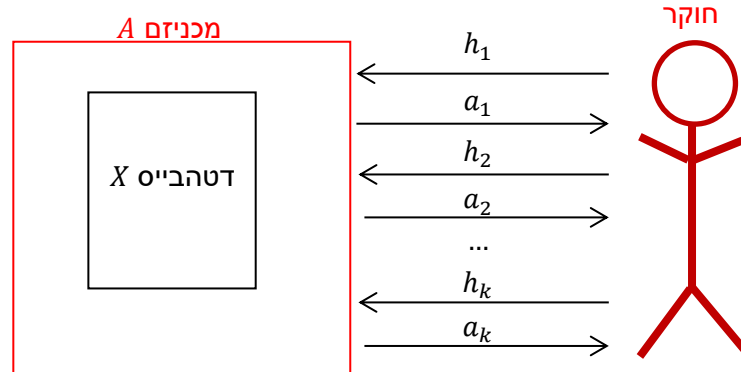


הרצאה 13: Private Multiplicative Weights (PMW)

Textbook: Cynthia Dwork and Aaron Roth. The Algorithmic Foundations of Differential Privacy

מרצה: אורי שטמר

תזכורת – מענה על שאלות אדפטיביות:



- נתון דטהבייס $X \in D^n$
- סדרה של שאלות ספירה h_1, h_2, \dots, h_k מגיעות אחת אחת. כל h_i היא פונקציה $h_i: D \rightarrow \{0,1\}$ והערך של h_i על דטהבייס X הוא $h_i \triangleq \frac{1}{|X|} \sum_{x \in X} h_i(x)$
- לאחר כל שאלת h_i עלינו להחזיר תשובה a_i

$$| \underbrace{h_i(X)}_{\substack{\text{הערך של} \\ \text{השאלת } h_i \\ \text{על הדטהבייס } X}} - \underbrace{a_i}_{\substack{\text{התשובה שלנו} \\ \text{לשאלת מספר } i}} | \leq \underbrace{\alpha}_{\substack{\text{פרמטר שגיאה} \\ \text{למשל } \alpha=0.1}}$$

המטרה: לכל $i \in [k]$ מתקיים

השאלה: על כמה שאלות נוכל לענות (כלומר מהו k) כפונקציה של $\alpha, \varepsilon, \delta, n$?

משפט (PMW): קיים אלגוריתם (ε, δ) -פרטי אשר בהינתן דטהבייס $X \in D^n$ עונה על k שאלות ספירה אדפטיביות עם דיוק α (בה"ג), עבור

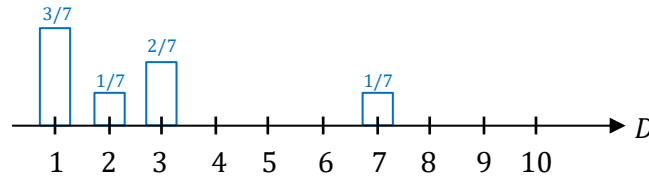
$$n \gtrsim \frac{\sqrt{\log |D| \cdot \log \frac{1}{\delta} \cdot \log k}}{\alpha^2 \varepsilon}$$

(האלג' רץ בזמן $\text{poly}(n, |D|)$ לכל שאלת לא)

הוכחה: יהיה לנו נח לחשוב על הדטהבייס X כעל התפלגות מעל D :

$$\forall d \in D : X(d) = \frac{|\{x \in X : x = d\}|}{n} = \frac{\text{מספר המופעים של } d \text{ ב } X}{n}$$

דוגמה בציור: אם הדומיין הוא $D = \{1,2,3, \dots, 10\}$ אזי עבור דטהבייס $X = (1,1,7,3,1,3,2)$ נקבל ייצוג



נשים לב שעם הסימון הזה, עבור שאילתת ספירה f נוכל לרשום:

$$f(X) = \frac{|\{x \in X : f(x) = 1\}|}{n} = \frac{\sum_{d \in D} |\{x \in X : x = d\}| \cdot f(d)}{n} =$$

$$= \sum_{d \in D} \frac{|\{x \in X : x = d\}|}{n} \cdot f(d) = \sum_{d \in D} X(d) \cdot f(d) = \mathbb{E}_{d \sim X}[f(d)]$$

באלגוריתם אנחנו נתחזק בצורה פרטית "דטהבייס סינטטי" \hat{X} (שגם אותו נייצג כהתפלגות מעל D).
 כשמגיעה שאילתא, נשווה את התשובה לשאילתא לפי X ולפי \hat{X} .
 אם התשובות קרובות אז נענה לפי \hat{X} .
 אחרת נעדכן את \hat{X} כדי שיהיה "קרוב יותר" ל- X .
 בניתוח הפרטיות אנו נראה שמספר העדכונים יהיה קטן ונשלם בקומפוזיציה רק על העדכונים.

האלגוריתם:

(1) אתחל $\hat{X} =$ ההתפלגות האחידה מעל הדומיין D .

(2) בצע לכל היותר $\frac{O(\log|D|)}{\alpha^2}$ פעמים (לולאה חיצונית)

(א) חשב $\hat{a} \leftarrow \frac{\alpha}{2} + \text{Lap}\left(\frac{1}{\varepsilon_0 \cdot n}\right)$ (כאשר ε_0 הוא פרמטר שנקבע בהמשך)

(ב) בצע לולאה פנימית:

(i) קבל את השאילתא הבאה f

(ii) חשב $v \leftarrow \text{Lap}\left(\frac{1}{\varepsilon_0 \cdot n}\right)$

(iii) אם $|f(X) - f(\hat{X})| + v < \hat{a}$

אזי החזר תשובה $a = f(\hat{X})$ והמשך עם הלולאה הפנימית

אחרת החזר תשובה $a = f(X) + \text{Lap}\left(\frac{1}{\varepsilon_0 \cdot n}\right)$ וצא מהלולאה הפנימית

(ג) עדכן את \hat{X} באופן הבא:

(i) לכל $d \in D$ חשב

$$g(d) = \begin{cases} \hat{X}(d) \cdot \exp\left(\frac{\alpha}{8} \cdot f(d)\right) & , a > f(\hat{X}) \\ \hat{X}(d) \cdot \exp\left(-\frac{\alpha}{8} \cdot f(d)\right) & , a < f(\hat{X}) \end{cases}$$

(ii) לכל $d \in D$ עדכן

$$\hat{X}(d) = \frac{g(d)}{\sum_{d' \in D} g(d')}$$

(ד) המשך עם הלולאה החיצונית

עלינו להראות 2 דברים:

1. האלגוריתם משמר פרטיות – קל
2. בה"ג התשובות שהאלג' מחזיר מדוייקות עד כדי α – יותר מסובך

מדוע האלגוריתם משמר פרטיות?

בצעדים (א)+(ב) של הלולאה החיצונית מבצעים בדיוק את אלגוריתם *AboveThreshold* (כאשר במידה ויוצאים מהלולאה הפנימית אנו מבצעים חישוב נוסף עם המ.לפלאס).
צעד (ג) מעדכן את \hat{X} ללא גישה נוספת לקלט X (מעבר למה שחישבנו בצורה פרטית בצעדים הקודמים).
לכן כל איטרציה של הלולאה החיצונית משמרת ε_0 -פ"ד.

לפי קומפוזיציה על פני $\frac{\log|D|}{\alpha^2} \approx$ החזרות, כל האלגוריתם משמר $\left(\sqrt{\frac{\log|D| \cdot \log\frac{1}{\delta}}{\alpha^2}} \cdot \varepsilon_0, \delta\right)$ -פ"ד.
לכן, אם נבחר פרמטר $\varepsilon_0 \approx \frac{\varepsilon \cdot \alpha}{\sqrt{\log|D| \cdot \log\frac{1}{\delta}}}$ אז כל האלגוריתם ישמר (ε, δ) -פ"ד.

מדוע האלגוריתם מחזיר תשובות מדויקות?

לאורך הריצה דוגמים לכל היותר $3k$ רעשים מ- $\text{Lap}\left(\frac{1}{\varepsilon_0 \cdot n}\right)$.
לפי תכונות התפלגות לפלאס, בה"ג, כל הרעשים האלה (בערך מוחלט) קטנים מ- $O\left(\frac{\log k}{\varepsilon_0 \cdot n}\right)$.
עבור

$$n \gtrsim \frac{\log k}{\varepsilon_0 \cdot \alpha} \approx \frac{\sqrt{|\log D| \cdot \log \frac{1}{\delta} \cdot \log k}}{\alpha^2 \varepsilon}$$

נקבל שכל הרעשים (בערך מוחלט) קטנים מ- $\alpha/8$. במקרה כזה נקבל שכל התשובות שהאלגוריתם מחזיר מדויקות עד כדי $\pm \frac{3\alpha}{4}$ (לפי אי-שוויון המשולש).

אנחנו עדיין צריכים להראות שהאלגוריתם לא יעצור באמצע הריצה (כלומר שהאלג' לא יעצור לפני שקיבל k שאילתות).

טענה: בהנחה שכל הדגימות מ- $\text{Lap}\left(\frac{1}{\varepsilon_0 \cdot n}\right)$ קטנות בערך מוחלט מ- $\alpha/8$, הלולאה החיצונית תתבצע לכל היותר $O\left(\frac{|\log D|}{\alpha^2}\right)$ פעמים.

רעיון ההוכחה:

נזכור שאנו חושבים על X ועל \hat{X} כעל התפלגויות מעל D .
נגדיר מדד למרחק בין התפלגויות שנסמנו $\text{KL}(X \parallel \hat{X})$ (אי שלילי)
בתחילת הריצה יתקיים $\text{KL}(X \parallel \hat{X}) \leq \log |D|$.
נראה שלאחר כל פעולת עדכון של \hat{X} (כלומר לאחר כל ביצוע של שלב (ג) באלג') מתקיים שהמרחק בין X ל- \hat{X} קטן בלפחות $\Omega(\alpha^2)$.
מסקנה: יתכנו לכל היותר $\frac{\log |D|}{\alpha^2} \approx$ צעדי עדכון

הגדרה:

עבור התפלגויות X, \hat{X} מעל D נגדיר את המדד הבא (הנקרא *Kullback–Leibler divergence*):

$$\text{KL}(X \parallel \hat{X}) = \sum_{d \in D} X(d) \cdot \log \left(\frac{X(d)}{\hat{X}(d)} \right)$$

עובדה 1: לכל זוג התפלגויות X, \hat{X} מעל D מתקיים $\text{KL}(X \parallel \hat{X}) \geq 0$.

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

$$\sum_i a_i \cdot \log \left(\frac{a_i}{b_i} \right) \geq \left(\sum_i a_i \right) \cdot \log \left(\frac{\sum_i a_i}{\sum_i b_i} \right)$$

ואמנם, בעזרת אי-שוויון זה נקבל

$$\text{KL}(X \parallel \hat{X}) = \sum_{d \in D} X(d) \cdot \log \left(\frac{X(d)}{\hat{X}(d)} \right) \geq \left(\sum_{d \in D} X(d) \right) \cdot \log \left(\frac{\sum_{d \in D} X(d)}{\sum_{d \in D} \hat{X}(d)} \right) = \left(\sum_{d \in D} X(d) \right) \cdot \log \left(\frac{1}{1} \right) = 0$$

עובדה 2: אם \hat{X} היא ההתפלגות האחידה מעל D , אזי לכל התפלגות X מתקיים

$$\text{KL}(X \parallel \hat{X}) \leq \log|D|$$

הוכחה:

$$\begin{aligned} \text{KL}(X \parallel \hat{X}) &= \sum_{d \in D} X(d) \cdot \log\left(\frac{X(d)}{\hat{X}(d)}\right) = \sum_{d \in D} X(d) \cdot \log(|D| \cdot X(d)) = \\ &= \underbrace{\sum_{d \in D} X(d) \cdot \log|D|}_{=\log|D|} + \underbrace{\sum_{d \in D} X(d) \cdot \log(X(d))}_{\leq 0} \leq \log|D| \end{aligned}$$

הוכחת הטענה:

נניח כי בצענו עדכון בשלב (ג) ועברנו מ- \hat{X} ל- \hat{X}' .
 המטרה שלנו היא להראות כי

$$\text{KL}(X \parallel \hat{X}) - \text{KL}(X \parallel \hat{X}') \geq \alpha^2$$

זה בכמה המרחק ל X מתכווץ.
 אני רוצה להראות שזה גדול.

העדכון בשלב (ג) מתבצע באחד מ-2 אופנים (תלוי אם $a > f(\hat{X})$ או לא).
 נניח למשל כי $a > f(\hat{X})$ (המקרה השני סימטרי), ונחשב:

$$\begin{aligned} \text{KL}(X \parallel \hat{X}) - \text{KL}(X \parallel \hat{X}') &= \sum_{d \in D} X(d) \cdot \log\left(\frac{X(d)}{\hat{X}(d)}\right) - \sum_{d \in D} X(d) \cdot \log\left(\frac{X(d)}{\hat{X}'(d)}\right) \\ &= \sum_{d \in D} X(d) \cdot \log\left(\frac{\hat{X}'(d)}{\hat{X}(d)}\right) = \sum_{d \in D} X(d) \cdot \log\left(\frac{g(d) / \sum_{\ell \in D} g(\ell)}{\hat{X}(d)}\right) = \\ &= \left[\sum_{d \in D} X(d) \cdot \log\left(\frac{g(d)}{\hat{X}(d)}\right) \right] - \log\left(\sum_{\ell \in D} g(\ell)\right) \\ &= \left[\sum_{d \in D} X(d) \cdot \log\left(\exp\left(\frac{\alpha}{8} \cdot f(d)\right)\right) \right] - \log\left(\sum_{\ell \in D} \hat{X}(\ell) \cdot \exp\left(\frac{\alpha}{8} \cdot f(\ell)\right)\right) = ((1)) \end{aligned}$$

נזכור כי לכל $y \leq 1$ מתקיים $e^y \leq 1 + y + y^2$. לכן,

$$\exp\left(\frac{\alpha}{8} \cdot f(\ell)\right) \leq 1 + \frac{\alpha}{8} \cdot f(\ell) + \frac{\alpha^2}{64} \cdot \underbrace{f^2(\ell)}_{\leq 1}$$

ולכן

$$((1)) \geq \left[\sum_{d \in D} X(d) \cdot \log\left(\exp\left(\frac{\alpha}{8} \cdot f(d)\right)\right) \right] - \log\left(\sum_{\ell \in D} \hat{X}(\ell) \cdot \left(1 + \frac{\alpha}{8} \cdot f(\ell) + \frac{\alpha^2}{64}\right)\right)$$

$$\begin{aligned}
&= \left[\sum_{d \in D} X(d) \cdot \frac{\alpha}{8} \cdot f(d) \right] - \log \left(\sum_{\ell \in D} \hat{X}(\ell) \cdot \left(1 + \frac{\alpha}{8} \cdot f(\ell) + \frac{\alpha^2}{64} \right) \right) \\
&= \frac{\alpha}{8} \cdot f(X) - \log \left(1 + \frac{\alpha}{8} \cdot f(\hat{X}) + \frac{\alpha^2}{64} \right) \\
&\quad \quad \quad \forall y : \log(1+y) \leq y \\
&\geq \frac{\alpha}{8} \cdot \underbrace{(f(X) - f(\hat{X}))}_{\geq \alpha/4} - \frac{\alpha^2}{64} \geq \frac{\alpha^2}{64} \\
&\quad \quad \quad \text{(נסביר עכשיו)}
\end{aligned}$$

מדוע $f(X) - f(\hat{X}) \geq \alpha/4$?

אנחנו מניחים כי כל הרעשים לאורך הריצה (בערך מוחלט) הם לכל היותר $\alpha/8$. לכן, כאשר מתבצע עדכון מתקיים

$$|f(X) - f(\hat{X})| + v > \hat{\alpha}$$

⇓

$$|f(X) - f(\hat{X})| > \frac{\alpha}{2} - \frac{\alpha}{8} - \frac{\alpha}{8} = \frac{\alpha}{4}$$

בנוסף, אנחנו מתנחים את המקרה בו

$$f(X) + \frac{\alpha}{8} \underbrace{\geq}_{\substack{\text{לפי האלגוריתם} \\ \alpha = f(X) + \text{Noise} \\ \text{והרעש חסום ע"י } \frac{\alpha}{8}}} a \underbrace{>}_{\substack{\text{המקרה אותו} \\ \text{אנו מנתחים}}} f(\hat{X})$$

⇓

$$f(\hat{X}) - f(X) \leq \frac{\alpha}{8}$$

כלומר $f(\hat{X}) - f(X) \leq \alpha/8$ וגם $|f(X) - f(\hat{X})| > \alpha/4$

מסקנה:

$$f(X) - f(\hat{X}) > \frac{\alpha}{4}$$

מ.ש.ל.