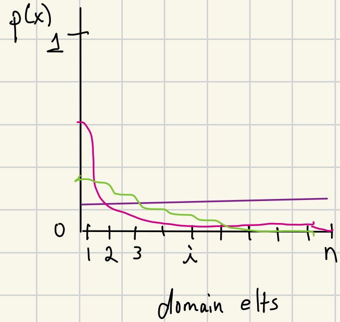
Learning & testing monotone distributions

Monotone distributions (over totally ordered domain)

total order

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



Monotonicity tester:

with prob $\geq 1-\delta$

- if p monotone decreasing, output PASS
- if p ε-far in L_1 from any mon dec dist q, output FAIL

Pset: lower bnd $\Omega(\sqrt{n})$ samples

Useful Tool: Birgé Decomposition + Flattening

different decomposition than H.W.
<u>OBLIVIOUS</u> (not depend on p)

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

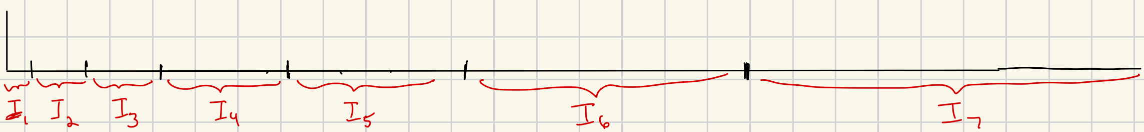
$$I_1^\varepsilon, I_2^\varepsilon, \ldots, I_\ell^\varepsilon \quad \text{s.t.}$$

$$|I_{k+1}^\varepsilon| = \lfloor (1+\varepsilon)^k \rfloor$$

will drop ε from notation since fixed by algorithm

Note that
$$|I_1^\varepsilon| = |I_2^\varepsilon| = \ldots = 1$$
$$|I_a^\varepsilon| = |I_{a+1}^\varepsilon| = \ldots = 2$$
$$\vdots$$

$\Theta(1/\varepsilon)$ such intervals due to "floor"
but then at some point the exponential "takes off"



Birge Decomposition

$|I_{k+1}| = \lfloor (1+\varepsilon)^k \rfloor$

<u>def.</u> "flattened distribution": given q

$\forall$ intervals $1 \le j \le \ell$, $\forall\, i \in I_j$

$$\tilde{q}(i) = \frac{q(I_j)}{|I_j|}$$

$\leftarrow$ total wt of interval

$\leftarrow$ # of domain elts in interval

<u>Note</u> $\tilde{q}(I_j) = q(I_j)$

but $\tilde{q}$ spreads weight uniformly over I_j

hopefully few of these { intervals in which q not close to uniform

<u>Birge's Thm</u>: If q is monotone decreasing then $\|\tilde{q} - q\|_1 < O(\varepsilon)$

<u>Corr</u>: " " " ε-close to " " " "

Testing algorithm: (1st try)

- Take m samples S' of q. ← how many?
- For each Birge partition I_j: ← parameter ε/c!

$$S_j \leftarrow S' \cap I_j$$

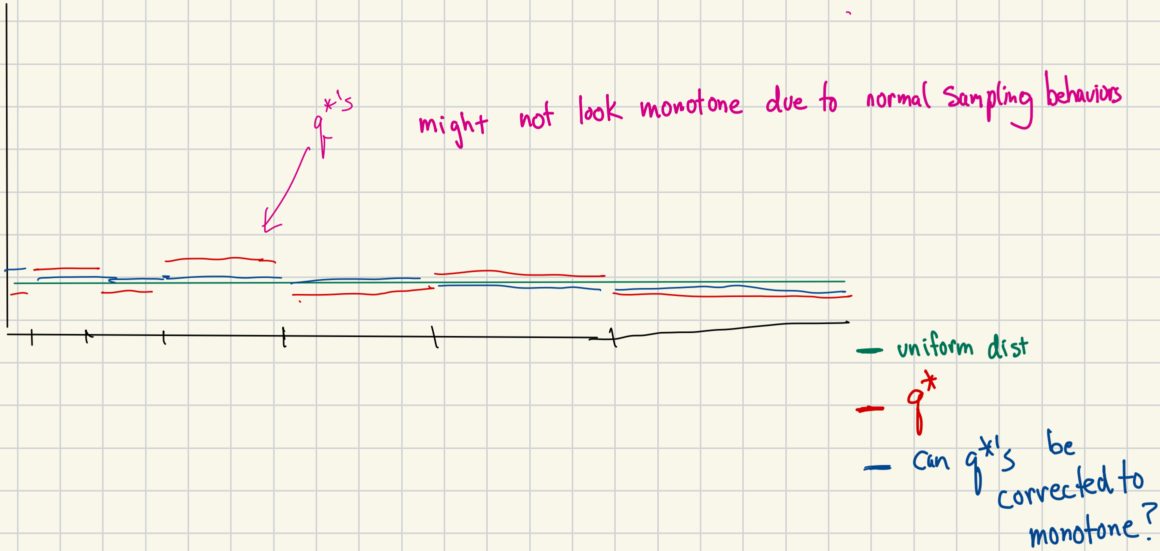
$$\hat{w}_j \leftarrow \frac{|S_j'|}{m} \quad \leftarrow \text{estimate of } q(I_j)$$

- Define q^*: $\forall i \in I_j$, $q^*(i) = \dfrac{\hat{w}_j}{|I_j|}$

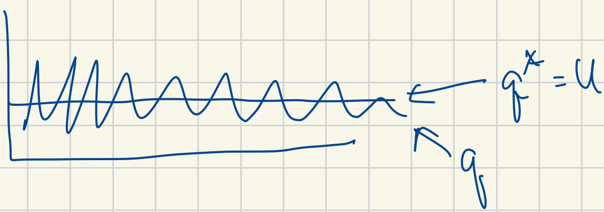
- Check that q^*'s are monotone decreasing

why is this a bad algorithm?

Problem 1 Sampling errors for $\hat{w}_j$'s



Problem 2 what if q not monotone within buckets?



Testing algorithm: "Testing by learning"

how many?

- Take m samples S' of q.
- For each Birge partition I_j: ← parameter $\frac{\varepsilon}{c}$!

$$S_j \leftarrow S' \cap I_j$$

$$\hat{w}_j \leftarrow \frac{|S_j'|}{m}$$ ← estimate of $q(I_j)$

- Define q^*: $\forall\, i \in I_j$, $q^*(i) = \frac{\hat{w}_j}{|I_j|}$

(A) • Use LP on $\hat{w}_j$'s to verify that q^* is $\frac{\varepsilon}{c}$ close to monotone

if no, Fail + halt

no new samples needed

this is LP in $O(\log n)$ vars

(B) • Test that L_1-dist of q + q^* is $< \frac{\varepsilon}{c}$

if no, Fail + halt

else accept

how to do this? even if q monotone, q + q^* are only close

how do we pass all "good" (monotone) q?

previous algorithms are not tolerant

Correctness (high level) (q monotone $\Rightarrow$ ~~tester~~ tester passes whp)

- if q monotone then $\tilde{q}$ monotone

  & Birgé $\Rightarrow$ $\|q-\tilde{q}\|_1 < \frac{\varepsilon}{c'}$

- Since $\hat{w}_j'$s are close to $q(I_j)$ $\leftarrow$ via Chernoff bnd argument

  $\Rightarrow \|\tilde{q} - q^*\|_1 < \frac{\varepsilon}{c'}$

- so q^* is $\frac{\varepsilon}{c'}$-close to monotone

- $\|q - q^*\|_1 < 2 \cdot \frac{\varepsilon}{c'}$ by $\Delta \neq$

difficulty Can distinguish $q = q^*$ from $\|q-q^*\|_1 > \varepsilon$ in $O(\sqrt{n})$ samples, but here we would need to distinguish $\|q-q^*\|_1 < \varepsilon'$ from $\|q-q^*\|_1 > \varepsilon$ in $O(\sqrt{n})$ samples. If q, q^* arbitrary, need $\Omega(n/\log n)$ samples.

in particular q^* close to monotone $\rightarrow$ but q is monotone in "PASS" case & q^* is "never too crazy", so we can do it !!!

Correctness (high level) to show: q ε-far from monotone $\Rightarrow$ tester fails whp

equivalent

show contrapositive: tester passes whp $\Rightarrow$ q ε-close to monotone

- tester passes (A) $\Rightarrow$ q^* is $\frac{\varepsilon}{c}$-close to monotone
- tester passes (B) $\Rightarrow$ $\|q^* - q\|_1 < \varepsilon/c$

$\Rightarrow$ q is $\frac{2\varepsilon}{c}$-close to monotone

via $\Delta \neq$ $\blacksquare$

What do we need from (B)?

- if q monotone & q^* close to monotone (passes (A))
  then pass whp
- if q far from monotone + q^* close to monotone
  then fail whp

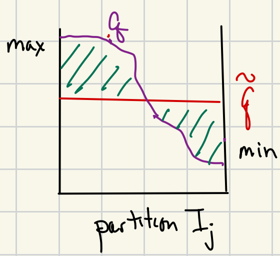
in this case the distance comes from inside buckets

idea inside buckets should be close to uniform, estimate collision prob.

Birge's Thm: If q is (~~ε-close to~~) monotone decreasing then $\|\tilde{q} - q\|_1 < O(\varepsilon)$

Proof of Birge's Thm

error in partition:



gross upper bnd on error:

$\leq (\max - \min) \cdot$ partition length

Type of Intervals:

if have any short intervals then there are $\geq \frac{1}{\varepsilon}$ size 1 intervals

- Size 1 intervals $|I_j| = 1$ — no error on these
- Short intervals $|I_j| < 1/\varepsilon$
- Long intervals $|I_j| \geq 1/\varepsilon$

if have any of these, max prob $\leq \varepsilon$

why?

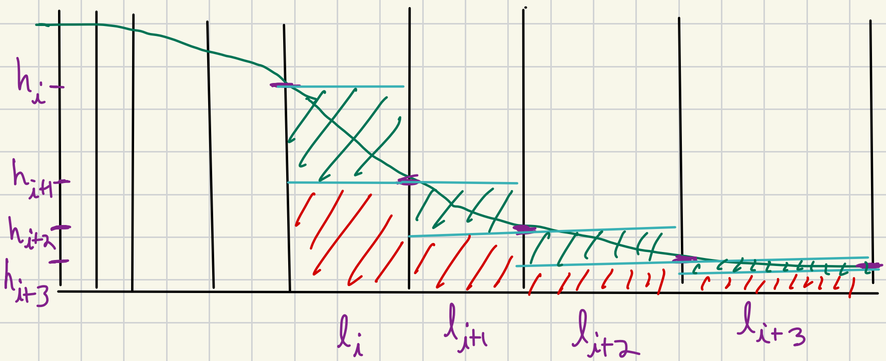
- # size 1 intervals $\geq \frac{1}{\varepsilon}$ (by partitioning)
- last size 1 interval has prob $\leq \varepsilon$ (min wt)

why? if last size 1 interval has wt $> \varepsilon$ then all previous size intervals have wt $> \varepsilon$

$\Rightarrow$ total wt of size 1 intervals $> \frac{1}{\varepsilon} \cdot \varepsilon > 1$ $\rightarrow\!\leftarrow$

Bounding $\sum_{\substack{\text{long} \\ \text{intervals}}} |I_j| (\text{max}-\text{min})$:

green rectangles upper bnd error



$$l_i = \lfloor (1+\varepsilon)^i \rfloor \qquad l_{i+1} - l_i = (1+\varepsilon)^{i+1} - (1+\varepsilon)^i \approx \varepsilon (1+\varepsilon)^i$$

$$\text{error} \le (h_i - h_{i+1}) \cdot l_i + (h_{i+1} - h_{i+2}) l_{i+1} + (h_{i+2} - h_{i+3}) l_{i+2} + \dots \quad \text{(green)}$$

$$\le h_i l_i + h_{i+1}(l_{i+1} - l_i) + h_{i+2}(\underbrace{l_{i+2} - l_{i+1}}_{\approx \varepsilon \cdot l_{i+1}}) + \dots \quad \text{algebra rewrite}$$

all h_i's $\le \varepsilon$

$\approx \varepsilon \cdot l_{i+1}$ by the way we partitioned

$$\le \varepsilon \left[l_i + \underbrace{\sum h_i l_{i-1}}_{} \right]$$

get rid of when bound short intervals.

area of red rectangles which is upper bounded by q so sum ≤ 1

Slight change of perspective:

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

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

Birgé's thm $\Rightarrow$ can learn monotone distributions

to w/in ε L_1-error

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