שיעור 4 - 21-04-2025

בסעיף זה נלמד מספר אי שוויונות שימושיים.

1.4.1. ***אי-שוויוני ריכוז* - Concentration inequalities**

אי-שוויוני ריכוז אלו מספקים חסמים על ההסתברות שמשתנה מקרי יסטה מערכו הצפוי (או ערך אחר כגון החציון) במידה מסוימת. במילים אחרות, אי-שוויוני ריכוז מאפשרים לנו לכמת עד כמה משתנה מקרי "מרוכז" סביב ערך מסוים. אי-שוויוני ריכוז הם כלים חשובים מאוד בתורת ההסתברות, סטטיסטיקה, מדעי הנתונים (בפרט באלגוריתמים הסתברותיים ולמידת מכונה) ותחומים רבים אחרים. הם מאפשרים להוכיח תוצאות על התנהגות של משתנים מקריים ושל פונקציות שלהם, ולנתח את ההסתברות לאירועים חריגים.

דוגמאות מוכרות לאי-שוויוני ריכוז כוללות:

- **אי-שוויון מרקוב** *(Markov's inequality):* מספק חסם על ההסתברות של משתנה מקרי אי-שלילי להיות גדול מערך מסוים, בהינתן התוחלת שלו.
- **אי-שוויון צ'בישב** *(Chebyshev's inequality):* מספק חסם על ההסתברות של משתנה מקרי להיות רחוק מהתוחלת שלו במידה מסוימת, בהינתן התוחלת והשונות שלו.
- חסמי צ'רנוף *(Chernoff bounds):* מספקים חסמים אקספוננציאליים על ההסתברות שסכום של משתנים מקריים בלתי תלויים (לרוב משתני ברנולי) יסטה באופן משמעותי מהתוחלת שלו. חסמים אלה חזקים יותר מאי-שוויונות מרקוב וצ'בישב במקרים רבים.
- אי-שוויון הופדינג *(Hoeffding's inequality):* דומה לחסמי צ'רנוף אך חל על סכום של משתנים מקריים חסומים.
- אי-שוויון אזומה-הופדינג *(Azuma-Hoeffding inequality):* הכללה של אי-שוויון הופדינג עבור מרטינגלים עם הבדלים חסומים.

משפט 1.10 **אי-שוויון מרקוב:** *(*Markov's inequality)

עבור משתנה מקרי $X \geq 0$.

$$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}X}{x}, \quad \forall x > 0$$

הוכחה:

יהיה משתנה מקרי $X \geq 0$, אזי עבור $x > 0$

$$\mathbb{E}X = \mathbb{E}\left(I_{\{X\geq x\}}X\right) + \mathbb{E}\left(I_{\{X<x\}}X\right) \geq \mathbb{E}\left(I_{\{X\geq x\}}X\right) \geq \mathbb{E}\left(I_{\{X\geq x\}}\right)x = x\,\mathbb{P}(X \geq x), \qquad \forall\, x > 0$$

האי השוויון הראשון נכון מכיוון ו- $X \geq 0$, האי שוויון השני נכון מכיוון ו- $X \geq x$, והשוויון האחרון נובע מכך ש-x הוא ערך קבוע.

עתה נעביר אגפים ונקבל את הנדרש.

מסקנה:

אם ל-X יש בנוסף מומנטים סופיים, $\mathbb{E}|X|^P < \infty$ עבור $P \in \mathbb{N}$, אז זנב ההסתברות יורד בקצב שהוא לפחות פולינומי.
זאת משום שלפי אי שוויון מרקוב:

$$\mathbb{E}|X|^p = \mathbb{E}\left(I_{\{X\geq x\}}|X|^p\right) + \mathbb{E}\left(I_{\{X<x\}}|X|^p\right) \geq \mathbb{E}\left(I_{\{X\geq x\}}|X|^p\right) \geq \mathbb{E}\left(I_{\{X\geq x\}}\right)|x|^p = |x|^p\,\mathbb{P}(X \geq x), \quad \forall x > 0$$

ולכן לכל $x > 0$ גדול כרצוננו $\mathbb{P}(X \geq x) \leq |x|^{-p}\,\mathbb{E}|X|^p$.

תוצאה 1.11 **אי-שוויון צ'בישב**: (Chebyshev's inequality)

יהיה X משתנה מקרי עם $E|X|^2 < \infty$, אזי

$$\mathbb{P}(|X - \mathbb{E}X| \geq x) \leq \frac{\mathbb{V}ar(X)}{x^2} \quad .$$

הוכחה:

נסמן $Y = (X - \mathbb{E}X)^2$. מכיוון שמדובר בחזקה ריבועית, Y הוא משתנה מקרי אי שלילי. לכל $\forall x > 0$, לפי אי שוויון מרקוב,

$$\mathbb{P}(Y \geq x^2) \leq \frac{\mathbb{E}(Y)}{x^2} \quad \Rightarrow \quad \mathbb{P}((X - \mathbb{E}X)^2 \geq x^2) \leq \frac{\mathbb{E}((X - \mathbb{E}X)^2)}{x^2} = \frac{\mathbb{V}ar(X)}{x^2} \quad \Rightarrow$$

$$\mathbb{P}(|X - \mathbb{E}X| \geq x) \leq \frac{\mathbb{V}ar(X)}{x^2}$$

המסקנה האחרונה נובעת מכך ש- $x > 0$ ולכן $\mathbb{P}(|X - \mathbb{E}X| \geq x) = \mathbb{P}((X - \mathbb{E}X)^2 \geq x^2)$. כנדרש.

חשוב לציין כי אי שוויון צ'ביצ'ב אינו דורש ידע על ההתפלגות, רק על קיום המומנט/ים.

מסקנה:

יהיו $X_1, \ldots, X_n$ משתנים מקריים שווי התפלגות לא מתואמים, ויהי $\bar{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$. אפילו אם ההתפלגות של X_j ידועה, בדרך כלל קשה למצוא את ההתפלגות של הממוצע $\bar{X}_n$.
אולם אם $\mathbb{E}X_j^2 < \infty$, אזי לפי אי שיוויון צ'ביצ'ב נקבל כי

$$\mathbb{P}(|\bar{X}_n - \mathbb{E}X_1| \geq a) \leq \frac{1}{a^2}\mathbb{V}ar(\bar{X}_n) = \frac{1}{a^2}\frac{1}{n}\mathbb{V}ar(X_1)$$

בפרט, המשמעות היא שההתפלגות של $\bar{X}_n$ מרוכזת סביב $\mathbb{E}X_1$ לפחות בקצב n.

1.4.2 **תוחלת וקמירות** ***(Expectation and convexity)***

<u>הגדרה 1.12</u>:

פונקציה $g: I \to \mathbb{R}$ היא <u>קמורה</u> על האינטרוול $I \subseteq \mathbb{R}$ אם לכל $x, y \in I$,

$$g(\lambda x + (1-\lambda)y) \le \lambda g(x) + (1-\lambda)g(y) \quad , \qquad \forall \lambda \in [0,1]$$

אם אי השוויון הפוך אז הפונקציה נקראת <u>קעורה</u>. ברור כי אם $g(\cdot)$ קמורה אז $-g(\cdot)$ קעורה, ולהיפך. ולכן כל תכונה של פונקציה קמורה מורחבת לפונקציה קעורה עם ההתאמות הנדרשות.

קמירות היא תכונה שימושית מאד.

<u>למה 1.13</u>:

פונקציה $g(\cdot)$ היא קמורה אם ורק אם קיימת פונקציה $\nu(x)$, כך ש

$$g(y) - g(x) \ge \nu(x)(y-x) \ , \ \ \forall \, x, y \in I$$

<u>הוכחה</u>:

נניח כי האי שוויון מתקיים, אז לכל $\forall \, x, y, z \in I$

$$g(y) - g(x) \ge \nu(x)(y-x)$$

$$g(z) - g(x) \ge \nu(x)(z-x)$$

נכפיל את המשוואה הראשונה ב-λ ואת השנייה ב- $(1-\lambda)$ ונחבר את שתי המשוואות, נקבל כי

$$\lambda g(y) + (1-\lambda)g(z) - g(x) \ge \nu(x)(\lambda y + (1-\lambda)z - x)$$

והקמירות מתקבלת אם ניקח $x \coloneqq \lambda y + (1-\lambda)z$. נציב את x ונקבל ב]פרט שאגף ימין מתאפס:

$$\lambda g(y) + (1-\lambda)g(z) - g(\lambda y + (1-\lambda)z) \ge 0$$

עתה נעביר אגף ונקבל את אי השוויון.

מצד שני, אם $g(\cdot)$ קמורה, אז לכל $\lambda \in (0,1)$, מהגדרת הקמירות נקבל

$$\frac{g(x+\lambda(y-x)) - g(x)}{\lambda} = \frac{g((1-\lambda)x + \lambda y) - g(x)}{\lambda} \le \frac{(1-\lambda)g(x) + \lambda g(y) - g(x)}{\lambda} = g(y) - g(x).$$

עתה אם $g(\cdot)$ גזירה, ניתן ל-λ לשאוף ל-0 ונקבל

$$g(y) - g(x) \ge g'(x)(x-y)$$

כלומר קיבלנו את אh השוויון הנדרש עם $\nu(x) := g'(x)$.*

* כזכור, נגזרת של פונקציה בנקודה x מוגדרת $\lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$.

גיאומטרית אי השוויון אומר כי הגרף של פונקציה קמורה שוכב מעל המשיק לגרף בכל נקודה.

פונקציות קמורות לא בהכרח גזירות, אבל אפשר להראות יש להן נגזרת מימין ונגזרת משמאל. לכן הטענה נותרת תקפה אם ניקח למשל, נגזרת מימין כאשר ניתן ל-λ לרדת ל-0 מימין - $\lambda \searrow 0$.

דוגמא 1.14: נניח $g(\cdot)$ גזירה פעמיים, אזי

$$g(y)-g(x)=\int_x^y g'(u)du=\int_x^y g'(x)du+\int_x^y \big(g'(u)-g'(x)\big)du=$$
$$g'(x)(y-x)+\int_x^y \big(g'(u)-g'(x)\big)du=$$
$$g'(x)(y-x)+\int_x^y\int_x^u g''(\mathrm{v})\,d\mathrm{v}du$$

לפי למה 1.13, $g(\cdot)$ קמורה אם ורק אם הביטוי הימני הוא חיובי לכל x ו-y , מה שאומר כי $g''(\mathrm{v})\ge 0$ לכל $\mathrm{v}\in I$.

זוהי דרך נוחה להוכיח קמירות. למשל $p\ge 2$ $g(x)=x^p$, $g(x)=e^x$ ועוד.

משפט 1.15 ***אי שוויון ג'נסן*** (.(Jensen inequality

תהי $g(\cdot)$ פונקציה קמורה ויהי X משתנה מקרי עם תוחלת סופית, אזי

$$\mathbb{E}g(X)\ge g(\mathbb{E}X)\ .$$

הוכחה:

מלמה 1.13 נקבל כי קיימת פונקציה $\nu(x)$ כך ש

$$g(X)-g(\mathbb{E}X)\ge \nu(\mathbb{E}X)(X-\mathbb{E}X)$$

עתה ניקח תוחלת משני צידי האי שוויון ונקבל

$$\mathbb{E}[g(X)-g(\mathbb{E}X)]=\mathbb{E}g(X)-g(\mathbb{E}X)\ge \mathbb{E}[\nu(\mathbb{E}X)(X-\mathbb{E}X)]=\nu(\mathbb{E}X)\mathbb{E}[(X-\mathbb{E}X)]=0$$
$$\mathbb{E}g(X)-g(\mathbb{E}X)\ge 0 \quad\Longrightarrow\quad \mathbb{E}g(X)\ge g(\mathbb{E}X)$$

כנדרש.

אי שוויון ג'נסן משמש להוכחת אי שוויוניות חשובים נוספים.

משפט 1.16 ***אי שוויון ליאפונוב*** (Lyapunov inequality)

יהיה X משתנה מקרי. עבור $p\ge q$,

$$(\mathbb{E}|X|^q)^{1/q}\le(\mathbb{E}|X|^p)^{1/p}$$

הפונקציה $r<1$ $g(x)=x^r$ קעורה על $\mathbb{R}_+$., מכיוון ש: $g''^{(x)}=r(r-1)x^{r-2}<0$ לכל $x>0$.

יהי $r\coloneqq {}^{q}\!/_{p}<1$, אזי לפי אי שוויון ג'נסן על פונקציות קמורות:

$$\mathbb{E}|X|^q=\mathbb{E}(|X|^p)^r\le(\mathbb{E}|X|^p)^r$$

נעלה את שני צידי האי שוויון בחזקת $1/q$ ונקבל את אי השוויון הנדרש:

$$(\mathbb{E}|X|^q)^{1/q}\le(\mathbb{E}|X|^p)^{r\cdot 1/q}=(\mathbb{E}|X|^p)^{1/p}\quad .$$

מסקנה 1.17

יהיה X משתנה מקרי כך ש- $\mathbb{E}|X|^p < \infty$ עבור $p > 0$ כלשהו, אזי $\mathbb{E}|X|^q < \infty$ לכל $q \le p$.

הוכחה: נובע ישירות מאי שוויון ליאפונוב.

בשל קוצר הזמן, אין אפשרות להציג את השימושים שנעשים בחומר הנלמד במהלך השיעור.

הקישור הבא מוביל לשאלה שהפניתי למודל perplexity AI תוכלו למצוא דוגמאות והפניות למקורות המדגימים את השימושים הרבים שנעשים באי שוויון ג'נסן בכלכלה, תורת האינפורמציה, אופטימיזציה ועוד, ולהמשיך ולחקור את הנושא בעזרת המודל והשאלות הקשורות שהוא מציע.

https://www.perplexity.ai/search/what-are-the-uses-of-of-jensen-uTJ3oIALRGCAxF8OIp_.Yw#0

במסמך "What are the uses of Jensen Inequality" אפשר לקרוא סיכום ראשוני של תשובות המודל.

תרגיל (לא הוצג בשיעור)

הוכיחו כי עבור $x > 0$ $\mathbb{E}[ln(X)] \le ln[\mathbb{E}X]$. **(דוגמא ליעילות לוגריתמית)**

תשובה:

הפונקציה הלוגריתמית היא פונקציה קעורה (נגזרת שנייה שלילית), ולכן אי השוויון נובע מאי שוויון ג'נסן עבור פונקציות קעורות.

סעיף 1.5: דגימה של משתנה מקרי

דגימת משתנה מקרי משמעותה להסתכל על מימוש מסוים של הניסוי המקרי. חבילות תוכנה כמו R, פייתון וכו' כוללות בתוכן פונקציות לדגימה מהתפלגויות שונות.

למשל ב-R ,

דגימה מהתפלגות אחידה סטנדרטית runif(n, min = 0, max = 1) $n=1$

דגימה מהתפלגות נורמלית סטנדרטית *rnorm(n, mean = 0, sd = 1)*

דגימה מהתפלגות מעריכית עם $\lambda = 1$ *rexp(n, rate = 1)*

דגימה מהתפלגות גיאומטרית *rgeom(n, prob)*

וכדומה

במקרים בהם התוכננה בה אנו משתמשים אינה כוללת פונקציה לדגימת ההתפלגות בה אנו מעוניינים אנו יכולים להשתמש בהתפלגות האחידה הסטנדרטית על מנת לדגום מההתפלגות הרצויה.

למה 1.8

יהי $V \sim U(0,1)$ ו- F פונקציית התפלגות מצטברת בדידה לגמרי, עם האטומים $x \in \{x_1, x_2, \dots\}$. אזי משתנה המקרי

$$X = \sum_{j \ge 1} x_j \, I\{F(x_j -) < V \le F(x_j)\} ,$$

הוא בעל ההתפלגות המצטברת הרצויה F.

הוכחה:

נשים לב כי בכל מימוש של המשתנה המקרי V יש בדיוק אינדיקטור יחיד המקבל את הערך 1 (השאר מתאפסים)

$$\mathbb{P}\left(X = x_j\right) = \mathbb{P}\ \left(F\left(x_j -\right) < V \leq F\left(x_j\right)\right) = F\left(x_j\right) - F\left(x_j -\right)$$

ולכן X מתפלג לפי F, כנדרש.

למה 1.9

יהי $V \sim U(0,1)$ ו- F פונקציית התפלגות מצטברת רציפה לגמרי עולה ממש.

1. אזי המשתנה המקרי

$$X = F^{-1}(V)$$

מתפלג לפי ההתפלגות המצטברת F.

2. יהיה X משתנה מקרי רציף לגמרי עם פונקציית התפלגות מצטברת עולה ממש G. אזי המשתנה המקרי

$$V = G(X)$$

מתפלג $U(0,1)$.

הוכחה:

1. לכל $x \in \mathbb{R}$, פונקציית ההתפלגות המצטברת של X היא

$$\mathbb{P}_X(X \leq x) = \mathbb{P}_V(F^{-1}(V) \leq x) = \mathbb{P}_V(F(F^{-1}(V)) \leq F(x)) = \mathbb{P}_V(V \leq F(x)) = F(x)$$

2. לכל $\mathrm{v} \in (0,1)$,

$$\mathbb{P}_V(V \leq \mathrm{v}) = \mathbb{P}_X(G(X) \leq \mathrm{v}) = \mathbb{P}_X(X \leq \mathrm{G}^{-1}(\mathrm{v}) = G(\mathrm{G}^{-1}(\mathrm{v})) = \mathrm{v}$$

קיבלנו איפה את ההפלגות האחידה.

הערה – הלמה בנוסח קצת אחר ניתנה בתרגיל 1 שאלה 3.

דוגמא 1.20:

ניצור דגימה מהתפלגות מעריכית עם פרמטר $\lambda > 0$ בעזרת התפלגות אחידה. פונקציית ההתפלגות המצטברת של ההתפלגות המעריכית היא $F(x) = 1 - e^{-\lambda x}$ רציפה ומונוטונית עולה ממש. $F^{-1}(y) = \frac{1}{\lambda} log \frac{1}{1-y}$. לכן אם נדגום $\mathrm{V} \sim U(0,1)$ אזי $X = \frac{1}{\lambda} log \frac{1}{1-\mathrm{V}} \sim Exp(\lambda)$.