07.06.2023 ***ניתוח מידע אדפטיבי***

הרצאה 7: יציבות + אדפטיביות בסטרימינג

Source: Lecture notes by Aaron Roth and Adam Smith

מרצה: אורי שטמר

<u>אבחנה 1:</u> יהי $\mathcal{A}$ אלגוריתם (ε,δ)-DP-יציב אשר פועל על דטהבייס בגודל n. יהי $\mathcal{B}$ אלגוריתם אשר פועל על T דטהבייסים בגודל n כ"א, מריץ על כל אחד מהם את אלגוריתם $\mathcal{A}$ ומחזיר את T הפלטים שהוא מקבל. אזי $\mathcal{B}$ מקיים (ε,δ)-DP-יציבות.

<u>הוכחה:</u> תרגיל

<u>משפט 2:</u> יהי W מכניזם (ε,δ)-DP-יציב אשר מקבל כקלט מדגם S בגודל $n=\Omega\left(\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right)$ ופולט שאילתא סטטיסטית q. אזי, לכל התפלגות $\mathcal{D}$ מתקיים

$$\Pr_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}\left[|q(S)-q(\mathcal{D})|>10\varepsilon\right]<\frac{\delta}{\varepsilon}$$

<u>"הוכחה" של משפט 2 עם הרבה פרטים חסרים:</u>

התנאים של משפט 2 הם:

- W הוא מכניזם (ε,δ)-DP-יציב אשר מקבל כקלט מדגם S ופולט שאילתא סטטיסטית q
- S הוא מדגם המכיל n דגימות iid מהתפלגות כלשהי $\mathcal{D}$
- אנחנו מסמנים $q\leftarrow W(S)$

<u>מטרה:</u> להראות <u>שבהסתברות גבוהה</u> מתקיים $q(S)\approx q(\mathcal{D})$

<u>התוכנית:</u> נתחיל מלהראות חסם בתוחלת ואז נחזק אותו קצת ואח"כ נהפוך אותו לחסם בהסתברות גבוהה.

<u>טענה א:</u>

$$\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}[q(S)]\approx\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}[q(\mathcal{D})]$$

<u>הסבר:</u> נסתכל על 2 הניסויים הבאים:

$S=(x_1,\dots,x_n)\sim\mathcal{D}^n$ $z\sim\mathcal{D}$ $i\in_R\{1,2,\dots,n\}$ $q\leftarrow W(S\setminus\{x_i\}\cup\{z\})$ Return $q(x_i)$	$\approx$ (בגלל DP-יציבות)	$S=(x_1,\dots,x_n)\sim\mathcal{D}^n$ $z\sim\mathcal{D}$ $i\in_R\{1,2,\dots,n\}$ $q\leftarrow W(S)$ Return $q(x_i)$
כאן הפלט הוא q מופעלת על איבר אקראי מההתפלגות $\mathcal{D}$ (בלתי תלוי ב- q). לכן בתוחלת על פני x_i מה שמוחזר כאן זה $q(\mathcal{D})$		כאן הפלט הוא הפעלה של q על איבר אקראי מהמדגם S. לכן בתוחלת על פני i מה שמוחזר כאן זה הממוצע האמפירי, כלומר $q(S)$

<u>טענה ב:</u> יהי B אלגוריתם הפועל על T מדגמים $\vec{S} = (S_1, \dots, S_T) \in (X^n)^T$ ומחזיר שאילתא $q: X \to \{0,1\}$ וגם מספר $1 \le t \le T$. אם B הוא (ε, δ)-DP-יציב אזי

$$\mathop{\mathbb{E}}_{\substack{\vec{S}\sim\mathcal{D}^n \\ (q,t)\leftarrow W(S)}} [q(S_t)] \approx \mathop{\mathbb{E}}_{\substack{\vec{S}\sim\mathcal{D}^n \\ (q,t)\leftarrow W(S)}} [q(\mathcal{D})]$$

ההוכחה של טענה ב דומה להוכחה של טענה א (אבל צריך לעקוב אחרי כל T המדגמים...)

<u>טענה ג (=משפט 2):</u> יהי W אלגוריתם (ε, δ)-*DP*-יציב הפולט שאילתא $q: X \to \{0,1\}$.
יהי $S \sim \mathcal{D}^n$ ונסמן $q \leftarrow W(S)$. אזי <u>בהסתברות גבוהה</u> מתקיים
$q(S) \approx q(\mathcal{D})$

<u>הסבר:</u> נניח בשלילה שבהסתברות מסוימת β מתקיים שאלגוריתם $W(S)$ פולט q כך ש- $|q(S) - q(\mathcal{D})|$ גדול.
נשתמש ב- W כדי לבנות אלגוריתם יציב B שסותר את טענה ב:

<u>אלגוריתם B:</u> קלט $S_1, \dots, S_T$ כאשר כל $S_t \sim \mathcal{D}^n$, עבור $T \approx \frac{1}{\beta}$.

(1) לכל $1 \le t \le T$ הרץ $q_t \leftarrow W(S_t)$

* לפי ההנחה בשלילה על W ולפי בחירת T, בהסתברות גבוהה קיים אינדקס t כך ש- $|q_t(S_t) - q_t(\mathcal{D})|$ גדול.

(2) בחר t כנ"ל בעזרת אלגוריתם NoisyAgrmax והחזר (q_t, t).

* בהסתברות גבוהה נקבל ש- $|q_t(S_t) - q_t(\mathcal{D})|$ הוא גדול וזה יסתור את החסם בתוחלת של טענה ב.

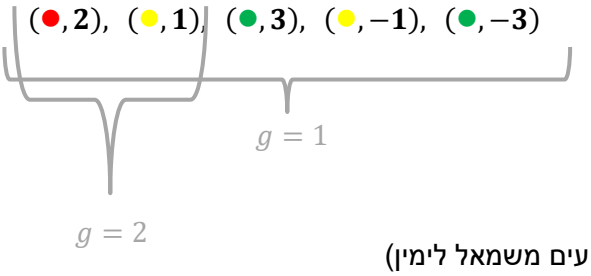
אפשר להראות שאלגוריתם B הוא *DP*-יציב (עם בערך אותם פרמטרים כמו W) ואפשר להראות שבהסתברות גבוהה (וגם בתוחלת) הערך האמפירי של השאילתא שהוא מחזיר רחוק מאוד מהערך שלה על פני ההתפלגות.
זה סותר את טענה ב ולכן מפריך את ההנחה בשלילה שקיים W כזה.

Adaptive Streaming

<u>הגדרה 1 (מודל הסטרימינג הלא-אדפטיבי):</u>

- A stream of length $\boldsymbol{m}$ over domain $[\boldsymbol{n}]$ is a sequence of updates $\big((\boldsymbol{u_1}, \boldsymbol{\Delta_1}), \dots, (\boldsymbol{u_m}, \boldsymbol{\Delta_m})\big)$
- Here $\boldsymbol{u_i} \in [\boldsymbol{n}]$ is the $\boldsymbol{i}$th item and $\boldsymbol{\Delta_i} \in \mathbb{Z}$ is its weight
- Let $\boldsymbol{g}: ([\boldsymbol{n}] \times \mathbb{Z})^* \to \mathbb{R}$ be a function
- At time $\boldsymbol{i}$ we obtain $(\boldsymbol{u_i}, \boldsymbol{\Delta_i})$ and need to output $\boldsymbol{z_i} \in (\boldsymbol{1} \pm \boldsymbol{\alpha}) \cdot \boldsymbol{g}\big((\boldsymbol{u_1}, \boldsymbol{\Delta_1}), \dots, (\boldsymbol{u_i}, \boldsymbol{\Delta_i})\big)$
- **Requirement: Sublinear space (we assume $\boldsymbol{n} \gg \boldsymbol{m}$)**

<u>דוגמה פשוטה:</u> נניח שהפונקציה g סופרת את מספר האיברים השונים ב stream. אזי,



(כאן האיברים ב stream מגיעים משמאל לימין)

<u>דוגמה יותר רצינית:</u>

- *Every item in the stream is a pair* $(\boldsymbol{u_i}, \boldsymbol{\Delta_i})$ *where* $\boldsymbol{u_i} \in \mathbb{R}^{\boldsymbol{n}}$ *is a standard basis vector and* $\boldsymbol{\Delta_i} \in \mathbb{R}$ *is its weight*
- *At every time step* $\boldsymbol{i}$*, the goal is to estimate* $\left\|\boldsymbol{f}^{(\boldsymbol{i})}\right\|_{\boldsymbol{2}}^{\boldsymbol{2}}$ *for* $\boldsymbol{f}^{(\boldsymbol{i})} = \boldsymbol{\Delta_1} \cdot \boldsymbol{u_1} + \cdots + \boldsymbol{\Delta_i} \cdot \boldsymbol{u_i}$

<u>האלגוריתם:</u>

1. *Let* $\boldsymbol{A}$ *be* $\boldsymbol{t}\boldsymbol{x}\boldsymbol{n}$ *matrix with entries uniformly in* $\{\pm\boldsymbol{1}\}$
2. *Initiate* $\boldsymbol{y} = \vec{\boldsymbol{0}} \in \mathbb{R}^{\boldsymbol{t}}$
3. *For* $\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{m}$ *do:*
   - *Obtain the next update vector* $\boldsymbol{v_i} = \boldsymbol{\Delta_i} \cdot \boldsymbol{u_i}$
   - *Let* $\boldsymbol{y} \leftarrow \boldsymbol{y} + \boldsymbol{A} \cdot \boldsymbol{v_i}$
   - *Output estimation* $\boldsymbol{z_i} = \frac{\boldsymbol{1}}{\boldsymbol{t}} \cdot \|\boldsymbol{y}\|_{\boldsymbol{2}}^{\boldsymbol{2}}$

<u>הניתוח:</u>

נסמן ב- a_ℓ את השורה ה-ℓ במטריצה A.

נשים לב שמתקיים:

$$z_i = \frac{1}{t} \cdot \|A \cdot v_1 + \cdots + A \cdot v_i\|_2^2 = \frac{1}{t} \cdot \left\|A \cdot f^{(i)}\right\|_2^2 = \frac{\left(a_1 \cdot f^{(i)}\right)^2 + \cdots + \left(a_t \cdot f^{(i)}\right)^2}{t}$$

כעת, לכל קביעה של $f \in \mathbb{R}^n$ ולכל $\ell \in [t]$ מתקיים

$$\mathbb{E}[a_\ell \cdot f] = \mathbb{E}\left[\sum_{j \in [n]} a_{\ell,j} \cdot f_j\right] = \sum_{j \in [n]} \mathbb{E}\left[a_{\ell,j}\right] \cdot f_j = 0$$

ובנוסף,

$$\mathbb{E}[(a_\ell \cdot f)^2] = \mathbb{E}\left[\left(\sum_{j\in[n]} a_{\ell,j}\cdot f_j\right)^2\right] \underset{(pairwise)}{=} \sum_{j\in[n]} \mathbb{E}\left[\left(a_{\ell,j}\cdot f_j\right)^2\right] = \sum_{j\in[n]} \mathbb{E}\left[a_{\ell,j}^2\right]\cdot f_j^2 = \sum_{j\in[n]} f_j^2 = \|f\|_2^2$$

כאשר המעבר המסומן באדום נכון כי עבור זוגות $j \neq j'$ נקבל שמתקיים $\mathbb{E}[a_{\ell,j}\cdot a_{\ell,j'}] = \mathbb{E}[a_{\ell,j}]\cdot\mathbb{E}[a_{\ell,j'}] = 0$ בגלל אי-תלות.

כלומר אנחנו רואים שבתוחלת, $(a_\ell \cdot f)^2$ נותן לנו בדיוק את מה שאנחנו רוצים. אבל מהי השונות? נזכור שלפי הגדרה:

$$\mathrm{Var}[(a_\ell \cdot f)^2] = \mathbb{E}[(a_\ell \cdot f)^4] - (\mathbb{E}[(a_\ell \cdot f)^2])^2$$

אז נחשב כ"א משני הביטויים האלה:

$$\mathbb{E}[(a_\ell \cdot f)^4] = \mathbb{E}\left[\left(\sum_{j\in[n]} a_{\ell,j}\cdot f_j\right)^4\right] \underset{(4wise)}{=} \sum_{j\in[n]} f_j^4 + 6\sum_{j\neq j'} f_j^2\cdot f_{j'}^2$$

$$(\mathbb{E}[(a_\ell \cdot f)^2])^2 = \left(\sum_{j\in[n]} f_j^2\right)^2 = \sum_{j\in[n]} f_j^4 + 2\sum_{j\neq j'} f_j^2\cdot f_{j'}^2$$

ולכן אנחנו מקבלים שמתקיים

$$\mathrm{Var}[(a_\ell \cdot f)^2] = 4\sum_{j\neq j'} f_j^2\cdot f_{j'}^2 \leq 2\left(\sum_{j\in[n]} f_j^2\right)^2 = 2\cdot\|f\|_2^4$$

וזה רק עבור שורה אחת. האלגוריתם שלנו מחזיר את המיצוע על פני t השורות. המיצוע הזה לא משנה את העובדה שבתוחלת אנחנו מחזירים את הדבר הנכון, אבל את השונות זה מקטין בצורה משמעותית. נניח שהשורות במטריצה הן בלתי תלויות. אזי:

$$\mathrm{Var}\left[\frac{\sum_{\ell=1}^t (a_\ell \cdot f)^2}{t}\right] = \frac{1}{t^2}\cdot\mathrm{Var}\left[\sum_{\ell=1}^t (a_\ell \cdot f)^2\right] \underset{\substack{independent\\ rows}}{=} \frac{1}{t^2}\cdot\sum_{\ell=1}^t \mathrm{Var}[(a_\ell \cdot f)^2] \leq \frac{2\cdot\|f\|_2^4}{t}$$

לכן, לפי אי-שוויון צ'בישב, לכל $\alpha > 0$ מתקיים:

$$\Pr\left[\left|\frac{\sum_{\ell=1}^t (a_\ell \cdot f)^2}{t} - \|f\|_2^2\right| > \alpha\cdot\|f\|_2^2\right] \leq \frac{\mathrm{Var}\left[\frac{\sum_{\ell=1}^t (a_\ell \cdot f)^2}{t}\right]}{\alpha^2\cdot\|f\|_2^4} \leq \frac{2}{t\cdot\alpha^2}$$

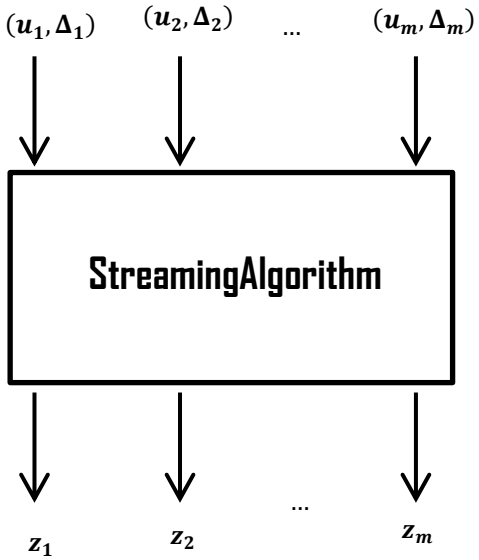
נבחר $t = \frac{2}{\alpha^2\beta}$ ונקבל

$$\Pr\left[\left|\frac{\sum_{\ell=1}^t (a_\ell \cdot f)^2}{t} - \|f\|_2^2\right| > \alpha\cdot\|f\|_2^2\right] \leq \beta$$

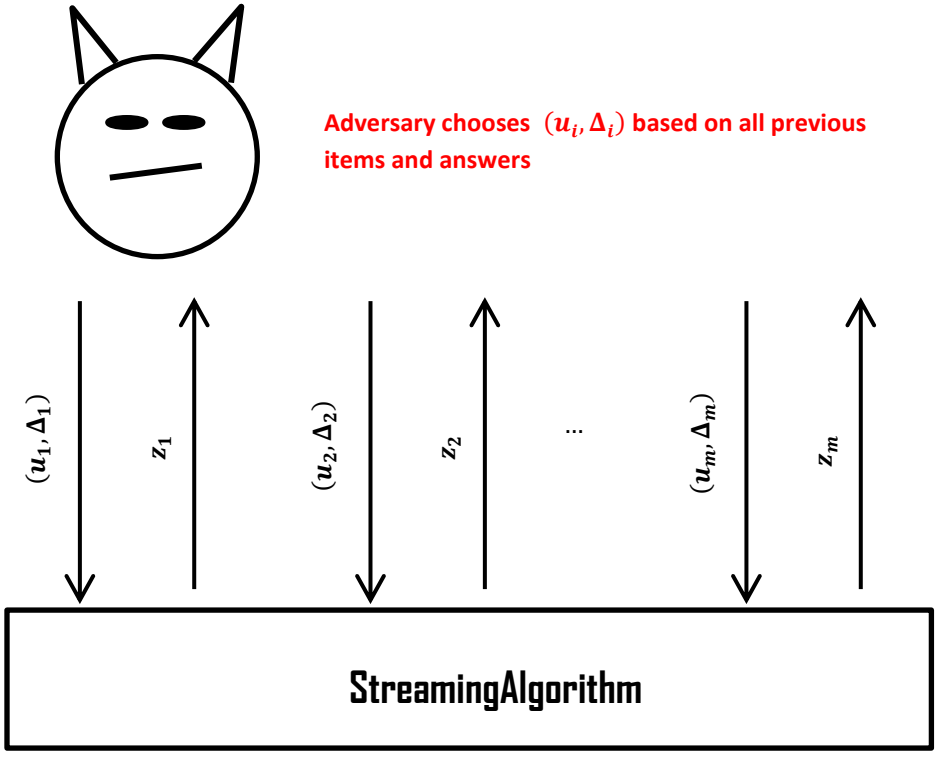
כלומר, כל אחת מהתשובות שהאלגוריתם מחזיר מדוייקת עד כדי טעות כפלית של $(1\pm\alpha)$ בהסתברות לפחות $1-\beta$.

שימו לב: בשביל הניתוח היינו צריכים להניח שהווקטור f <u>נקבע</u> מראש. באופן כללי, במודל ה streaming הלא-אדפטיבי אנחנו מניחים שכל ה stream נקבע מראש (אבל האלגוריתם מקבל את האיברים אחד אחד). בציור:

$$\big((u_1,\Delta_1),\ldots,(u_m,\Delta_m)\big) = \textit{fixed stream (unknown to the algorithm)}$$



במודל האדפטיבי לעומת זאת, האיברים ב stream לא נקבעים מראש. הם נקבעים תוך כדי ריצה על ידי "יריב" שרואה את התשובות שהאלגוריתם החזיר עד עכשיו. בציור:



<u>מה הקושי עכשיו?</u> האיברים ב stream נבחרים על סמך התשובות הקודמות של האלגוריתם ולכן יכולות להיות תלויות באקראיות הפנימית של האלגוריתם. התלויות האלה משבשות את ההבטחות של רוב אלגוריתמי ה streaming הקיימים.

<u>הגדרה (מודל ה streaming האדפטיבי):</u>

- Fix a function $\boldsymbol{g}$: $([\boldsymbol{n}] \times \mathbb{Z})^* \to \mathbb{R}$
- Two-player game between a (randomized) **StreamingAlgorithm** and an **Adversary**
- In the $\boldsymbol{i}$th round:
  1. The **Adversary** chooses an update $(\boldsymbol{u_i}, \boldsymbol{\Delta_i})$ for the stream, which can depend on all previous stream updates and outputs of **StreamingAlgorithm**
  2. The **StreamingAlgorithm** processes the new update and outputs its current response $\boldsymbol{z_i}$
- The goal of the **Adversary** is to make the **StreamingAlgorithm** output an incorrect response $\boldsymbol{z_i}$ at some point $\boldsymbol{i}$

Adversary for the AMS Sketch

Recall AMS sketch

- Random matrix $\boldsymbol{A} \in \{\pm\mathbf{1}\}^{\boldsymbol{t}\times\boldsymbol{n}}$
- After the $\boldsymbol{i}$th update, respond with $\frac{1}{t}\left\|\boldsymbol{A}\cdot\boldsymbol{f}^{(i)}\right\|_2^2 = \left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot\boldsymbol{f}^{(i)}\right\|_2^2$ where
$$\boldsymbol{f}^{(i)} = \boldsymbol{\Delta_1}\cdot\boldsymbol{u_1} + \cdots + \boldsymbol{\Delta_i}\cdot\boldsymbol{u_i}$$

The attack

- Set $\boldsymbol{w} \leftarrow \boldsymbol{C}\cdot\sqrt{\boldsymbol{t}}\cdot\boldsymbol{e_1}$
- For $\boldsymbol{i} = \mathbf{2}, \mathbf{3}, \ldots, \boldsymbol{m} = \boldsymbol{O(t)}$ do
  1. $\mathbf{old} \leftarrow \left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot\boldsymbol{w}\right\|_2^2$
  2. $\boldsymbol{w} \leftarrow \boldsymbol{w} + \boldsymbol{e_i}$
  3. $\mathbf{new} \leftarrow \left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot\boldsymbol{w}\right\|_2^2$
  4. If $\mathbf{new} > \mathbf{old}$ then $\boldsymbol{w} \leftarrow \boldsymbol{w} - \boldsymbol{e_i}$

Analysis

- At all times $\|\boldsymbol{w}\|_2^2 \geq \boldsymbol{C^2}\cdot\boldsymbol{t}$ by init
  $\Longrightarrow$ Suffices to show that $\left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot\boldsymbol{w}\right\|_2^2$ drops below $\boldsymbol{C^2/2}\cdot\boldsymbol{t}$
- $\mathbf{new_i} = \left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot(\boldsymbol{w}+\boldsymbol{e_i})\right\|_2^2 = \left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot\boldsymbol{w}\right\|_2^2 + \left\|\frac{1}{\sqrt{t}}\boldsymbol{A}\cdot\boldsymbol{e_i}\right\|_2^2 + \mathbf{2}\left\langle\frac{1}{\sqrt{t}}\boldsymbol{Aw},\ \frac{1}{\sqrt{t}}\boldsymbol{Ae_i}\right\rangle$
$$= \mathbf{old_i} + \mathbf{1} + \mathbf{2}\left\langle\frac{1}{\sqrt{t}}\boldsymbol{Aw},\ \frac{1}{\sqrt{t}}\boldsymbol{Ae_i}\right\rangle$$
- So, $\mathbf{new_i} - \mathbf{old_i} \approx \mathbf{2}\left\langle\frac{1}{\sqrt{t}}\boldsymbol{Aw},\ \frac{1}{\sqrt{t}}\boldsymbol{Ae_i}\right\rangle$

This inner product is symmetric, and is "negative enough" with constant probability, because Ae_i is the i^{th} column of A which is uniform on $\{\pm 1\}$ even conditioned on Aw (which is independent of the i^{th} column of A).