הרצאה 1

1 אי שוויונות ריכוז:

משפט 1.1: אי שוויון מרקוב: יהי X משתנה מקרי (מ"מ) אי-שלילי. אזי לכל ממשי t מתקיים:

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}$$

באינטואיציה: אם מ"מ מקיים תנאים מסוימים, אז ההסתברות שהוא הרבה יותר מהתוחלת שלו, קטנה. כלומר, הוא "כנראה" "בערך" התוחלת שלו.

אפשר לכתוב גם: $\mathbb{P}\left(X \geq \lambda \mathbb{E}(X)\right) \leq \frac{1}{\lambda}$. כלומר ההסתברות שX גדול מ-למדא פעמים התוחלת, הוא 1 חלקי למדא. ככל שלמדא גדל, ההסתברות קטנה.

בכל שימוש של אי"ש מרקוב צריך לציין שהמשתנה המקרי הוא אי-שלילי!

לדוגמה: נניח שיש 100 סטודנטים בכיתה. אז מרחב המדגם הוא הסטודנטים, ונבחר סטודנט באופן מקרי ואחיד. נגדיר מ"מ: $X(\omega) = \text{grade}$, הציון של הסטודנט (בין 0-100). אז X הוא מ"מ אי"ש. נניח שהציון הממוצע הוא 50. נשאל, מה ההסתברות שסטודנט מסוים קיבל 100. אי"ש מרקוב אומר לנו שההסתברות שהציון **גדול או שווה 100** היא: $\mathbb{P}(X \geq 100) \leq \frac{50}{100} = \frac{1}{2}$, ומכיוון שהציונים הם לכל היותר 100, זה גם ההסתברות שהציון הוא בדיוק 100.

אי"ש מרקוב מספק חסם, והוא לא תמיד חסם הדוק. כלומר יכול להיות שההסתברות היא הרבה פחות מהחסם.

במקרה שלנו, זה חסם הדוק כי יכול להיות שההסתברות היא אכן חצי. (במקרה הזה זה אומר שחצי מהסטודנטים קיבלו 100, וזה אומר שהחצי השני קיבל 0).

הוכחת המשפט: יהי $t > 0$ ממשי כלשהו, ויהי I_t האינדיקטור למאורע: $X \geq t$. כלומר $I_t = 1$ אם $X \geq t$, ואחרת 0. נשים לב שלפי הגדרה מתקיים: (1) $X \geq t \cdot I_t$. (כי אם X קטן מ-t, האינדיקטור 0, וכי X אי"ש). ולכן:

$$t \cdot \mathbb{P}(X \geq t) = t \cdot \mathbb{P}(I_t = 1) = t \cdot \mathbb{E}(I_t) =^{\text{א}} \mathbb{E}(t \cdot I_t) \leq^{\text{ב}} \mathbb{E}(X)$$

א – לינאריות התוחלת.
ב – מונוטוניות התוחלת, ו- (1).

הערה 1.2: באופן כללי, אי"ש מרקוב הוא הכי טוב שאפשר. כלומר, עם הנתונים המוגדרים במשפט, יכול להיות מצב שבו החסם הזה הדוק ולכן כדי לקבל חסם קטן יותר, נצטרך יותר נתונים. מקרה אחד כזה הוא הדוגמה שהבאנו. באופן יותר כללי, לכל $k \geq 1$ ממשי נוכל להגדיר:

$$X_k \sim f(x) = \begin{cases} k, & \frac{1}{k} \\ 0, & 1 - \frac{1}{k} \end{cases}$$

ונקבל ש: $\mathbb{E}(X_k) = k \cdot \frac{1}{k} + 0 \cdot \left(1 - \frac{1}{k}\right) = 1$, כלומר: $\mathbb{P}(X_k \geq k) = \mathbb{P}(X_k = k) = \frac{1}{k} = \frac{\mathbb{E}(X)}{k}$. קיבלנו שאי"ש מרקוב נותן לנו חסם הדוק.

הערה: מתי נכון להגדיר אינדיקטור?
בדרך כלל כשיש לנו מ"מ שקל לבטא אותו כסכום של הרבה מ"מ פשוטים אחרים.

משפט 1.3: אי-שוויון צ'בישב: יהי X מ"מ עם שונות סופית. אז, לכל $t > 0$ ממשי, מתקיים:

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

כלומר, נסתכל על **המרחק** בין X לתוחלת שלו. מה ההסתברות שהמרחק הוא יותר מ- t? ההסתברות קטנה מהשונות חלקי t בריבוע. ככל ש t (המרחק) גדול יותר, ההסתברות קטנה. ככל שהשונות עולה (יש יותר תוצאות רחוקות מהתוחלת) ההסתברות גדלה.

תזכורת: אם השונות סופית, גם התוחלת קיימת וסופית.

צ'בישב נותן **ריכוז מידה.** כלומר, עד כמה כל התוצאות קרובות לממוצע. לדוגמה בסיפור עם הסטודנטים, אם הממוצע הוא 50 יכול להיות שכולם קיבלו 50 (ריכוז מושלם) או שחצי קיבלו 100 וחצי 0. צ'בישב משתמש בשונות כדי להגיד עד כמה רוב התוצאות נמצאות ב"חלון" כלשהו מסביב לממוצע.

הוכחה: מכיוון ש $\left(\mathrm{X}-\mathbb{E}(\mathrm{X})\right)^2$ הוא מ"מ אי-שלילי, נוכל להשתמש באי"ש מרקוב ולקבל:

$$\mathbb{P}(|\mathrm{X}-\mathbb{E}(\mathrm{X})| \geq \mathrm{t}) =^{ב} \mathbb{P}\left(\left(\mathrm{X}-\mathbb{E}(\mathrm{X})\right)^2 \geq \mathrm{t}^2\right) \leq \frac{\mathbb{E}\left(\left(\mathrm{X}-\mathbb{E}(\mathrm{X})\right)^2\right)}{\mathrm{t}^2} =^{א} \frac{\mathrm{Var}(\mathrm{X})}{\mathrm{t}^2}$$

א – הגדרת שונות.

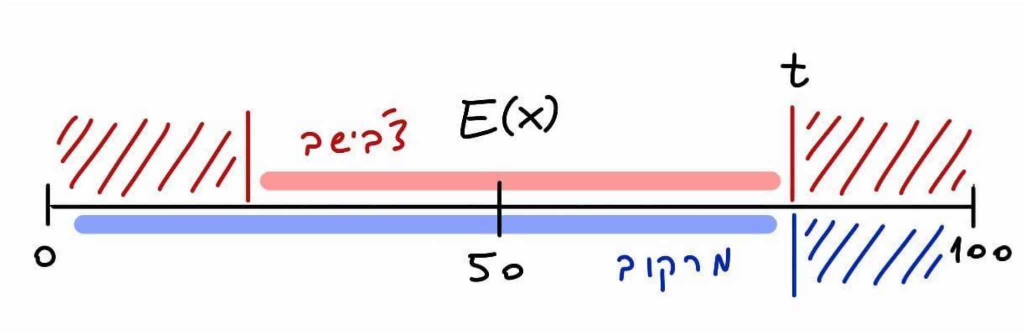
ב – כי $|\mathrm{X}-\mathbb{E}(\mathrm{X})| \geq \mathrm{t}$ אמ"מ $\left(\mathrm{X}-\mathbb{E}(\mathrm{X})\right)^2 \geq \mathrm{t}^2$.

הערה 1.4: עבור $\mathrm{t}=\lambda\sigma_{\mathrm{X}}$, כאשר $\lambda>0$ ממשי, אי"ש צ'בישב אומר ש:

$$\mathrm{P}(|\mathrm{X}\ -\mathrm{E}(\mathrm{X})| \geq\ \lambda\sigma\mathrm{X}) \leq \frac{\mathrm{Var}(\mathrm{X})}{\lambda^2 \cdot \mathrm{Var}(\mathrm{X})} = \frac{1}{\lambda^2}$$

כלומר, ההסתברות שX רחוק מהתוחלת ב-למדא סטיות תקן, קטנה ביחס של למדא בריבוע.

הערה: מתי נשתמש במרקוב ומתי בצ'בישב? צ'בישב חזק יותר, אבל הוא דורש יותר כדי להפעיל אותו. אם מרקוב מספיק אז נשתמש בו. אחרת נשתמש בצ'בישב.



עוד הערה: תמיד אפשר לקחת במרקוב t קטן יותר מהתוחלת, אבל אז נקבל חסם גדול יותר מ-1.

דוגמה 1: במפעל מסוים מיוצרים שולחנות. מספר השולחנות המיוצרים בכל חודש מיוצג ע"י מ"מ X, שהתוחלת שלו היא 100 והשונות 100. נרצה לקבל חסם עליון ל $\mathbb{P}(X \geq 120)$. מכיוון **שX אי שלילי,** אי"ש מרקוב ייתן לנו:

$$\mathbb{P}(X \geq 120) \leq \frac{\mathbb{E}(X)}{120} = \frac{100}{120} = \frac{5}{6}$$

מכיוון **שהשונות של X סופית** נוכל להשתמש גם באי"ש צ'בישב:

$$\mathbb{P}(X \geq 120) = \mathbb{P}(X-100 \geq 20) \leq^{א} \mathbb{P}(|X-100| \geq 20) \leq \frac{Var(X)}{20^2} = \frac{100}{400} = \frac{1}{4}$$

א – המאורע הראשון מוכל במאורע השני: $\mathbb{P}(|X-100| \geq 120) = \mathbb{P}(X \leq 80 \vee X \geq 120)$.

במעבר, החסם נהיה פחות הדוק, אבל הרווחנו את זה שהשתמשנו ביותר מהנתונים – זה שהשונות סופית.

דוגמה 2: נטיל מטבע הוגן 1000 פעמים באופן בת"ל. מ"מ X הוא מספר ההטלות שיצאו עץ (h). נשים לב ש $X\sim Bin(1000,\frac{1}{2})$ ולכן $\mathbb{E}(X)=500$, $Var(X)=1000\cdot\frac{1}{2}\left(1-\frac{1}{2}\right)=250$. אזי:

$$\mathbb{P}(450<X<550)=1-\mathbb{P}(X\leq 450\vee X\geq 550)=1-\mathbb{P}(|X-500|\geq 50)$$
$$=1-\underbrace{\mathbb{P}(|X-\mathbb{E}(X)|\geq 50)}_{\leq 1}\geq^{א} 1-\frac{Var(X)}{50^2}=1-\frac{250}{2500}=0.9$$

א – אם $t\leq s\leq 1$, אז $1-t\geq 1-s$.

באופן אינטואיטיבי, אנחנו יודעים ש"כנראה" נקבל בערך 500 פעמים עץ. בפועל, ההסתברות לקבל **בדיוק** 500 דווקא נמוכה ויש הסתברות חיובית כלשהי לקבל 0 או 1000. צ'בישב מפרמל את זה עם שני פרמטרים:
1) **רוחב החלון** – מה ההגדרה ל"בערך".
2) **ההסתברות** – מה נחשב "הסתברות גבוהה".

דוגמה 3: הרעיון: נגיד שאנחנו רוצים לבדוק שמטבע מסוים הוא הוגן, כלומר $p\approx 1/2$ בקירוב מסוים. נבצע אלגוריתם שנותן לנו מספר q (ההסתברות האמיתית לעץ), ונבדוק עד כמה q שונה מ- p. פורמלית:
נתון מטבע עם הסתברות p לעץ (h) בכל הטלה. נרצה לחשב את p בקירוב. כלומר, בהינתן $\varepsilon,\delta>0$, נרצה q ממשי כך ש: $|q-p|<\varepsilon$ מתקיים בהסתברות לפחות $1-\delta$. האלגוריתם: נקבע n מתאים עבור ה- ε,δ. אם מספר ההטלות שיצאו עץ הוא m, האלגוריתם פולט m/n. יהי X מספר ההטלות שיצאו עץ. נשים לב ש $X\sim Bin(n,p)$ ובפרט $\mathbb{E}(X)=np,\ \ Var(X)=np(1-p)$. השונות סופית אז אי"ש צ'בישב נותן:

$$\mathbb{P}\left(\left|\frac{X}{n}-p\right|\geq\varepsilon\right)=\mathbb{P}(|X-np|\geq\varepsilon n)=\mathbb{P}(|X-\mathbb{E}(X)|\geq\varepsilon n)\leq^{א}\frac{Var(X)}{\varepsilon^2 n^2}=\frac{p(1-p)}{\varepsilon^2 n}$$

נרצה שזה יהיה קטן מדלתא. מכיוון ש $p(1-p)\leq 1/2$, נציב:

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

כלומר, כל n שמקיים $n>\frac{1}{2\varepsilon^2\delta}$ ייתן לנו $\mathbb{P}\left(\left|\frac{X}{n}-p\right|\geq\varepsilon\right)<\delta$, כנדרש. כלומר קיבלנו ש *X/n* קרוב עד כדי אפסילון ל-p, וזה נותן לנו קירוב של p.

משפט 1.5: אי-שוויונות צ'רנוף: אי-שוויונות פחות כלליים אבל יותר חזקים:

עבור $X_1\dots X_n$ מ"מ בת"ל שמחזירים ערכים בטווח [0,1], נגדיר $X=\sum_{i=1}^n X_i$ ואז מתקיים:

א) לכל $t>0$ מתקיימים:

$$\mathbb{P}(X\geq\mathbb{E}(X)+t)\leq e^{-2t^2/n}$$
$$\mathbb{P}(X\leq\mathbb{E}(X)-t)\leq e^{-2t^2/n}$$

ב) לכל $\varepsilon>0$ מתקיים:

$$\mathbb{P}\big(X\leq(1-\varepsilon)\mathbb{E}(X)\big)\leq e^{-\varepsilon^2\mathbb{E}(X)/2}$$

ג) אם, בנוסף, $\varepsilon<3/2$ אז מתקיים:

$$\mathbb{P}\big(X\geq(1+\varepsilon)\mathbb{E}(X)\big)\leq e^{-\varepsilon^2\mathbb{E}(X)/3}$$

(בסעיף ג, מחלקים את המעריך ב-3 - כלומר החסם יותר הדוק).

נתבונן שוב בדוגמה 2, הפעם עם אי-שוויון א. מתקיים:

$$\mathbb{P}(X\leq 450)=\mathbb{P}(X\leq 500-50)\leq e^{-2\cdot\frac{50^2}{1000}}=e^{-5}$$

$$\mathbb{P}(X \geq 550) = \mathbb{P}(X \geq 500 + 50) \leq e^{-5}$$

ובסה"כ: $\mathbb{P}(450 < X < 550) \overset{א}{=} 1 - [\mathbb{P}(X \leq 450) + \mathbb{P}(X \geq 550)] \geq 1 - 2e^{-5} \approx 0.98652$

א – כי המאורעות זרים, ניתן לחבר.

ההוכחה של משפט 1.5 לא בחומר של הקורס. במקום זה, נוכיח משפט דומה אבל יותר פשוט:

משפט 1.6: אי-שוויונות צ'רנוף-הופדינג:

יהיו $X_1 \dots X_n$ מ"מ בת"ל כך ש: $\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = 1/2$ לכל $1 \leq i \leq n$. ויהי $X = \sum_{i=1}^{n} X_i$. אזי, לכל $t > 0$ מתקיים:

$$\mathbb{P}(X \geq t) \leq e^{-t^2/(2n)}$$

$$\mathbb{P}(X \leq -t) \leq e^{-t^2/(2n)}$$

(התוחלת היא 0).

החסם יותר טוב מאשר אי"ש צ'רנוף א, (פי 4 במעריך).