# הסתברות מ' (00094412) | הרצאה 12 ואחרונה

שם: איל

April 3, 2024

# נושאי השיעור: אי-שוויונות ומשפטי גבול/קירובים

## נושא ראשון - מוטיבציה לאי-שוויונות

- בחיים האמיתיים לא תמיד אפשר לחשב הסתברויות בצורה מדויקת ולכן משתמשים בחסמים (או קירובים) כדי לקבל תשובה מספיק טובה לשימושים פרקטיים.

**דוגמה 1.**

- מטילים קובייה הוגנת $n$ פעמים עבור $n$ גדול מאוד.
- מצאו את ההסתברות שממוצע ההטלות קטן מ-4.
- הניחו כי ההטלות בלתי תלויות.

**פיתרון:**

- נגדיר משתנים מקריים:
  - $X_K$ = תוצאת ההטלה ה-$k$.
- אזי ההסתברות שלו היא $P_{X_K}(x)=\begin{cases}\frac{1}{6} & x=1,\ldots,6\\ 0 & otherwise\end{cases}$
  - מכיוון ש $X_1,X_2,\ldots,X_n$ בלתי תלויים.
- כעת נחשב את סכום ההטלות:

$$S_n=\sum_{k=1}^{n}X_k$$

– **סימון:** ממוצע ההטלות יסומן $\overline{X}_n$:

$$\overline{X}_n = \frac{1}{n}S_n = \frac{1}{n} \cdot \sum_{k=1}^{n} X_k$$

- אנחנו מחפשים את $P\left(\overline{X}_n \leq 4\right)$.
- ניסיון ראשון:

$$P\left(\overline{X}_n \leq 4\right) = \sum_{x\ :\ x<4} P_{\overline{X}_n(x)}$$

– נמצא את הפונקציה $P_{\overline{X}_n}(x)$:

$$P_{\overline{X}_n}(x) = P\left(\overline{X}_n = x\right)$$

$$= P\left(\frac{1}{n}S_n = x\right)$$

$$= P_{S_n}(n \cdot x)$$

* כעת ננסה למצוא את $P_{S_n}(y)$:

$$P_{S_n}(y) = P\left(S_n = y\right)$$

$$= P\left(\sum_{k=1}^{n} X_k = y\right)$$

· זוהי פונקציית הסתברות של פונקציה של משתנה מקרי ולכן:

$$= \sum_{x_1,x_2,\ldots,x_n\in\{1,\ldots,6\},\ \sum x_k=y}^{\infty} P_{X_1,X_2,\ldots,X_n}\left(x_1,x_2,\ldots,x_n\right)$$

· המשתנים המקריים בלתי תלויים ולכן:

$$= \sum_{x_1,x_2,\ldots,x_n\in\{1,\ldots,6\},\ \sum x_k=y}^{\infty} P_{X_1}\left(x_1\right) \cdot P_{X_2}\left(x_2\right) \cdot \ldots \cdot P_{X_n}\left(x_n\right)$$

- נשים לב שזו דרך מסובכת כי אנחנו לא יודעים כמה איברים אנחנו סוכמים ולכן נרצה לחפש דרך פשוטה יותר.

- ניסיון שני: במקום לחשב באופן מדויק, נמצא חסמים על ההסתברות המבוקשת (נתמקד בעיקר בחסמים עליונים).

## נושא שני - אי שוויונות (חסמים)

**הגדרה 2.** אי שוויון (חֶסֶם) מרקוב Markov's Inequality

- יהא $X$ משתנה מקרי (כלשהו) עם תוחלת מוגדרת.
- אזי לכל $0<x$ מתקיים:

$$P\left(X\geq x\right)\leq\frac{E\left(\left|X\right|\right)}{x}$$

הוכחה.

- ראשית נשים לב ש-$P\left(X\geq x\right)\leq P\left(\left|X\right|\geq x\right)$ כי המאורע $P\left(\left|X\right|\geq x\right)$ מכיל את המאורע $P\left(X\geq x\right)$.
  - לכן נניח (בה"כ) ש-$X$ מקבל ערכים אי שליליים בלבד ונראה כי $P\left(\left|X\right|\geq x\right)\leq\frac{E\left(\left|X\right|\right)}{x}$ ואז נקבל $P\left(X\geq x\right)\leq\frac{E\left(\left|X\right|\right)}{x}$
- לפי הגדרת אינדיקטור, לכל $x>0$ מתקיים:

$$\left|X\right|\geq\left|X\right|\cdot\overbrace{I_{\left\{ \left|X\right|\geq x\right\} }}^{\in\left\{ 0,1\right\} }$$

  - בנוסף, אם $\left|X\right|<x$ אז נקבל $0$ באגף ימין ואם $\left|X\right|\geq x$ אז נקבל $\left|X\right|$ באגף ימין ולכן:

$$\left|X\right|\geq x\cdot I_{\left\{ \left|X\right|\geq x\right\} }$$

  - ניקח תוחלת משני האגפים (וראינו בעבר שאם $Z\geq W$ אז $E\left(Z\right)\geq E\left(W\right)$):

$$E\left(\left|X\right|\right)\geq E\left(x\cdot I_{\left\{ \left|X\right|\geq x\right\} }\right)$$

$$=x\cdot E\left(I_{\left\{ \left|X\right|\geq x\right\} }\right)$$

    * תוחלת של אינדיקטור $I_{\left\{ \left|X\right|\geq x\right\} }$ היא ההסתברות שהמאורע יקרה, כלומר $P\left(\left|X\right|\geq x\right)$
    * ולכן:

$$=x\cdot P\left(\left|X\right|\geq x\right)$$

* כלומר קיבלנו:

$$E\left(|X|\right)\geq x\cdot P\left(|X|\geq x\right)$$

· נחלק ב-$x$ ונקבל:

$$P\left(|X|\geq x\right)\leq\frac{E\left(|X|\right)}{x}$$

■

נחזור לדוגמה שהתחלנו איתה ונחשב חסם על ההסתברות:

**המשך דוגמה 1:**

- רצינו לחשב את $P\left(\overline{X}_n<4\right)$:

$$P\left(\overline{X}_n<4\right)=1-P\left(\overline{X}_n\geq4\right)$$

– נשתמש באי שוויון מרקוב כדי לקבל:

$$1-P\left(\overline{X}_n\geq4\right)\geq1-\frac{E\left(\left|\overline{X}_n\right|\right)}{4}$$

* ומכיוון ש-$\overline{X}_n$ חיובי ממש כי הוא ממוצע הטלות, מתקיים:

$$P\left(\overline{X}_n<4\right)\geq1-\frac{E\left(\overline{X}_n\right)}{4}$$

- נמצא את $E\left(\overline{X}_n\right)$:

$$E\left(\overline{X}_n\right)=E\left(\frac{1}{n}\cdot S_n\right)=\frac{1}{n}\cdot E\left(S_n\right)$$

$$=\frac{1}{n}\cdot E\left(\sum_{k=1}^{n}X_k\right)$$

$$= \frac{1}{n} \cdot \sum_{k=1}^{n} E\left(X_k\right)$$

$$= \frac{1}{n} \cdot \sum_{k=1}^{n} 3.5$$

$$= \frac{1}{n} \cdot n \cdot 3.5$$

$$= 3.5$$

– נציב זאת ונקבל:

$$P\left(\overline{X}_n < 4\right) \geq 1 - \frac{E\left(\overline{X}_n\right)}{4} = 1 - \frac{3.5}{4}$$

* כלומר:

$$P\left(\overline{X}_n < 4\right) \geq \frac{1}{8}$$

• נחפש חסם טוב יותר:

**הגדרה 3.** אי שוויון צ'בצ'יב Chebyshev (יותר טוב מאי-שוויון מרקוב אבל דורש יותר תנאים):

• יהא $X$ משתנה מקרי כלשהו עם תוחלת ושונות סופיים.

• לכל $0 < x$ מתקיים:

$$P\left(\left|X - E\left(X\right)\right| \geq x\right) \leq \frac{Var\left(X\right)}{x^2}$$

• או באופן שקול, לכל $0 < t$ מתקיים:

$$P\left(\left|X - E\left(X\right)\right| \geq t \cdot SD\left(X\right)\right) \leq \frac{1}{t^2}$$

הוכחה.

- יהי $0 < x$.
- נבחן את אגף שמאל ונעלה את שני האגפים שלו בריבוע:

$$P\left(\left|X-E\left(X\right)\right|\geq X\right)=P\left(\left(X-E\left(X\right)\right)^{2}\geq x^{2}\right)$$

  - לפי אי שוויון מרקוב:

$$P\left(\left(X-E\left(X\right)\right)^{2}\geq x^{2}\right)\leq\frac{E\left(\left(X-E\left(X\right)\right)^{2}\right)}{x^{2}}=\frac{Var\left(X\right)}{x^{2}}$$

- כעת נראה את הגרסה השקולה:
  - יהי $0 < t$.
    * נבחן שוב את אגף שמאל $P\left(\left|X-E\left(X\right)\right|\geq t\cdot SD\left(X\right)\right)$:
      · לפי החלק הראשון נקבל עם $x=t\cdot SD\left(X\right)$:

$$P\left(\left|X-E\left(X\right)\right|\geq t\cdot SD\left(X\right)\right)=P\left(\left(X-E\left(X\right)\right)^{2}\geq\left(t\cdot SD\left(X\right)\right)^{2}\right)\leq\frac{Var\left(X\right)}{\left(t\cdot SD\left(X\right)\right)^{2}}=\frac{1}{t^{2}}$$

■

**נחזור לדוגמה 1:**

- כאמור, חיפשנו את

$$P\left(\overline{X}_{n}<4\right)=1-P\left(\overline{X}_{n}\geq4\right)$$

  - בנוסף, מצאנו ש $E\left(\overline{X}_{n}\right)=3.5$ ולכן אם נחסר אותו משני האגפים של אי השוויון בהסתברות נקבל:

$$P\left(\overline{X}_{n}\geq4\right)=P\left(\overline{X}_{n}-E\left(\overline{X}_{n}\right)\geq4-\underbrace{E\left(\overline{X}_{n}\right)}_{=0.5}\right)=P\left(\overline{X}_{n}-E\left(\overline{X}_{n}\right)\geq\frac{1}{2}\right)$$

    * ראינו שמתקיים:

$$P\left(\overline{X}_{n}-E\left(\overline{X}_{n}\right)\geq\frac{1}{2}\right)\leq P\left(\left|\overline{X}_{n}-E\left(\overline{X}_{n}\right)\right|\geq\frac{1}{2}\right)$$

* נשתמש באי-שוויון צ'בצ'יב עבור $\overline{X}_n$ ונקבל:

$$\leq \frac{Var\left(\overline{X}_n\right)}{\left(\frac{1}{2}\right)^2}$$

· נמצא את $Var\left(\overline{X}_n\right)$:

$$Var\left(\overline{X}_n\right) = Var\left(\frac{1}{n}\cdot S_n\right) = \frac{1}{n^2}Var\left(S_n\right)$$

$$= \frac{1}{n^2}Var\left(\sum_{k=1}^{n}X_k\right)$$

· כל ה-$X_k$ בלתי תלויים ולכן:

$$= \frac{1}{n^2}\sum_{k=1}^{n}Var\left(X_k\right)$$

$$= \frac{1}{n^2}\cdot n\cdot Var\left(X_1\right)$$

$$= \frac{1}{n}\cdot 3$$

* ולכן:

$$P\left(\overline{X}_n < 4\right) \leq 1 - \frac{Var\left(\overline{X}_n\right)}{\left(\frac{1}{2}\right)^2} = 1 - \frac{3}{n}\cdot 4 = 1 - \frac{12}{n}$$

• לסיכום:

$$P\left(\overline{X}_n < 4\right) \leq 1 - \frac{12}{n}$$

**דוגמה 4.**

• הראו שממוצע ההטלות הולך ומתקרב לתוחלת של הטלה בודדת באופן הבא:

– הראו שלכל $\varepsilon > 0$ ההסתברות שהממוצע יהיה רחוק מ-3.5 בלפחות $\varepsilon$ שואף ל-0 כאשר מספר ההטלות שואף ל-$\infty$.

**פיתרון:**

- צריך להראות שלכל $\varepsilon > 0$ מתקיים:

$$P\left(\left|\overline{X}_n - 3.5\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

- נציב $E\left(\overline{X}_n\right) = 3.5$:

$$P\left(\left|\overline{X}_n - 3.5\right| > \varepsilon\right) = P\left(\left|\overline{X}_n - E\left(\overline{X}_n\right)\right| > \varepsilon\right)$$

– לפי אי שוויון צ'יבצ'יב מתקיים:

$$P\left(\left|\overline{X}_n - E\left(\overline{X}_n\right)\right| > \varepsilon\right) \le \frac{\overbrace{Var\left(X\right)}^{=\frac{3}{n}}}{\varepsilon^2} = \frac{3}{n\cdot\varepsilon^2}$$

* נשים לב כי לכל $\varepsilon > 0$ מתקיים $\frac{3}{n\cdot\varepsilon^2} \xrightarrow[n\to\infty]{} 0$

# נושא שלישי - סכומים וממוצעים של משתנים מקריים בלתי תלויים שווי התפלגות (iid - identically independently distributed)

- יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים שווי התפלגות עם תוחלת $\mu \in \mathbb{R}$ ושונות $\sigma^2 > 0$ מוגדרות.
- נגדיר:
  – סכום המשתנים המקריים:

$$S_n = \sum_{k=1}^{n} X_k$$

  – ממוצע המשתנה המקרי:

$$\overline{X}_n = \frac{1}{n}\cdot S_n$$

- מכיוון שהמשתנים המקריים הם שווי התפלגות אז יש להם אותה תוחלת ולכן התוחלת של $S_n$ היא:

$$E\left(S_n\right) = E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E\left(X_k\right) = n\cdot\mu$$

• התוחלת של $\overline{X}_n$ היא:

$$E\left(\overline{X}_n\right) = E\left(\frac{1}{n}\cdot S_n\right) = \frac{1}{n}\cdot E\left(S_n\right) = \mu$$

• השונות של $S_n$ היא:

$$Var\left(S_n\right) = Var\left(\sum_{k=1}^{n} X_k\right)$$

– ומכיוון ש$X_k$ בלתי תלויים לכל $k$, מתקיים:

$$= \sum_{k=1}^{n} Var\left(X_k\right)$$

$$= n\cdot\sigma^2$$

• השונות של $\overline{X}_n$ היא:

$$Var\left(\overline{X}_n\right) = Var\left(\frac{1}{n}\cdot S_n\right) = \frac{1}{n^2}Var\left(S_n\right) = \frac{\sigma^2}{n}$$

הערה 5. לממוצע $\overline{X}_n$ יש תוחלת $\mu$ כמו של כל $X_K$ אבל השונות הולכת וקטנה ככל ש-$n$ גדל. כלומר $\overline{X}_n$ "מתקרב" להיות משתנה מקרי קבוע $\mu$ ככל ש-$n$ גדל.
נפרמל את המונח "מתקרב":

**הגדרה 6.** התכנסות חלשה לקבוע (התכנסות בהסתברות):

• יהיו $Y_1, Y_2, \ldots$ משתנים מקריים כלשהם (לא בהכרח מוגדרים על אותו מרחב מדגם) ויהא $Y$ משתנה מקרי קבוע
• נאמר כי הסדרה $\left(Y_n\right)_{n\geq 1}$ מתכנסת באופן חלש ל- $Y$ כאשר $n\to\infty$ אם לכל $\varepsilon > 0$ מתקיים:

$$P\left(\left|Y_n - y\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

• באופן שקול לפי הגדרת הגבול:

– לכל $\varepsilon_1, \varepsilon_2 > 0$ קיים $n_0 \in \mathbb{N}$ כך שלכל $n > n_0$ מתקיים:

$$P\left(\left|Y_n - y\right| > \varepsilon_1\right) < \varepsilon_2$$

- <u>סימון</u>: $Y_n \underset{n\to\infty}{\Rightarrow} Y$

**משפט 7.** החוק החלש של המספרים הגדולים.

- יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי תלויים שווי התפלגות עם תוחלת $\mu \in \mathbb{R}$ מוגדרת.
- אזי:

$$\overline{X}_n \underset{n\to\infty}{\Rightarrow} \mu$$

  – כלומר לכל $\varepsilon > 0$ מתקיים:

$$P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) \xrightarrow[n\to\infty]{} 0$$

הוכחה.

- לשם פשטות, נניח שגם השונות של $X_1, X_2, \ldots$ מוגדרת.
  – כלומר $\sigma^2 = Var\left(X_k\right)$
- ראינו שמתקיים:

$$P\left(\left|\overline{X}_n - \mu\right| > \varepsilon\right) = P\left(\left|\overline{X}_n - E\left(\overline{X}_n\right)\right| > \varepsilon\right)$$

$$= P\left(\left(\overline{X}_n - E\left(\overline{X}_n\right)\right)^2 > \varepsilon^2\right)$$

  – ואז לפי אי-שוויון צ'יבצ'יב:

$$\le \frac{Var\left(\overline{X}_n\right)}{\varepsilon^2} = \frac{\frac{\sigma^2}{n}}{\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

■

**דוגמה 8.**

- ברשותכם מטבע עם הסתברות $p \in (0,1)$ לקבל "ראש".
- מציעים להטיל את המטבע $n$ פעמים ולתת כניחוש ל-$p$ את שיעור ה"ראש" שיתקבל.

- כמה הטלות צריך לעשות כדי שהניחוש יהיה קרוב ל-$p$ בלכל היותר $\delta = 0.1$ בהסתברות לפחות $1-\varepsilon$ עבור $\varepsilon = 0.05$.
- הניחו כי ההטלות בלתי תלויות.

**פיתרון:**

- נגדיר:

1. $S_n$ = מספר ה"ראש" שהתקבלו ב-$n$ ההטלות. (מתפלג בינומית)
2. $\overline{X}_n = \frac{1}{n} \cdot S_n$ = שיעור הראשים שהתקבלו ב-$n$ ההטלות.

- אנחנו מחפשים $n$ מינימלי כך שלכל $p$ יתקיים:

$$P\left(\left|\overline{X}_n - p\right| > \delta\right) \leq \varepsilon$$

- נשים לב כי $S_n$ הוא סכום של אינדיקטורים:
  - לכן נגדיר $X_k$ להיות האינדיקטורים של האם יצא "ראש" בהטלה ה-$k$:

$$X_k = \begin{cases} 1 & head\ is\ k_{th}\ roll \\ 0 & otherwise \end{cases}$$

  - אזי מכיוון שכל האינדיקטורים הם בלתי תלויים ושווי התפלגות מתקיים:

$$Ber\left(p\right) \sim X_1, X_2, \ldots, X_n$$

  - וגם מתקיים $S_n = \sum_{k=1}^{n} X_k$.
  - נחשב את $\mu$ (תוחלת של משתנה מקרי ברנולי):

$$\mu = E\left[X_k\right] = p$$

  - נחשב את $\sigma^2$:

$$\sigma^2 = Var\left(X_k\right) = p\left(1-p\right)$$

- קיבלנו:

$$E\left(\overline{X}_n\right) = \mu = p$$

$$Var\left(\overline{X}_n\right)=\frac{\sigma^2}{n}=\frac{p\cdot(1-p)}{n}$$

– לכן לפי אי שוויון צ'ביצ'יב מתקיים:

$$P\left(\left|\overline{X}_n-p\right|>\delta\right)=P\left(\left|\overline{X}_n-E\left(\overline{X}_n\right)\right|>\delta\right)$$

$$\leq\frac{Var\left(\overline{X}_n\right)}{\delta^2}=\frac{\frac{p(1-p)}{n}}{\delta^2}$$

– כאמור, מחפשים $n$ מינימלי ולכן נדרוש:

$$\frac{\frac{p(1-p)}{n}}{\delta^2}<\varepsilon$$

$$n\geq\frac{p\left(1-p\right)}{\delta^2\varepsilon}$$

* מכיוון ש-$p$ לא ידוע, ניקח $p$ מקסימלי, כלומר $\arg_p\min p\left(1-p\right)=\frac{1}{2}$:

$$\frac{p\left(1-p\right)}{\delta^2\varepsilon}=\frac{1}{4}\cdot\frac{1}{\delta^2\varepsilon}$$

* לכן נקבל:

$$n=\left\lceil\frac{1}{4\cdot\delta^2\cdot\varepsilon}\right\rceil$$

## נושא רביעי - משפט הגבול המרכזי

ראינו שאם $X_1,X_2,\ldots$ הם בלתי תלויים ושווי התפלגות עם תוחלת ושונות מוגדרות אז:

- $E\left(\overline{X}_n\right)=\mu$
- $Var\left(\overline{X}_n\right)=\frac{\sigma^2}{n}\iff SD\left(\overline{X}_n\right)=\frac{\sigma}{\sqrt{n}}$
  – באופן לא פורמלי: (סטיות מסדר גודל של $\frac{\sigma}{\sqrt{n}}$) $\overline{X}_n=\mu+$

- נרצה להבין איך הסטיות האלה מסדר גודל של $\frac{\sigma}{\sqrt{n}}$ מתנהגות.
  - לכן ננרמל את המשתנה המקרי, כלומר נחסר ממנו את התוחלת ונחלק בסטיית התקן:

$$\hat{X}_n := \frac{\overline{X}_n - E\left(\overline{X}_n\right)}{SD\left(\overline{X}_n\right)}$$

* נחשב את התוחלת והשונות של $\hat{X}_n$:

$$E\left(\hat{X}_n\right) = E\left(\frac{\overline{X}_n - E\left(\overline{X}_n\right)}{SD\left(\overline{X}_n\right)}\right) \overbrace{=}^{linearity} \frac{E\left(\overline{X}_n\right) - E\left(\overline{X}_n\right)}{SD\left(\overline{X}_n\right)} = 0$$

$$Var\left(\hat{X}_n\right) = \frac{Var\left(\overline{X}_n\right)}{\left(SD\left(\overline{X}_n\right)\right)^2} = 1$$

· כלומר אחרי שנרמלנו, נקבל שתמיד התוחלת היא 0 והשונות היא 1.
· משפט הגבול המרכזי יגיד ש$\hat{X}_n$ "שואף" להתפלג נורמלית עם פרמטרים $0, 1$.

**הגדרה 9.** התכנסות חלשה למשתנה מקרי רציף:

- יהיו $Y_1, Y_2, \ldots$ משתנים מקריים כלשהם ו-$Y$ משתנה מקרי <u>רציף</u> (לא בהכרח על אותו מרחב מדגם).
- נאמר כי הסדרה $\left(Y_n\right)_{n\geq 1}$ מתכנסת באופן חלש ל-$Y$ כאשר $n \to \infty$ אם לכל $y \in \mathbb{R}$ מתקיים:

$$F_{Y_n}\left(y\right) \xrightarrow[n\to\infty]{} F_Y\left(y\right)$$

- באופן מפורש:
  - לכל $y \in \mathbb{R}$ מתקיים:

$$\lim_{n\to\infty} P\left(Y_n \leq y\right) = P\left(Y \leq y\right)$$

**משפט 10.** משפט הגבול המרכזי (המשפט הוא מרכזי, לא הגבול).

- יהיו $X_1, X_2$ משתנים מקריים בלתי תלויים ושווי התפלגות עם תוחלת ושונות מוגדרים.
- אזי:

$$\hat{X}_n \underset{n\to\infty}{\Rightarrow} Z$$

- כאשר $Z \sim \mathcal{N}(0,1)$.

- באופן מפורש:
  - לכל $y \in \mathbb{R}$ מתקיים:

$$P\left(\hat{X}_n \leq y\right) \xrightarrow[n\to\infty]{} \Phi(y)$$

**המשמעות אינטואיטיבית של המשפט:**

- משפט הגבול המרכזי אומר שאם $n$ מספיק גדול אז $\mathcal{N}(0,1) \overbrace{\approx}^{almost\ \ disributed} \hat{X}_n$.
- תזכורת: $\hat{X}_n := \frac{\overline{X}_n - E\left(\overline{X}_n\right)}{SD\left(\overline{X}_n\right)}$.
  - לכן:

$$\overline{X}_n = \mu + \frac{\sigma}{\sqrt{n}} \cdot \hat{X}_n$$

- תזכורת:
  - אם $Z \sim \mathcal{N}(0,1)$ אז $aZ + b \sim \mathcal{N}\left(b, a^2\right)$
- ולכן אצלנו עבור $n$ מספיק גדול, נקבל בקירוב:

$$\overline{X}_n \approx \mathcal{N}\left(\overbrace{\mu}^{=E\left(\overline{X}_n\right)}, \overbrace{\frac{\sigma^2}{n}}^{=Var\left(\overline{X}_n\right)}\right)$$

- אם נרצה לחשב את ההתפלגות של $S_n$ אז נשתמש בקשר $S_n = n \cdot \overline{X}_n$ ונקבל:

$$S_n \approx \mathcal{N}\left(\overbrace{n \cdot \mu}^{=E(S_n)}, \overbrace{\sigma^2 \cdot n}^{=Var(S_n)}\right)$$

- כלומר, עבור $n$ מספיק גדול, הממוצע והסכום של $X_1, X_2, \ldots$ (משתנים מקריים בלתי תלויים ושווי התפלגות) מתנהגים כאילו הם מתפלג נורמלי עם הפרמטרים המתאימים.

**דוגמה 11.**

- בהמשך לדוגמאות הקודמות, השתמשו במשפט הגבול המרכזי כדי לקרב את ההסתברות שהממוצע של $n=10^5$ הטלות קובייה בלתי תלויות יפול בקטע $[3.49, 3.51]$.

**פיתרון:**

- נחפש את $P\left(3.49 \le \overline{X}_n \le 3.51\right)$.

- לפי המשמעות האינטואיטיבית של משפט הגבול המרכזי מתקיים $$\overline{X}_n \approx \mathcal{N}\left(\overbrace{\mu}^{=3.5}, \overbrace{\frac{\sigma^2}{n}}^{=\frac{3}{10^5}}\right)$$

  – לכן:

$$P\left(3.49 \le \overline{X}_n \le 3.51\right) \approx P\left(\mathcal{N}\left(3.5, \frac{3}{10^5}\right) \in [3.49, 3.51]\right)$$

$$= \Phi\left(\frac{3.51-3.5}{\sqrt{\frac{3}{10^5}}}\right) - \Phi\left(\frac{3.49-3.5}{\sqrt{\frac{3}{10^5}}}\right)$$

$$= 0.972$$

## נושא חמישי - קירוב נורמלי ופואסוני להתפלגות הבינומית

- ראינו בדוגמה מספר 8 שאם $S_n \sim Bin\left(n,p\right)$ אז $S_n = \sum_{k=1}^n X_k$ כאשר $X_1, X_2, \ldots, X_n \sim Ber\left(p\right)$ הם משתנים מקריים בלתי תלויים ושווי התפלגות.

  – לכן, לפי המשמעות של משפט הגבול המרכזי מתקיים עבור $n$ גדול מאוד:

$$S_n \approx \mathcal{N}\left(\overbrace{n \cdot p}^{=n \cdot \mu}, \overbrace{n \cdot p\left(1-p\right)}^{n \cdot \sigma^2}\right)$$

- מצד שני, כשלמדנו את הקירוב הפואסוני להתפלגות הבינומית ראינו:

  – אם $S_n \sim Bin\left(n, p=\frac{\lambda}{n}\right)$ אז לכל $i \in \mathbb{R}$ מתקיים:

$$P\left(S_n = y\right) \xrightarrow[n\to\infty]{} P\left(W = y\right)$$

* כאשר $W\sim Pois\left(\lambda\right)$.

- באופן אינטואיטיבי, $S_n\approx Pois\left(\lambda=n\cdot p\right)$
- במילים אחרות, עבור $n$ גדול מאוד יש שתי דרכים לקרב משתנה מקרי בינומי:
  - קירוב נורמלי אומר:

$$Bin\left(n,p\right)\approx\mathcal{N}\left(np,np\left(1-p\right)\right)$$

* עבור $p$ קבוע ו-$n$ גדול. מסמנים $p>\frac{10}{n}$.

  - והקירוב הפואסוני אומר:

$$Bin\left(n,p\right)\approx Pois\left(np\right)$$

* כאשר $p=\frac{\lambda}{n}$, כלומר $p$ הוא סדר גודל של $\frac{1}{n}$. מסומן ב-$p\leq\frac{10}{n}$

- אז מי מהקירובים נכון? לא נכנסנו לזה בהרצאה.
  - אם ירצו שנשתמש בקירוב במבחן יהיו צריכים להגיד לנו באיזה קירוב להשתמש.

**הגדרה 12. התכנסות חלשה במקרה שכל המשתנים המקריים מקבליים ערכים ב-$\mathbb{N}\cup\{0\}$**

- יהיו $Y_1,Y_2,\ldots$ וגם $Y$ משתנים מקריים בדידים המקבלים ערכים בתחום $\mathbb{N}\cup\{0\}$ בלבד.
- נאמר כי $\left(Y_n\right)_{n\geq1}$ מתכנסת באופן חלש ל-$Y$ כאשר $n\to\infty$ אם:

$$P_{Y_n}\left(y\right)\xrightarrow[n\to\infty]{}P_Y\left(y\right)$$

  - בכל $y\in\mathbb{N}\cup\{0\}$.
- הסימון הוא אותו סימון כמו של ההתכנסויות לפני כן, $\underset{n\to\infty}{\Rightarrow}$.

**משפט 13.** קירוב פואסוני למשתנה מקרי בינומי:

- אם $S_n\sim Bin\left(n,\frac{\lambda}{n}\right)$ עבור $\lambda>0$ ו-$W\sim Pois\left(\lambda\right)$
  - אז $S_n\underset{n\to\infty}{\Rightarrow}W$.

- נראה אי שוויון נוסף שהוא חסם טוב יותר מאי-השוויונות הקודמים אבל הוא דורש יותר תנאים:

**משפט 14. חסם צ'רנוף / הופצינג / קרמר.**

- יהא $X$ משתנה מקרי כלשהו.
- אז לכל $x \in \mathbb{R}$ מתקיים:

$$P\left(X > x\right) \leq \inf_{s>0}\left\{\frac{E\left(e^{s\cdot X}\right)}{e^{s\cdot x}}\right\}$$

  – בתנאי שהתוחלת מוגדרת.

הוכחה.

- יהי $s > 0$:

$$P\left(X > x\right) = P\left(s\cdot X > s\cdot x\right)$$

  – נעלה את שני האגפים של אי השוויון ב-$e$:

$$= P\left(e^{s\cdot X} > e^{s\cdot x}\right)$$

    * ולפי אי שוויון מרקוב מתקיים:

$$\leq \frac{E\left(e^{s\cdot X}\right)}{e^{s\cdot x}}$$

  – הראנו את זה לכל $s$ ובפרט עבור $0 < s$ שהוא האינפימום.

■

- מזה אפשר להוכיח את חסם צ'רנוף:

**משפט 15.** חסם צ'רנוף.

  – יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים שווי התפלגות המקבלים ערכים בקטע $[a,b]$ בלבד.

    * אז לכל $x \in \mathbb{R}$ מתקיים:

$$P\left(\left|\overline{X}_n - E\left(\overline{X}_n\right)\right| \geq x\right) \leq 2\cdot e^{-\frac{2\cdot n\cdot x^2}{(b-a)}}$$

**נחזור לדוגמה עם המטבעות ונשתמש בחסם צ'רנוף:**

- עם אי שוויון צ'ביצ'יב קיבלנו חסם

$$P\left(\left|\overline{X}_n-E\left(\overline{X}_n\right)\right|\geq x\right)\leq\frac{\sigma^2}{x^2\cdot n}$$

**דוגמה 16.**

- חזרו על הדוגמה עם המטבע ו-$p$ לא ידוע, תוך שימוש באי שוויון צ'רנוף.

**פיתרון:**

- הגדרנו את $\overline{X}_n$ להיות שיעור ה"ראש" וראינו:

$$\overline{X}_n=\frac{1}{n}\cdot\sum_{k=1}^{n}X_k$$

  - עבור $X_k\sim Ber\left(p\right)$

- אזי $X_k$ מקבל ערכים בקטע $[0,1]$ ומתקיים $E\left(X_k\right)=p=E\left(\overline{X}_n\right)$
- לכן נקבל:

$$P\left(\left|\overline{X}_n-p\right|\geq\delta\right)$$

$$=P\left(\left|\overline{X}_n-E\left(\overline{X}_n\right)\right|\geq\delta\right)$$

  - ולפי אי שוויון צ'רנוף מתקיים:

$$\leq2\cdot e^{-2\frac{n\cdot\sigma^2}{1-0}}$$

  - נדרוש:

$$\leq2\cdot e^{-2\frac{n\cdot\sigma^2}{1}}\leq\frac{\varepsilon}{2}$$

    * נקבל:

$$-2n\delta^2\leq\ln\left(\frac{\varepsilon}{2}\right)$$

- כלומר:

$$n \geq \left\lceil \frac{\ln\left(\frac{\varepsilon}{2}\right)}{2 \cdot \delta^2} \right\rceil$$