Lecture 15

Learning & Testing Distributions:

Monotonicity
Monotone distributions (over totally ordered domain)

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

Monotonicity tester:
- if $p$ monotone decreasing, output PASS
- if $p$ $\epsilon$-far in $L_1$ from any mon dec dist $q$, output FAIL

with probability $\geq 1-\delta$
Useful Tool: Birge Decomposition & Flattening

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

$I_1^\varepsilon, I_2^\varepsilon, \ldots, I_l^\varepsilon$ st.

$|I_{k+1}^\varepsilon| = \left[(1+\varepsilon)^k \right]$  

$\leftarrow$ will drop $\varepsilon$ in notation since $\varepsilon$ is fixed by algorithm

Note that

$|I_1^\varepsilon| = |I_2^\varepsilon| = \ldots = 1 \quad \leftarrow \Theta(\frac{1}{\varepsilon})$ intervals

$|I_a^\varepsilon| = |I_{a+1}^\varepsilon| = \ldots = 2$

but then at some point the exponential "takes off"
Def. "flattened distribution":

\[
\forall \text{ intervals } 1 \leq j \leq l, \quad \forall i \in I_j
\]

\[
\tilde{q} (i) = \frac{q(I_j)}{|I_j|} \leq \text{total weight of interval}
\]

\[
\leq \text{# els in interval}
\]

\[
\text{all els in interval assigned same weight}
\]

\[
\text{Note: } q(I_j) = \tilde{q}(I_j)
\]

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

Corr: \( \tilde{q} \) is \( \varepsilon \)-close to \( q \) and \( \| \tilde{q} - q \|_1 \) is \( O(\varepsilon) \)
Birge’s Thm: If \( q \) is \((\varepsilon\text{-close to})\) monotone decreasing then \( \| \tilde{q} - q \|_1 = O(\varepsilon) \)

Testing algorithm:

- Take \( M \) samples \( S \) of \( q \).
- For each Birge partition \( I_j \):
  \[ S_j = S \cap I_j \]
  \[ \tilde{\omega}_j = \frac{n_j}{m} \]
- Define \( q^* \) \( \forall i \in I_j \): \( q^*(i) = \frac{\tilde{\omega}_j}{|I_j|} \)
- Use LP on \( \tilde{\omega}_j 's \) to verify that \( q^* \) is \( \varepsilon \)-close to monotone
  - if no, Fail & halt
- Test that \( L_1 \)-dist of \( q + q^* \) is \( < \frac{\varepsilon}{c} \)
  - if no, Fail & halt
  - else, accept
Birge's Thm: If \( q \) is \( \varepsilon \)-close to monotone decreasing then \( \| \tilde{q} - q \|_1 \leq O(\varepsilon) \)

Correctness (high level)

- If \( q \) monotone then \( \tilde{q} \) is monotone
  \( (\varepsilon \leq \frac{\varepsilon}{c}) \)
- Since \( \hat{w}_i \)'s are close to \( q(I_j) \) \leftarrow \text{Chernoff argument}
  \( \| \tilde{q} - q^* \|_1 \leq \frac{\varepsilon}{c} \)
- So \( q^* \) is \( \frac{\varepsilon}{c} \)-close to monotone.
- \( \| q - q^* \|_1 \leq 2 \cdot \frac{\varepsilon}{c} \) by \( \Delta \hat{w} \)

Testing algorithm:

- Take \( m \) samples \( S \) of \( q \).
- For each Birge partition \( I_j \):
  \( S_j = S \setminus I_j \)
  \( \hat{w}_j = \frac{|S_j|}{m} \)
- Define \( \tilde{q}^* \) where \( \forall i \in I_j, \tilde{q}^*(i) = \frac{\hat{w}_j}{|I_j|} \)
- Verify that \( \tilde{q}^* \) is \( \varepsilon \) close to monotone (no samples)
- Test that \( L_1 \)-dist of \( q \) and \( q^* \) is \( \leq \frac{\varepsilon}{c} \)

difficulty:
we can distinguish \( \| q - q^* \|_1 = 0 \) in \( O(\sqrt{n}) \) samples
but we don't can't in general distinguish \( \| q - q^* \|_1 < \frac{\varepsilon}{c} \) from \( \| q - q^* \|_1 > \varepsilon \) in \( O(\sqrt{n}) \) samples

Luckily: this is a special case since we know \( \tilde{q} \) is monotone!
Birge's Thm: If \( q \) is \((\varepsilon, \text{close})\)-monotone decreasing then
\[
\| \hat{q} - q \|_1 < O(\varepsilon)
\]

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

Show contrapositive: tester passes whp \( \Rightarrow q \) \( \varepsilon \)-close to monotone.

- Tester passes \( \Rightarrow q^* \frac{\varepsilon}{\varepsilon^2} \)-close to monotone.
- Tester passes \( \Rightarrow \| q - q^* \|_1 < \frac{\varepsilon}{\varepsilon^2} \)
  \( \Rightarrow q \) is \( 2\varepsilon \frac{\varepsilon}{\varepsilon^2} \)-close to monotone.

Testing algorithm:

- Take \( m \) samples \( S' \) of \( q \).
- For each Birge partition \( I_j \):
  \[
  S_j = S' \cap I_j,
  \quad \hat{w}_j = \frac{|S_j|}{m}
  \]
- Define \( \hat{q}^* = \sum_{j \in I_j} q^*(x) \frac{\hat{w}_j}{|I_j|} \)
- Verify that \( \hat{q}^* \) is \( \varepsilon \)-close to monotone (no sampled).
- Test that \( L_1 \)-dist of \( q \) and \( \hat{q}^* \) is \( < \frac{\varepsilon}{\varepsilon^2} \).
Birge's Thm: If \( q \) is (\( \varepsilon \)-close)monotone decreasing then \( \| \tilde{q} - q \| < O(\varepsilon) \)

Proof of Birge's Thm

error in partition:

\[
\begin{align*}
\min \quad & \tilde{q} \\
\max \quad & \tilde{q}
\end{align*}
\]

gross upper bnd on error:

\[ \leq (\max - \min) \cdot \text{partition length} \]

Type of Intervals:

- Size 1 intervals
- Short intervals
- Long intervals

Total error \[ \leq \sum_{j=1}^{l} |I_j| \cdot (\max \text{ prob in } I_j - \min \text{ prob in } I_j) \]

\[ = \sum_{\text{size 1 intervals}} 1 \cdot 0 + \sum_{\text{short intervals}} |I_j| \cdot (\max - \min) + \sum_{\text{long intervals}} |I_j| \cdot (\max - \min) \]

\( \Rightarrow \) total wt \( \frac{1}{\varepsilon} \cdot \lambda < 1 \)

\( \Rightarrow \lambda < \varepsilon \)
Bounding \[ \sum |I_j| (\text{max-min}) : \]

- Green rectangles = upper bound on error
- Upper bnd on error
- \( \forall i \in I_j, q_i(x) = \frac{q(x)}{|I_j|} \)

Error \[ \leq (h_i - h_{i+1}) l_i + (h_{i+1} - h_{i+2}) l_{i+1} + (h_{i+2} - h_{i+3}) l_{i+2} + \ldots \]

- \( \leq h_i l_i + h_{i+1} (l_{i+1} - l_i) + h_{i+2} (l_{i+2} - l_{i+1}) + \ldots \)
- All \( h_i \)'s in this area are \( \leq \varepsilon \)!
- All \( h_i \)'s in this positive \( \approx \varepsilon \cdot l_{i+1} \)

\[ \leq \varepsilon \left( l_i + \sum h_i l_{i-1} \right) \]

- Get rid of red rectangles which is upper bounded by \( p \), so sum \( \leq 1 \)

\[ \text{area of short intervals} \]

\[ \text{Bounding} \]

\[ \text{Birge Flattening} \]
\[ h_{i-1}, h_{i+1}, h_{i+2}, h_{i+3} \]

\[ l_i, l_{i+1}, l_{i+2}, l_{i+3} \]
Slight change of perspective:

if we know \( q \) is monotone, can we learn it?

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

Birge's Thm \( \Rightarrow \) can learn monotone distributions
to within \( \epsilon \) \( L_1 \)-error
in \( \Theta \left( \frac{1}{\epsilon^2 \log n} \right) \) samples.
Testing algorithm:

1. Take $m$ samples $S'$ of $q$.
2. For each Birge partition $I_j$:
   - $S_j := S' \cap I_j$
   - $n_j := |S_j| \quad \hat{w}_j := |S_j| / m$
3. Define $q^* \forall i \in I_j$ $q^*(i) = \frac{\hat{w}_j}{|I_j|}$
4. Verify that $q^*$ is $\varepsilon$-close to monotone
5. Test that $L_1$-dist of $q + q^*$ is $< \frac{\varepsilon}{2}$