

12 (00094412) הסתברות מ' | תרגול

שם: איל

April 4, 2024

נושא התרגול: אי שוויונות והתכנסויות של משתנים מקריים

נושא ראשון - אי שוויונות:

אי שוויון מרקוב:

יהי X משתנה מקרי אי שלילי, אזי לכל $a > 0$ מתקיים כי:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

הערה: במקרה ש $E[X] > a$ מקבלים אי שוויון טריוויאלי.

• כלומר אם המשתנה המקרי שלנו אי שלילי אז לכל מספר חיובי a אפשר לחסום את ההסתברות $P(X \geq a)$.

אי שוויון צ'ביצ'ב:

יהי X משתנה מקרי עם תוחלת $E[X] = \mu$ ושונות $Var(X) = \sigma^2$ סופיות. אזי לכל $a > 0$ מתקיים כי:

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

צורה אחרת לרישום בעזרת סטיות תקן:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- החסם הזה הוא קצת יותר הדוק והוא מנצל את העובדה שהשונות של X היא סופית.
- הגרסה השקולה היא עבור $k > 0$ כלשהו.

שאלה 1:

- יהי משתנה מקרי X שמסמל הכנסה (נניח שהוא רציף ואי שלילי)
- נתון שההכנסה ממוצעת היא 10,000, הכוונה היא ש: $E[X] = 10,000$.
- נתון שסטית התקן היא $SD(X) = 8,000$.

צ"ל: מצאו חסם עליון ל- $P(X > 50,000)$

- א. לפי מרקוב
- ב. לפי צ'ביצ'ב

1. א. פיתרון:

- נמצא חסם ל- $P(X > 50,000)$ בעזרת אי שוויון מרקוב:

– X רציף ולכן:

$$P(X \geq 50,000) = P(X > 50,000)$$

* לפי אי שוויון מרקוב:

$$P(X \geq 50,000) = \frac{E[X]}{50,000} = \frac{10,000}{50,000} = 0.2$$

1. ב. פיתרון:

• נמצא חסם ל- $P(X > 50,000)$ בעזרת אי שוויון צ'ביצ'ב:

– נחסר את הממוצע:

$$P(X \geq 50,000) = P(X - E[X] \geq 50,000 - E[X])$$

$$= P(X - 10,000 \geq 40,000)$$

– נשים לב כי לפי הגדרת הערך המוחלט מקבלים מאורעות זרים, כלומר:

$$P(|X - 10,000| \geq 40,000) = P(X - 10,000 \geq 40,000) + P(X - 10,000 \leq -40,000)$$

* נתון כי X אי שלילי ולכן $P(X - 10,000 \leq -40,000) = 0$. כלומר:

$$P(|X - 10,000| \geq 40,000) = P(X - 10,000 \geq 40,000)$$

– נציב זאת ונקבל:

$$P(X \geq 50,000) = P(|X - 10,000| \geq 40,000)$$

– ולפי אי שוויון מרקוב מתקיים:

$$P(|X - 10,000| \geq 40,000) \leq \frac{(8,000)^2}{40,000} = \frac{1}{25} = 0.04$$

• נשים לב כי בעזרת צ'ביצ'ב קיבלנו חסם הדוק יותר.

• בנוסף, נשים לב ששני החסמים לא הניחו שהמשתנה המקרי מתפלג בצורה כלשהי.

– זו בערך הרמה של המבחן.

נושא שני - התכנסות חלשה של משתנים מקריים:

התכנסות חלשה

נאמר של סדרה של מ"מ $(X_n)_{n>0}$ מתכנסת בהתכנסות חלשה למשתנה מקרי X אם אחד מהתנאים הבאים מתקיים, כתלות בסוג המשתנים.

מקרה 1: אם X_n ו- X הינם משתנים מקררים בדידים המקבלים ערכים טבעיים

$$X_n \xrightarrow[n \rightarrow \infty]{} X \quad \text{if} \quad P_{X_n}(k) \xrightarrow[n \rightarrow \infty]{} P_X(k), \forall k \in N.$$

מקרה 2: אם X הינו משתנה מקרי רציף

$$X_n \xrightarrow[n \rightarrow \infty]{} X \quad \text{if} \quad F_{X_n}(x) \xrightarrow[n \rightarrow \infty]{} F_X(x), \forall x \in R.$$

מקרה 3: אם X הינו קבוע. כלומר $P(X = x_0) = 1$ עבור $x_0 \in \mathbb{R}$

$$X_n \xrightarrow[n \rightarrow \infty]{} X \quad \text{if} \quad P(|X_n - x_0| > \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0, \forall \varepsilon > 0.$$

• במילים אחרות, ממירים את ההתכנסות של הסדרה $(X_n)_{n>0}$ ל- X בהתכנסות של פונקציות.

נושא שלישי - החוק החלש של המספרים הגדולים:

החוק החלש של המספרים הגדולים:

יהיו $X_1, X_2, X_3 \dots$ מ"מ ב"ת ושווי התפלגות בעלי תוחלת μ .

יהי $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ ממוצע המשתנים הללו עד למקום ה- n .

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{} \mu \text{ אזי}$$

כלומר, לכל $\varepsilon > 0$ מתקיים:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

משמעות: ככל שהמדגם גדל הממוצע מתקרב לתוחלת כרצוננו בהסתברות שואפת ל-1.

• \bar{X}_n נקרא גם הממוצע האמפירי, כי הוא סכום המשתנים המקריים חלקי n .

תרגיל 2:

יהיו X_1, X_2, X_3, \dots משתנים מקריים בלתי תלויים ובעלי פונקציית ההסתברות הבאה:

2	3	4	5	x
0.4	0.3	0.2	0.1	$P_X(x)$

א. מצאו את $E(X_1), E(\bar{X}_n), Var(X_1), Var(\bar{X}_n)$

פיתרון 2. א.

• נמצא את $E(X_1)$:

$$E(X_1) = \sum_x x \cdot P_{X_1}(x) = 0.1 \cdot 5 + 0.2 \cdot 4 + 0.3 \cdot 3 + 0.4 \cdot 2 = 3$$

• נמצא $Var(X_1)$:

$$Var(X_1) = \overbrace{E(X_1^2)}^{=10} - \overbrace{(E(X_1))^2}^9 = 1$$

• נמצא את $E(\bar{X}_n)$:

$$E(\bar{X}_n) = E\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right) = \frac{1}{n} \cdot \sum_{i=1}^n E(X_i) = E(X_i) = 3$$

– נשים לב שזה מוכיח שאם המשתנים המקריים בלתי תלויים אז $E(\bar{X}_n) = E(X_i)$ לכל $1 \leq i \leq n$.

• נמצא $Var(\bar{X}_n)$:

$$Var(\bar{X}_n) = Var\left(\frac{1}{n} \cdot \sum_{i=1}^n X_i\right)$$

– ומכיוון שהמשתנים המקריים בלתי תלויים מתקיים:

$$Var(\bar{X}_n) = \frac{1}{n^2} \cdot \sum_{i=1}^n \overbrace{Var(X_i)}^{=1} = \frac{1}{n}$$

2. ב. העזרו באי שוויון צ'ביצ'ב על מנת למצוא את גודל המדגם המינימלי n הדרוש, עבורו ההסתברות שהממוצע סוטה מהתוחלת בכלל היותר 0.01 היא לפחות 95%.

2. ב. פיתרון:

$$P\left(\left|\bar{X}_n - \underbrace{E(\bar{X}_n)}_{\mu}\right| < 0.01\right) \geq 0.95$$

• אנחנו מחפשים את $0.95 \geq P\left(\left|\bar{X}_n - \underbrace{E(\bar{X}_n)}_{\mu}\right| < 0.01\right)$.

• נשים לב שהסימן של אי השוויון הפוך ולכן ניקח משלים:

$$1 - P(|\bar{X}_n - \mu| < 0.01) = P(|\bar{X}_n - 3| \geq 0.01) \leq 1 - 0.95 = 0.05$$

– מאי שוויון צ'ביצ'ב מתקיים:

$$P(|\bar{X}_n - 3| \geq 0.01) \leq \frac{Var(\bar{X}_n)}{(0.01)^2}$$

* הראנו בסעיף א' $Var(\bar{X}_n) = \frac{1}{n}$ ולכן:

$$P(|\bar{X}_n - 3| \geq 0.01) \leq \frac{\frac{1}{n}}{(0.01)^2} = \frac{10,000}{n}$$

* נדרוש: $\frac{\frac{1}{n}}{(0.01)^2} < 0.05$ וזה מתקיים עבור $n \geq 20,000$.

ג. צ"ל: החוק החלש של המספרים הגדולים מתקיים בשאלה.

פיתרון 2. ג.

• אנחנו צריכים להראות שהממוצע אכן שואף לתוחלת עבור $n \rightarrow \infty$.

• יהי $\varepsilon > 0$.

• בסעיף ב' מצאנו:

$$0 \leq P(|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon) \leq \frac{\frac{1}{n}}{\varepsilon^2} = \frac{1}{n \cdot \varepsilon^2}$$

– מתקיים $0 \leq P(|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{\frac{1}{n \cdot \varepsilon^2}} 0$ וגם

* ולכן (לפי משפט הסנדוויץ' מאינפי) נקבל:

$$P(|\bar{X}_n - E(\bar{X}_n)| \geq \varepsilon) \xrightarrow[n \rightarrow \infty]{} 0$$

נושא רביעי - משפט הגבול המרכזי:

משפט הגבול המרכזי:

יהיו X_1, X_2, \dots מ"מ ב"ת שווי התפלגות (i.i.d) עם תוחלת μ ושונות σ^2 .

נגדיר $S_n = \sum_{i=1}^n X_i$ ו- $\bar{X}_n = \frac{S_n}{n}$, להיות הסכום שלהם עד למקום ה- n וממוצעם.

עבור $Z \sim N(0,1)$, מתקיים

$$\boxed{\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{n \rightarrow \infty} Z}$$

בפרט, מתקיים גם

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \xrightarrow{n \rightarrow \infty} \Phi(x)$$

$$P(\bar{X}_n \leq \mu + x\sigma/\sqrt{n}) = P(S_n \leq n\mu + x\sigma\sqrt{n}) \xrightarrow{n \rightarrow \infty} \Phi(x).$$

משמעות:

עבור n מספיק גדול \bar{X}_n בקירוב מתפלג כמו משתנה מקרי נורמלי מהתפלגות $N\left(\mu, \frac{\sigma^2}{n}\right)$

ו- S_n מתפלג בקירוב כמו משתנה מקרי נורמלי מהתפלגות $N(n\mu, n\sigma^2)$.

• כלומר:

\bar{X}_n - מתכנס בהתכנסות חלשה למשתנה מקרי $Y \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.

S_n - מתכנס בהתכנסות חלשה למשתנה מקרי $W \sim \mathcal{N}(n \cdot \mu, n \cdot \sigma^2)$.

• זאת אומרת:

$$P\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq x\right) = P\left(\frac{S_n - n \cdot \mu}{\sigma\sqrt{n}} \leq x\right) \xrightarrow{n \rightarrow \infty} \Phi(x)$$

שאלה 3:

בבנק מסוים לקוחות שונים באים לבצע הפקדות באופן ב"ת.

הסכום בש"ח שכל לקוח מפקיד הוא מ"מ $U_i \sim \text{Uni}[0, 100]$

עפ"י מדיניות הבנק, כל כספר צריך לקרוא למנהל כדי לרוקן את הקופה לאחר 90 הפקדות של לקוחות.

בנוסף, הסכום המירבי שיכול להימצא בקופה הינו 5000 ש"ח.

א. עפ"י משפט הגבול המרכזי, מה (בקירוב) ההסתברות שהמנהל מרוקן מהקופה סכום שהינו גבוה מ-5,000 ש"ח?

ב. נניח שהוחלט לקבוע מחדש את המספר n (במקום 90) של הלקוחות המפקידים אצל כספר מסוים לפני ריקון הקופה. מצא את הערך n המקסימלי המקיים כי המדיניות תופר בהסתברות קטנה מ-0.01.

ג. (עבודה עצמית) חזרו על הסעיף הקודם תוך שימוש באי-שוויון צ'ביצ'ב.

פיתרון 3. א.

• נראה שמתקיימים תנאי משפט הגבול המרכזי:

1. U_i הם משתנים מקריים בלתי תלויים עם אותה התפלגות (אחידה)

2. התוחלת מוגדרת לכל i כי לפי דף הנוסחאות מתקיים $E[U_i] = 50$.

3. השונות מוגדרת לכל i כי לפי הנוסחה של Var של משתנה המתפלג אחיד מתקיים $Var(U_i) = \frac{100^2}{12}$.

• נסמן

$$S_n = \sum_{i=1}^{n=90} U_i$$

• לפי משפט הגבול המרכזי, מכיוון ש- S_n הוא סכום של משתנים מקריים בלתי תלויים ושווי התפלגות עם תוחלת ושונות מוגדרים, מתקיים:

$$S_n \approx \mathcal{N}(n \cdot E[U_i], n \cdot Var(U_i))$$

– במקרה שלנו $n = 90$ ולכן:

$$S \approx \mathcal{N}(90 \cdot E[U_i], 90 \cdot Var(U_i))$$

• נחשב את $P(S > 5000)$.

– נשתמש במשלים כדי לקבל:

$$P(S > 5000) = 1 - P(S \leq 5000)$$

– נצטרך לנרמל את S כדי להשתמש ב- Φ :

$$= 1 - P \left(\overbrace{\frac{S - 90 \cdot E[U_i]}{\sqrt{90 \cdot Var(U_i)}}}^{\approx \mathcal{N}(0,1)} \leq \frac{5,000 - 90 \cdot E[U_i]}{\sqrt{90 \cdot Var(U_i)}} \right)$$

$$= 1 - \Phi(1.827) = 0.0339$$

פיתרון 3. ב.

• כעת כבר לא מגבילים ל-90 לקוחות אלא n לקוחות וההסתברות שהמדיניות כן צריכה להיות קטנה מ-0.01

– המאורע בו המדיניות מופרת הוא $\{S_n > 5000\}$
 – כלומר, צריך לדרוש:

$$P(S_n > 5000) \leq 0.01$$

• בסעיף א' ראינו:

$$P(S_n > 5000) = 1 - P(S_n \leq 5000)$$

– וכן ראינו לפי משפט הגבול המרכזי:

$$S_n \approx \mathcal{N}(n \cdot E[U_i], n \cdot Var(U_i))$$

* ננרמל את S_n כדי לקבל:

$$= 1 - P \left(\overbrace{\frac{S_n - n \cdot E[U_i]}{\sqrt{n \cdot Var(U_i)}}}^{\approx \mathcal{N}(0,1)} \leq \frac{5,000 - n \cdot E[U_i]}{\sqrt{n \cdot Var(U_i)}} \right)$$

• מכיוון שכעת $\frac{S_n - n \cdot E[U_i]}{\sqrt{n \cdot Var(U_i)}} \approx \mathcal{N}(0,1)$, מתקיים:

$$= 1 - \Phi \left(\frac{5,000 - n \cdot E[U_i]}{\sqrt{n \cdot Var(U_i)}} \right)$$

· נדרוש :

$$1 - \Phi \left(\frac{5,000 - n \cdot E[U_i]}{\sqrt{n \cdot Var(U_i)}} \right) \leq 0.01$$

$$\Leftrightarrow$$

$$\Phi \left(\frac{5,000 - n \cdot E[U_i]}{\sqrt{n \cdot Var(U_i)}} \right) \leq 0.99$$

· **הערה:** מכיוון ש- Φ היא CDF אז היא בהכרח פונקציה מונוטונית עולה.

· מתקיים $2.328 = \Phi^{-1}(0.99)$ ולכן :

$$\frac{5,000 - n \cdot \overbrace{E[U_i]}^{=50}}{\sqrt{n \cdot \underbrace{Var(U_i)}_{=833.33}}} \geq 2.328$$

· נקבל משוואה ריבועית והתשובה היא $0 \leq n \leq 87.43$, כלומר ה- n המקסימלי הוא $n = 87$.

שאלה 4:

n בנים ו- $2n$ בנות משתתפים במסיבת ריקודים.

כל בת בוחרת באקראי ובאופן ב"ת את אחד הבנים להיות בין הזוג שלה (בן יכול להיבחר ע"י יותר מבת אחת או לא להיבחר כלל).
א. מצא את תוחלת פרופורציית הבנים שלא נבחרו.

ב. (עבודה עצמית) הוכיחו כי שונות פרופורציית הבנים שלא נבחרו שואפת לאפס כאשר $n \rightarrow \infty$.

ג. הראו שפרופורציית מספר הבנים שלא נבחרו על ידי אף בת מתכנסת בצורה חלשה ל- e^{-2} .

פיתרון 4. א.

• קודם כל צריך להבין מה משמעות הביטוי "פרופורציית הבנים שלא נבחרו":

– נסמן ב- X משתנה מקרי שמציין את מספר הבנים שלא נבחרו.

– כעת פרופורציית הבנים שלא נבחרו הוא מספר הבנים שלא נבחרו חלקי מספר הבנים הכולל:

$$\frac{X}{n}$$

• כלומר צ"ל: $E \left[\frac{X}{n} \right]$

• נחשב את התוחלת בשיטת האינדקטורים:

– נגדיר אינדקטור I_i - הבן ה- i לא נבחר.
* כעת יתקיים

$$X = \sum_{i=1}^n I_i$$

• מלינאריות התוחלת מתקיים:

$$\begin{aligned} E\left[\frac{X}{n}\right] &= \frac{1}{n} \cdot E[X] = \frac{1}{n} \cdot E\left[\sum_{i=1}^n I_i\right] = \frac{1}{n} \cdot \sum_{i=1}^n E[I_i] \\ &= \frac{1}{n} \cdot \sum_{i=1}^n P(i_{th} \text{ boy was not picked}) \end{aligned}$$

– ומכיוון שהבחירות של הבנות הן בלתי תלויות זו בזו, מתקיים לכל i :

$$P(i_{th} \text{ boy was not picked}) = \frac{n-1}{n}$$

– ולכן:

$$= \frac{1}{n} \cdot \sum_{i=1}^n P(i_{th} \text{ boy was not picked}) = \frac{1}{n} \cdot \sum_{i=1}^n \left(\frac{n-1}{n}\right)^{2n} = \left(\frac{n-1}{n}\right)^{2n}$$

פיתרון 4. ב.

• נראה $Var\left(\frac{X}{n}\right) \xrightarrow{n \rightarrow \infty} 0$

• תרגול עצמי.

פיתרון 4. ג.

• נראה שפורפורציית מספר הבנים שלא נבחרו על ידי אף בת מתכנסת בצורה חלשה ל- e^{-2} .

– כלומר המשתנה המקרי $\frac{X}{n}$ מתכנס באופן חלש ל- e^{-2} .

• הערת אגב:

– בסעיף א' הוכחנו ש $E\left[\frac{X}{n}\right] = \left(\frac{n-1}{n}\right)^{2n}$
* אפשר לשים לב ש $E\left[\frac{X}{n}\right] \xrightarrow{n \rightarrow \infty} e^{-2}$

* כלומר רוצים שנראה ש $\frac{X}{n}$ מתכנס לאותו מספר שהתוחלת שלו, $E\left[\frac{X}{n}\right]$, מתכנסת אליו.
 • (מוזכיר את החוק החלש של המספרים הגדולים אבל לא בדיוק אותו דבר)

• נסמן $L = e^{-2}$.

• נסמן $X_n = \frac{X}{n} = \frac{1}{n} \cdot \sum_{i=1}^n I_i$

• יהי $\varepsilon > 0$.

• ניקח את המרחק $|X_n - L|$ ונוסיף ונחסיר את $E(X_n)$ כדי לקבל:

$$|X_n - L| = |X_n - E[X_n] + E[X_n] - L|$$

– ולפי אי שוויון המשולש מתקיים:

$$\leq |X_n - E[X_n]| + |E[X_n] - L|$$

– נשים לב כי $E\left[\frac{X}{n}\right] = \left(\frac{n-1}{n}\right)^{2n} = \left(1 - \frac{1}{n}\right)^{2n} \xrightarrow{n \rightarrow \infty} e^{-2}$ ולכן מתקיים $E[X_n] \xrightarrow{n \rightarrow \infty} e^{-2}$
 – מכיוון ש $E[X_n] \xrightarrow{n \rightarrow \infty} e^{-2}$ אז קיים n גדול מספיק כך ש:

$$\left| L - E\left(\widehat{\frac{X}{n}}_{X_n}\right) \right| \leq \frac{\varepsilon}{2}$$

* כלומר עבור n גדול מספיק, מתקיים:

$$|X_n - L| \leq |X_n - E[X_n]| + \frac{\varepsilon}{2}$$

• המטרה שלנו היא להראות $P(|X_n - L| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$

• מכיוון ש $|X_n - L| \leq |X_n - E[X_n]| + \frac{\varepsilon}{2}$ מתקיים:

$$\{|X_n - L| > \varepsilon\} \Rightarrow \left\{|X_n - E[X_n]| + \frac{\varepsilon}{2} \geq \varepsilon\right\} = \left\{|X_n - E[X_n]| \geq \frac{\varepsilon}{2}\right\}$$

• זאת מכיוון שאם $|X_n - L| > \varepsilon$ אז $|X_n - E[X_n]| + \frac{\varepsilon}{2} \geq \varepsilon$

• מכיוון שהמאורע $\{|X_n - L| > \varepsilon\}$ גורר את המאורע $\{|X_n - E[X_n]| \geq \frac{\varepsilon}{2}\}$, אז מתקיים שהיחס בין ההסתברויות הוא:

$$0 \leq P(|X_n - L| > \varepsilon) \leq P\left(|X_n - E[X_n]| \geq \frac{\varepsilon}{2}\right)$$

· לפי אי שוויון צ'ביצ'ב מתקיים:

$$P\left(|X_n - E[X_n]| \geq \frac{\varepsilon}{2}\right) \leq \frac{\text{Var}(X_n)}{\left(\frac{\varepsilon}{2}\right)^2}$$

· בסעיף ב' ראינו ש- $\text{Var}(X_n) \xrightarrow{n \rightarrow \infty} 0$ ולכן:

$$\frac{\text{Var}(X_n)}{\left(\frac{\varepsilon}{2}\right)^2} \xrightarrow{n \rightarrow \infty} 0$$

· לפי משפט הסנדוויץ' מתקיים:

$$0 \leq P(|X_n - L| > \varepsilon) \leq P\left(|X_n - E[X_n]| \geq \frac{\varepsilon}{2}\right) \xrightarrow{n \rightarrow \infty} 0$$

· כלומר, כנדרש:

$$P(|X_n - L| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$