ניתוח מידע אדפטיבי 28.06.2023

הרצאה 11: עוד תוצאות שליליות + מעינה לשאילתות ע"י דגימה

Source: Lecture notes by Aaron Roth and Adam Smith

מרצה: אורי שטמר

נראה איך להשתמש בתוצאות אי-האפשרות מהשיעור שעבר גם כדי לקבל תוצאה שלילית עבור adaptive streaming. ההבדל מתוצאת אי-האפשרות שראינו בשיעורים 8-9 היא שעכשיו נקבל תוצאה שלילית עבור בעיית סטרימינג שמודגרת ע"י פונקציה ממשית.

התוכנית: אנחנו נראה בעיית סטרימינג שאפשר לפתור אותה בקלות במודל הלא-אדפטיבי, אבל אי אפשר לפתור אותה בצורה טובה במודל האדפטיבי. ספציפית, נראה שאם אפשר לפתור אותה בצורה טובה במודל האדפטיבי אז זה גורר קיום של מכניזם יעיל שעונה על שאילתות אדפטיביות בצורה טובה מידי, מה שאנחנו יודעים שלא יכול להיות.

בעיית SADA:

- כל עדכון s_i הוא או נקודה $s_i = p_i \in X$ או פונקציה $s_i = h_i: X \to \{0,1\}$.
- המטרה: בכל שלב i, אחרי שנקבל את העדכון s_i, עלינו להחזיר קירוב לממוצע של הפונקציה האחרונה שקיבלנו על המולטיסט שמכיל את כל הנקודות שקיבלנו בסטרים.

בצורה פורמלית,

- יהי $S = (s_1, s_2, \dots, s_i)$ פרפיקס של הסטרים
- נסמן ב- f_S את הפונקציה האחרונה שמופיעה ב- S
- נסמן ב- $P_S = \{s_j : j \in [i] \,\&\, s_j \in X\}$ את המולטיסט שמכילה את כל הנקודות שמופיעות ב- S.
- נגדיר את פונקציית המטרה באופן הבא:

$$g(S) = \frac{1}{|P_S|} \sum_{x \in P_S} f_S(x)$$

באופן אינטואיטיבי, בעולם הלא-אדפטיבי קל לפתור את בעיית ה SADA כי נוכל בכל שלב להעריך את הממוצע של f_S בעזרת מדגם אקראי קטן מתוך P_S. כלומר, אנחנו לא צריכים לזכור את כל המולטיסט P_S (שהוא יכול להיותר מאוד גדול) כדי להעריך את פונקציית המטרה. מספיק לנו לזכור מדגם קטן ממנו.

משפט (לא פורמלי): קיים אלג' (יעיל חישובית) לבעיית ה SADA במודל הלא-אדפטיבי השתמש בזיכרון $O(\text{polylog}(m))$

סקיצת הוכחה:

בעצם כל מה שאנחנו צריכים לעשות זה להבין איך אלגוריתם סטרימינג יכול לתחזק מדגם אקראי מתוך הנקודות שהופיעו בסטרים. ברגע שנדע לעשות את זה אז נוכל לענות על השאילתות בעזרת המדגם ונקבל דיוק (במודל הלא-אדפטיבי) בעזרת צ'רנוף.

יתרה מכך, מספיק להראות אלג' סטרימינג שמתחזק דגימה בודדת אקראית לאורך הסטרים, כי אם נדע לעשות את זה אז פשוט נריץ אותו במקביל y פעמים כדי לקבל מדגם בגודל y.

Reservoir Sampling [Vitter, 1985]:

1. Init $s = \perp$
2. When the next item u_i is read, toss a biased coin and with probability $1/i$ let $s = u_i$ (note that we need to maintain both the element s and a counter for i)
3. Output s

משפט: נניח שהאיברים בסטרים מגיעים מדומיין בגודל n, ספציפית $u_i \in [n]$ ונניח שאורך הסטרים הוא m. אזי האלגוריתם צורך זיכרון $O(\log(n)+\log(m))$ והפלט הוא איבר יוניפורמי מהסטרים, כלומר לכל $i \in [m]$ מתקיים ש- $\Pr[s = u_i] = \frac{1}{m}$.

הוכחה: ההוכחה באינדוקציה על $i \in [m]$

נניח כי המשפט מתקיים לאחר j הצעדים הראשונים. כלומר, לאחר j הצעדים הראשונים, לכל $i \in [j]$ מתקיים ש-

$$\Pr[s = u_i] = \frac{1}{j}$$

נראה שזה נכון גם לאחר הצעד הבא:

- לפי הבניה, עבור u_{j+1} מתקיים $\Pr\left[s = u_{j+1}\right] = \frac{1}{j+1}$
- ועבור $i \in [j]$ מתקיים:

$$\Pr[s = u_i] = \Pr\left[\begin{matrix} s = u_i \\ \text{after } j \text{ steps}\end{matrix}\right] \cdot \Pr[u_i \text{ survives}] = \frac{1}{j} \cdot \left(1 - \frac{1}{j+1}\right) = \frac{1}{j} \cdot \frac{j}{j+1} = \frac{1}{j+1}$$

מ.ש.ל.

לעומת זאת, כפי שנראה עכשיו, במודל האדפטיבי זאת בעיה קשה:

משפט (לא פורמלי): כל אלג' (יעיל חישובית) לבעיית ה SADA במודל האדפטיבי חייב להשתמש בזיכרון $\Omega(\text{poly}(m))$

הקושי נובע מכך שאין לנו את היכולת לענות בצורה יעילה על הרבה שאילתות אדפטיביות בעזרת מדגם קטן. כפי שראינו, אם אנחנו רוצים לענות על הרבה שאילתות אדפטיביות אז המדגם שלנו חייב להיות גדול ולכן באופן אינטואיטיבי גם הזיכרון שלנו יהיה צריך להיות גדול בשביל שנוכל לפתור את בעיית ה SADA. פורמלית נראה זאת ע"י רדוקציה: אם קיים אלג' סטרימינג אדפטיבי שפותר את בעיית ה SADA "טוב מידי" אז נוכל לבנות ממנו מכניזם שעונה על שאילתות "בצורה טובה מדי", מה שלא יתכן.

אז נניח שיש לנו אלג' $\mathcal{A}$ שפותר את בעיית ה SADA בעולם האדפטיבי. על מנת לפשט את ההוכחה, נניח כי $\mathcal{A}$ הוא דטרמניסטי. נבנה מ- $\mathcal{A}$ מכניזם $\mathcal{M}$ שעונה על שאילתות באופן הבא:

Algorithm $\boldsymbol{\mathcal{M}}$

Input: Dataset $\boldsymbol{P} = (\boldsymbol{p_1}, \dots, \boldsymbol{p_n}) \in \boldsymbol{X^n}$ containing $\boldsymbol{n}$ elements from the domain $\boldsymbol{X}$
Algorithm used: An adversarially robust streaming algorithm $\boldsymbol{\mathcal{A}}$ for the SADA problem

1. For $\boldsymbol{p} \in \boldsymbol{P}$ feed the update $\boldsymbol{p}$ to $\boldsymbol{\mathcal{A}}$
2. For $\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \dots, \boldsymbol{k}$:
   a) Obtain next function $\boldsymbol{f_i}: \boldsymbol{X} \to \{\boldsymbol{0}, \boldsymbol{1}\}$
   b) Feed the update $\boldsymbol{f_i}$ to $\boldsymbol{\mathcal{A}}$ and obtain an answer $\boldsymbol{z_i}$
   c) Output $\boldsymbol{z_i}$

<u>ננסה לנתח את האלג' הזה:</u>

- נקבע התפלגות $\mathfrak{D}$ מעל X ונחשוב על הריצה של $\mathcal{M}$ על מדגם P שמכיל דגימות ב"ת מ- $\mathfrak{D}$.
- לפי הבניה, אם אלג' $\mathcal{A}$ מצליח לפתור את בעיית ה *SADA* אז בכל שלב i מתקיים:

$$z_i \approx \frac{1}{|P|}\sum_{p\in P} f_i(p)$$

- זה טוב, אבל זה לא בדיוק מה שאנחנו רוצים... אנחנו רוצים לדאוג שהתשובות של $\mathcal{M}$ יהיו מדוייקות ביחס להתפלגות $\mathfrak{D}$ ולאו דוקא ביחס לממוצע האמפירי על P!

<u>דיון:</u> איך נשלים את ההוכחה?

מענה על שאילתות ע"י subsampling

תזכורת: בהרצאה 6 ראינו את המשפטים הבאים:

<u>משפט:</u> יהי $M: X^* \to F$ מכניזם. אם לכל $S, S' \in X^*$ שכנים מתקיים

$$\Pr_{y\leftarrow M(S)}\left[\ \ln\left(\frac{\Pr[M(S)=y]}{\Pr[M(S')=y]}\right) > \varepsilon\ \right] < \delta$$

אז M מקיים (ε,δ)-*DP*-יציבות, כלומר לכל $S, S' \in X^*$ שכנים ולכל $B \subseteq F$ מתקיים

$$\Pr[M(S)\in B] \le e^{\varepsilon}\cdot \Pr[M(S')\in B] + \delta$$

למעשה גם הכיוון ההפוך של המשפט הזה נכון, כלומר השקילות הזאת היא אם ורק אם.

<u>משפט:</u> יהי $M: X^* \to F$ מכניזם המקיים מקיים $(\varepsilon, 0)$-*DP*-יציבות. אזי לכל $S, S' \in X^*$ שכנים מתקיים

$$\mathbb{E}_{y\leftarrow M(S)}\left[\ \ln\left(\frac{\Pr[M(S)=y]}{\Pr[M(S')=y]}\right)\ \right] \lesssim \varepsilon^2$$

נחליש קצת את הגדרת היציבות שאנחנו עובדים איתה. נראה שהגרסה המוחלשת עדיין יהיה מספיק טובה כדי להבטיח הכללה אדפטיבית. בנוסף, מכיוון שהיא תהייה יותר חלשה, נוכל לתכנן אלג' שעומדים בהגדרה המוחלשת (אבל לא עומדים בהגדרה המלאה של DP-יציבות)

<u>סימון:</u>
עבור שני משתנים מקריים P,Q נסמן

$$D(P|Q)=\mathbb{E}_{y\leftarrow P}\left[\ln\left(\frac{\Pr[P=y]}{\Pr[Q=y]}\right)\right]$$

אז עבור מכניזם M שהוא $(\varepsilon,0)$-DP-יציב, לכל S,S' שכנים מתקיים $D\big(M(S)\big|M(S')\big)\lesssim\varepsilon^2$

<u>הגדרה:</u> מכניזם $\mathcal{A}:X^n\to F$ מקיים ε-*ALKL*-יציבות אם קיים מכניזם $\mathcal{A}':X^{n-1}\to F$ כך שלכל $S\in X^n$ מתקיים

$$\frac{1}{n}\sum_{i\in[n]}D\big(\mathcal{A}(S)\big|\mathcal{A}'(S_{-i})\big)\le\varepsilon$$

כאשר S_{-i} מתקבל מ S על ידי מחיקת האיבר ה-i.

נשים לב: אם מכניזם M מקיים ε-DP-יציבות אזי בפרט הוא מקיים $\approx\varepsilon^2$-ALKL-יציבות. הכיוון השני לא בהכרח נכון.

<u>המטרות הנוכחיות שלנו:</u>

- מטרה 1: להראות שיש אלגוריתמים "מועילים" שמקיימים ALKL-יציבות אבל לא DP-יציבות
- מטרה 2: להראות ש ALKL-יציבות עדיין מספיק כדי להבטיח הכללה אדפטיבית

מטרה 1 – answering queries by subsampling

נסתכל על שני המכניזמים הבאים

Mechanism $\mathcal{A}'$	Mechanism $\mathcal{A}$
Input: Dataset $S=(x_1,\dots,x_{n-1})\in X^{n-1}$ Given a query $q:X\to\{0,1\}$ w.p. $1/n^2$: output 0/1 w.p. ½ each otherwise sample $i\in[n-1]$ and return $q(x_i)$	Input: Dataset $S=(x_1,\dots,x_n)\in X^n$ Given a query $q:X\to\{0,1\}$ sample $i\in[n]$ and return $q(x_i)$

<u>משפט:</u> מכניזם $\mathcal{A}$ מקיים ε-ALKL-יציבות עבור $\varepsilon=O\left(\frac{1}{n^2}\right)$. כלומר, לכל קביעה של השאילתא q ולכל קביעה של הדטהבייס S מתקיים

$$\frac{1}{n}\sum_{i\in[n]}D\big(\mathcal{A}(S)\big|\mathcal{A}'(S_{-i})\big)\le\varepsilon$$

טענת קומפוזיציה ללא הוכחה: קמפוזיציה של ε מכניזמים, כ"א משמר ε-ALKL-יציבות, משמרת $k\varepsilon$-ALKL-יציבות. לכן, המשפט הנ"ל אומר לנו שנוכל להפעיל את אלג' הדגימה $\approx n^2$ פעמים לפני "שתגמר" לנו היציבות.

הוכחה:

נקבע $S = (x_1, \dots, x_n)$ ונקבע פרדיקט $q: X \to \{0,1\}$.

נסמן ב k את מספר ה-1-ים ב S. כלומר $k = \sum_{i=1}^{n} \mathbb{1}\{q(x_i) = 1\}$

נחלק לשלושה מקרים:

$$k \le \frac{n}{3} \quad , \quad k \in \left[\frac{n}{3}, \frac{2n}{3}\right] \quad , \quad k \ge \frac{2n}{3}$$

מקרה א: $\boldsymbol{k \in \left[\frac{n}{3}, \frac{2n}{3}\right]}$

לכל $i \in [n]$ מתקיים

$$\Pr[\mathcal{A}'(S_{-i}, q) = 0] \in \left(1 \pm O\left(\frac{1}{n}\right)\right) \cdot \Pr[\mathcal{A}(S, q) = 0]$$

וגם

$$\Pr[\mathcal{A}'(S_{-i}, q) = 1] \in \left(1 \pm O\left(\frac{1}{n}\right)\right) \cdot \Pr[\mathcal{A}(S, q) = 1]$$

כלומר במקרה זה אנחנו מקבלים בערך $\frac{1}{n}$-DP-יציבות ולכן בערך $\frac{1}{n^2}$-ALKL-יציבות.

מקרה ב: $\boldsymbol{k \le \frac{n}{3}}$

טענת עזר 1: נניח שיש לי 2 מ"מ Y, Z ואני מגדיר 2 מ"מ חדשים:

$$Y' = \begin{cases} Y & , \quad \text{w.p. } \alpha \\ 1 & , \quad \text{else} \end{cases} \quad , \quad Z' = \begin{cases} Z & , \quad \text{w.p. } \alpha \\ 1 & , \quad \text{else} \end{cases}$$

אזי

$$D(Y'|Z') = \alpha \cdot D(Y|Z)$$

הוכחה:

$$D(Y'|Z') = \mathbb{E}_{y \leftarrow Y'}\left[\ln\left(\frac{\Pr[Y' = y]}{\Pr[Z' = y]}\right)\right] = (1-\alpha) \cdot \ln\left(\frac{1-\alpha}{1-\alpha}\right) + \alpha \cdot \mathbb{E}_{y \leftarrow Y}\left[\ln\left(\frac{\alpha \cdot \Pr[Y = y]}{\alpha \cdot \Pr[Z = y]}\right)\right]$$

$$= \mathbb{E}_{y \leftarrow Y}\left[\ln\left(\frac{\Pr[Y = y]}{\Pr[Z = y]}\right)\right] = D(Y|Z)$$

טענת עזר 2:

יהי $p \in (0,1)$ ויהי $\Delta \in \left(-\frac{1}{10}, +\frac{1}{10}\right)$ כך ש- $p \ge 2|\Delta|$ וכך ש- $2p + \Delta \in (0,1)$. אז מתקיים

$$D\left(Ber(p) \,\middle|\, Ber(p + \Delta)\right) \le O\left(\frac{1}{p} \cdot \Delta^2\right)$$

הוכחה:

נגדיר

$$Y \sim Ber\left(\frac{p}{2p+\Delta}\right) \quad , \quad Z' \sim Ber\left(\frac{p+\Delta}{2p+\Delta}\right)$$

$$Y' \sim Ber(p) \quad , \quad Z' \sim Ber(p+\Delta)$$

ונשים לב שמתקיים

$$Y' = \begin{cases} Y & , \quad \text{w.p. } 2p+\Delta \\ 0 & , \quad \text{else} \end{cases} \qquad , \qquad Z' = \begin{cases} Z & , \quad \text{w.p. } 2p+\Delta \\ 0 & , \quad \text{else} \end{cases}$$

כי אז

$$\Pr[Y' = 1] = (2p+\Delta) \cdot \frac{p}{2p+\Delta} = p$$

$$\Pr[Z' = 1] = (2p+\Delta) \cdot \frac{p+\Delta}{2p+\Delta} = p+\Delta$$

ובנוסף, בדומה למקרה א,

$$\Pr[Y = 1] \in \left(1 \pm O\left(\frac{\Delta}{p}\right)\right) \cdot \Pr[Z = 1]$$

וגם

$$\Pr[Y = 0] \in \left(1 \pm O\left(\frac{\Delta}{p}\right)\right) \cdot \Pr[Z = 0]$$

כלומר במקרה זה אנחנו מקבלים בערך $\frac{\Delta}{p}$-DP-יציבות ולכן

$$D(Y|Z) \lesssim \frac{\Delta^2}{p^2}$$

כעת, לפי טענת עזר 1 מתקיים

$$D\big(Ber(p) \,\big|\, Ber(p+\Delta)\big) = D(Y'|Z') = (2p+\Delta) \cdot D(Y|Z) \approx \frac{1}{p} \cdot \Delta^2$$

מ.ש.ל. טענת עזר 2

נחזור לניתוח של מקרה ב:

data outcome	$\boldsymbol{S}$	$\boldsymbol{S_{i-1}}$ **when** $\boldsymbol{i \le k}$	$\boldsymbol{S_{i-1}}$ **when** $\boldsymbol{i > k}$
1	$\frac{k}{n}$	$\frac{k-1}{n-1} \cdot \left(1 - \frac{1}{n^2}\right) + \frac{1}{2n^2}$	$\frac{k}{n-1} \cdot \left(1 - \frac{1}{n^2}\right) + \frac{1}{2n^2}$
0	$\frac{n-k}{n}$	$\frac{n-k}{n-1} \cdot \left(1 - \frac{1}{n^2}\right) + \frac{1}{2n^2}$	$\frac{n-k-1}{n-1} \cdot \left(1 - \frac{1}{n^2}\right) + \frac{1}{2n^2}$

ולכן

$$\frac{1}{n}\sum_{i\in[n]} D\big(\mathcal{A}(S)\big|\mathcal{A}'(S_{-i})\big) =$$

$$= \frac{1}{n}\cdot\left[k\cdot D\left(Ber\left(\frac{k}{n}\right)\middle|Ber\left(\frac{k-1}{n-1}\cdot\left(1-\frac{1}{n^2}\right)+\frac{1}{2n^2}\right)\right)+(n-k)\cdot D\left(Ber\left(\frac{k}{n}\right)\middle|Ber\left(\frac{n-k-1}{n-1}\cdot\left(1-\frac{1}{n^2}\right)+\frac{1}{2n^2}\right)\right)\right]$$

$$\approx \frac{1}{n}\cdot\left[k\cdot D\left(Ber\left(\frac{k}{n}\right)\middle|Ber\left(\frac{k}{n}\pm\frac{n-k}{n^2}\right)\right)+(n-k)\cdot D\left(Ber\left(\frac{k}{n}\right)\middle|Ber\left(\frac{k}{n}\pm\frac{k}{n^2}\right)\right)\right]$$

$$\underbrace{\leq}_{\text{ט.ע. }2} \frac{1}{n}\cdot\left[k\cdot O\left(\frac{n}{k}\cdot\left(\frac{n-k}{n^2}\right)^2\right)+(n-k)\cdot O\left(\frac{n}{k}\cdot\left(\frac{k}{n^2}\right)^2\right)\right]$$

$$\approx \frac{1}{n^2}+\frac{1}{n^2}\approx\frac{1}{n^2}$$

הערה: הנחנו כאן ש- $k\geq 2$ כדי שיתקיים $\frac{k}{n}>2\cdot\frac{n-k}{n^2}$ כדי שנוכל להשתמש בטענת עזר 2. זה לא באמת משנה הרבה

מטרה 1 – להראות ש ALKL גורר הכללה אדפטיבית

<u>משפט:</u> יהי W מכניזם ε-ALKL-יציב אשר מקבל כקלט מדגם $S\in X^n$ ופולט שאילתא סטטיסטית q. אזי, לכל התפלגות $\mathcal{D}$ מעל X מתקיים

$$\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}\Big[q(S)-q(\mathcal{D})\Big]<2\sqrt{\varepsilon}$$

ברמת האינטואיציה, זה דומה למשהו שראינו עבור DP-יציבות: נסתכל על 2 הניסויים הבאים

$S=(x_1,\dots,x_n)\sim\mathcal{D}^n$ $i\in_R\{1,2,\dots,n\}$ $q\leftarrow W(S\setminus\{x_i\})$ Return $q(x_i)$	$\approx$ (בגלל יציבות)	$S=(x_1,\dots,x_n)\sim\mathcal{D}^n$ $i\in_R\{1,2,\dots,n\}$ $q\leftarrow W(S)$ 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 $q(S)$

הוכחה:

$$\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}\left[q(S)-q(\mathcal{D})\right]=\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}\left[\mathop{\mathbb{E}}_{i\sim[n]}[q(x_i)]\right]-\mathop{\mathbb{E}}_{\substack{S\sim\mathcal{D}^n\\ q\leftarrow W(S)}}\left[\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]$$

$$=\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S)}[q(x_i)]-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S)}\left[\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]$$

$$=\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S)}[q(x_i)]-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S)}\left[\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]$$
$$-\left\{\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S_{-i})}[q(x_i)]-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S_{-i})}[q(x_i)]\right\}$$

$$=\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S)}[q(x_i)]-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S)}\left[\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]$$
$$-\left\{\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S_{-i})}[q(x_i)]-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\ \mathop{\mathbb{E}}_{q\leftarrow W(S_{-i})}\ \mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right\}$$

$$=\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\left[\mathop{\mathbb{E}}_{q\leftarrow W(S)}[q(x_i)]-\mathop{\mathbb{E}}_{q\leftarrow W(S_{-i})}[q(x_i)]\right]$$
$$-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\left[\mathop{\mathbb{E}}_{q\leftarrow W(S)}\left[\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]-\mathop{\mathbb{E}}_{q\leftarrow W(S_{-i})}\left[\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]\right]$$

$$=\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\left[\sum_q(\Pr[W(S)=q]-\Pr[W(S_{-i})=q])\cdot q(x_i)\right]$$
$$-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\left[\sum_q(\Pr[W(S)=q]-\Pr[W(S_{-i})=q])\cdot\mathop{\mathbb{E}}_{x\sim\mathcal{D}}[q(x)]\right]$$

$$\leq\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\left[\sum_q(\Pr[W(S)=q]-\Pr[W(S_{-i})=q])\cdot 1\right]$$
$$-\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\left[\sum_q(\Pr[W(S)=q]-\Pr[W(S_{-i})=q])\cdot 1\right]$$

$$\leq 2\cdot\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\mathrm{TV}\big(W(S),W(S_{-i})\big)\underbrace{\leq}_{TV(p,q)\leq\sqrt{D(p|q)}}2\cdot\mathop{\mathbb{E}}_{S\sim\mathcal{D}^n}\ \mathop{\mathbb{E}}_{i\sim[n]}\sqrt{D\big(W(S)\big|W(S_{-i})\big)}\leq 2\sqrt{\varepsilon}$$

מ.ש.ל.

הערה: המשפט האחרון מראה רק חסם בתוחלת. צריך עוד להפוך אותו לחסם הסתברותי כדי שנוכל להשתמש בו (כמו שרצינו עם DP-יציבות). במקרה של ALKL המעבר לחסם ההסתברותי פחות "חלק" ממה שראינו דרך DP-יציבות, אבל אפשרי.