17.05.2023 ***ניתוח מידע אדפטיבי***

# הרצאה 6: יציבות

*Source: Lecture notes by Aaron Roth and Adam Smith*

מרצה: אורי שטמר

נניח שיש לנו $k$ מכניזמים $M_1, M_2, \dots, M_k$ שכל אחד מהם הוא $(\varepsilon, \delta)$-DP-יציב (כל אחד מהם מקבל דטהבייס + פרמטר). נחשוב על המכניזם $\vec{M}$ אשר מקבל מבצע קומפוזיציה אדפטיבית של כל המכניזמים האלה. כלומר:

$$\vec{M}(S) = M_k\left(S, M_{k-1}\left(S, M_{k-2}\left(S, M_{k-3}(S, \dots)\right)\right)\right)$$

**<u>משפט 1:</u>** יהיו $0 < \varepsilon, \delta \leq 1$. יהי $\vec{M}$ מכניזם המבצע $k$ הפעלות אדפטיביות של מכניזמים שכל אחד הוא $(\varepsilon, \delta)$-DP-יציב (ללא גישה נוספת לדטהבייס). אזי $\vec{M}$ הוא

$$\text{DP-יציב-}\left(2k \cdot \varepsilon^2 + \sqrt{2k \ln\left(\frac{1}{k\delta}\right)} \cdot \varepsilon \;\;,\;\; 2k\delta\right)$$

לשם פשטות, אנחנו נוכיח רק את הגרסה מוחלשת הבאה של משפט הקומפיזיציה

**<u>משפט 2:</u>** יהיו $0 < \varepsilon, \delta \leq 1$. יהי $\vec{M}$ מכניזם המבצע $k$ הפעלות אדפטיביות של מכניזמים שכל אחד הוא $(\varepsilon, \mathbf{0})$-DP-יציב (ללא גישה נוספת לדטהבייס). אזי לכל $\hat{\delta} > 0$ המכניזם $\vec{M}$ הוא $(\hat{\varepsilon}, \hat{\delta})$-DP-יציב עבור

$$\hat{\varepsilon} = 2k\varepsilon^2 + \sqrt{2k \cdot \ln\left(1/\hat{\delta}\right)}\,\varepsilon$$

**<u>הוכחה:</u>**

נקבע זוג מדגמים שכנים $S, S'$. עלינו להראות כי לכל מאורע $B$ מתקיים

$$\Pr\left[\vec{M}(S) \in B\right] \leq e^{\hat{\varepsilon}} \cdot \Pr[M(S') \in B] + \hat{\delta}$$

נסמן:
$Y_1, \dots, Y_k =$ משתנים מקריים המיצגים את הפלטים של $M_1, \dots, M_k$ במהלך הריצה של $\vec{M}(S)$.

$Y_1', \dots, Y_k' =$ משתנים מקריים המיצגים את הפלטים של $M_1, \dots, M_k$ במהלך הריצה של $\vec{M}(S')$.

בנוסף נסמן

$$V = (Y_1, \dots, Y_k) \quad , \quad V' = (Y_1', \dots, Y_k')$$

בסימונים האלה, מספיק להוכיח שלכל מאורע $B$ מתקיים

$$\Pr[V \in B] \leq e^{\hat{\varepsilon}} \cdot \Pr[V' \in B] + \hat{\delta}$$

**<u>טענת עזר 3:</u>** מספיק להוכיח כי

$$\Pr_{v \leftarrow V}\left[\;\ln\left(\frac{\Pr[V = v]}{\Pr[V' = v]}\right) > \hat{\varepsilon}\;\right] < \hat{\delta}$$

**הוכחת טענת עזר 3:**

נגדיר

$$W = \left\{ v \in \mathrm{Supp}(V) \ : \ \ln\left(\frac{\Pr[V=v]}{\Pr[V'=v]}\right) > \hat{\varepsilon} \right\}$$

לפי התנאים של טענת העזר, מתקיים

$$\Pr[V \in W] < \hat{\delta}$$

נקבע קבוצה $B$. כעת,

$$\Pr[V \in B] = \Pr[V \in B \cap W] + \Pr[V \in B \setminus W] \leq \hat{\delta} + e^{\hat{\varepsilon}} \cdot \Pr[V' \in B \setminus W] \leq \hat{\delta} + e^{\hat{\varepsilon}} \cdot \Pr[V' \in B]$$

מ.ש.ל. (טענת העזר)

נחזור להוכחת המשפט.

נקבע סדרה $v = (y_1, y_2, \dots, y_k)$ של פלטים עבור $M_1, \dots, M_k$ ונחשב

$$\ln\left(\frac{\Pr[V=v]}{\Pr[V'=v]}\right) = \ln\left(\prod_{i=1}^{k} \frac{\Pr[Y_i = y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}]}{\Pr[Y_i' = y_i | Y_1' = y_1, Y_2' = y_2, \dots, Y_{i-1}' = y_{i-1}]}\right)$$

$$= \sum_{i=1}^{k} \ln\left(\frac{\Pr[Y_i = y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}]}{\Pr[Y_i' = y_i | Y_1' = y_1, Y_2' = y_2, \dots, Y_{i-1}' = y_{i-1}]}\right)$$

$$\triangleq \sum_{i=1}^{k} c_i(y_1, \dots, y_{i-1}, y_i)$$

אז הצגנו את $\ln\left(\frac{\Pr[V=v]}{\Pr[V'=v]}\right)$ כסכום של $k$ משתנים מקריים $C_1, C_2, \dots, C_k$ כאשר כל $C_i$ כזה הוא משתנה מקרי המקבל את הערך $c_i(y_1, \dots, y_{i-1}, y_i)$.

עכשיו אנחנו רוצים להפעיל את אי-שוויון אזומה על המשתנים המקריים האלה.

הערה: האקראיות היא מעל הגרלת $V = (Y_1, \dots, Y_k)$, כלומר אנחנו רוצים לחסום את

$$\Pr_{v \leftarrow V}\left[ \ \ln\left(\frac{\Pr[V=v]}{\Pr[V'=v]}\right) > \hat{\varepsilon} \ \right] = \Pr_{v \leftarrow V}\left[ \ \sum_{i=1}^{k} c_i(y_1, \dots, y_{i-1}, y_i) > \hat{\varepsilon} \ \right]$$

כדי להפעיל את אזומה אנחנו צריכים לחסום אם $|C_i|$ ואנחנו צריכים לנתח תוחלת.
נזכור כי לפי הגדרה,

$$c_i(y_1, y_2, \dots, y_i) = \ln\left(\frac{\Pr[Y_i = y_i | Y_1 = y_1, Y_2 = y_2, \dots, Y_{i-1} = y_{i-1}]}{\Pr[Y_i' = y_i | Y_1' = y_1, Y_2' = y_2, \dots, Y_{i-1}' = y_{i-1}]}\right) = ((1))$$

כאשר $Y_j, Y_j'$ הם משתנים מקריים המייצגים את התשובה ה $j$ במהלך הריצה על $S, S'$ בהתאמה.

נזכור כי ברגע שקבענו את התשובות $y_1, y_2, \dots, y_{i-1}$ אז זה קובע את הפרמטר $p_i = p_i(y_1, y_2, \dots, y_{i-1})$ שמוזן להפעלה של המכניזם $M_i$.

לכן,

$$((1)) = \ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]}\right) = ((2))$$

כעת נזכור שהמכניזם $M_i$ הוא $(\varepsilon, 0)$-*DP*-יציב ולכן, לפי הגדרת *DP*-יציבות מתקיים

$$((2)) \in [-\varepsilon, \varepsilon]$$

בפרט, קיבלנו כי לכל $y_1, y_2, \dots, y_i$ מתקיים

$$|c_i(y_1, y_2, \dots, y_i)| \leq \varepsilon$$

ולכן

$$\Pr[|C_i| \leq \varepsilon] = 1$$

כעת נחשב את התוחלת המותנית:

$$\mathbb{E}\left[ C_i \,\middle|\, C_1 = c_1, \dots, C_{i-1} = c_{i-1} \right] = \mathbb{E}_{v \leftarrow V}\left[ c_i(y_1, y_2, \dots, y_i) \,\middle|\, C_1 = c_1, \dots, C_{i-1} = c_{i-1} \right]$$

$$\underbrace{=}_{\substack{\text{לפי החשבון שעשינו} \\ \text{במשוואות } ((2)),((1))}} \mathbb{E}_{v \leftarrow V}\left[ \ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]}\right) \,\middle|\, C_1 = c_1, \dots, C_{i-1} = c_{i-1} \right]$$

$$\underbrace{=}_{\substack{\text{משפט} \\ \text{התוחלת} \\ \text{השלמה}}} \sum_{y_1, \dots, y_{i-1}} \mathbb{E}_{v \leftarrow V}\left[ \ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]}\right) \,\middle|\, y_1, \dots, y_{i-1} \right] \cdot \Pr_{v \leftarrow V}\left[ y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right]$$

$$= \sum_{y_1, \dots, y_{i-1}} \mathbb{E}_{y_i \leftarrow Y_i}\left[ \ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]}\right) \,\middle|\, y_1, \dots, y_{i-1} \right] \cdot \Pr_{v \leftarrow V}\left[ y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right] = ((3))$$

כמו שאמרנו קודם, ברגע שקבענו את $y_1, \dots, y_{i-1}$ אז זה קובע את הפרמטר $p_i = p_i(y_1, y_2, \dots, y_{i-1})$ וקובע את ההתפלגות של $Y_i$ להיות הפלט של $M_i(S, p_i)$. לכן,

$$((3)) = \sum_{y_1, \dots, y_{i-1}} \mathbb{E}_{y_i \leftarrow M_i(S, p_i)}\left[ \ln\left(\frac{\Pr[M_i(S, p_i) = y_i]}{\Pr[M_i(S', p_i) = y_i]}\right) \right] \cdot \Pr_{v \leftarrow V}\left[ y_1, \dots, y_{i-1} \,\middle|\, c_1, \dots, c_{i-1} \right] = ((4))$$

**<u>טענת עזר 4 (נוכיח עוד רגע):</u>** לכל קביעה של $p_i$ מתקיים

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] \leq 2\cdot\varepsilon^2$$

בעזרת טענת עזר 4 נקבל

$$((4)) \leq \sum_{y_1,\dots,y_{i-1}} 2\cdot\varepsilon^2 \cdot \Pr_{v\leftarrow V}\left[y_1,\dots,y_{i-1} \,\middle|\, c_1,\dots,c_{i-1}\right] = 2\cdot\varepsilon^2$$

איפה אנחנו בהוכחה של משפט 2?

אנחנו רוצים להוכיח ש-

$$\Pr_{v\leftarrow V}\left[\sum_{i=1}^{k} c_i(y_1,\dots,y_{i-1},y_i) > \hat{\varepsilon}\right] \leq \hat{\delta}$$

והראינו כי לכל $i$ מתקיים $|C_i| \leq \varepsilon$ וגם $\mathbb{E}\left[C_i \,\middle|\, C_1=c_1,\dots,C_{i-1}=c_{i-1}\right] \leq 2\varepsilon^2$.

לכן, לפי אזומה, לכל $z>0$ מתקיים

$$\Pr_{v\leftarrow V}\left[\sum_{i=1}^{k} c_i(y_1,\dots,y_{i-1},y_i) > k2\varepsilon^2 + z\sqrt{k}\varepsilon\right] \leq \exp\left(-\frac{z^2}{2}\right)$$

עבור $\hat{\delta}>0$ נבחר $z=\sqrt{2\cdot\ln(1/\hat{\delta})}$ ונסמן ונקבל $\hat{\varepsilon} = 2k\varepsilon^2 + \sqrt{2k\cdot\ln(1/\hat{\delta})}\,\varepsilon$

$$\Pr_{v\leftarrow V}\left[\sum_{i=1}^{k} c_i(y_1,\dots,y_{i-1},y_i) > \hat{\varepsilon}\right] \leq \hat{\delta}$$

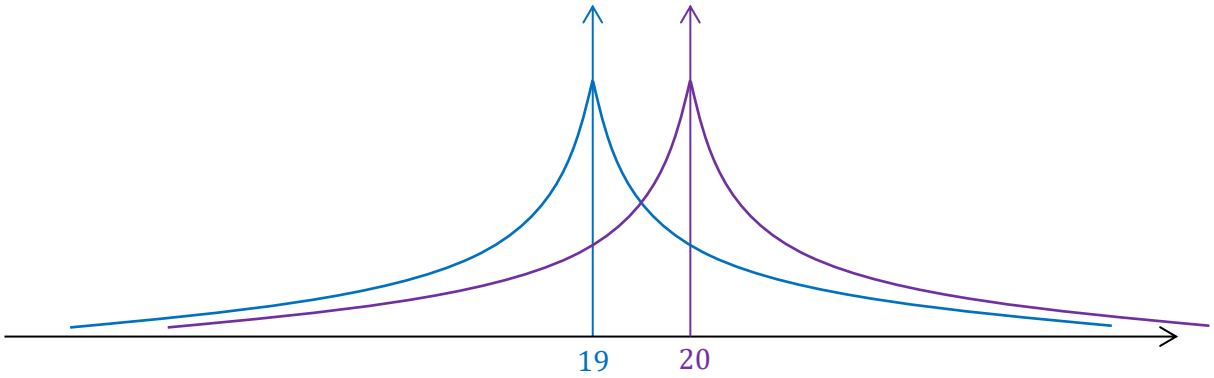
מ.ש.ל. (משפט 2)

נותר להוכיח את טענת עזר 4.

**<u>אינטואיציה לגבי קומפוזיציה:</u>**

למה שהיציבות תתדרדר רק כמו $\sqrt{k}$ ולא כמו $k$ ?

נניח שאנחנו מריצים $k$ פעמים את המכניזם הלפלאסי על קלט $S$ או על קלט שכן $S'$ עם אותה שאילתא עם ערך 19 על $S$ ועם ערך 20 על $S'$. בכל סיבוב נקבל דגימה מאחת מ- 2 ההתפלגויות הבאות:




- נניח שאומרים לי שהריצו את המ.לפלאס על אחד מ- $S, S'$ (נניח הערך האמיתי כל פעם הוא או 19 או 20), אבל אני לא יודע על מי מהם הריצו.
- אני רוצה לנסות לנחש האם הקלט הוא $S$ או $S'$
- פלט > *19.5* "גורם לי לחשוב" ש- $S'$ יותר סביר
- פלט < *19.5* "גורם לי לחשוב" ש- $S$ יותר סביר
- אבל לפעמים נקבל פלט < *19.5* גם על $S'$, מה שרק מטעה אותי בניסיון להבחין בין $S$ ל- $S'$ (כלומר לפעמים גם "נרוויח יציבות")

**עוד אינטואיציה:** אז נניח שאנחנו מריצים את המ.לפלאס על $S$ (כלומר דוגמים מההתפלגות הכחולה בציור הנ"ל). ברור שההסתברות שנקבל פלט קטן מ 19 היא בדיוק $1/2$. בנוסף, ההסתברות לקבל פלט בין 19 ל *19.5* היא בערך $\varepsilon$ (אפשר לחשב את זה עם אינטגרל על פונק' הצפיפות). כלומר, ההסתברות שנקבל פלט באיזור בו הגרף הכחול גבוה מהסגול היא בערך $\frac{1}{2}+\varepsilon$, ובהסתברות בערך $\frac{1}{2}-\varepsilon$ נקבל פלט "שמטעה אותנו" שעבורו הגרף הסגול גבוה יותר.
מה יהיה מספר הפעמים שנקבל פלט "שלא מטעה אותנו"? זה משתנה מקרי עם התפלגות בינומית עם הסתברות הצלחה $\frac{1}{2}+\varepsilon$. התוחלת של מספר ההצלחות מתוך $k$ דגימות היא $\frac{k}{2}+\varepsilon k$, כלומר בתוחלת יש באמת ייתרון מסויים לטובת "פלטים שלא מטעים אותנו". אבל סטיית התקן של משתנה מקרי כזה היא בערך $\frac{\sqrt{k}}{2}$. לכן אם $\varepsilon k \gg \frac{\sqrt{k}}{2}$ אז הייתרון שיש לנו "נבלע" בתוך סטיית התקן. כלומר, כל עוד $\varepsilon \ll \frac{1}{2\sqrt{k}}$ אז יהיה לנו מאוד קשה להבחין בין ההתפלגות הסגולה לכחולה.

**אינטואיציה לגבי טענת עזר 4:** בהוכחה הסתכלנו על המשתנה המקרי שמוגדר כך: נדגום נקודה $y$ לפי ההתפלגות הכחולה, ואז נסתכל על לוג היחס בין הגובה של הקו הכחול לגובה של הקו הסגול. אנחנו כבר יודעים שלוג היחס הזה תמיד חסום ע"י $\pm\varepsilon$. אבל דיי ברור שהתוחלת של המשתנה המקרי הזה חייבת להיות הרבה יותר קטנה מ $\varepsilon$ (כי לפעמים המספר הזה אפילו שלילי...). אפשר להראות שהתוחלת כאן היא בערך $\varepsilon^2$.

נעשה את החשבון עבור המקרה של שני הלפלסים. כלומר נראה למה התוחלת קטנה מ $\varepsilon^2 \approx$ עבור הציור עם שני הגרפים כאשר אנחנו מניחים שהדטהבייס האמיתי הוא $X$ עם ערך *19*:

- $\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) = \varepsilon$ אם דוגמים ערך $y_i \leq 19$
- $\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) = -\varepsilon$ אם דוגמים ערך $y_i \geq 20$

- באמצע הוא בין לבין.. נניח לרעתנו שמתקיים $\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)=\varepsilon$ גם עבור $19 \le y_i \le 20$
- נשים לב שההסתברות להיות באמצע היא קטנה:

$$\Pr_{y_i \leftarrow M_i(S,p_i)}[19 \le y_i \le 20] = \int_0^1 \frac{\varepsilon}{2}\cdot \exp(-\varepsilon \cdot y)\,dy = \left[\frac{\varepsilon}{2}\cdot\frac{\exp(-\varepsilon\cdot y)}{-\varepsilon}\right]_0^1$$
$$= \frac{1}{2}(1-e^{-\varepsilon}) \approx \frac{1}{2}\big(1-(1-\varepsilon)\big) = \frac{\varepsilon}{2}$$

- ולכן

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] \le \Pr[y_i \le 19]\cdot\varepsilon + \Pr[19 < y_i < 20]\cdot\varepsilon + \Pr[y_i > 20]\cdot(-\varepsilon)$$
$$\approx \frac{1}{2}\varepsilon + \frac{\varepsilon}{2}\varepsilon - \left(\frac{1}{2}-\frac{\varepsilon}{2}\right)\varepsilon = \varepsilon^2$$

**<u>הוכחה פורמלית של טענת עזר 4:</u>**

תחילה נשים לב שמתקיים

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] = \sum_{y_i}\Pr[M_i(S,p_i)=y_i]\cdot\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \ge 0$$

הסבר: זה נובע מאי-שיוויון *log-sum* שאומר שאם $a_1,\dots,a_n,b_1,\dots,b_n$ הם מספרים אי-שליליים אזי

$$\sum_i a_i\cdot\ln\left(\frac{a_i}{b_i}\right) \ge \left(\sum_i a_i\right)\cdot\ln\left(\frac{\sum_i a_i}{\sum_i b_i}\right)$$

באותו אופן, גם מתקיים $\mathbb{E}_{y_i \leftarrow M_i(S',p_i)}\left[\ln\left(\frac{\Pr[M_i(S',p_i)=y_i]}{\Pr[M_i(S,p_i)=y_i]}\right)\right] \ge 0$

כעת,

$$\mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] \le$$

$$\le \mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] + \mathbb{E}_{y_i \leftarrow M_i(S',p_i)}\left[\ln\left(\frac{\Pr[M_i(S',p_i)=y_i]}{\Pr[M_i(S,p_i)=y_i]}\right)\right]$$

$$= \mathbb{E}_{y_i \leftarrow M_i(S,p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right] - \mathbb{E}_{y_i \leftarrow M_i(S',p_i)}\left[\ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)\right]$$

$$= \sum_{y_i} \Pr_{M_i(S,p_i)}[y_i] \cdot \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) - \sum_{y_i} \Pr_{M_i(S',p_i)}[y_i] \cdot \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right)$$

$$= \sum_{y_i} \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \cdot \left( \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right)$$

$$\leq \sum_{y_i} \left| \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right| \cdot \left| \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right|$$

$$\leq \max_{y_i \in \mathrm{Supp}(Y_i)} \left| \ln\left(\frac{\Pr[M_i(S,p_i)=y_i]}{\Pr[M_i(S',p_i)=y_i]}\right) \right| \cdot \sum_{y_i} \left| \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right|$$

$$\leq \varepsilon \cdot \sum_{y_i} \left| \Pr_{M_i(S,p_i)}[y_i] - \Pr_{M_i(S',p_i)}[y_i] \right| \underbrace{=}_{\text{ראו תזכורת למטה}} \varepsilon \cdot 2 \cdot \mathrm{SD}\left( M_i(S,p_i) \ , \ M_i(S',p_i) \right)$$

$$= 2\varepsilon \cdot \max_{T \subseteq \mathrm{Supp}(Y_i)} \left| \Pr_{M_i(S,p_i)}[y_i \in T] - \Pr_{M_i(S',p_i)}[y_i \in T] \right| \underbrace{\leq}_{\substack{\text{נובע כי } M_i \\ \text{הוא } DP \text{ יציב} \\ \text{(הסבר נוסף למטה)}}} 2\varepsilon \cdot (1 - e^{-\varepsilon}) \leq 2\varepsilon^2$$

תזכורת: המרחק הסטטיסטי $\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2)$ הוא מדד למרחק בין 2 התפלגויות $\mathcal{D}_1, \mathcal{D}_2$ ויש לו 2 הגדרות שקולות:

$$\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \sum_{y} \left| \Pr_{x \sim \mathcal{D}_1}[x = y] - \Pr_{x \sim \mathcal{D}_2}[x = y] \right| = \max_{T} \left| \Pr_{x \sim \mathcal{D}_1}[x \in T] - \Pr_{x \sim \mathcal{D}_2}[x \in T] \right|$$

הסבר נוסף למעבר בשורה האחרונה: נסתכל על מאורע $T$ הממקסם את הביטוי ונניח בה"כ כי
$$\Pr_{M_i(S,p_i)}[y_i \in T] \geq \Pr_{M_i(S',p_i)}[y_i \in T]$$

אזי, מכיוון ש- $M_i$ הוא $(\varepsilon, 0)$-$DP$-יציב נקבל

$$\left| \Pr_{M_i(S,p_i)}[y_i \in T] - \Pr_{M_i(S',p_i)}[y_i \in T] \right| = \Pr_{M_i(S,p_i)}[y_i \in T] - \Pr_{M_i(S',p_i)}[y_i \in T] \leq$$

$$\leq \Pr_{M_i(S,p_i)}[y_i \in T] - e^{-\varepsilon} \cdot \Pr_{M_i(S,p_i)}[y_i \in T] = (1 - e^{-\varepsilon}) \cdot \Pr_{M_i(S,p_i)}[y_i \in T] \leq (1 - e^{-\varepsilon})$$

מ.ש.ל. (טענת עזר 4). זה מסיים את ההוכחה של משפט הקומפוזיציה (משפט 2).

**<u>מסקנה 5:</u>** יהיו $0 < \varepsilon, \delta < 1$ ויהי $k \in \mathbb{N}$. נסמן $\gamma = \frac{\varepsilon}{\sqrt{8k \cdot \ln(1/\delta)}}$. אזי המכניזם $\vec{M}_{\text{Lap}}^{1/(\gamma n)}$ אשר מוגדר להיות הקומפוזיציה של $k$ הפעלות של $M_{\text{Lap}}^{1/(\gamma n)}$ הוא $(\varepsilon, \delta)$-DP-יציב (עבור $k$ שאילתות).

הנקודה החשובה כאן היא שפרמטר היציבות מתדרדר רק כמו $\sqrt{k}$ ולא לניארית עם $k$. כלומר, בצענו $k$ הפעלות של מכניזמים שכל אחד מהם הוא (בערך) $\frac{\varepsilon}{\sqrt{k}}$ יציב, וקיבלנו מכניזם שהוא $\varepsilon$ יציב.

**<u>טענה 6:</u>** יהיו $0 < \beta, \varepsilon, \delta < 1$, יהי $k \in \mathbb{N}$, ונסמן $\gamma = \frac{\varepsilon}{\sqrt{8k \cdot \ln(1/\delta)}}$. המכניזם $\vec{M}_{\text{Lap}}^{1/(\gamma n)}$ הוא $(\alpha, \beta)$-מדוייק אמפירית עבור $k$ שאילתות בהינתן מדגם בגודל $n$, כאשר

$$\alpha = \frac{1}{\gamma n} \cdot \ln\left(\frac{k}{\beta}\right) = O\left(\frac{\sqrt{k}}{\varepsilon n} \cdot \sqrt{\ln\left(\frac{1}{\delta}\right)} \cdot \ln\left(\frac{k}{\beta}\right)\right)$$

**<u>הוכחה:</u>**

נזכור שעבור כל שאילתא $q_i$, המכניזם $\vec{M}_{\text{Lap}}^{1/(\gamma n)}$ מחזיר תשובה $q_i(S) + Y_i$ עבור $Y_i \sim \text{Lap}\left(\frac{1}{\gamma n}\right)$. לכן, כדי לחסום את השגיאה האמפירית מספיק לחסום את $\max_i |Y_i|$.

נקבע אינדקס $i$ מסויים. כפי שראינו בשיעור הקודם, עבור $\alpha \geq \frac{1}{\gamma n} \cdot \ln\left(\frac{k}{\beta}\right)$ מתקיים:

$$\Pr[|Y_i| > \alpha] = \exp(-\gamma n \alpha) \leq \frac{\beta}{k}$$

לכן, לפי חסם האיחוד

$$\Pr\left[\max_i |Y_i| > \alpha\right] \leq \beta$$

מ.ש.ל. (טענה 6)

אז מה קיבלנו? המכניזם $\vec{M}_{\text{Lap}}^{1/(\gamma n)}$ הוא $(\varepsilon, \delta)$-DP-יציב (עבור $k$ שאילתות) והוא גם $(\alpha, \beta)$-מדוייק אמפירית עבור

$$\alpha = O\left(\frac{\sqrt{k}}{\varepsilon n} \cdot \sqrt{\ln\left(\frac{1}{\delta}\right)} \cdot \ln\left(\frac{k}{\beta}\right)\right)$$

הנקודה החשובה כאן היא שהשגיאה מתנהגת רק כמו $\sqrt{\frac{k}{n}} \approx \frac{\sqrt{k}}{n}$ ולא כמו $\frac{k}{n}$. זה אומר שמספר השאילתות $k$ יכול להיות כמעט $n^2$ ועדיין נוכל להבטיח דיוק (אמפירי) לא טריוויאלי.

כדי להשלים את התמונה אנחנו צריכים להראות שאם מכניזם הוא גם *DP*-יציב וגם מדוייק-אמפירית אז הוא מבטיח דיוק-סטטיסטי. לצורך כך אנחנו צריכים להכיר מכניזם *DP*-יציב נוסף

## מכניזם NoisyArgmax

נניח שיש לנו אנליסט שקובע $T$ שאילתות ומעוניין לזהות אחת מהן עם ערך גבוה (בעזרת מכניזם שהוא *DP*-יציב). אפשרות אחת שעומדת בפני האנליסט היא להעריך את <u>כל</u> אחת מ- $T$ השאילתות האלה, נניח בעזרת מכניזם הלפלס, ולאחר מכן לבחור אחת שההערכה שלה הייתה גבוהה. אבל זה בזבזני כי האנליסט לא מעוניין ללמוד את הערך של כל השאילתות האלה...

**NoisyAgrmax**

1. Accept $T$ queries $q_1, q_2, \dots, q_T$
2. For every $1 \le t \le T$ let $\hat{a}_t = q_t(S) + Y_t$ where $Y_t \sim \text{Lap}\left(\frac{2}{\varepsilon n}\right)$
3. Return $t$ such that $\hat{a}_t \ge \hat{a}_j$ for every $j$

אנחנו מניחים שהשאילתות $q_1, \dots, q_T$ שניתנות לאלגוריתם NoisyAgrmax הם שאילתות עם רגישות $1/n$
(תזכורת: לפונקציה $q$ יש רגישות $1/n$ אם שינוי של איבר אחד מהמדגם יכול לשנות את הערך האמפירי של $q$ בלכל היותר $1/n$. למשל, שאילתא סטטיסטית היא שאילתא עם רגישות $1/n$)

**<u>משפט 7:</u>** מכניזם NoisyAgrmax הוא $(\varepsilon, 0)$-*DP*-יציב

**<u>הוכחה:</u>**

נקבע את השאילתות $q_1, \dots, q_T$, נקבע שני מדגמים שכנים $S, S'$, ונקבע איבר $b \in \{1,2, \dots, T\}$.

עלינו להראות שמתקיים

$$\Pr[\text{NoisyAgrmax}(S) = b] \le e^{\varepsilon} \cdot \Pr[\text{NoisyAgrmax}(S') = b]$$

נזכור כי לכל שאילתא $q_j$ אנחנו מוסיפים רעש $Y_t \sim \text{Lap}\left(\frac{2}{\varepsilon n}\right)$.
נסמן ב- $\vec{Y}_{-b}$ את <u>כל</u> הרעשים <u>חוץ</u> מהרעש $Y_b$.

מספיר להראות כי לכל קביעה $\vec{y}_{-b}$ של $\vec{Y}_{-b}$ מתקיים

$$\Pr[\text{NoisyAgrmax}(S) = b | \vec{y}_{-b}] \le e^{\varepsilon} \cdot \Pr[\text{NoisyAgrmax}(S') = b | \vec{y}_{-b}]$$

כי אז

$$\Pr[\text{NoisyAgrmax}(S) = b] = \sum_{\vec{y}_{-b}} \Pr[\vec{y}_{-b}] \cdot \Pr[\text{NoisyAgrmax}(S) = b | \vec{y}_{-b}] \le \cdots$$

אז נקבע וקטור רעש $\vec{y}_{-b}$. עכשיו מתקיים

$$\Pr[\text{NoisyAgrmax}(S) = b | \vec{y}_{-b}] = \Pr\left[q_b(S) + \text{Lap}\left(\frac{2}{\varepsilon n}\right) > \max_{j \ne b}\{q_j(S) + y_j\}\right]$$

$$= \Pr\left[q_b(S) + \frac{1}{n} + \text{Lap}\left(\frac{2}{\varepsilon n}\right) > \max_{j \ne b}\left\{q_j(S) + \frac{1}{n} + y_j\right\}\right]$$

$$\le \Pr\left[q_b(S) + \frac{1}{n} + \text{Lap}\left(\frac{2}{\varepsilon n}\right) > \max_{j \ne b}\{q_j(S') + y_j\}\right]$$

$$\le \Pr\left[q_b(S') + \frac{2}{n} + \text{Lap}\left(\frac{2}{\varepsilon n}\right) > \max_{j \ne b}\{q_j(S') + y_j\}\right]$$

$$= \Pr\left[\text{Lap}\left(\frac{2}{n}, \frac{2}{\varepsilon n}\right) > \max_{j \ne b}\{q_j(S') + y_j\} - q_b(S')\right] = ((1))$$

נזכור שפונקציות הצפיפות $h_{\mu,\lambda}$ ו- $h_{0,\lambda}$ של $\mathrm{Lap}(0,\lambda)$ ושל $\mathrm{Lap}(\mu,\lambda)$, בהתאמה, מקיימות

$$e^{-\mu/\lambda} \le \frac{h_{0,\lambda}(x)}{h_{\mu,\lambda}(x)} \le e^{\mu/\lambda}$$

ולכן

$$((1)) \le e^{\varepsilon} \cdot \Pr\left[\mathrm{Lap}\left(\frac{2}{\varepsilon n}\right) > \max_{j\neq b}\{q_j(S') + y_j\} - q_b(S')\right]$$

$$= e^{\varepsilon} \cdot \Pr\left[q_b(S') + \mathrm{Lap}\left(\frac{2}{\varepsilon n}\right) > \max_{j\neq b}\{q_j(S') + y_j\}\right] = e^{\varepsilon} \cdot \Pr[\mathrm{NoisyAgrmax}(S') = b|\vec{y}_{-b}]$$

**<u>טענה 8:</u>** אלגוריתם $NoisyAgrmax$ מחזיר אינדקס $t$ כך שבהסתברות לפחות $1-\beta$ מתקיים

$$q_t(S) \ge \max_j\{q_j(S)\} - O\left(\frac{1}{\varepsilon n} \cdot \ln\frac{T}{\beta}\right)$$

(ההוכחה של טענה 8 דומה להוכחה של טענה 6)

## DP-יציבות + דיוק-אמפירי גוררים דיוק-סטטיסטי

**<u>משפט 9:</u>** יהי $M$ מכניזם אשר בהינתן מדגם בגודל $n = \Omega\left(\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right)$ עונה על $k$ שאילתות סטטיסטיות אדפטיביות כך ש-

א. $M$ הוא $(\varepsilon,\delta)$-DP-יציב
ב. $M$ הוא $(\alpha,\beta)$-מדוייק-אמפירית

אזי $M$ הוא $\left(\alpha+10\varepsilon,\ \beta+\frac{k\delta}{\varepsilon}\right)$-מדוייק-סטטיסטית.

**<u>מסקנה (משפט 9 + מכניזם הלפלס):</u>**
קיים מכניזם <u>יעיל חישובית</u> אשר בהינתן מדגם בגודל $n \gtrsim \frac{\sqrt{k}}{\alpha^2}$ עונה על $k$ שאילתות אדפטיביות עם דיוק סטטיסטי $\alpha$.

בדומה למה שעשינו עם הנושא של דחיסת-טרנסקריפט, כדי להוכיח את משפט 9 אנחנו נוכיח את המשפט הבא:

**<u>משפט 10:</u>** יהי $M$ מכניזם $(\varepsilon,\delta)$-DP-יציב עבור $k$ שאילתות סטטיסטיות הפועל על מדגם בגודל $n = \Omega\left(\frac{1}{\varepsilon^2}\log\frac{1}{\delta}\right)$.
אזי לכל אנליסט $A$, לכל התפלגות $\mathcal{D}$, ולכל $1 \le i \le k$ מתקיים

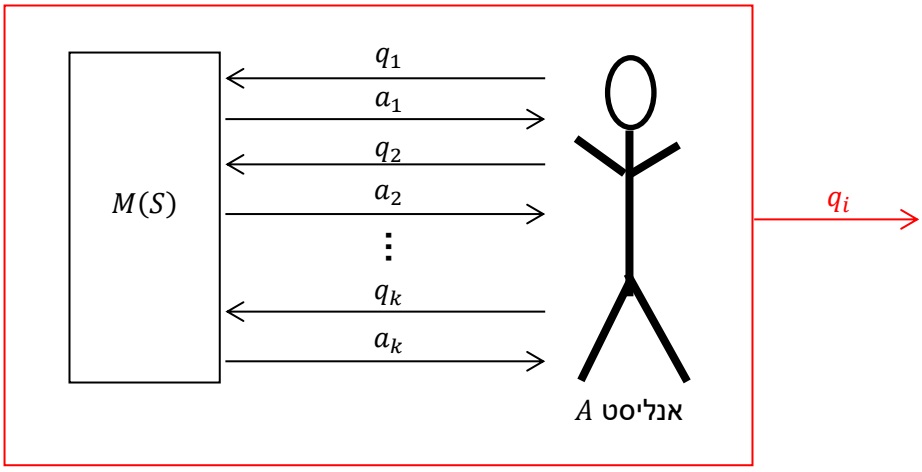
$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}\left[|q_i(S) - q_i(\mathcal{D})| > 10\varepsilon\right] < \frac{\delta}{\varepsilon}$$

בדומה למה שראינו בנושא של דחיסת-טרנסקריפט, משפט 9 נובע ממשפט 10 ע"י שימוש בחסם האיחוד ובאי-שוויון המשולש.

**<u>לקראת ההוכחה של משפט 10:</u>**

נקבע אנליסט $A$, נקבע התפלגות $\mathcal{D}$ ונקבע $1 \leq i \leq k$. נסמן ב- $MA_i$ את המכניזם אשר בהינתן קלט $S$ מסמלץ את האינטראקציה בין $A$ ל- $M$ ובסיום האינטראקציה פולט את <u>השאילתא</u> ה-$i$-ית $q_i$ שהאנליסט שאל במהלך הריצה.

בציור:




$MA_i(S)$

זהו מכניזם אשר מקבל כקלט מדגם $S$ ופולט שאילתא סטטיסטית $q_i$. בנוסף, מכיוון ש- $M$ הוא DP-יציב, אנחנו יודעים שכל הטרנסקריפט בין $A$ ל- $M$ "לא רגיש" לשינויי של דגימה בודדת מהמדגם $S$. בפרט, גם ההתפלגות מעל $q_i$ לא רגישה לשינויי של דגימה בודדת מהמדגם $S$. כלומר, $MA_i$ הוא בעצמו DP-יציב. נח לנו לחשוב עליו כי אין שום אנליסט שמדבר איתו. הוא פשוט לוקח מדגם ואז פולט שאילתא.

כעת, במקום להוכיח את משפט 10, נוכל להוכיח את המשפט הבא:

**<u>משפט 11:</u>** יהי $W$ מכניזם $(\varepsilon, \delta)$-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

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

נשים לב שמשפט 11 הוא פשוט יותר ממשפט 10: אין כאן מישהו ששואל שאילתות ואין כאן סיבובים. רק מכניזם $W$ שלוקח מדגם ומחזיר שאילתא. בנוסף, נשים לב שמשפט 10 נובע ממשפט 11, כי כמו שאמרנו, $MA_i$ עומד בדרישות של משפט 11 ולכן התוצאה מתקיימת עבור השאילתא ה-i-ית.