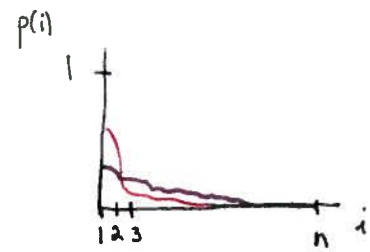
Lecture 15:

Testing monotonicity of distributions

Testing & Learning Monotone Distributions (over totally ordered domain)



Def. p over $[n]$ is "monotone decreasing" if $\forall\ i \in [n-1]$ $\quad p(i) \geq p(i+1)$

Monotonicity Tester:

- if p monotone increasing, Pass with prob $\geq 3/4$
- if p ε-far in L_1 dist from mon increasing, Fail with prob $\geq 3/4$

Useful tool: "Birge Decomposition"

(note: this is a different decomposition than in homework (upcoming) in particular, it is oblivious!)

decompose domain $1..n$ into $\ell = \Theta\left(\frac{\log \varepsilon n}{\varepsilon}\right) \approx \Theta\left(\frac{\log n}{\varepsilon}\right)$ intervals

$I_1^\varepsilon, I_2^\varepsilon, \ldots, I_\ell^\varepsilon$ s.t.

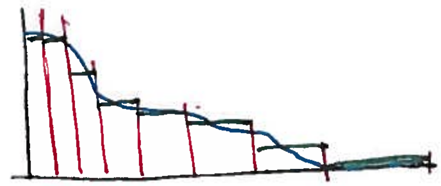
$|I_{k+1}^\varepsilon| = \lfloor (1+\varepsilon)^i \rfloor$ ← will drop ε in notation once it's fixed

$|I_1^\varepsilon| = |I_2^\varepsilon| = \ldots = 1$

$|I_a^\varepsilon| = |I_{a+1}^\varepsilon| = \ldots = 2$ but then at some point the sizes grow exponentially

$\vdots$

define "flattened distribution"



$\forall\ 1 \le j \le \ell$

$\forall\ i \in I_j$ $\qquad \tilde{q}(i) = \dfrac{q(I_j)}{|I_j|}$ ← assign all elements in same interval the same probability

note: $q(I_j) = \tilde{q}(I_j)$

Birge's Thm if q mon decreasing then $\|\tilde{q} - q\|_1 < \varepsilon$

Corr'l if q ε-close to mon decreasing then $\|\tilde{q} - q\|_1 < O(\varepsilon)$

Testing Algorithm:

how can we do this? $\tilde{q}$ isn't even if q monotone, problem exactly uniform. See from next hw set.

Take samples of q

do uniformity test for each partition (using samples that fell in it):
(if not enough samples then pass)

"FAIL" if $> \varepsilon$ fraction of weight is in partitions that fail.

$w_j \leftarrow$ # samples that fell in partition j

use LP to verify w close to monotone
(that is dist which gives wt w_j to each Birge bucket + uniform within each bucket)

note this is LP on $O(\log n)$ vars

How many samples?

for each partition with enough weight, say $\frac{\varepsilon}{\log n}$, need $\frac{\sqrt{n}}{\varepsilon^2}$ samples

need $\frac{\sqrt{n} \cdot \log n}{\varepsilon^3}$ for each one

$\approx O(\sqrt{n}\ \text{polylog}\, n \cdot \text{poly}\, \frac{1}{\varepsilon})$

need another $\log\log n$ for union bound

(note: this can be improved!!)

Last step:

Sampling error might make w_j's look non monotone

difficulty

purple is not monotone
but is close

good thing: only $\frac{\log n}{\sqrt{\varepsilon}}$ variables!

Can be solved via brute force
LP (actually quite efficient)
⋮

So: monotone p likely to pass
ε-far from monotone p: either (1) non uniform in buckets
or (2) w far from monotone

{ more details on next page

Slightly changing perspective...

What if we know dist q is monotone, can we learn it?

Yes! use sampling to estimate $\tilde{q}_\varepsilon(I_j)$'s

Birge's Thm ⇒ Can learn monotone distributions to w/in $\le \varepsilon$ L_1 error
in $\Theta(\frac{1}{\varepsilon^3} \log n)$ samples.

Correctness: (high level idea)

q monotone:

- Birge $\Rightarrow \|q - \tilde{q}\| < \varepsilon' < \varepsilon$

- Claim: (ignoring partitions with max wt $\frac{\varepsilon}{2n}$) for $\underbrace{\leq \frac{\log n}{\varepsilon}}_{\text{"Bad partitions" are } < \varepsilon \text{ fraction of all partitions}}$ partitions, the $\frac{\text{min value of } q}{\text{max value of } q} \leq \varepsilon'$

An issue: total # of "bad" partitions is small, but also need total weight to be small or need to fix algorithm

- for "good" partitions, uniformity test likely to pass [see HW in future]

- also, w is close to $\tilde{q}$ which is monotone, so w is close to monotone

q ε-far from monotone:

- assume q "likely to pass"

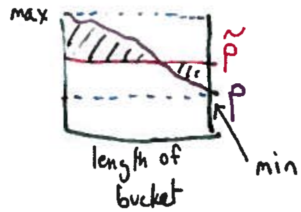
$\Rightarrow$ most Birge buckets close to uniform + w close to monotone $\Big\}$ correct to monotone $\tilde{q}$ which is uniform on buckets s.t. average bucket wt. is monotone

$|q - \tilde{q}|$ is small + $\tilde{q}$ close to monotone $\Rightarrow$ q close to monotone

thus ε-far $\Rightarrow$ likely to fail ∎

Proof of Birge's Thm :

Error in bucket



gross upper bound on error:
$\leq (\text{max} - \text{min}) \cdot \text{bucket length}$

Partition of Intervals:

- Size 1 Intervals $|I_j| = 1$ ← if we have any short intervals, there are $\Omega(1/\varepsilon)$ of these; if not, we can learn the distribution. no error on these!
- Short Intervals $|I_j| < 1/\varepsilon$
- Long Intervals $|I_j| \geq 1/\varepsilon$

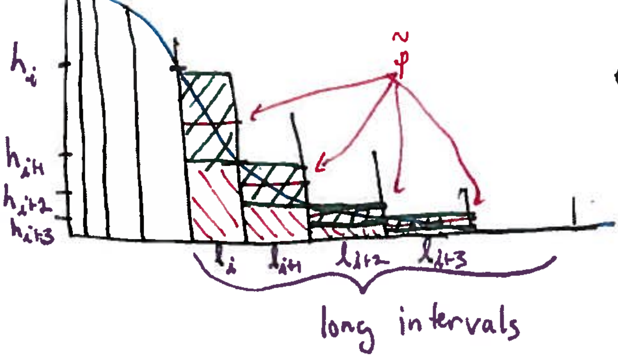
← if we have these then max prob $\leq \varepsilon$ (since # size 1 intervals is $\Omega(1/\varepsilon)$)

↑ therefore min size 1 interval has prob $\leq \varepsilon$ which upper bounds later probabilities too since p is monoton

$$\text{total error} \leq \sum_{j=1}^{\ell} |I_j| \cdot (\text{max prob in } I_j - \text{min prob in } I_j)$$

$$= \underbrace{\sum_{\substack{\text{size 1}\\ \text{intervals}}} 1 \cdot 0}_{\substack{0 \\ \text{since no difference}}} + \underbrace{\sum_{\substack{\text{short}\\ \text{intervals}}} |I_j| (\text{max} - \text{min})}_{\text{omitted: idea is bound similarly to the long intervals but need to group together intervals of same size}} + \underbrace{\sum_{\substack{\text{long}\\ \text{intervals}}} |I_j| (\text{max} - \text{min})}_{\text{see below}}$$

Picture for long intervals:



green rectangles = upper bnd on error

$$\text{error} \leq (h_i - h_{i+1}) l_i + (h_{i+1} - h_{i+2}) l_{i+1} + (h_{i+2} - h_{i+3}) l_{i+2} + \ldots$$

$$= \underbrace{h_i l_i}_{\text{all } h_i\text{'s in this area are } < \varepsilon!} + h_{i+1}(l_{i+1} - l_i) + h_{i+2} \underbrace{(l_{i+2} - l_{i+1})}_{\text{positive, } + \approx \varepsilon \cdot l_{i+1} \text{ by way that we partitioned}} + h_{i+3}(l_{i+3} - l_{i+2})$$

$$\leq \varepsilon \Big[\underbrace{l_i}_{\text{get rid of this when bounding short intervals}} + \underbrace{\sum h_i l_{i-1}}_{\text{this is area of red rectangles, which is upper bounded by } p \text{ so sum is } \leq 1} \Big]$$