varepsilon'$ far from h_1
  but ε'-close to h_2

  whp $\alpha < a_2 + \varepsilon' + \frac{\varepsilon'}{2}$

  $\Rightarrow$ return h_2 whp

- if h_1 or h_2 $> 10\varepsilon'$ from but not $12\varepsilon'$ far
  if $p_1 - p_2 \le 5\varepsilon'$ then declares draw, so neither are declared "winner"

  else $\|h_1 - h_2\| > 9\varepsilon'$ far

  + similar reasoning shows that medium far will not win (in fact, will lose)

The Cover Method

a method for learning distributions

def $\mathcal{C}^{\mathcal{D}}$ is a ε-cover of $\mathcal{D}$ if $\forall\, p \in \mathcal{D}$ $\exists\, q \in \mathcal{C}^{\mathcal{D}}$ s.t. $\|p-q\|_1 \le \varepsilon$

$\mathcal{C}^{\mathcal{D}}$: set of distributions (smaller)

$\mathcal{D}$: set of distributions (big)

Why useful? - allows us to "approx" $\mathcal{D}$

hopefully $\mathcal{C}$ is much smaller than $\mathcal{D}$

note $\mathcal{C}$ not unique

Thm $\exists$ algorithm, given $p \in \mathcal{D}$, which takes $O\left(\frac{1}{\varepsilon^2}\log|\mathcal{C}^{\mathcal{D}}|\right)$ samples of p + outputs $h \in \mathcal{C}^{\mathcal{D}}$ s.t. $\|h-p\|_1 \le 6\varepsilon$ with prob $\ge 9/10$

Pf.

since $p \in \mathcal{D}$, $\exists\, q \in \mathcal{C}^{\mathcal{D}}$ s.t. $\|p-q\|_1 \le \delta$

(but there could be more than 1) $\Leftarrow$ we just need to find one, not even required to return p

will run Choose on p with every pair $q_1, q_2 \in \mathcal{C}^{\mathcal{D}}$

if q doesn't win all of its "matches" then it loses to someone that is not so bad

Furthermore can show that whp there is a q' s.t.

q' $\boxed{\text{wins}}$ or $\boxed{\text{ties}}$ all matches.

The cover method

Example 1: learning distribution of a coin

domain = $\{0,1\}$

need to learn bias

Here $\mathcal{D} = \mathbb{R}$

describes bias of coin

if use $\mathcal{C} = \{ 0, \frac{1}{k}, \frac{2}{k}, \dots, \frac{k-1}{k}, 1 \}$

then $\forall$ bias p, let $\frac{i}{k} \leq p \leq \frac{i+1}{k}$

then picking $\tilde{p} = \frac{i}{k}$ gives $\|p-\tilde{p}\|_1 = \left|\frac{i}{k} - p\right| + \left|\left(1-\frac{i+1}{k}\right) - (1-p)\right|$

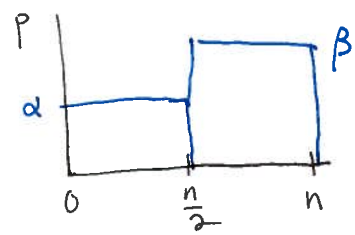
$\leq \frac{2}{k}$

so using $k = \Theta\left(\frac{1}{\varepsilon}\right)$ gives $\|p-\tilde{p}\|_1 \leq \varepsilon$

$|\mathcal{C}| = k+1$, # samples needed by cover method is $O\left(\frac{1}{\varepsilon^2} \cdot \log \frac{1}{\varepsilon}\right)$

$= \Theta\left(\frac{1}{\varepsilon}\right)$

Example 2: 2-bucket distributions



now need to specify α and β

so $\mathcal{C} = \left\{ \left(\frac{i}{k}, \frac{j}{k}\right) \,\middle|\, i, j \in \{0 \dots k\} \right\}$

$|\mathcal{C}| = \Theta\left(\left(\frac{1}{\varepsilon}\right)^2\right)$

samples is $O\left(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}\right)$

Example 3: monotone distributions

Birge $\Rightarrow$ $\mathcal{C} = \left\{ \left(\frac{i_1}{k}, \dots, \frac{i_{(\log n/\varepsilon)}}{k}\right) \,\middle|\, i_1, i_2, \dots \in \{0 \dots k\} \right\}$

$|\mathcal{C}| = \Theta\left(\frac{1}{\varepsilon}^{(\log n)/\varepsilon}\right)$ $\Rightarrow$ # samples is $O\left(\frac{1}{\varepsilon^3} \log n \cdot \log \frac{1}{\varepsilon}\right)$