17.11.2022 ***פרטיות מידע***

הרצאה 4: חסמי צ'רנוף והופדינג

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

מרצה: אורי שטמר

צעד אחורה:
כלים בסיסיים מתורת ההסתברות (לא בהכרח קשורים לפרטיות)

נחשוב על הבעיה הפשוטה הבאה. נניח שיש לנו מטבע לא הוגן, כלומר ההסתברות לקבל "עץ" לא שווה להסתברות לקבל "פלי". אנחנו יכולים להטיל את המטבע כמה פעמים שנרצה ולראות את התוצאה. איך נוכל ללמוד מהי ההסתברות לקבל "עץ" בהטלת המטבע הזה?

פורמלית:
ישנה התפלגות $\text{Bernoulli}(p)$ עבור פרמטר p לא ידוע לנו. כלומר אם $X \sim \text{Bernoulli}(p)$ אזי $\Pr[X=1]=p$ ו- $\Pr[X_i=0]=1-p$. יש לנו "דטהבייס" $x=(x_1,x_2,\dots,x_n)$ המכיל n דגימות בלתי תלויות מההתפלגות הזאת (כלומר מכיל את התוצאות של n הטלות מטבע). איך נוכל להשתמש ב x כדי ללמוד מהו p?

אפשרות אחת היא להשתמש בממוצע האמפירי $\hat{p}=\frac{1}{n}\sum_{i=1}^{n}x_i$ בתור ההערכה שלנו ל- p.

כמה זה מדוייק? איך בכלל ניתן לכמת את "רמת הדיוק" כאן?

דוגמה:
נניח ש- $p=1/2$ ונניח ש- $n=1000$. למרות <u>שהתוחלת</u> של $\hat{p}$ שווה בדיוק ל- p, <u>ההסתברות</u> שבאמת נקבל $\hat{p}=1/2$ בדיוק היא יחסי קטנה, ספציפית לכל היותר $\sqrt{2/\pi n}$ (למעשה, אפשר להתחכם ולבחור את p להיות אי-רציונלי, ובמקרה זה ההסברות ש- $\hat{p}=p$ היא בדיוק אפס). כלומר, הטענה הבאה **<u>שגוייה:</u>** ככל ש- n גדל, ההסתברות ש- $\hat{p}$ יפגע בדיוק ב- p עולה.

מה כן אפשר להגיד?
- אנחנו יכולים להיות "די בטוחים" ש- $\hat{p}$ יהיה בין 0.45 ל- 0.55 (ההסתברות זה לא קורה היא ≥ 0.02)
- אנחנו יכולים להיות אפילו "יותר בטוחים" שזה יהיה בין 0.44 ל- 0.56 (ההסתברות שזה לא קורה היא לכל היותר 0.0015)

טיעונים כאלה נקראים high probability bounds או confidence intervals.

עכשיו אנחנו מעוניינים להבין איך אפשר <u>להוכיח</u> טענות כאלה, כלומר איך אפשר להוכיח אמירות מהצורה

ההסתברות ש $|\hat{p}-p|$ יהיה יותר מ- A היא לכל היותר B

כאשר ההסתברות היא מעל הטלת המטבעות, כלומר מעל הגרלת "הדטהבייס" שלנו x, וכאשר נהייה מעוניינים ש- A,B יהיו קטנים ככל האפשר.

בשביל זה אנחנו צריכים להיזכר ברקע מהסתברות.

<u>משפט [אי-שוויון מרקוב]:</u> לכל משתנה מקרי אי שלילי Y ולכל $a>0$ מתקיים

$$\Pr[Y \geq a] \leq \frac{\mathbb{E}[Y]}{a}$$

<u>הוכחה:</u> נוכיח עבור משתנה מקרי Y בדיד:

$$\mathbb{E}[Y] = \sum_{y} y \cdot \Pr[Y = y] \geq \sum_{y \geq a} y \cdot \Pr[Y = y] \geq \sum_{y \geq a} a \cdot \Pr[Y = y] = a \cdot \sum_{y \geq a} \Pr[Y = y] = a \cdot \Pr[Y \geq a]$$

בדוגמה שלנו עם המטבע, אנחנו יודעים ש- $\mathbb{E}[\hat{p}] = p$ ולכן אי-שוויון מרקוב נותן לנו איזשהו מידע (חלש) לגבי הקשר בין p ו- $\hat{p}$. אבל זה עוד לא מספיק כדי לקבל high probability bounds כמו שרצינו. עדיין, אי-שוויון מרקוב הוא כלי מאוד שימושי. בפרט, הוא מאפשר לנו להוכיח את אי-שוויון צ'בישב:

<u>משפט [אי-שוויון צ'בישב]:</u> לכל משתנה מקרי Y עם תוחלת $\mu = \mathbb{E}[Y]$ ושונות $\sigma^2 = \mathrm{Var}(Y) = \mathbb{E}[(Y-\mu)^2]$ ולכל $a > 0$ מתקיים:

$$\Pr[|Y - \mu| \geq a\sigma] \leq \frac{1}{a^2}$$

<u>הוכחה:</u>

$$\Pr[|Y - \mu| \geq a\sigma] = \Pr[(Y-\mu)^2 \geq a^2\sigma^2] \leq \frac{\mathbb{E}[(Y-\mu)^2]}{a^2\sigma^2} = \frac{1}{a^2}$$

כאשר אי השוויון נובע מאי-שוויון מרקוב.

זה כבר מספיק כדי לתת לנו איזשהו חסם הסתברותי לא טריוויאלי לגבי המטבע שלנו. כלומר אנחנו רוצים להשתמש באי-שוויון צ'בישב כדי לקבל חסם הסתברותי על $|\hat{p} - p|$. בשביל זה אנחנו צריכים לנתח את השונות של $\hat{p}$. בשביל זה אנחנו צריכים להיזכר בתכונות של שונות ותוחלת.

<u>משפט:</u> עבור משתנים מקריים <u>בלתי תלויים</u> מתקיים שתוחלת של מכפלה שווה למכפלת התוחלות, כלומר

$$\mathbb{E}[X_1 \cdot X_2 \cdots X_n] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2] \cdots \mathbb{E}[X_n]$$

<u>הוכחה:</u> נוכיח עבור משתנים מקריים X, Y בדידים ובלתי תלויים

$$\mathbb{E}[X \cdot Y] = \sum_{x.y} \Pr[X = x, Y = y] \cdot xy = \sum_{x.y} \Pr[X = x] \cdot \Pr[Y = y] \cdot xy$$
$$= \left(\sum_{x} \Pr[X = x] \cdot x\right) \cdot \left(\sum_{y} \Pr[Y = y] \cdot y\right) = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

<u>משפט:</u> עבור משתנה מקרי Y ועבור $a > 0$ מתקיים

$$\mathrm{Var}(aY) = a^2 \cdot \mathrm{Var}(Y)$$

<u>הוכחה:</u> נסמן $\mu = \mathbb{E}[Y]$. אזי

$$\mathrm{Var}(aY) = \mathbb{E}[(aY - a\mu)^2] = a^2\mathbb{E}[(Y-\mu)^2] = a^2 \cdot \mathrm{Var}(Y)$$

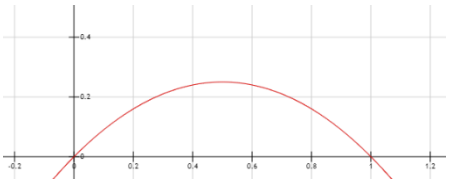
משפט: עבור זוג משתנים מקריים בלתי תלויים Y_1, Y_2 מתקיים

$$\mathrm{Var}(Y_1 + Y_2) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2)$$

"הוכחה": נובע מהגדרת השונות ומהעובדה (שהוכחנו) שעבור משתנים מקריים בלתי תלויים מתקיים שתוחלת של מכפלה שווה למכפלת התוחלות.

נחזור למטבע שלנו. אנחנו רוצים לנתח את השונות של $\hat{p}$. הזכרו כי הגדרנו $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$ כאשר של X_i הוא משתנה מקרי ברנולי (עם פרמטר p). עבור כל X_i כזה מתקיים ש-

$$Var(X_i) = \mathbb{E}[(X_i - p)^2] = \underbrace{\mathbb{E}[X_i^2]}_{\substack{X_i \text{ is a bit} \\ \text{and so } X_i^2 = X_i \\ \text{and } \mathbb{E}[X_i^2] = \mathbb{E}[x_i] = p}} - 2p\mathbb{E}[X_i] + p^2 = p - 2p^2 + p^2 = p(1-p)$$



עובדה (ראו גרף משמאל):
לכל $0 \le p \le 1$ מתקיים ש- $p(1-p) \le \frac{1}{4}$. לכן $Var(X_i) \le \frac{1}{4}$.

מסקנה:

$$Var(\hat{p}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \le \frac{1}{4n}$$

לכן, נוכל להפעיל את אי-שוויון צ'בישב על $\hat{p}$ ולקבל שלכל $a > 0$ מתקיים

$$\Pr\left[|\hat{p} - p| \ge \frac{a}{2\sqrt{n}}\right] \le \frac{1}{a^2}$$

עבור פרמטר $\beta > 0$ נסמן $a = \sqrt{\frac{1}{\beta}}$ ונקבל

$$\Pr\left[|\hat{p} - p| \ge \sqrt{\frac{1/\beta}{4n}}\right] \le \beta$$

או במילים אחרות, בהסתברות לפחות $1 - \beta$ מתקיים ש $|\hat{p} - p| \le \sqrt{\frac{1/\beta}{4n}}$

צורה אחרת להסתכל על זה: נניח שעבור פרמטרים $\alpha, \beta > 0$ מסויימים, אני רוצה להבטיח שבהסתברות לפחות $(1-\beta)$ ההערכה שלנו $\hat{p}$ תהייה קרובה ל- p עד כדי טעות α. מהו גודל המדגם n שאני צריך?
כדי לענות על השאלה הזאת בעזרת אי-השוויון האחרון, נדרוש $\sqrt{\frac{1/\beta}{4n}} \le \alpha$ $|\hat{p} - p|$. נפתור עבור n ונקבל שמספיק לדרוש ש-

$$n \ge \frac{1/\beta}{4\alpha^2}$$

זה מצויין. זה נותן לנו high probability bound כמו שרצינו. אבל אנחנו לא רוצים להסתפק בזה. בדרישה הנ"ל מתקיים ש- n צריך לגדול לינראית עם $1/\beta$, שזה לא כל כך טוב אם אנחנו רוצים ש- β יהיה פצפון (כלומר אם אנחנו רוצים להיות "סופר בטוחים" בנכונות השערוך שלנו.

מסתבר שבמקרה הזה אפשר לקבל הבטחות הרבה יותר טובות בעזרת מה שנקרא חסמי צ'רנוף/הופדינג.

<u>משפט [חסמי צ'רנוף והופדינג]:</u>
יהיו $X_1, X_2, \dots, X_n$ משתנים מקריים בלתי תלויים כאשר לכל i מתקיים $\Pr[X_i=1] = p$ ו- $\Pr[X_i=0] = 1-p$,
עבור פרמטר $0 < p < 1$. תוחלת סכום המשתנים היא $\mathbb{E}[\sum_{i=1}^n X_i] = p \cdot n$. אזי מתקיים:

(א) לכל $0 < \alpha < 1$ מתקיים $\Pr[\sum_{i=1}^n X_i \geq (1+\alpha) \cdot pn] < \exp(-\alpha^2 pn/4)$

(ב) לכל $0 < \alpha < 1$ מתקיים $\Pr[\sum_{i=1}^n X_i \leq (1-\alpha) \cdot pn] < \exp(-\alpha^2 pn/2)$

עבור $B > A > 0$ יהיו $X_1, X_2, \dots, X_n \in [A, B]$ משתנים מקריים בלתי תלויים ונסמן $\sum_{i=1}^n \mathbb{E}[X_i] = \mu$. אזי:

(ג) לכל $\alpha > 0$ מתקיים $\Pr[|\sum_{i=1}^n X_i - \mu| \geq \alpha] \leq 2\exp\left(-\frac{2\alpha^2}{n \cdot (B-A)^2}\right)$

לפני שנדבר על ההוכחה של המשפט הזה, נראה מה הוא נותן לנו עבור הדוגמה שלנו עם המטבע. למשל, בעזרת (ג) אנחנו מקבלים ש

$$\Pr[|\hat{p} - p| \geq \delta] \leq 2\exp(-2\delta^2 n)$$

כדי להבטיח ש- $2\exp(-2\alpha^2 n)$ יהיה לכל היותר β (עבור פרמטר $\beta > 0$ כלשהו), מספיק לדאוג ש $n \geq \frac{\ln\left(\frac{2}{\beta}\right)}{2\alpha^2}$.
שימו לב שעכשיו התלות של n ב $1/\beta$ היא לוגריתמית. זה אומר "שבמחיר" נמוך יחסית מבחית גודל המדגם n אנחנו יכולים לדאוג ש- β יהיה פצפון.

אז אנחנו יודעים להעריך את התוחלת של מטבע לא ידוע. באופן אולי מתפיע, זה כלי סופר שימושי. למשל, נניח שהדטה שלנו מכיל נתונים רפואיים של אנשים, שנדגמו באקראי מאוכלוסיה מסויימת. בנוסף, נניח שיש לנו בדיקה מסויימת אשר בהינתן המידע הרפואי של אדם, אומרת אם הוא חולה במחלה מסויימת או לא. בעזרת אותה שיטה נוכל להעריך, מתוך הדטה, את אחוז החולים בכלל האוכלוסיה.

<u>הוכחת חסמי צ'רנוף/הופדינג:</u> אנחנו נוכיח רק את (א). ההוכחות של (ב),(ג) דומות. יהיו $X_1, X_2, \dots, X_n$ משתנים מקריים בלתי תלויים כאשר לכל i מתקיים $\Pr[X_i = 1] = p$ ו- $\Pr[X_i = 0] = 1-p$ ויהי $0 < \alpha < 1$. עלינו להראות ש-

$$\Pr\left[\sum_{i=1}^n X_i \geq (1+\alpha) \cdot pn\right] < \exp(-\alpha^2 pn/4)$$

נסמן $t = (1+\alpha)pn$ ונסמן $c = \alpha/2$. (נשים לב שמכיוון ש- $0 < \alpha < 1$ אזי $0 < c < 1/2$).
נחשב:

$$\Pr[\sum X_i \geq t] = \Pr[c \cdot \sum X_i \geq c \cdot t] = \Pr\left[e^{c \cdot \sum X_i} \geq e^{c \cdot t}\right] = ((1))$$

כעת לפי אי-שוויון מרקוב נקבל ש

$$((1)) \leq e^{-c \cdot t} \cdot \mathbb{E}\left[e^{c \cdot \sum X_i}\right] = e^{-c \cdot t} \cdot \mathbb{E}[e^{c \cdot X_1} \cdot e^{c \cdot X_2} \cdots e^{c \cdot X_n}] = ((2))$$

כעת מכיוון שהמשתנים $X_1, \dots, X_n$ הם בלתי תלויים נקבל ש

$$((2)) = e^{-c \cdot t} \cdot \mathbb{E}[e^{c \cdot X_1}] \cdot \mathbb{E}[e^{c \cdot X_2}] \cdots \mathbb{E}[e^{c \cdot X_n}] = ((3))$$

ומכיוון ש- $X_1, \dots, X_n$ מתפלגים אותו הדבר מתקיים $\mathbb{E}[e^{c \cdot X_1}] = \mathbb{E}[e^{c \cdot X_2}] = \cdots = \mathbb{E}[e^{c \cdot X_n}]$ ולכן

$$((3)) = e^{-c\cdot t}\cdot\left(\mathbb{E}[e^{c\cdot X_1}]\right)^n$$

כלומר קיבלנו ש

$$\Pr[\sum X_i \geq t] \leq e^{-c\cdot t}\cdot\left(\mathbb{E}[e^{c\cdot X_1}]\right)^n$$

כעת נשים לב ש-

$$\mathbb{E}[e^{c\cdot X_1}] = p\cdot e^c + (1-p)e^0 = p\cdot e^c + 1 - p \leq p(1+c+c^2) + 1 - p = 1 + p(c+c^2) \leq e^{p(c+c^2)}$$

כאשר אי-השוויון הראשון הוא לפי הנוסחה $e^z \leq 1 + z + z^2$ המתקיימת לכל $z \leq 1$ וכאשר אי-השוויון השני הוא לפי הנוסחה $1 + z \leq e^z$ המתקיימת לכל $z \in \mathbb{R}$.

נציב זאת בחשבון הקודם שעשינו ונקבל

$$\Pr[\sum X_i \geq t] \leq e^{-c\cdot t}\cdot\left(\mathbb{E}[e^{c\cdot X_1}]\right)^n \leq e^{-c\cdot t}\cdot\left(e^{p(c+c^2)}\right)^n = e^{-c\cdot t}\cdot e^{pn(c+c^2)} = ((4))$$

נזכור שבחרנו $t = (1+\alpha)pn$ ולכן נקבל

$$((4)) = e^{-c\cdot(1+\alpha)pn}\cdot e^{pn(c+c^2)} = e^{-cpn(\alpha-c)} = ((5))$$

נזכור שבחרנו $c = \alpha/2$ ולכן

$$((5)) = e^{-\alpha^2 pn/4}$$

מ.ש.ל.

<u>מה קרה בהוכחה הזאת?</u>

אי-שוויון מרקוב אומר שמשתנה מקרי אי שלילי לא יכול לסטות <u>בהרבה</u> מהתוחלת שלו. אבל בחסם צ'רנוף רצינו להראות ש- $\sum X_i$ לא יכול לסטות אפילו <u>בקצת</u> מהתוחלת שלו. לכן במקום להפעיל את אי-שוויון מרקוב ישירות על $\sum X_i$ חשבנו על המשתנה המקרי $\exp(\sum X_i)$.

למה זה טוב? עכשיו מרקוב אומר לנו שההסתברות ש- $\exp(\sum X_i)$ יחרוג <u>בהרבה</u> מהתוחלת שלו היא קטנה וזאת אומרת שההסתברות ש- $\sum X_i$ יחרוג מהתוחלת שלו <u>אפילו בקצת</u> היא קטנה (כי אם $\sum X_i$ חורג אפילו קצת אז $\exp(\sum X_i)$ חורג הרבה...)

בנוסף, בזכות העובדה ש- $X_1, \ldots, X_n$ הם בלתי תלויים יכולנו לנתח את התוחלת של $\exp(\sum X_i)$ כי התוחלת הזאת התפצלה לנו למכפלה של תוחלות.