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:

$$t \cdot \mathbb{P}(X \geq t) = t \cdot \mathbb{P}(I_t = 1) = t \cdot \mathbb{E}(I_t) =^{\aleph} \mathbb{E}(t \cdot I_t) \leq^{\beth} \mathbb{E}(X)$$

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:

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

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

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

א – הגדרת שונות.

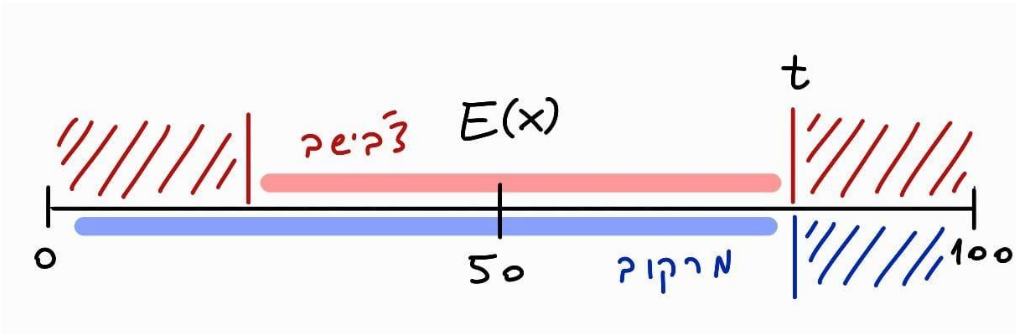
ב – כי $|X-\mathbb{E}(X)| \geq t$ אמ"מ $\left(X-\mathbb{E}(X)\right)^2 \geq t^2$.

הערה 1.4: עבור $t=\lambda\sigma_X$, כאשר $\lambda>0$ ממשי, אי"ש צ'בישב אומר ש:

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

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:

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

א – אם $t\le s\le 1$, אז $1-t\ge 1-s$.

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:

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

נרצה שזה יהיה קטן מדלתא. מכיוון ש $p(1-p)\le 1/2$, 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 $\mathbb{P}\left(\left|\frac{X}{n}-p\right|\ge\varepsilon\right)<\delta$, 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:

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

ב) לכל $\varepsilon>0$ מתקיים:

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

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

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

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

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

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

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

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

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).

הרצאה 2

1 אי-שוויונות ריכוז ומשפטי גבול

משפט 1.1: אי שוויונות צ'רנוף-הופדינג:

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

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

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

(התוחלת היא 0).

החסם פחות טוב מאשר אי"ש צ'רנוף א, (פי 4 במעריך) אבל ההוכחה הרבה יותר פשוטה.

הרעיון המרכזי של ההוכחה: אנחנו רוצים לחסום את ההסתברות של המאורע $\mathbb{P}(X \ge t)$ ע"י אי"ש מרקוב. אבל X הוא לא אי"ש. אז עוברים למשתנה מקרי אחר, $e^{\lambda X}$ שהוא תמיד חיובי. אז נבצע: $\mathbb{P}(X \ge t)=\mathbb{P}(e^{\lambda X} \ge e^{\lambda t})$ וזה יאפשר לנו להשתמש במרקוב ולחסום את ההסתברות עם התוחלת (של המ"מ החדש) ולחסום עם חסם עליון על התוחלת: $\mathbb{P}\left(e^{\lambda X} \ge e^{\lambda t}\right) \le \frac{\mathbb{E}(e^{\lambda X})}{e^{\lambda t}} \le$ $e^{\frac{\lambda^2 n}{2}-\lambda t}$, ומשם יהיה נוח להגיע לחסם של המשפט.

ההוכחה משתמשת בטענה הבאה:

למה 1.2: לכל מספר ממשי $\lambda > 0$ מתקיים:

$$\frac{e^{\lambda}+e^{-\lambda}}{2} \le e^{\lambda^2/2}$$

ההוכחה בעזרת טורי טיילור של שלושת הפונקציות:

$$e^{\lambda}+e^{-\lambda}=\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}+\sum_{n=0}^{\infty}\frac{\overbrace{(-1)^n\lambda^n}^{(-\lambda)^n}}{n!} \overset{\text{ב}}{=} 2\sum_{n=0}^{\infty}\frac{\lambda^{2n}}{(2n)!} \overset{\text{א}}{\le} 2\sum_{n=0}^{\infty}\frac{\lambda^{2n}}{n!\cdot 2^n} \overset{\text{ג}}{=} 2\sum_{n=0}^{\infty}\frac{\left(\frac{\lambda^2}{2}\right)^n}{n!}=2e^{\lambda^2/2}$$

א – כי $(2n)! \ge n!\cdot 2^n$ מתקיים לכל n שלם טבעי[1].
ב – כל האיברים עם n אי זוגי מצטמצמים (חיובי בראשון, שלילי בשני). אז זה בעצם כל ה-n הזוגיים, פעמיים.
ג – הביטוי השמאלי זה פשוט טור טיילור של הפונקציה השלישית.

הוכחת משפט 1.1: נוכיח ש: $\mathbb{P}(X \ge t) \le e^{-t^2/2n}$. לכל שלם ממשי $\lambda>0$, n, $1 \le i \le n$ מתקיים:

$$\mathbb{E}\left(e^{\lambda X_i}\right)=\frac{e^{\lambda}+e^{-\lambda}}{2} \le e^{\lambda^2/2}$$

ע"פ הגדרת התוחלת, וע"פ טענת העזר. מכיוון שכל ה- X_i בת"ל ע"פ ההנחה, גם המ"מ $e^{\lambda X_i}$ הם בת"ל. אז:

$$\mathbb{E}\left(e^{\lambda X}\right)=\mathbb{E}\left(e^{\lambda \sum_{i=1}^{n} X_i}\right) \overset{\text{א}}{=} \mathbb{E}\left(\prod_{i=1}^{n} e^{\lambda X_i}\right) \overset{\text{ב}}{=} \prod_{i=1}^{n}\mathbb{E}\left(e^{\lambda X_i}\right) \overset{\text{ג}}{\le} e^{\lambda^2 n/2}$$

א – כי $e^{a+b}=e^a \cdot e^b$.
ב – כי כל ה- $e^{\lambda X_i}$ בת"ל.
ג – ע"פ טענת העזר.

ולכן: (1)

[1] אפשר להוכיח אלגברית, אבל גם: $(2n)!/n!\,2^n$ זה מספר הדרכים לחלק 2n אנשים ל-n זוגות. אז זה מספר שלם חיובי.

$$\mathbb{P}(X \geq t) =^{א} \mathbb{P}\left(e^{\lambda X} \geq e^{\lambda t}\right) \leq^{ב} \frac{\mathbb{E}\left(e^{\lambda X}\right)}{e^{\lambda t}} \leq e^{\frac{\lambda^2 n}{2} - \lambda t}$$

א – כי $f(x) = e^{\lambda x}$ היא פונקציה מונוטונית עולה. אז אם $X \geq t$, גם $e^{\lambda X} \geq e^{\lambda t}$.
ב – אי"ש מרקוב, מכיוון שהמ"מ החדש הוא אי-שלילי.

אנחנו רוצים למצוא את החסם הטוב ביותר. מכיוון שאי השוויון מתקיים לכל למדא חיובי, אנחנו רוצים להקטין את:

$$g(\lambda) = e^{\frac{\lambda^2 n}{2} - \lambda t}$$

נגזור לפי למדא ונקבל:

$$g'(\lambda) = (\lambda n - t)e^{\frac{\lambda^2 n}{2} - \lambda t}$$

מכיוון שהחלק של e תמיד חיובי, 0 מתקבל רק כאשר $\lambda = t/n$. וניתן לראות שזו נקודת מינימום כי לכל $\varepsilon > 0$ מתקיים:
$g'(t/n + \varepsilon) > 0$, $g'(t/n - \varepsilon) < 0$. כלומר הפונקציה יורדת לפני ועולה אחרי. נציב $\lambda = t/n$ ב-(1):

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

כנדרש.

1.1 משפטי גבולות:

משפט 1.3: החוק החלש של המספרים הגדולים.

יהי $\{X_i\}_{i=1}^{\infty}$ סדרה של מ"מ בת"ל, לכולם אותה התפלגות ותוחלת סופית μ. אזי, לכל $\varepsilon > 0$ מתקיים:

$$\lim_{n \to \infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

באופן אינטואיטיבי, ה"ממוצע" של המ"מ שואף לתוחלת. יותר פורמלית: ככל שסכום יותר מ"מ, ההסתברות שההפרש בין הממוצע שלהם לתוחלת גדול מאפסילון, שואפת לאפס.

ההבדל בין זה לצ'בישב: צ'בישב הוא תוצאה כמותית - חוסם במספר קבוע עבור מספר מסוים של מ"מ. החוק הזה הוא תוצאה איכותית – ככל שיהיו יותר מ"מ, נתכנס למשהו מסוים.

נוכיח מקרה פרטי של המשפט, שבו יש לכל המ"מ שונות סופית σ^2:

לפי לינאריות התוחלת, מתקיים לכל n טבעי: (2)

$$\mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}{n} = \mu$$

ומכיוון שכולם בת"ל, מתקיים: (3)

$$Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\overbrace{Var(X_1)}^{\sigma^2} + \cdots + Var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

ולכן, לכל $\varepsilon > 0$ מתקיים ע"פ אי"ש צ'בישב:

$$\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) =^{2} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right)\right| \geq \varepsilon\right)$$
$$\leq^{א} \frac{Var\left(\frac{X_1 + \cdots + X_n}{n}\right)}{\varepsilon^2} =^{3} \frac{\sigma^2}{\varepsilon^2 n}$$

א – אי"ש צ'בישב.

ובסה"כ:

$$0 \leq \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{\varepsilon^2 n} \overset{א}{=} 0$$

א – כי σ סופי. (דרשנו שונות סופית).

ולכן:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right) = 0$$

לסיכום, המשפט חלש יותר מצ'בישב במובן מסוים, כי הוא לא נותן חסם ספציפי. הוא רק אומר שאם נלך מספיק רחוק אז זה יסתדר.

משפט 1.4: החוק החזק של המספרים הגדולים.

יהי $\{X_i\}_{i=1}^{\infty}$ סדרה של מ"מ בת"ל, לכולם אותה התפלגות, תוחלת סופית μ, ושונות סופית. אזי, לכל $\varepsilon > 0$ מתקיים:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right) = 1$$

כלומר, ההסתברות שהממוצע שואף לתוחלת כאשר n שואף לאינסוף, היא 1. לכאורה כל מה שעשינו זה להחליף בין ההסתברות לגבול, אבל זאת החלפה לא טריוויאלית והיא משנה את המשמעות. בשני החוקים נטען שהסדרה של המ"מ מתכנסת ל-μ. ההבדל הוא בצורת ההתכנסות. כדי להסביר את זה, נגדיר שתי צורות התכנסות:

הגדרה 1.5: נאמר שסדרה $\{X_n\}_{n=1}^{\infty}$ של מ"מ **מתכנסת בהסתברות** למ"מ X כאשר n שואף לאינסוף, אם לכל $\varepsilon > 0$ מתקיים:

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$$

את החלק השמאלי נכתוב בדרך כלל ככה: $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon)$

כלומר, ההסתברות שההפרש בין X_n ל- X גדול מאפסילון, שואפת לאפס כאשר n שואף לאינסוף. נסמן:

$$X_n \xrightarrow{p} X$$

הגדרה 1.6: נאמר שסדרה $\{X_n\}_{n=1}^{\infty}$ של מ"מ **מתכנסת כמעט בוודאות** למ"מ X כאשר n שואף לאינסוף, אם מתקיים:

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

אם מרחב המדגם סופי, זה אומר שלכל מאורע, כאשר n שואף לאינסוף, $X_n(\omega) = X(\omega)$. אם מרחב המדגם אינסופי אז זה קורה ב"כמעט כולם".

כלומר, ההסתברות שנקבל מאורע שעבורו הסדרה שואפת לX, היא 1. נסמן:

$$X_n \xrightarrow{a.s} X$$

באופן כללי, התכנסות "כמעט בוודאות" גוררת התכנסות בהסתברות. בכיוון ההפוך זה לא תמיד נכון, לדוגמה:

יהי $X \equiv 0$ מ"מ ותהי $\{X_n\}_{n=1}^{\infty}$ סדרה בת"ל של מ"מ כך ש $\mathbb{P}(X_n = 1) = 1/n$, $\mathbb{P}(X_n = 0) = 1 - 1/n$ לכל n טבעי. יהי $\varepsilon > 0$ כלשהו, אז מתקיים:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \lim_{n\to\infty} \mathbb{P}(X_n = 1) = \lim_{n\to\infty} 1/n = 0$$

כלומר, $X_n \xrightarrow{p} X$.

מצד שני, נב"ש ש $X_n \xrightarrow{a.s} X$. ע"פ הגדרה 1.6 זה אומר שבהסתברות 1, לכל אפסילון חיובי מתקיים $|X_n - X| \leq \varepsilon$ לכל n החל ממקום מסוים. כלומר לפי ההתפלגות של ה-X_n, בהסתברות 1 מתקיים $X_n = 0$ לכל n החל ממקום מסוים. אבל, לכל m טבעי מתקיים שההסתברות ש $X_n = 0$ לכל $n > m$ היא:

$$\mathbb{P}(\forall n > m, X_n = 0) \overset{\text{א}}{=} \prod_{n=m}^{\infty}(1-\frac{1}{n}) \overset{\text{ב}}{\leq} \prod_{n=m}^{\infty} e^{-1/n} = e^{\overbrace{-\sum_{n=m}^{\infty} 1/n}^{-\infty}} = 0$$

א - מכיוון שהמ"מ כולם בת"ל.
ב – כי $1 - x \leq e^{-x}$.

בסתירה לכך שזה קורה בהסתברות 1.

דוגמה 1: נטיל מטבע n פעמים באופן בת"ל. בהרצאה הקודמת ראינו כלים למציאת על ההסתברות שנקבל עץ הרבה יותר או פחות מחצי מהפעמים. הכלים האלה נתנו תוצאה כמותית – לכל n מסוים, נתנו חסם מסוים. משפטים 1.3, 1.4 נותנים תוצאה איכותית – כלומר, לא נותנים חסם מסוים עבור אף n אלא מבטיחים תוצאה מסוימת עבור n מספיק גדול.
החוק החלש אומר שעבור כל אפסילון ודלתא חיוביים, אם נטיל את המטבע מספיק פעמים אז בהסתברות לפחות $1 - \delta$ נקבל עץ לפחות $(1/2 - \varepsilon)n$ ולכל היותר $(1/2 + \varepsilon)n$ פעמים.

יש עוד דוגמאות בקובץ המקורי.

הרצאה 3

1 מבוא לשיטה ההסתברותית

שימוש בכלים הסתברותיים כדי להוכיח טענות דטרמיניסטיות. לדוגמה: אם נרצה להוכיח שקיים אובייקט עם תכונה מסוימת, נוכל לספור את כל האובייקטים, לספור את האובייקטים ה"רעים" ולהראות שיש פחות רעים מאשר אובייקטים בכלל. בשיטה ההסתברותית, נחשב מה ההסתברות לבחור אובייקט טוב ונראה שההסתברות חיובית.

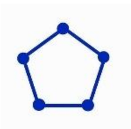
כלים שנשתמש בהם הם: לינאריות התוחלת, והעובדה שתמיד קיים ω כך ש $X(\omega) \le \mathbb{E}(X)$, וגם ω כך ש $X(\omega) \ge \mathbb{E}(X)$. (אם הכל קטן ממש מהתוחלת, התוחלת הייתה קטנה יותר).

1.1 חסם תחתון למספרי רמזי:

עבור שני טבעיים k, ℓ, **מספר רמזי** $R(k,\ell)$ הוא הטבעי n הקטן ביותר כך שלכל גרף על n קודקודים יש קליקה בגודל k או קבוצה בת"ל בגודל ℓ. כלומר, $R(k,\ell) > n$ אם **קיים** גרף על n קודקודים שאין בו קליקה בגודל k או קבוצה בת"ל בגודל ℓ.

(אינטואיציה: אם אין לי הרבה צלעות, תהיה לי קבוצה בת"ל בגודל ℓ. ואם נוסיף מספיק צלעות כדי שלא תהיה קבוצה בת"ל בגודל הזה, תהיה קליקה בגודל k. עבור כל שני מספרים כאלה, אם ניקח n מספיק גדול אז זה יתקיים לכל גרף).

דוגמאות עבור מספרי רמזי "אלכסוניים" – כאשר $k = \ell$.



$R(1,1) = 1$ – עם קודקוד אחד יש גרף יחיד. אז "לכל" גרף יש קליקה או קבוצה בת"ל בגודל 1.
$R(2,2) = 2$ – עם שני קודקודים, יש שני גרפים: מלא או ריק. בכל אחד יש את אחת הדרישות.
$R(3,3) = 6$ – נדגים שזה חייב להיות יותר מ-5: בגרף בתמונה אין אף קליקה או קבוצה בת"ל בגודל 3.
$R(4,4) = 18$ – הסיפור של הניסוי הסוציולוגי: בכל קבוצה של לפחות 18 אנשים, יהיו 4 אנשים שכולם מכירים את כולם או 4 אנשים שאף אחד לא מכיר את השני. זו בעצם תופעה מתמטית ולא סוציולוגית.
$R(5,5) = ?$ – לא ידוע. המספרים פשוט גדולים מדי.
באופן כללי, $R(t,t) \le 4^t$.

משפט 1.1: אם $\binom{n}{t}2^{1-\binom{t}{2}} < 1$, אז $R(t,t) > n$.

כלומר, בהינתן t, נוכל לקבל חסם תחתון עבור n.

הוכחה: נקרא לגרף גרף **טוב** אם אין בו קליקה או קבוצה בת"ל בגודל t. גרף שיש בו אחד מהם ייקרא **רע**.
עבור n נתון, נבנה גרף G על $[n]$ בצורה הבאה: לכל $1 \le i < j \le n$ (כל זוג קודקודים שונים), נטיל מטבע הוגן שייקבע אם יש ביניהם צלע. כל ההטלות בת"ל.
נראה שההסתברות שהגרף המתקבל הוא טוב היא חיובית ממש,
וזה יוכיח שקיים גרף טוב על n קודקודים, כלומר $R(t,t) > n$.
יהיו $S_1 \dots S_{\binom{n}{t}}$ כל תתי הקבוצות של $[n]$ בגודל t.
לכל $1 \le i \le \binom{n}{t}$ נגדיר את המאורע A_i: 'הקבוצה $G[S_i]$ היא קליקה או קבוצה בת"ל'.
כלומר כל ההטלות יצאו עץ או כולם פלי. מתקיים לכל $1 \le i \le \binom{n}{t}$: $\mathbb{P}(A_i) = 2 \cdot 2^{-\binom{t}{2}}$.
צריך שנקבל עץ או פלי (לכן כפול 2), באופן עקבי $\binom{t}{2}$ פעמים. כל פעם הסתברות חצי. ולכן:

$$\mathbb{P}(G \text{ is bad}) \overset{א}{=} \mathbb{P}\left(\bigcup_{i=1}^{\binom{n}{t}} A_i\right) \overset{ב}{\le} \sum_{i=1}^{\binom{n}{t}} \mathbb{P}(A_i) \overset{ג}{=} \binom{n}{t} 2^{1-\binom{t}{2}} \overset{ד}{<} 1$$

א – מספיק שאחת הקבוצות מקיימת את התנאי.
ב – חסם איחוד.
ג – לפי החישוב לעיל.
ד – הנחת הטענה.

הנחנו ש $\binom{n}{t}2^{1-\binom{t}{2}} < 1$, וחישבנו (בלי קשר) את ההסתברות לגרף רע. וראינו שזה יוצא אותו דבר. כלומר אם ההנחה מתקיימת, אז יש הסתברות פחות מ-1 לקבל גרף רע, כלומר הסתברות חיובית לקבל גרף טוב. כלומר צריך יותר מ- n קודקודים כדי שלא יהיה גרף טוב – או במילים אחרות, $R(t,t) > n$.

מסקנה 1.2: לכל $t > 4$ מתקיים $R(t,t) > \lfloor 2^{t/2} \rfloor$. הוכחה: יהי $n = \lfloor 2^{t/2} \rfloor$. ע"פ משפט 1.1, מספיק להוכיח שמתקיים $\binom{n}{t}2^{1-\binom{t}{2}} < 1$. ואכן מתקיים:

$$\binom{n}{t}2^{1-\binom{t}{2}} =^{א} \frac{n(n-1)\cdot\ldots\cdot(n-t+1)}{t!}2^{1-\frac{t(t-1)}{2}}$$

$$\leq^{ב} \frac{n^t}{t!}2^{1-\frac{t^2-t}{2}} \leq^{ג} \frac{\left(2^{\frac{t}{2}}\right)^t \cdot 2^{1-\frac{t^2}{2}+\frac{t}{2}}}{t!}$$

$$=^{ד} \frac{2^{\frac{t^2}{2}+1-\frac{t^2}{2}+\frac{t}{2}}}{t!} =^{ה} \frac{2^{1+\frac{t}{2}}}{t!} \leq^{ז} \frac{2^{t-1}}{t!}$$

$$=^{ח} \prod_{i=2}^{t}\frac{2}{i} <^{ט} 1$$

א – פתיחת הchoose

ב – במונה יש t פעמים משהו שקטן או שווה n

ג – הצד הימני במונה זה המשך פתיחה של השבר. הצד השמאלי זה הצבה של הערך שהגדרנו ל-n.

ד – פתיחת סוגריים.

ה – צמצום.

ז – כי $1 + \frac{t}{2} \leq t - 1$, מתקיים לכל $t \geq 4$.

ח – במכנה כתוב בעצם: $2 \cdot \ldots \cdot (t-1)\ t$.

ט – כי כל מכנה הוא לפחות 2.

1.2 טורניר ותכונת S_k

טורניר על n קודקודים הוא הגרף השלם על n קודקודים, שלכל צלע נתנו כיוון.

כיוון צלע $u \to v$ מתאר "ניצחון" של u על v. יהי $T = (V, E)$ טורניר, תהי $A \subseteq V$, ויהי $u \in V \setminus A$.

נאמר ש u **שולטת** על A אם לכל $v \in A$, כיוון הצלע uv הוא $u \to v$. נסמן: $\overrightarrow{uv} \in E$.

בהינתן k טבעי, נאמר ש T מקיים את S_k אם לכל $A \subseteq V$ בגודל k יש קודקוד $u \in V \setminus A$ ששולט על A.

כלומר לכל קבוצה בגודל k שנבחר, יש קודקוד אחר ששולט עליהם.

טענה 1.3: אם $\binom{n}{k}\left(1-2^{-k}\right)^{n-k} < 1$, אז יש טורניר על n קודקודים שמקיים את S_k.

הוכחה: יהי $T = (V, E)$ טורניר אקראי על $V = \{1,2,\ldots,n\}$,

ולכל $1 \leq u < v \leq n$ נטיל מטבע הוגן (כל ההטלות בת"ל). אם יצא עץ, הכיוון יהיה $\overrightarrow{uv}$ ואחרת בכיוון השני.

כלומר $\overrightarrow{uv} \in E$ מתקיים בהסתברות 1/2.

תהי $A \subseteq V$ כלשהי מגודל k, ויהי קודקוד $x \in V \setminus A$ כלשהו.

ההסתברות ש $\overrightarrow{xa} \in E$ לכל $a \in A$ היא 2^{-k}. כלומר ההסתברות שזה לא מתקיים היא $1 - 2^{-k}$.

ההסתברות שאין אף קודקוד $x \in V \setminus A$ ששולט על A היא $\left(1-2^{-k}\right)^{n-k}$.

(כי צריך שההטלה תצא פלי לכל אחד מהקודקודים שלא בA – יש $n-k$ כאלה. וכל ההטלות בת"ל).

אנחנו רוצים לדעת מה קורה עם **כל** הקבוצות בגודל k. יש $\binom{n}{k}$ כאלה.

נבצע חסם איחוד ונקבל שההסתברות שT לא מקיים את S_k היא לכל היותר $\binom{n}{k}\left(1-2^{-k}\right)^{n-k}$.
כלומר אם זה קטן ממש מ1, אז ההסתברות שT כן מקיים את S_k היא חיובית ממש, כנדרש.

1.3 קבוצות שלטות קטנות בגרפים:

יהי גרף $G=(V,E)$. תת קבוצה $S\subseteq V$ נקראת קבוצה **שלטת** בגרף אם לכל קודקוד שלא בS קיים שכן אחד לפחות בS. (V עצמה שלטת באופן ריק). המטרה שלנו היא להוכיח את קיומה של קבוצה שלטת קטנה.

משפט 1.4: יהי $G=(V,E)$ גרף על n קודקודים, עם דרגה מינימלית $\delta>1$.
אז קיימת בG קבוצה שלטת עם לכל היותר $n\frac{1+\ln(\delta+1)}{\delta+1}$ קודקודים.

הוכחה: יהי $p=\frac{\ln(\delta+1)}{\delta+1}$. נבנה קבוצה אקראית X בצורה הבאה:
נתחיל עם $X=\emptyset$, ולכל $u\in V$ נוסיף את u ל-X בהסתברות p, כאשר כל הבחירות בת"ל.
נשים לב שמתקיים $|X|\sim Bin(n,p)$ ובפרט, $\mathbb{E}(|X|)=np$.
תהי $Y\subseteq V\setminus X$ קבוצת הקודקודים שאין להם שכן בX. לכל $v\in V$, יהי Y_v האינדיקטור למאורע $v\in Y$. אזי:

$$\mathbb{E}(Y_v)=\mathbb{P}(Y_v=1)=\mathbb{P}(\text{niether v nor its neighbors are in X})=(1-p)^{\deg_G(v)+1}\le(1-p)^{\delta+1}$$

כי צריך **לא לצרף** את הקודקוד לX מספר הפעמים של הדרגה של v, ועוד 1.
יהי $S=X\cup Y$. בבירור, S שלטת על G. (כי הוספנו לקבוצה את כל הקודקודים שלא היה להם שכן בX. אז עכשיו כל קודקוד, או שהוא נמצא בS או שיש לו שכן בX). ומתקיים:

$$\mathbb{E}(|S|)\overset{\text{א}}{=}\mathbb{E}(|X|)+\mathbb{E}(|Y|)\overset{\text{ב}}{=}np+\sum_{v\in V}\mathbb{E}(Y_v)\overset{\text{ג}}{\le}np+n(1-p)^{\delta+1}\overset{\text{ד}}{\le}$$
$$np+ne^{-p(\delta+1)}\overset{\text{ה}}{=}n\left(\frac{\ln(\delta+1)}{\delta+1}+e^{-\ln(\delta+1)}\right)\overset{\text{ו}}{=}\frac{\ln(\delta+1)+1}{\delta+1}n$$

א – איחוד זר.
ב – לינאריות התוחלת של Y, והגדרת התוחלת של X.
ג – החסם בחישוב לעיל, כפול מספר הקודקודים בV.
ד – $1-p\le e^{-p}$.
ה – נוציא n החוצה, ונציב את הערך שהגדרנו ל p. בחזקה של e זה מצטמצם.
ז – e בחזקת מינוס לן זה 1 חלקי e בחזקת לן, זה מצטמצם ונשאר רק הדלתא ועוד 1.

בסה"כ הוכחנו שזה התוחלת של S. בוודאות יש קבוצה S אחת שהגודל שלה הוא לכל היותר התוחלת (הדוגמה של הממוצע מתחילת השיעור) אז קיימת קבוצה S (שהיא שלטת) בגודל לכל היותר מה שיצא. כנדרש.

הרצאה 4

1 שיטות המומנט הראשון והשני

יהי X מ"מ אי-שלילי שמחזיר ערכים שלמים (נקרא גם מ"מ **סופר**). נניח שהתוחלת שואפת ל0 כאשר n שואף לאינסוף (עבור פרמטר n כלשהו). אז מי-שוויון מרקוב, נקבל:

$$\mathbb{P}(X \neq 0) = \mathbb{P}(X \geq 1) \leq \frac{\mathbb{E}(X)}{1} = \mathbb{E}(X) \Rightarrow \lim_{n\to\infty} \mathbb{P}(X=0) = 1$$

בעצם קיבלנו שX הוא "כמעט תמיד" 0.

השימוש הזה באי-שוויון מרקוב נקרא **שיטת המומנט הראשון**. (שימוש בתוחלת כדי להסיק דברים על ההסתברות).

עכשיו, נניח ש $\mathbb{E}(X)$ שואף לאחד או אפילו לאינסוף כאשר n שואף לאינסוף. האם זה אומר ש $\lim_{n\to\infty} \mathbb{P}(X=0) = 0$? מסתבר שלא בהכרח, לדוגמה:

לכל טבעי n יהי X_n מ"מ כך ש $\mathbb{P}(X_n = n^2) = 1/n$, $\mathbb{P}(X=0) = 1 - 1/n$. אזי התוחלת היא אינסוף:

$$\lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} n = \infty$$

אבל בבירור ניתן לראות ש:

$$\lim_{n\to\infty} \mathbb{P}(X=0) = 1$$

שזה הגיוני, כי ככל ש-n עולה, ההסתברות לקבל את n^2 קטנה.

איזה עוד נתונים נצטרך כדי שכן נוכל להסיק ש $\lim_{n\to\infty} \mathbb{P}(X=0) = 0$? אנחנו רוצים להסיק משהו על X לפי התוחלת שלו. אבל מסתבר שX לא חייב להיות תמיד קרוב לתוחלת. בדיוק בשביל זה הגדרנו שונות – זה אומר לנו עד כמה X מתרחק מהתוחלת. אם נפעיל את אי-שוויון צ'בישב, נקבל:

$$\mathbb{P}(X=0) \leq \underbrace{\mathbb{P}\Big(|X - \mathbb{E}(X)| \geq \mathbb{E}(X)\Big)}_{X \geq 2\mathbb{E}(X) \text{ or } x \leq 0} \leq \frac{\mathrm{Var}(X)}{\big(\mathbb{E}(X)\big)^2}$$

אז אם הביטוי הימני שואף לאפס כאשר n שואף לאינסוף, זה יגיד לנו שההסתברות שX הוא אפס גם שואפת לאפס. אפשר לסמן את זה כך:

$$\mathrm{Var}(X) = o\left(\big(\mathbb{E}(X)\big)^2\right)$$

(כלומר, השונות "קטנה בהרבה" מהתוחלת בריבוע).

השימוש הזה באי-שוויון צ'בישב נקרא **שיטת המומנט השני**. (שימוש בתוחלת בריבוע).

באופן כללי, כדי לבדוק אם אובייקט קיים:
נייצר מ"מ X שסופר מופעים של האובייקט הזה,
ונרצה להחליט האם X שווה 0 (כלומר אין אובייקט) או גדול מ0 (כלומר יש).
כדי להראות שהוא שווה 0 נשתמש במומנט הראשון – נראה שהתוחלת שווה 0.
בגלל שX לא יכול להיות שלילי, זב אומר שX תמיד 0.
כדי להראות שזה גדול מאפס, נשתמש במומנט השני – נראה שההסתברות שX שווה 0 קטנה ממשהו ששואף לאפס.

1.1 הופעת משולש:

טענה 1.1: נבנה גרף G על $[n]$ בצורה הבאה: לכל $1 \le i < j \le n$ נטיל מטבע **לא הוגן**, כאשר כל ההטלות בת"ל. אם יצא עץ (הסתברות p), נוסיף את הצלע ij ל-G. נגדיר את המאורע $A_G =$ "יש משולש בG". צ"ל:

א) אם $p = o(1/n)$, אז $\lim_{n\to\infty} \mathbb{P}(A_G) = 0$

ב) אם $p = \omega(1/n)$, אז $\lim_{n\to\infty} \mathbb{P}(A_G) = 1$

כלומר, אם p הרבה יותר קטן מ $1/n$, אז ההסתברות למשולש שואפת לאפס כאשר n שואף לאינסוף.
ואם p הרבה יותר גדול מ- $1/n$, אז ההסתברות למשולש שואפת ל-1 כאשר n שואף לאינסוף.

תזכורת: $f = o(g) \Leftrightarrow g = \omega(f)$

הוכחת א: נגדיר $t = \binom{n}{3}$ ויהיו $A_1, \dots, A_t$ כל תתי הקבוצות של $[n]$ בגודל 3.
לכל $1 \le i \le t$, נגדיר את X_i האינדיקטור למאורע "הצלעות של A_i יוצרות משולש בG".
יהי $X = \sum_{i=1}^{t} X_i$, אז X סופר את מספר המשולשים בG. מתקיים:

$$\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = p^3$$

לכל $1 \le i \le t$. לפי לינאריות התוחלת נקבל:

$$\mathbb{E}(X) = \sum_{i=1}^{t} \mathbb{E}(X_i) = \binom{n}{3} p^3$$

בפרט: אם $p = o(1/n)$ (כלומר $\frac{p}{1/n} \to 0$),
אז $\lim_{n\to\infty} \mathbb{E}(X) = 0$ (כי $\binom{n}{3} \approx n^3$, אז $(np)^3 = o(1)$).

אז לפי המומנט הראשון, ההסתברות שיש בG משולש שואפת לאפס כאשר n שואף לאינסוף. (התוחלת שואפת לאפס, ודרך אי-שוויון מרקוב נקבל שההסתברות שX גדול מ0 שואפת לאפס). מש"ל א.

הוכחת ב: נניח $p = \omega(1/n)$, נובע מכך ש $\lim_{n\to\infty} \mathbb{E}(X) = \infty$.
נשתמש בשיטת המומנט השני: נשתמש באי-שוויון צ'בישב כדי לחסום את ההסתברות ש $\mathbb{P}(X = 0)$.
בשביל זה נצטרך לחסום את השונות של X. נשתמש בנוסחה לשונות של סכום: (2)

$$\mathrm{Var}(X) = \sum_{i=1}^{t} \mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le t} \mathrm{Cov}(X_i, X_j)$$

ראשית, נשים לב שמתקיים: (3) לכל $1 \le i \le t$

$$\mathrm{Var}(X_i) = \mathbb{E}(X_i^2) - \left(\mathbb{E}(X_i)\right)^2 \le \mathbb{E}(X_i) = p^3$$

נבחר $1 \le i < j \le t$ כלשהם ונגדיר $\ell = |A_i \cap A_j|$.
נשים לב שבגלל ש $i \ne j$, מתקיים $\ell \le 2$.
אם $\ell \le 1$, זה אומר ש X_i, X_j נקבעו ע"י קבוצות זרות של הטלות
(כי יש להן לכל היותר קודקוד אחד משותף. אז אין להן צלעות משותפות) ולכן הם בת"ל.
בפרט, $\mathrm{Cov}(X_i, X_j) = 0$.
נניח עכשיו ש $\ell = 2$. אז מתקיים: (4)

$$\mathrm{Cov}(X_i, X_j) = \mathbb{E}(X_i, X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \mathbb{P}(X_i = 1, X_j = 1) - p^6 = p^5 - p^6 \le p^5$$

נשלב את (2), (3), (4):

$$\mathrm{Var}(X) \leq^{\text{א}} n^3p^3 + 2n^4p^5 =^{\text{ב}} o(n^6p^6) =^{\text{ג}} o\left(\left(\mathbb{E}(X)\right)^2\right)$$

א – חסם עליון לt, כפול השונות של X_i.
וחסם עליון למספר הדרכים לבחור 2 משולשים: אם יש צלע אחת או 2 משותפות צריך לבחור רק 4 קודקודים,
אז זה חסם עליון לזה. ואם אין צלעות משותפות השונות המשותפת היא 0. והשונות המשותפת היא לכל היותר p^5.

ב – בגלל ההנחה ש $p = \omega(1/n)$.

ג – בגלל החסם שמצאנו על התוחלת.

בסך הכל, נקבל שההסתברות שאין משולש שואפת לאפס. כלומר ההסתברות שיש משולש שואפת ל1, כנדרש.

1.2 סכומים זרים:

נאמר שלקבוצה $\{x_1, \dots, x_k\}$ יש **סכומים זרים** אם לכל אחת מ 2^k תתי הקבוצות יש סכום שונה.
עבור מספר טבעי n, יהי $f(n)$ השלם הגדול ביותר k שעבורו קיימים שלמים $1 \leq x_1, \dots, x_k \leq n$ כך שלקבוצה $\{x_1, \dots, x_k\}$ יש סכומים זרים. (כלומר, עבור כל n, מה הקבוצה הכי גדולה שיש לה סכומים זרים).
לדוגמה $\{2^i : 0 \leq i \leq \lfloor \log_2 n \rfloor\}$. זה מראה שמתקיים $f(n) \geq 1 + \lfloor \log_2 n \rfloor$.
אנחנו נראה שהחסם התחתון הזה הוא הדוק מבחינה אסימפטוטית.

טענה 1.2: $f(n) \leq \log_2 n + \frac{1}{2}\log_2 \log_2 n + O(1)$.

הוכחה: אנחנו רוצים למצוא חסם עליון ל $f(n)$. נעשה את זה כך:
נמציא מ"מ שמייצג סכום תת קבוצה,
נמצא את התוחלת והשונות,
נשתמש בצ'בישב כדי לקבל חסם על ההסתברות של ההפרש בין X לתוחלת,
ונשתמש בתכונת הסכומים הזרים כדי לקבל חסם שתלוי ב $f(n)$:

יהי $k = f(n)$ ויהיו $1 \leq x_1 \dots x_k \leq n$ שלמים כך ש $\{x_1 \cdots x_k\}$ בעלת סכומים זרים.
יהיו $I_1 \dots I_k$ מ"מ בת"ל כך ש $\mathbb{P}(I_i = 0) = \mathbb{P}(I_i = 1) = 1/2$ לכל $1 \leq i \leq k$.
יהי $X = \sum_{i=1}^{k} x_i I_i$. לפי לינאריות התוחלת, נקבל:

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{k} x_i I_i\right) = \sum_{i=1}^{k} x_i \cdot \mathbb{E}(I_i) = \frac{x_1 + \cdots + x_k}{2}$$

האינדיקטורים בעצם נותנים לנו סכום אקראי של תת קבוצה של $\{x_1 \cdots x_k\}$. ורואים שהתוחלת היא בעצם הממוצע.

בנוסף, בגלל שכל האינדיקטורים בת"ל, מתקיים: (5)

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{k} x_i I_i\right) =^{\text{א}} \sum_{i=1}^{k} x_i^2 \cdot \mathrm{Var}(I_i) =^{\text{ב}} \frac{x_1^2 + \cdots + x_k^2}{4} \leq^{\text{ג}} \frac{n^2 k}{4}$$

א – סקלר יוצא מהשונות בחזקת 2. וכל הcovariance הם 0 כי המ"מ בת"ל.
ב - $I_i \sim \mathrm{Ber}(1/2)$, אז השונות היא 1/4.
ג – כל $X_i \leq n$.

עכשיו נוכל להשתמש בצ'בישב: נציב $t = 2\sqrt{\mathrm{Var}(X)}$ (כדי שיצטמצם עם ה Var בצ'בישב) ונקבל:

$$\mathbb{P}(|X - \mathbb{E}(X)| > t) \leq \frac{\mathrm{Var}(X)}{t^2} = \frac{1}{4}$$

ולכן: (6) $\mathbb{P}(|X-\mathbb{E}(X)|\le t)\ge 1-\frac{1}{4}=\frac{3}{4}$.

מצד שני, בגלל שלפי ההנחה כל הסכומים של תתי הקבוצות זרים,
ההסתברות ש $X-\mathbb{E}(X)=s$ עבור s ממשי כלשהו היא או 0 או 2^{-k}
(כי יש בדיוק 2^k סכומים שונים, אז יש בדיוק את המספר הזה של הפרשים אפשריים.
אז אם s הוא לא אחד ההפרשים, אין סיכוי שנגיע אליו, לא משנה מה X יוצא).
בפרט, מתקיים $\mathbb{P}(X-\mathbb{E}(X)=s)\le 2^{-k}$ לכל $s\in[-t,t]$.
יש לכל היותר $2t+1$ ערכים בטווח הזה שעבורם $\mathbb{P}(x-\mathbb{E}(X)=s)>0$,
(ערכים שיכולים להיות ההפרש. אם הסכום של כל האיקסים היה זוגי, אז התוחלת שלמה וזה מספר הערכים השלמים בטווח. ואם הסכום היה אי-זוגי, התוחלת היא משהו ועוד חצי אבל עדיין שיש את אותו מספר של הפרשים אפשריים). אז חסם איחוד ייתן לנו: (7)

$$\mathbb{P}(|X-\mathbb{E}(X)|\le t)\le^{\text{א}} 2^{-k}(2t+1)\le^{\text{ב}} 2^{-k}(2n\sqrt{k}+1)$$

א – חסם על ההסתברות כפול מספר הערכים האפשריים.

ב – חסם עליון על t. ניזכר שt מוגדר לפי השונות, ויש את החסם העליון על השונות (5).

מצאנו חסם עליון ותחתון להסתברות. אם נשווה את (6), (7) נקבל אי שוויון של k: (8)

$$\frac{3}{4}\le^{\text{א}} 2^{-k}\left(2n\sqrt{k}+1\right)\Rightarrow^{\text{ב}}\frac{3}{4}\cdot 2^k\le 2n\sqrt{k}+1\Rightarrow^{\text{ג}}\frac{2^k}{\sqrt{k}}\le Cn$$

א – שילוב של 6, 7.

ב – נכפול את שני האגפים ב 2^k

ג – נחלק ב $\sqrt{k}$, ו"נבלע" את ה 1+ ע"י הקבוע C.

עבור $C>0$ כלשהו. עכשיו אנחנו רוצים לטעון שk קטן יותר ממשהו. איך נעשה את זה?
אנחנו רוצים לפתוח את הצד השמאלי של האי-שוויון האחרון, אבל ה2 בחזקת k והשורש מקשים.
אנחנו יודעים שזה כנראה ייצא משהו עם לוג.
הטריק הוא להציב במקום k משהו דומה אבל יותר גדול ממה שרוצים להוכיח, ונראה שאי השוויון לא מתקיים.
וזה מוכיח שk צריך להיות קטן יותר ממה שהצבנו.
אז נציב את אותה הנוסחה, אבל במקום ה $O(1)$ נשים משהו גדול.
נציב את $k=L>O(1)$. במונה, נפרק את החזקה כדי שיהיה יותר נוח:

$$\frac{2^k}{\sqrt{k}}=^{\text{א}}\frac{\overbrace{2^{\log_2 n}}^{=n}\cdot 2^{\overbrace{\frac{1}{2}\log_2\log_2 n}^{=\log_2(\log_2 n)^{\frac{1}{2}}}}\cdot 2^L}{\sqrt{\log_2 n+\frac{1}{2}\log_2\log_2 n+L}}\ge^{\text{ב}}\frac{n\cdot\sqrt{\log_2 n}\cdot 2^L}{\sqrt{2\log_2 n}}=^{\text{ג}} L'\cdot n>Cn$$

א – נציב את מה שרוצים להוכיח.

ב – המונה שווה. המכנה גדול יותר (כי חצי לוג של לוג זה הרבה יותר קטן, עבור n גדול).

ג – נבחר 'L מתאים – חישוב של ה2^L ביחד עם השורש 2 שנשאר במכנה. עבור L גדול, ברור שזה יותר גדול מC, כי C קבוע קטן כלשהו. גם אם L לא הרבה יותר גדול מC, 2 בחזקת L זה עדיין יותר גדול.

כלומר כדי שאי השוויון 8 יתקיים, צריך שk יהיה קטן ממה שהצבנו.

מכאן נובע ש $k=f(n)\le\log_2 n+\frac{1}{2}\log_2\log_2 n+O(1)$, כנדרש.

הרצאה 5

1 מבוא לגרפים מקריים:

בהרצאה 3 השתמשנו בגרף שייצרנו בצורה מקרית כדי להוכיח חסם תחתון על מספרי רמזי. נראה עוד שימושים ושיטות:

1.1 מודל בסיסי על גרפים מקריים:

הגדרה 1.1: מודל גרף מקרי ארדוש-רניי $G(n,m)$ הוא מרחב ההסתברות שבו $\Omega = \{G = ([n], E) : |E| = m\}$, (כל הגרפים על ה-n קודקודים האלה, שיש להם בדיוק m צלעות), $\mathbb{P}$ היא ההתפלגות האחידה, כלומר:

$$\mathbb{P}(G) = \binom{\binom{n}{2}}{m}^{-1}, \quad \text{לכל } G \in \Omega.$$

כי יש $\binom{n}{2}$ צלעות אפשריות, ונבחר m מתוכן. נשים לב ששני גרפים עם אותו מספר צלעות (אפילו באותו מבנה, נגיד גרפים עם צלע אחת) הם גרפים שונים, אם זו צלע בין קודקודים שונים.

המודל הזה מאפשר לנו לחקור תכונות של גרף עם מספר גדול של קודקודים. הבעיה היא שיש בו הרבה דברים שהם לא בת"ל ולכן לא נוח להשתמש בו. לדוגמה אם יש חלק בגרף שיש בו הרבה צלעות; בגלל שקבענו את מספר הצלעות, זה משפיע על ההסתברות לצלעות במקומות האחרים. המודל הבא יותר נוח:

הגדרה 1.2: מודל גרף מקרי בינומי $G(n,p)$ הוא מרחב הסתברות שבו Ω היא קבוצת כל הגרפים על $[n]$, ופונקציית ההסתברות היא:

$$\mathbb{P}(G) = p^{|E(G)|}(1-p)^{\binom{n}{2}-|E(G)|}$$

כלומר, אם שההסתברות לקיום צלע מסוימת היא p. אז ההסתברות לגרף ספציפי היא ההסתברות שכל אחד מהצלעות הקיימות נבחרה, ואף אחד מהאחרות לא נבחרה.

הגדרה: "**תכונה**" של גרף היא פשוט תת-קבוצה של גרפים. לדוגמה, כל הגרפים שקיים בהם מעגל.
אם Q היא תכונה של גרף ומתקיים $G \in Q$, נאמר שG מקיימת את Q. תכונה Q נקראת **מונוטונית עולה** אם היא סגורה להוספת צלעות – כלומר, מהרגע שהיא קיימת בגרף, כל עוד רק נוסיף צלעות התכונה תישמר.
פורמלית: אם $G \in Q$ ו- $G \subseteq H$ גוררים ש- $H \in Q$.
לדוגמה, קיום מעגל היא תכונה מונוטונית עולה. קיום מעגל פשוט היא לא.

הערה 1.3:

א) דרך סטנדרטית לייצור גרף במודל הזה היא: לכל $1 \le i < j \le n$ נטיל מטבע שיוצא עץ בהסתברות p.
כל ההטלות בת"ל. הצלעות של G הן כל הצלעות שעבור הקודקודים שלהן יצא עץ.

ב) $\mathbb{P}(G)$ תלוי רק במספר הצלעות בG (ולא במבנה שלו או משהו אחר).
בפרט, לכל $0 \le m \le \binom{n}{2}$, הפונקציה אחידה על כל הגרפים על $[n]$ עם m צלעות.

ג) עבור $m \approx \binom{n}{2}p$, נצפה ש $G(n,p), G(n,m)$ יתנהגו בצורה דומה.
אמנם יש הבדלים בין המודלים, לצרכים שלנו בקורס הם מהותית אותו דבר.
(אפשר להוכיח את זה פורמלית אבל לא נעשה את זה בקורס הזה).
בפרט, ניתן להוכיח שמתקיים: עבור כל תכונת גרף מונוטונית עולה Q:

$$\lim_{n\to\infty} \mathbb{P}(G(n,m) \in Q) = 1 \Leftrightarrow \lim_{n\to\infty} \mathbb{P}(G(n,p) \in Q) = 1$$

בפועל, זה נותן לנו לעבוד רק עם $G(n,p)$, שהוא יותר נוח, ולהסיק דברים על $G(n,m)$.

ד) מרחב ההסתברות $G(n,1/2)$ הוא מ"ה אחיד מעל כל הגרפים על קבוצת הקודקודים $[n]$.
לכן, לכל גרף עם תכונה Q, אם נוכיח שמתקיים:

$$\lim_{n\to\infty} \mathbb{P}(G(n,1/2) \in Q) = 1$$

זה בפועל מוכיח ש"כמעט" כל גרף גדול מקיים את Q.

דוגמה 1: נראה של"כמעט" כל גרף גדול יש קוטר לכל היותר 2. קוטר של גרף – המרחק המקסימלי בין כל שני קודקודים. (קוטר 2 משמעותו שלכל שני קודקודים יש או צלע משותפת או שכן משותף). נסמן את התכונה D_2. לפי הערה 1.3 ד, מספיק להוכיח שמתקיים:

$$\lim_{n\to\infty} \mathbb{P}(G(n,1/2) \notin D_2) = 0$$

יהי $G \sim G(n,1/2)$ ולכל $1 \le i < j \le n$, נסמן A_{ij} את המאורע $\mathrm{dist}_G(i,j) > 2$.
לכל $z \in V(G) \setminus \{i,j\}$, ההסתברות ש z הוא שכן משותף של שניהם היא $\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$
ולכן ההסתברות ש z הוא לא שכן משותף של שניהם היא $\frac{3}{4}$.
לכן, ההסתברות שאין להם שכן משותף היא $(3/4)^{n-2}$.
בפרט, $\mathbb{P}(A_{ij}) \le (3/4)^{n-2}$ לכל שני קודקודים. נוכל להסיק שמתקיים:

$$\mathbb{P}(G \notin D_2) =^{\text{א}} \mathbb{P}\left(\bigcup_{1\le i<j\le n} A_{ij}\right) \le^{\text{ב}} \sum_{1\le i<j\le n} \mathbb{P}(A_{ij}) \le^{\text{ג}} \binom{n}{2}\left(\frac{3}{4}\right)^{n-2} =^{\text{ד}} o(1)$$

א – מספיק שזוג אחד לא מקיים את התכונה.
ב – חסם איחוד.
ג – מספר הזוגות, והחסם על ההסתברות שמצאנו.
ד – כי $(3/4)^{n-2}$ פונקציה מעריכית, שואפת לאפס הרבה יותר מהר מאשר $\binom{n}{2} \in O(n^2)$ שהיא בפועל פולינום.

אז מהערה 1.3ג נקבל ש $\lim_{n\to\infty} \mathbb{P}(G(n,m) \notin D_2) = 0$, עבור $m = \left\lceil \frac{n(n-1)}{4} \right\rceil$.
אפשר להוכיח את זה ישירות (ולא דרך המודל השני) אבל זה קשה יותר.

1.2 חשיפה הדרגתית, ומונוטוניות ב $G(n,p)$

לפעמים שימושי לבנות גרף בשלבים. נגיע לשלב מסוים שנוח לנו, נסיק משהו לגבי הגרף ואז נמשיך. מה שעשינו מקודם, הטלנו מטבע לכל צלע אפשרית. אפשר לעשות את זה בשלבים – לעצור בשלב מסוים, להסיק משהו ואז להמשיך. עכשיו אנחנו רוצים לעשות משהו אחר: אנחנו רוצים להטיל את **כל** המטבעות, הטלה "חלקית":

הצעה 1.4: יהי k שלם חיובי, ונניח שיש לנו: $0 \le p, p_1, \dots p_k \le 1$ שמקיימים: $1 - p = \prod_{i=1}^{k}(1-p_i)$.
אז $G(n,p)$ ו- $\bigcup_{i=1}^{k} G(n,p_i)$ מייצגים את אותו מרחב הסתברות. כלומר לגרף מסוים, ההסתברות לקבל אותו שווה בשני המרחבים. כלומר אם נסיק משהו על גרף באחד המרחבים, המסקנה תקפה גם למרחב השני.

הדרך שקיבלנו את המרחב השני היא: הטלנו את כל ה-$\binom{n}{2}$ מטבעות, בהסתברות שונה מה-p המקורי. וחזרנו על אותה פעולה עבור כל p_i. כאילו הפרדנו כל מטבע ל-k מטבעות שונים. הרעיון העיקרי הוא שלא סתם נטיל את כולם: נטיל את הראשון ונסיק מסקנות, נטיל את השני ונסיק, וכו'.

הוכחה: מרחב המדגם זהה (כל הגרפים על [n]).
אנחנו צריכים רק להראות שבהינתן גרף מסוים, ההסתברות לקבל אותו שווה בשני המרחבים.
נוכיח שלכל זוג קודקודים, ההסתברות לקיום צלע שווה בשני המודלים:
יהיו $1 \le i < j \le n$ כלשהם. נשים לב שבשני המודלים, קיום כל אחת מהצלעות הם מאורעות בלתי תלויים.
בנוסף, ההסתברות ש ij היא **לא** צלע היא $\prod_{i=1}^{k}(1-p_i) = 1-p$, שווה בשני המודלים.

1.4 הוא כלי פשוט אבל שימושי. בפרט, הוא יכול להוכיח את המונוטוניות של $G(n,p)$ ב-p.

הצעה 1.5: מונוטוניות. תהי Q תכונה מונוטונית עולה, ויהיו $0 \le p_1 \le p_2 \le 1$, כאשר $p_1 < 1$.
אזי $\mathbb{P}(G(n,p_1) \in Q) \le \mathbb{P}(G(n,p_2) \in Q)$.
כלומר, אם ההסתברות לצלע גדולה יותר, ההסתברות לתכונה מונוטונית עולה גדולה יותר.

הוכחה: נגדיר: $p_0 = 1 - \frac{1-p_2}{1-p_1}$ ונשים לב שמתקיים: $0 \le p_0 \le 1$, וגם: $1-p_2 = (1-p_0)(1-p_1)$.
לכל $i \in \{0,1,2\}$ יהי $G_i \sim G(n,p_i)$. בבירור מתקיים $G_1 \subseteq G_0 \cup G_1$.

אז מכיוון שQ מונוטונית עולה, מתקיים:

$$\mathbb{P}(G_1 \in Q) \leq^{א} \mathbb{P}(G_0 \cup G_1 \in Q) =^{ב} \mathbb{P}(G_2 \in Q)$$

א – חסם איחוד
ב – לפי 1.4, $G_2 \sim G_0 \cup G_1$

מה שבעצם עשינו זה שייצרנו את G_1, ראינו שהוא מקיים את התכונה שרצינו, ואז הוספנו לו צלעות וזה נתן לנו את G_2. ומכיוון שהתכונה מונוטונית עולה, הוספת צלעות לא תבטל אותה.

1.3 גרפים עם מותן גדולה ומספר כרומטי גדול

לפני שנגדיר ונוכיח את המסקנות מהחלק הקודם, נצטרך מספר הגדרות וטענות עזר:

הגדרה 1.6: **המותן** של גרף G, נסמן: $g(G)$, הוא אורך המעגל הקצר ביותר ב-G.
אם אין מעגלים נגדיר $g(G) = \infty$.

הגדרה 1.7: יהי $G = (V, E)$ גרף, ותהי $S \subseteq V$. S תיקרא בת"ל אם אין בה צלעות. **מספר אי-התלות** של G הוא גודל הקבוצה הבת"ל הגדולה ביותר ב-G. נסמן אותו $\alpha(G)$.

הגדרה 1.8: **המספר הכרומטי** של G הוא השלם החיובי k הקטן ביותר שעבורו קיימת פונקציה $c : V \to \{1 \dots k\}$,
כך שאם $uv \in E$, אז $c(u) \neq c(v)$. נקרא גם "k-צביעה" של G.
המשמעות היא פשוט שלכל שני קודקודים שמחוברים ע"י צלע יש צבעים שונים. נסמן אותו $\chi(G)$.
K צביעה של גרף זו פשוט חלוקה של הקודקודים ל-k קבוצות בת"ל.

נשים לב שאם יש בתוך G עותק של K_d (קליקה על d קודקודים), אז $\chi(G) \geq d$.
טבעי שנשאל האם גם ההפך מתקיים, כלומר שאם $\chi(G) \geq d$ אומר ש- $K_d \subseteq G$.
זה מתקיים באופן טריוויאלי עבור $d \in \{1,2\}$, אבל כפי שנראה בדוגמה, לא מתקיים באופן כללי:

משפט 1.9: (ארדוש 1959). לכל k, ℓ שלמים קיים גרף G כך שמתקיים: $\chi(G) > k$ וגם $g(G) > \ell$.
כלומר, מספר כרומטי גדול לא חוסם את המותן (יכול להיות שלא נוכל לצבוע עם מספר מסוים של צבעים גם אם המעגל הכי קטן הוא גדול – ונובע מכך בקלות שאין קליקה).
וגם, מותן גדולה לא מגבילה את המספר הכרומטי (תמיד נוכל למצוא גרף שאי אפשר לצבוע עם k צבעים. צביעה היא לא תכונה לוקאלית – היא תלויה בכל הגרף, וצביעה בחלק מסוים לא מבטיחה שום דבר לגבי הצביעה של כל הגרף).

לפני שנוכיח את 1.9, נוכיח חסם תחתון על המספר הכרומטי של גרף:

טענה 1.10: יהי G גרף על n קודקודים. אזי $\chi(G) \geq n/\alpha(G)$.

הוכחה: יהי $k = \chi(G)$, ותהי $c : V \to \{1 \dots k\}$ k-צביעה של G.
לכל $1 \leq i \leq k$ נגדיר: $A_i = \{u \in V | c(u) = i\}$. כלומר, קבוצת הקודקודים שצבועים בצבע i.
נשים לב ש $A_1 \cup \dots \cup A_k$ היא חלוקה של V, ושכל ה-A_i בת"ל. אזי:

$$n = |V| = \sum_{i=1}^{k} |A_i| \leq k \cdot \alpha(G)$$

כי הגודל של כל אחת מהקבוצות הוא לכל היותר הגודל של $\alpha(G)$, הקבוצה הבת"ל הכי גדולה. מסקנה:
$\chi(G) = k \geq n/\alpha(G)$, כנדרש.

הוכחת 1.9: יהי $\theta = 1/2\ell$, ויהי $G \sim G(n, p)$, כאשר n "מספיק גדול"[1] וכאשר $p = n^{\theta - 1}$. נקרא למעגל ב-G **קצר** אם הוא באורך לכל היותר ℓ. יהי X משתנה מקרי שסופר כמה מעגלים קצרים יש. אזי מתקיים:

[1] פורמלית, n הוא סדרה אינסופית ששואפת לאינסוף. אז פשוט נגיד שניקח את ה-n שמספיק גדול שהטענה תתקיים.

$$\mathbb{E}(X) =^{א} \sum_{i=3}^{\ell} \frac{n(n-1)\cdot \ldots \cdot (n-i+1)}{2i} \cdot p^i \leq^{ב} \sum_{i=3}^{\ell} n^i \cdot p^i =^{ג} \sum_{i=3}^{\ell} n^{\theta i} =^{ד} O\left(n^{\theta \ell}\right) =^{ה} o(n)$$

א – עבור כל אורך מעגל: בחירת i קודקודים עם חשיבות לסדר, וחלוקה בספירה הכפולה – אפשר לבחור קודקוד התחלתי למעגל ב-i דרכים, ו"כיוון" למעגל בשני דרכים. וכל אחד כפול p בחזקת מספר הקודקודים.

ב – המונה הוא מכפלה של i גורמים, שכל אחד קטן שווה n.

ג – במקום p נציב $n^{\theta-1}$. אז כשמעלים את זה בחזקת i נקבל $n^{\theta i-i}$, וזה כפול n^i. אז זה מצטמצם.

ד – האיבר הכי גדול בסכום הוא $n^{\theta\ell}$, ומחברים ℓ איברים. אז ניקח את האיבר הגדול ℓ פעמים וזה חסם עליון לסכום, ובגלל ש ℓ קבוע הוא יוצא ב-O.

ה – בגלל ש $\theta < 1/\ell$, אנחנו מעלים את n בחזקה קטנה מ1.

אז מאי-שוויון מרקוב נקבל: (1) $\mathbb{P}(X \geq n/2) = o(1)$.

מה עושים מכאן: אנחנו לא רוצים מעגלים קצרים, ואנחנו רוצים שמספר הצביעה יהיה גבוה.
אז אנחנו נרצה להגריל גרף כזה ששני התנאים האלה מתקיימים.
כדי להראות שמספר הצביעה גבוה, נראה חסם עליון על גודל הקבוצה הבת"ל המקסימלית.
ולפי טענה 1.10, זה נותן חסם תחתון למספר הכרומטי.
הבעיה היא שלא קיבלנו שאין בכלל מעגלים קצרים, קיבלנו שיש מעט.
אז מה שנעשה זה שנראה שההסתברות של שני המאורעות שואפת לאפס,
ולכן ההסתברות שלפחות אחד מהם קורה גם שואפת לאפס (חסם איחוד).
כלומר, בהסתברות גבוה קיבלנו גרף שיש בו מעט מעגלים ואין בו קבוצה בת"ל גדולה.
קיבלנו שגרף מקרי כמעט נתן לנו את מה שרצינו,
עכשיו "נתקן" את הגרף בצורה שלא תהרוס את התכונה של מספר הצביעה ונגרום לכך שאין מעגלים קצרים.
ונקבל שהגרף מקיים את התכונות שרצינו.
הטריק העיקרי הוא: התחלנו עם גרף מקרי, ונתקן אותו בצורה דטרמיניסטית כדי לקבל את מה שרוצים.

נחזור להוכחה: בעצם יש פה סדרה של גרפים, כי n הוא טכנית סדרה ו- p תלוי ב- n. קיבלנו שעבור n מספיק גדול, התוחלת של מספר המעגלים הקצרים היא קטנה ולכן עם אי"ש מרקוב הראנו שההסתברות שיש לפחות n/2 מעגלים שואפת לאפס. בחלק הבא, נראה שבהסתברות גבוהה, $\alpha(G)$ לא גדול.

נציב $t = \lceil 3 \ln n/p \rceil$ ונשים לב שמתקיים: (2)

$$\mathbb{P}(\alpha(G) \geq t) \leq^{א} \binom{n}{t} (1-p)^{\binom{t}{2}} \leq^{ב} n^t e^{-\frac{pt(t-1)}{2}} \leq^{ג} \left(n \cdot e^{-\frac{p(t-1)}{2}}\right)^t \leq^{ד} \left(n \cdot e^{-1.4 \ln n}\right)^t =^{ה} o(1)$$

א – חסם איחוד: ההסברות שיש קבוצה בת"ל בגודל לפחות t, היא ההסתברות שלכל קבוצה בגודל t שנבחר (יש $\binom{n}{t}$ כאלו), כל הצלעות לא קיימות – כלומר הסתברות $(1-p)$, $\binom{t}{2}$ פעמים. בגלל שבמודל הזה אין תלות בין כל הצלעות.

ב – נחסום את $\binom{n}{t}$ על ידי n^t, ונחסום את $1-p$ על ידי e^{-p}. ונפתח את $\binom{t}{2}$.

ג – נוציא את הt החוצה.

ד – ה-t,n נשארו אותו דבר. ניזכר מה הצבנו ב-p. אנחנו רוצים להגדיל את הביטוי הכללי, כלומר להגדיל את החזקה. בגלל שיש שם מינוס, נקטין את מה שכתוב. אז היינו רוצים לעשות: (ה-3 הוא פשוט קבוע גדול מ-2):

$$e^{-\frac{p}{2}(t-1)} \leq e^{-\frac{p}{2}\cdot\frac{3\ln n}{p}} = n^{-\frac{3}{2}}$$

הבעיה היא שלא התייחסנו לזה שהיה כתוב $(t-1)$. אז נכפיל את כל החזקה ב-0.99: כאשר n שואף לאינסוף, זה מוריד הרבה יותר מ- $p/2$, שזה מה שצריך:

$$e^{-\frac{p}{2}(t-1)} \leq e^{-\frac{p}{2}\cdot\frac{3\ln n}{p}\cdot 0.99} = n^{-\frac{3}{2}\cdot 0.99} \leq n^{-1.4}$$

ה – כי זה n כפול מספר ששואף לאפס (e בחזקת מינוס אינסוף), בחזקת t.

נחבר את (1), (2) ונקבל: $\mathbb{P}\left(X \geq \frac{n}{2} \text{ or } \alpha(G) \geq t\right) = o(1)$ כמו שכתבנו בהסבר למעלה.
בפרט, קיים גרף H על n קודקודים שיש בו לכל היותר $n/2$ מעגלים קצרים,
ושהקבוצה הבת"ל הכי גדולה שלו היא בגודל לפחות t.
עכשיו נתקן את הגרף: נרצה שלא יהיו מעגלים קצרים.
אם נמחק קודקוד אחד (לפחות) מכל מעגל קצר,
נקבל שאין מעגלים קצרים בכלל, ונישאר עם לפחות $n/2$ קודקודים בגרף.
ובגרף החדש מתקיים: $g(H') > \ell$ וגם $\alpha(H') \leq t$
(כי לא מחקנו צלעות, אז מה שלא היה בת"ל לא יהפוך להיות בת"ל). נסיק ש:

$$\chi(H') \geq^{\text{א}} \frac{|V(H')|}{\alpha(H')} \geq^{\text{ב}} \frac{n/2}{\left\lceil 3\ln\frac{n}{p}\right\rceil} \geq^{\text{ג}} n^{\theta/2} >^{\text{ד}} k$$

א - לפי טענה 1.10

ב – לפי התהליך שבנינו את הגרף

ג - $\frac{n^{\theta}}{12\ln n} = \frac{np}{12\ln n} = \frac{\frac{1}{2}\left(\frac{n}{2}\right)}{3\ln\frac{n}{p}} \geq \frac{n/2}{\left\lceil 3\ln\frac{n}{p}\right\rceil}$ החצי באי"ש הראשון זה כדי לפצות על זה שהורדנו את הערך עליון, וכל השאר זה אלגברה ונציב את מה שקבענו ל-p.

ד – n שואף לאינסוף, אז בחזקה גדולה מ-1 הוא גדול מכל קבוע.

בסה"כ הוכחנו שהמותן של הגרף גדולה מ- ℓ, ושמספר הצביעה גדול מ- k. בפרט, הוכחנו יותר מזה – הוכחנו שמספר הצביעה שואף לאינסוף.

הרצאה 6

1 פונקציות סף בגרפים מקריים

בהרצאה 4 הוכחנו ש:

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \text{ contains a triangle}) = \begin{cases} 0, & \text{if } p = o(1/n) \\ 1, & \text{if } p = \omega(1/n) \end{cases}$$

כלומר, כל עוד p יותר קטן מ- $1/n$ (בסדר גודל), ההסתברות למעגל שואפת לאפס. ואם p גדול יותר מ- $1/n$, ההסתברות שואפת ל1. אז אם נצייר גרף של ההסתברות למשולש כפונקציה של p, אנחנו יודעים את הצורה הכללית של הגרף: משמאל ל-$1/n$ הוא קרוב לאפס, מימין ל-$1/n$ הוא קרוב ל-1. סוג של מדרגה. בסביבה הקרובה של $1/n$ אנחנו לא יודעים בדיוק מה מוגדר.

נרחיב את התופעה לכל תכונה מונוטונית עולה:

הגדרה 1.1: פונקציה $p_0(n)$ תיקרא **פונקציית סף** עבור תכונה מונוטונית עולה Q אם מתקיים:

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \in Q) = \begin{cases} 0, & \text{if } p = o(p_0) \\ 1, & \text{if } p = \omega(p_0) \end{cases}$$

כלומר היא הנקודה שבה הקפיצה מתרחשת.

משפט 1.2: לכל תכונה לא-טריוויאלית ומונוטונית עולה יש סף. לא נוכיח את המשפט בקורס הזה.
(תכונה לא טריוויאלית היא תכונה שקיימת בכל גרף או לא קיימת באף גרף)
המשפט הזה נותן לנו את "חוק אפס – אחד": עבור n מספיק גדול, ההסתברות לתכונה היא בפועל 0 או 1 (למעט סביבה כלשהי של הסף). לפעמים נוכל גם להקטין את החלון הזה.

הגדרה 1.3: סף של תכונה מונוטונית עולה Q ייקרא **חד** אם לכל $\varepsilon > 0$ מתקיים:

$$\lim_{n\to\infty} \mathbb{P}(G(n,p) \in Q) = \begin{cases} 0, & \text{if } p \le (1-\varepsilon)p_0 \\ 1, & \text{if } p \ge (1+\varepsilon)p_0 \end{cases}$$

אחרת, הסף ייקרא **גס**.

1.1 דרגה מינימלית וקשירות של גרפים מקריים

משפט 1.4: $\frac{\ln n}{n}$ הוא סף חד עבור דרגה מינימלית חיובית וקשירות של $G(n,p)$.

הוכחה: יהי $\varepsilon > 0$, יהי n מספיק גדול, ויהי $G \sim G(n,p)$. מכיוון שלכל גרף קשיר יש דרגה מינימלית חיובית, מספיק להוכיח:

1) אם $p \le \frac{(1-\varepsilon)\ln n}{n}$, אז $\lim_{n\to\infty} \mathbb{P}(\delta(G) \ge 1) = 0$
2) אם $p \ge \frac{(1+\varepsilon)\ln n}{n}$ אז $\lim_{n\to\infty} \mathbb{P}(G \text{ is connected}) = 1$

כי כל אחד יגרור גם את השני בכיוון שצריך.

הוכחת 1: צריך להוכיח שקיים משהו – זה מתאים לשיטת המומנט השני. נניח ש $p \le \frac{(1-\varepsilon)\ln n}{n}$. לכל $1 \le j \le n$ יהי I_j האינדיקטור למאורע "j הוא קודקוד מבודד". אז $X = \sum_{j=1}^{n} I_j$ סופר את מספר הקודקודים המבודדים. אנחנו נרצה להראות שהתוחלת שואפת לאינסוף. מתקיים:

$$\mathbb{E}(X) \overset{\text{א}}{=} \sum_{j=1}^{n} \mathbb{E}(I_j) \overset{\text{ב}}{=} \sum_{j=1}^{n} \mathbb{P}(I_j = 1) \overset{\text{ג}}{=} n(1-p)^{n-1}$$

א – לינאריות התוחלת
ב – תוחלת של אינדיקטור
ג – ההסתברות שקודקוד מסוים יהיה מבודד, כפול n קודקודים.

עכשיו, בגלל ה-p שבחרנו, מתקיים: (1)

$$\lim_{n\to\infty}\mathbb{E}(X) =^{א} \lim_{n\to\infty} n(1-p)^{n-1} \geq^{ב} \lim_{n\to\infty} ne^{-(p+p^2)(n-1)} \geq^{ג}$$
$$\lim_{n\to\infty} ne^{-\frac{(1-\varepsilon/2)\ln n}{n}\cdot n} =^{ד} \lim_{n\to\infty} n^{1-(1-\varepsilon/2)} =^{ה} \lim_{n\to\infty} n^{\varepsilon/2} = \infty$$

א – נציב את מה שמצאנו

ב – אנחנו רגילים להשתמש באי-השוויון $1-p \leq e^{-p}$, אבל פה אנחנו צריכים את הכיוון ההפוך. אז נשתמש בזה שהם קרובים: $1-p \approx e^{-p}$, ונתקן. פיתוח טור טיילור של e^{-x+x^2} מראה ש $(1-p)^{n-1} \geq e^{-(p+p^2)(n-1)}$.

ג – היינו רוצים להגיד שמה שיש בחזקה זה np. זה לא בדיוק מה שכתוב אבל זה קרוב, כי n שואף לאינסוף אז מינוס 1 זה זניח, ו-p שואף לאפס אז להוסיף p^2 זה זניח. אם היה כתוב את זה, אז היה מתקיים:

$$np \leq n\cdot\frac{(1-\varepsilon)\ln n}{n} = (1-\varepsilon)\ln n$$

אבל יש שם משהו קצת יותר גדול מ- np, אז כדי לתקן נכפול את הכל ב $\left(1+\frac{\varepsilon}{10}\right)$ ואז זה כן גדול מ- np, אבל עדיין קטן מ- $(1-\varepsilon/2)\ln n$:

$$\left(1+\frac{\varepsilon}{10}\right)np \leq n\cdot\frac{\left(1+\frac{\varepsilon}{10}\right)(1-\varepsilon)\ln n}{n} = \left(1+\frac{\varepsilon}{10}\right)(1-\varepsilon)\ln n \leq \left(1-\frac{\varepsilon}{2}\right)\ln n$$

אז רצינו להקטין את הגבול, ובגלל שהחזקה במינוס זה אומר שצריך להגדיל את מה שיש בתוך הביטוי. במקום $p+p^2$ לקחנו את p שזה קטן יותר, אז במקום לכפול ב $(1-\varepsilon)$ כפלנו ב $(1-\varepsilon/2)$.

ד – ה-n מצטמצמים, e בחזקת ln מצטמצם.

ה – פתיחה של הסוגריים.

נחשב את השונות: לכל $1 \leq j \leq n$ מתקיים:

$$\mathrm{Var}(I_j) = \mathbb{E}(I_j^2) - \left(\mathbb{E}(I_j)\right)^2 \leq^{א} \mathbb{E}(I_j) = (1-p)^{n-1}$$

א – תוחלת של אינדיקטור, $\mathbb{E}(I_j) = \mathbb{E}(I_j^2)$

ומתקיים לכל $1 \leq i < j \leq n$:

$$\mathrm{Cov}(I_i, I_j) = \mathbb{E}(I_iI_j) - \mathbb{E}(I_i)\mathbb{E}(I_j) =^{א} \mathbb{P}(I_i = 1, I_j = 1) - (1-p)^{2n-2} =^{ב}$$
$$(1-p)^{2n-3} - (1-p)^{2n-2} =^{ג} p(1-p)^{2n-3}$$

א – אנחנו יודעים את התוחלת של כל אחד.

ב – מה ההסתברות של שני האינדיקטורים ביחד? עבור שני קודקודים, ההסתברות ששניהם מבודדים.
ההסתברות שהראשון מבודד: עבור $n-1$ קודקודים יצא $1-p$, ועבור הקודקוד השני צריך לבדוק רק $n-2$ קודקודים.

ג – גורם משותף.

ולכן,

$$\mathrm{Var}(X) = \sum_{j=1}^{n}\mathrm{Var}(I_j) + 2\sum_{1\leq i<j\leq n}\mathrm{Cov}(I_i, I_j) \leq^{א} n(1-p)^{n-1} + n^2p(1-p)^{2n-3}$$

א – לפי מה שמצאנו לעיל.

אז מתקיים:

$$\frac{\mathrm{Var}(X)}{\left(\mathbb{E}(X)\right)^2} \leq^{\text{א}} \frac{\mathbb{E}(X)+n^2p(1-p)^{2n-3}}{\left(\mathbb{E}(X)\right)^2} =^{\text{ב}} \frac{1}{\mathbb{E}(X)}+\frac{p}{1-p} =^{\text{ג}} o(1)$$

א – לפי מה שמצאנו

ב – נפרק את השבר, ונציב את מה שמצאנו וזה מצטמצם.

ג – כי התוחלת שואפת לאינסוף, ו-p שואף לאפס.

אז לפי המומנט השני, נקבל:

$$\lim_{n\to\infty}\mathbb{P}(\delta(G)\geq 1)=\lim_{n\to\infty}\mathbb{P}(X=0)=0$$

הוכחת 2: נניח ש $p \geq (1+\varepsilon)\ln n/n$. אם G לא קשיר, אז רכיב הקשירות הקטן ביותר שלו הוא בגודל k עבור: $1\leq k\leq \lfloor n/2\rfloor$. ולכן: (2)

$$\mathbb{P}(\text{G is disconnected}) \leq^{\text{א}} \sum_{k=1}^{n\backslash 2}\binom{n}{k}(1-p)^{k(n-k)} \leq^{\text{ב}} \sum_{k=1}^{\sqrt{n}}\binom{n}{k}e^{-pk(n-k)}+\sum_{k=\sqrt{n}}^{n\backslash 2}\binom{n}{k}e^{-pk(n-k)}$$

א – עבור כל גודל אפשרי של הרכיב הקשירות הקטן, נבחר שאין צלעות בין הרכיב לשאר הגרף.

ב – נפצל את הסכום לשניים. וגם $1-p\leq e^{-p}$

נתבונן בכל חלק של הסכום בנפרד. החלק הראשון: (3)

$$\sum_{k=1}^{\sqrt{n}}\binom{n}{k}e^{-pk(n-k)} \leq^{\text{א}} \sum_{k=1}^{\sqrt{n}}n^k e^{-pk(n-k)} = \sum_{k=1}^{\sqrt{n}}\left(ne^{-p(n-k)}\right)^k$$
$$\leq^{\text{ב}} \sum_{k=1}^{\sqrt{n}}\left(ne^{-\frac{(1+\varepsilon)\ln n}{n}(n-\sqrt{n})}\right)^k \leq^{\text{ג}} \sum_{k=1}^{\sqrt{n}}\left(ne^{-\left(1+\frac{\varepsilon}{2}\right)\ln n}\right)^k = \sum_{k=1}^{\sqrt{n}}n^{-\frac{\varepsilon k}{2}} \leq \sum_{k=1}^{\infty}n^{-\frac{\varepsilon k}{2}} = \frac{n^{-\varepsilon\backslash 2}}{1-n^{-\varepsilon\backslash 2}}$$

א – חסם עליון על $\binom{n}{k}$

ב – נציב במקום p. זה קטן יותר מp, ולכן אי השוויון מתקיים. וגם, k חסום ב $\sqrt{n}$, אז אם נרשום את השורש במקום k זה המקסימום, ובגלל שהחזקה שלילית זה בעצת מגדיל את הסכום.

ג – בגלל ש $\sqrt{n}=o(n)$, כאשר n שואף לאינסוף ההפרש שואף ל-n, אז אפשר "לשלם" בזה שנקטין את אפסילון.

החלק השני: (4)

$$\sum_{k=\sqrt{n}}^{n\backslash 2}\binom{n}{k}e^{-pk(n-k)} \leq^{\text{א}} \sum_{k=\sqrt{n}}^{n\backslash 2}\left(\frac{en}{k}\right)^k e^{-pk(n-k)} = \sum_{k=\sqrt{n}}^{n\backslash 2}\left(\frac{en}{k}\cdot e^{-p(n-k)}\right)^k \leq^{\text{ב}} \sum_{k=\sqrt{n}}^{n\backslash 2}\left(\frac{en}{\sqrt{n}}\cdot e^{-\frac{(1+\varepsilon)\ln n}{n}\cdot\left(n-\frac{n}{2}\right)}\right)^k$$
$$=^{\text{ג}} \sum_{k=\sqrt{n}}^{n\backslash 2}\left(en^{1\backslash 2}\cdot e^{-(1\backslash 2+\varepsilon\backslash 2)\ln n}\right)^k = \sum_{k=\sqrt{n}}^{n\backslash 2}\left(en^{-\varepsilon\backslash 2}\right)^k \leq \sum_{k=1}^{\infty}\left(en^{-\varepsilon\backslash 2}\right)^k = \frac{en^{-\varepsilon\backslash 2}}{1-en^{-\varepsilon\backslash 2}}$$

א - $\binom{n}{k}\leq\left(\frac{en}{k}\right)^k$

ב – כמו בסכום הקודם, נציב במקום p ונציב את ה-k המקסימלי. במכנה נציב את ה-k המינימלי.

ג – צמצום.

נחבר את 2,3,4 ונקבל:

$$\lim_{n\to\infty}\mathbb{P}(\text{G is disconnected}) \leq \lim_{n\to\infty}\frac{n^{-\varepsilon\backslash 2}}{1-n^{-\varepsilon\backslash 2}}+\lim_{n\to\infty}\frac{en^{-\varepsilon\backslash 2}}{1-en^{-\varepsilon\backslash 2}}=0+0=0$$

כנדרש.

הרצאה 7

1 מבוא לאלגוריתמים מקריים

אלגוריתם מקרי לעיתים יהיה פשוט או יעיל יותר מאלגוריתם דטרמיניסטי. נתאר שתי דוגמאות:

1.1 שוויון פולינומים

בהינתן שני פולינומים: אחת בצורת מכפלה והשנייה בצורה קנונית:

$$F(x) = \prod_{i=1}^{d} (x - a_i)\,, \qquad G(x) = x^d + \sum_{j=0}^{d-1} b_j x^j$$

נרצה לבדוק האם $F(x) \equiv G(x)$. דרך ישירה לעשות את זה היא להעביר את $F(x)$ לצור קנונית ולהשוות מקדמים.
אבל זה ידרוש $\Theta(d^2)$ פעולות כפל וחיבור. האלגוריתם המקרי הבא יהיה הרבה יותר מהיר:

אלגוריתם מקרי להשוואת פולינומים:

קלט: פולינומים $F(x), G(x)$ כמתואר לעיל.
פלט: $F(x) \equiv G(x)$ או $F(x) \not\equiv G(x)$.

1. נבחר שלם $r \in \{1,2, \dots 100d\}$ באופן מקרי ואחיד.
2. אם $F(r) \neq G(r)$, נחזיר $F(x) \not\equiv G(x)$. אחרת נחזיר $F(x) \equiv G(x)$.

סיבוכיות האלגוריתם: נניח שהבחירה בשלב 1 דורשת זמן קבוע $O(1)$. החישוב של $F(r)$, $G(r)$ דורש זמן $\Theta(d)$.
שזה סה"כ זמן הריצה של האלגוריתם.

ננתח את הנכונות: נניח ש $F(x) \equiv G(x)$. במקרה הזה, בוודאות $F(r) \neq G(r)$ ולכן נחזיר את התשובה הנכונה.
נניח עכשיו ש $F(x) \not\equiv G(x)$. אם $F(r) \neq G(r)$, נחזיר תשובה נכונה. אם $F(r) = G(r)$, נחזיר תשובה שגויה.
מה ההסתברות שזה יקרה?

אם $F(x) \not\equiv G(x)$, אז הפולינום $H(x) = F(x) - G(x)$ הוא לא פולינום האפס ולכן יש לכל היותר d שורשים.
ובפרט, לכל היותר d שורשים בקבוצה $\{1,2, \dots 100d\}$.
בגלל שבחרנו את r מתוך הקבוצה באופן אחיד, נובע ש $\mathbb{P}\big(F(r) = G(r)\big) \leq \frac{1}{100}$.

לסיכום, אם האלגוריתם החזיר שהם לא שווים, זה נכון.
אם האלגוריתם החזיר שהם שווים, זה נכון בהסתברות לפחות 0.99, מהיר יותר אבל אולי שגוי.
נרצה להקטין מאד את הסיכוי לטעות, לדוגמה על ידי הגדלת הקבוצה שמתוכה נבחר את r.
אבל זה עלול להשפיע על הסיבוכיות (הסיבוכיות של בחירה מתוך קבוצה יכולה להיות תלויה בגודל הקבוצה).
בנוסף, עבודה עם מספרים גדולים מאד יכולה להיות בעייתית עבור מחשבים.
ניתן פתרון יותר כללי ויותר טוב:

אלגוריתם מקרי משופר להשוואת פולינומים:

קלט: פולינומים $F(x), G(x)$ כמתואר לעיל, ומספר שלם חיובי k.
פלט: $F(x) \equiv G(x)$ או $F(x) \not\equiv G(x)$.

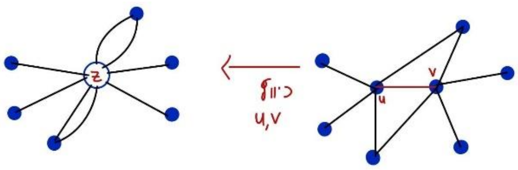
1. עבור $1 \leq i \leq k$:
   a. נבחר שלם $r_i \in \{1,2, \dots 2d\}$ באופן מקרי ואחיד.
   b. אם $F(r_i) \neq G(r_i)$, נעצור נחזיר $F(x) \not\equiv G(x)$.
2. נחזיר $F(x) \equiv G(x)$.

בניתוח דומה למה שעשינו מקודם, נקבל שהסיבוכיות היא $\Theta(kd)$ (שזה $\Theta(d)$ אם k הוא קבוע). בנוסף, אם האלגוריתם החזיר שהם שונים, זה בוודאות נכון. ואם האלגוריתם החזיר שהם שווים, זה נכון בהסתברות לפחות $1 - 2^{-k}$.

1.2 אלגוריתם חתך מינימלי רנדומלי

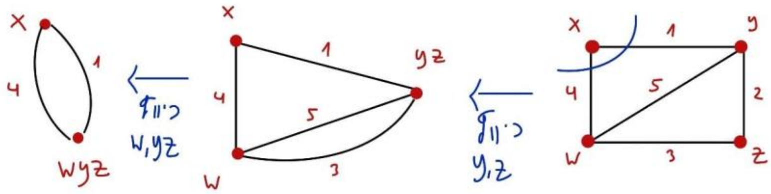
יהי $G=(V,E)$ גרף קשיר. **חתך** בG הוא קבוצה $A\subseteq E$ כך ש $G\setminus A$ לא קשיר.
נרצה למצוא חתך בגודל מינימלי. נתאר אלגוריתם מקרי פשוט, שמשתמש ב**כיווץ קשתות:**

הגדרות: **מולטיגרף** הוא גרף שמאפשר קיום של יותר מצלע אחת בין שני קודקודים (צלעות מקבילות).
לולאה היא צלע מקודקוד לעצמו.



בהינתן מולטיגרף בלי לולאות, **כיווץ** צלע uv נעשה על ידי איחוד u,v לקודקוד חדש z_{uv} ומחיקת כל הצלעות בין u ל-v. כל צלע שחיברה בין u או v לקודקוד אחר, עכשיו תחבר בין אותו קודקוד לקודקוד החדש z_{uv}. נקרא לגרף החדש $G\setminus uv$ ונשים לב שיכולות להיות בו צלעות מקבילות אבל אין בו לולאות.

אלגוריתם מקרי למציאת חתך מינימלי:



קלט: גרף קשיר G על n קודקודים.
פלט: חתך של G.

1. יהי $G_0=G$ מולטיגרף.
2. עבור $1\le i\le n-2$:
   a. נבחר צלע $e_i\in E(G_{i-1})$ באופן מקרי ואחיד.
   b. נגדיר את $G_i=G_{i-1}\setminus e_i$ (נכווץ את הצלע).
3. נחזיר את $E(G_{n-2})$.

מכיוון שכל צעד בלולאה לוקח זמן $O(n)$, זמן הריצה הכולל הוא $O(n^2)$.

הפלט תמיד יהיה 2 קודקודים עם צלעות ביניהן, והצלעות האלה מהוות חתך.

טענה 1.1: האלגוריתם מחזיר חתך מינימלי בהסתברות לפחות $\frac{2}{n(n-1)}=\binom{n}{2}^{-1}$. (זו לא הסתברות גבוהה).

הוכחה: יהי $A\subseteq E(G)$ חתך מינימלי כלשהו. נגדיר את k הגודל של החתך.
לכל $1\le i\le n-2$, נגדיר את E_i המאורע $e_i\notin A$. כלומר בסיבוב הזה, לא כיווצנו צלע ששייכת לחתך.
נשים לב ש:

$$\mathbb{P}(\text{the algorithm returns a min}-\text{cut})\ge\mathbb{P}(\text{the algorithm returns A})\overset{\aleph}{=}\mathbb{P}\left(\bigcap_{j=1}^{n-2}E_j\right)$$

$$=\mathbb{P}(E_1)\cdot\mathbb{P}(E_2|E_1)\cdot\dots\cdot\mathbb{P}\left(E_{n-2}\middle|\bigcap_{i=1}^{n-3}E_i\right)$$

א – כל פעם שמכווצים צלע, היא נמחקת. כלומר החתך זה כל הצלעות שלא מחקנו. אז אם החזרנו את A זה אומר שבכל שלב בחרנו צלע שלא שייכת ל-A.

ואכן, מכיוון ש A הוא חתך, אם נוריד אותו זה מחלק את V לשתי קבוצות: S, $V\setminus S$. שאין צלעות ביניהן.
אם נכווץ צלעות ששני הקודקודים שלהם שייכים לאחת הקבוצות $n-2$ פעמים,
אז S ו- $V\setminus S$ יהפכו כל אחת לקודקוד והצלעות שמחברות בין שני הקודקודים האלה הן בדיוק A.
נשאר להוכיח ש $\mathbb{P}\left(\bigcap_{j=1}^{n-2}E_j\right)\ge\binom{n}{2}^{-1}$.
מכיוון שהחתך המינימלי הוא בגודל k, כל חתך הוא בגודל לפחות k.
בפרט, הדרגה המינימלית היא לפחות k (כי אחרת נוכל פשוט לקחת רק את הקודקוד הזה).
ולכן, $|E(G)|\ge kn\setminus 2$. (הדרגה המינימלית כפול מספר הקודקודים, חלקי 2)
עכשיו נוכל לקבל את ההסתברות של E_1: מכיוון ש e_1 נבחרה באופן מקרי ואחיד מתוך $E(G)$, נקבל ש:

$$\mathbb{P}(E_1)=\frac{|E(G)\setminus A|}{|E(G)|}=\frac{|E(G)|-|A|}{|E(G)|}=1-\frac{|A|}{|E(G)|}\ge 1-\frac{k}{kn\setminus 2}=1-\frac{2}{n}$$

נשים לב ש $|V(G_1)| = n - 1$, ושהגודל של חתך מינימלי ב-G_1 הוא לפחות k
(כי חתך מינימלי של G_1 הוא גם חתך מינימלי של G). ולכן:

$$\mathbb{P}(E_2|E_1) \geq \frac{|E(G_1)\backslash A|}{|E(G_1)|} = \frac{|E(G_1)| - |A|}{|E(G_1)|} = 1 - \frac{|A|}{|E(G_1)|} \geq 1 - \frac{k}{k(n-1)\backslash 2} = 1 - \frac{2}{n-1}$$

ובאופן דומה, לכל $3 \leq i \leq n-2$:

$$\mathbb{P}\left(E_i \middle| \cap_{j=1}^{i-1} E_j\right) \geq 1 - \frac{k}{k(n-i+1)\backslash 2} = 1 - \frac{2}{n-i+1}$$

(1 פחות k חלקי החסם התחתון על מספר הצלעות).

בסה"כ נקבל ש:

$$\mathbb{P}\left(\bigcap_{j=1}^{n-2} E_j\right) = \mathbb{P}(E_1) \cdot \mathbb{P}(E_2|E_1) \cdot \dots \cdot \mathbb{P}\left(E_{n-2} \middle| \cap_{i=1}^{n-3} E_i\right) \geq \prod_{i=1}^{n-2}\left(1 - \frac{2}{n-i+1}\right) = \prod_{i=1}^{n-2}\left(\frac{n-i+1-2}{n-i+1}\right)$$
$$= \prod_{i=1}^{n-2}\left(\frac{n-i-1}{n-i+1}\right) = \left(\frac{n-2}{n}\right)\left(\frac{n-3}{n-1}\right)\left(\frac{n-4}{n-2}\right) \cdot \dots \cdot \left(\frac{3}{5}\right)\left(\frac{2}{4}\right)\left(\frac{1}{3}\right) =^{\text{א}} \frac{2}{n(n-1)} = \binom{n}{2}^{-1}$$

כמו באלגוריתם של הפולינומים, אם נחזור על האלגוריתם מספיק פעמים נקבל הסתברות גבוהה לתשובה נכונה.
לדוגמה, אם נריץ $n(n-1)\ln n$ פעמים ונחזיר את החתך הכי קטן שמצאנו, נקבל שההסתברות לטעות היא לכל היותר:

$$\left(1 - \frac{2}{n(n-1)}\right)^{n(n-1)\ln n} \leq^{\text{א}} e^{-2\ln n} = \frac{1}{n^2}$$

א - $1 - p \leq e^{-p}$

הסיבוכיות היא $O(n^4 \ln n)$

הרצאה 8

1 אלגוריתמים מקריים

1.1 מיון מהיר רנדומלי

אלגוריתם מיון מהיר רנדומלי - RandQS:

קלט: קבוצה $S = \{x_1 \dots x_n\}$ של מספרים ממשיים זרים.
פלט: האיברים של S בסדר ממוין.

1. אם $|S| \leq 1$, נחזיר את S.
2. נבחר ציר $p \in S$ באופן מקרי ואחיד.
3. נחלק את שאר האיברים של S לשתי קבוצות כך:
   a. $S_1 = \{x \in S : x < p\}$
   b. $S_2 = \{x \in S : x > p\}$
4. נחזיר: $\mathrm{RandQS}(S_1),\ p,\ \mathrm{RandQS}(S_2)$

משפט 1.1: זמן ריצה הצפוי הוא $\Theta(n \log n)$. בעצם, זמן הריצה הוא משתנה מקרי, וזה התוחלת שלו.

טענת עזר 1.2: נסתכל על המערך הממוין $y_1 < \cdots < y_n$, ניקח שני מספרים $y_i < y_j$.
התקיימה השוואה ביניהם אמ"מ אחד מהם היה האיבר הראשון שנבחר להיות ציר (pivot) מבין $y_i, y_{i+1} \cdots y_j$.

הוכחה: נשים לב שמתקיימת השוואה אמ"מ בזמן מסוים שניהם היו באותה קבוצה ואחד מהם נבחר בתור ציר.
יהי $y \leq k \leq j$ מספר כך ש y_k הוא המספר הראשון שנבחר בתור ציר מתוך $y_i \dots y_j$.
בהכרח קיים איבר כזה כי y_i, y_j נמצאים באותה קבוצה בהתחלה ובסוף לא.
אם k הוא i או j, תתקיים השוואה. אחרת, נשווה את שניהם ל- y_k.
מכיוון שאחד גדול ואחד קטן ממנו, הם יהיו בקבוצות נפרדות ואף פעם לא תתקיים השוואה ביניהם.

הוכחת 1.1: יהי X משתנה מקרי שסופר את המספר הכולל של השוואות. נשים לב שמספיק להראות ש $E(X) = \theta(n \log n)$.
אנחנו נוכיח ש: $E(X) = 2n \ln n + \theta(n)$.

לכל $1 \leq i < j \leq n$, יהי האינדיקטור $X_{ij} = 1$ אם השוונו בין y_i, y_j באלגוריתם. אחרת, 0.
נשים לב שלכל $1 \leq i < j \leq n$, הם מושווים לכל היותר פעם אחת (מסקנה מ1.2).
מכאן נובע שמספר ההשוואות הכולל הוא סכום האינדיקטורים של ההשוואה לכל זוג:

$$X = \sum_{1 \leq i < j \leq n} X_{ij} = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}$$

(נרוץ על כל i מ-1 עד n-1 ובתוך כל i, נרוץ על כל j מ-i עד n).
ולכן מלינאריות התוחלת, נקבל ש:

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\right) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}(X_{ij})$$

עכשיו, מטענה 1.2 נובע שמתקיים לכל $1 \leq i < j \leq n$:

$$\mathbb{E}(X_{ij}) = \mathbb{P}(X_{ij} = 1) = \frac{2}{j - i + 1}$$

(כי מתוך j-i+1 האיברים, ההשוואה תקרה רק אם בחרנו אחד מבין שני איברים). ולכן:

$$\mathbb{E}(X)=\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\mathbb{E}(X_{ij})\overset{\text{א}}{=}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\frac{2}{j-i+1}\overset{\text{ב}}{=}2\sum_{i=1}^{n-1}\sum_{k=2}^{n-i+1}\frac{1}{k}\overset{\text{ג}}{=}2\sum_{k=2}^{n}\sum_{i=1}^{n-k+1}\frac{1}{k}\overset{\text{ד}}{=}2\sum_{k=2}^{n}\frac{n+1-k}{k}=$$

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

$$=2(n+1)\sum_{k=2}^{n}\frac{1}{k}-2(n-1)\overset{\text{ה}}{=}2(n+1)\sum_{k=1}^{n}\frac{1}{k}-2n+2-2(n+1)$$

$$=(2n+2)\sum_{k=1}^{n}\frac{1}{k}-4n$$

א – לפי מה שנובע מטענה 1.2

ב – נציב $k \coloneqq j-1+1$. ה2 יוצא החוצה.

ג – החלפת סדר סכימה: במקום לסכום לפי שורות, נסכום לפי עמודות:

	K=2	K=3	...	K=n-1	K=n
i=1	1/2	1/3	...	1 / n-1	1 / n-2
i=2	1/2	1/3	...	1 / n-1	
…	…	…	...		
i=n-2	1/2	1/3			
i=n-1	1/2				
Sum:	n-1 / 2	n-2 / 3		2 / n-1	1 / n-2

ד – השבר בתוך הסיגמא לא תלוי ב-i. תלוי רק ב-k.

ה – נרצה שהסיגמא תתחיל מ-1. אז נוריד את מה שהוספנו – $2(n+1)$.

ניזכר שבאופן כללי: $\ln n \le \sum_{k=1}^{n}\frac{1}{k} \le \ln n + 1$, כלומר הסיגמה היא $\ln n$ ועוד קבוע. אז:

$$\mathbb{E}(X)=(2n+2)\big(\ln n+\Theta(1)\big)-4n=2n\ln n+\Theta(n)+2\ln n+\Theta(1)=2n\ln n+\Theta(n)$$

כנדרש.

אינטואיציה: זה הממוצע שמצאנו. ובגלל ש $\Omega(n\ln n)$ זה חסם תחתון למיון השוואתי, זה אומר שרוב הפעמים נקבל זמן ריצה קרוב לזה. כי אם הרבה פעמים היינו הרבה מעל זה, היינו צריכים גם הרבה פעמים שהרבה מתחת לזה כדי שזה יהיה הממוצע. אבל אין זמני ריצה שהם הרבה פחות מזה (כי זה חסם תחתון) אז חייב להיות שרוב זמני הריצה קרובים לזה.

טענה 1.3: לכל קבוצה בת n מספרים ממשיים שונים, זמן הריצה יהיה $\theta(n\ln n)$ בהסתברות של לפחות $1-\frac{1}{n}$:

הוכחה: בכל ריצה של האלגוריתם, נבנה עץ בינארי שמתאר את הריצה כך:
בכל קודקוד יש תת קבוצה של S, ואיבר p. הקבוצה S היא השורש.
אם קודקוד מתאים לתת-קבוצה 'S ואיבר p , אז הילד השמאלי S_1' יהיה כל האיברים שקטנים מ-p,
והילד הימני S_2' יהיה כל האיברים שגדולים מ-p. (נניח שהקבוצות לא ריקות. אם קבוצה ריקה זה התנאי בסיס).
החלוקה של 'S לשתי קבוצות לוקחת זמן לינארי לפי גודל הקבוצה $O(|S'|)$.
כדי להוכיח את הטענה, נוכיח שגובה העץ הוא $O(\ln n)$ בהסתברות לפחות $1-\frac{1}{n}$.
נוכיח תחילה שההסתברות שהמרחק בין עלה כלשהו לשורש הוא לפחות $24\ln n$, היא לכל היותר n^{-2}:

יהי P המסלול מעלה כלשהו לשורש. קודקוד ב-P נקרא קודקוד **טוב** אם מתקיים: $\max\{|S_1'|,|S_2'|\} \le \frac{2}{3}|S'|$.
(אם הקבוצה הגדולה היא לכל היותר 2/3 מגודל הקבוצה המקורית, כלומר החלוקה היא בערך באמצע).
אחרת הוא נקרא **רע**. אנחנו נמצא חסמים עליונים לקודקודים רעים ולקודקודים טובים, וככה נגביל את אורך המסלול וגובה העץ.

טענה 1.4: עבור n מספיק גדול, לכל היותר $3\ln n$ מהקודקודים הם טובים.

הוכחה: יהיו $v_1 \cdots v_t$ הקודקודים הטובים, לפי סדר הופעתם ב-P. (1 הכי קרוב לשורש).
לכל $1 \le i \le t$, נגדיר s_i הגודל של תת-הקבוצה של S שמתאימה ל-v_i.
לכל $1 \le i \le t-1$. מתקיים (1): $s_{i+1} \le \frac{2}{3} s_i$
(הגודל של הקבוצה בכל קודקוד הוא לכל היותר 2/3 הגודל של הקודקוד הטוב הקודם במסלול). ולכן:

$$1 \le s_t \le^{\text{א}} \left(\frac{2}{3}\right)^{t-1} n \Rightarrow^{\text{ב}} \left(\frac{3}{2}\right)^{t-1} \le n \Rightarrow^{\text{ג}} t-1 \le \log_{\frac{2}{3}} n \Rightarrow t \le \log_{\frac{2}{3}} n + 1 =^{\text{ד}} \frac{\ln n}{\ln(2/3)} + 1 \le 3 \ln n$$

א – כי (1) מתקיים בכל שלב בדרך.
ב – נתעלם מ s_t. נכפול את שני האגפים ב $\left(\frac{3}{2}\right)^{t-1}$.
ג – נוציא לוג לשני האגפים.
ד – נעבור בסיס לוג.

אינטואיטיבית, כל פיצול טוב מקדם אותי המון, ולכן לא יכולים להיות הרבה כאלה (כי אם יש הרבה פשוט נסיים קודם).

עכשיו, יהי 'P ה- $24 \ln n$ קודקודים הראשונים (קרובים לשורש) במסלול. (אם P קצר יותר, פשוט ניקח את כולו).
יהי X משתנה מקרי שסופר את מספר הקודקודים הרעים ב- 'P.
לכל $u \in P'$ נגדיר אינדיקטור $X_u = 1$ אם u רע, אחרת 0.
נשים לב שמתקיים:

1) $X = \sum_{u \in P'} X_u$,
2) כל ה- X_u בת"ל,
3) $\mathbb{P}(X_u = 1) \le 2/3$.

(כי כדי שקודקוד יהיה רע, צריך לבחור מה- 2/3 איברים הקיצוניים).

בפרט, מתקיים: $\mathbb{E}(X) \le \frac{2}{3}|P'| \le 16 \ln n$.
אנחנו יודעים את התוחלת של X, והוא סכום של אינדיקטורים בת"ל אז אפשר להשתמש באי-שוויון צ'רנוף:

$$\mathbb{P}(|P| \ge 24 \ln n) =^{\text{א}} \mathbb{P}(|P'| = 24 \ln n) \le^{\text{ב}} \mathbb{P}(X \ge 21 \ln n) \le \mathbb{P}(X \ge \mathbb{E}(X) + 5 \ln n) \le^{\text{ג}} e^{-2\frac{(5 \ln n)^2}{24 \ln n}}$$
$$= e^{-2 \cdot \frac{25 \cdot (\ln n)^2}{24 \ln n}} = e^{-2 \cdot \frac{25 \ln n}{24}} = \left(\frac{1}{e^{\ln n}}\right)^{2 \cdot \frac{25}{24}} \le \frac{1}{n^2}$$

א – אם P ארוך יותר מ $24 \ln n$, אז 'P יהיה בדיוק באורך הזה.
ב – מטענה 1.4, מתוך $24 \ln n$ קודקודים, לכל היותר $3 \ln n$ יכולים להיות טובים.
ג – אי"ש צ'רנוף.

אז מכיוון שבקבוצה יש n איברים אז יש לכל היותר n עלים, חסם איחוד נותן לנו שההסתברות שיש מסלול באורך לפחות $24 \ln n$ היא לכל היותר: $n \cdot \mathbb{P}(|P| \ge 24 \ln n) \le n \cdot n^{-2} = 1/n$, כנדרש.

2 סוגי אלגוריתמים הסתברותיים

אלגוריתם לאס וגאס: הפלט תמיד נכון. זמן הריצה הוא משתנה מקרי. (לדוגמה, המיון שראינו עכשיו).
ההגדרה הסטנדרטית כוללת דרישה שהתוחלת של זמן הריצה תהיה סופית.

אלגוריתם מונטה קרלו: הפלט יכול לטעות בהסתברות (בדרך כלל קטנה מאד). יש שני תתי-סוגים:

טעות חד-צדדית: אם יצא אמת (בה"כ), זה בוודאות נכון. אבל אם יצא שקר (בה"כ) יש הסתברות לטעות.
(שני האלגוריתמים בהרצאה הקודמת הם כאלה).

טעות דו-צדדית: בכל פלט יש הסתברות לטעות.

הרצאה 10

תזכורת מהרצאה 9: (לא היה השנה)

פונקציה μ היא $\sigma-\text{additive}$, (סיגמא-אדטיבית), אם היא מקיימת: $\mu(\cup_{n=1}^{\infty}A_n)=\Sigma_{n=1}^{\infty}\mu(A_n)$
(כלומר, ערך הפונקציה על איחוד בן-מניה שווה לסכום הערכים על כל איברי האיחוד).

סיגמא-אלגברה A על קבוצה X היא משפחה של תתי קבוצות של X שמקיימת:

1) $\emptyset\subseteq A$
2) אם $E\in A$ אז גם $E^C=X\setminus E\in A$ (המשפחה סגורה ללקיחת משלים).

מרחב הסתברות הוא שלישיה: $(\Omega,\mathcal{F},\mathbb{P})$ כאשר:

1) Ω היא קבוצה לא-ריקה. נקראת **מרחב המדגם.**
2) $\mathcal{F}\subseteq 2^{\Omega}$ היא $\sigma-\text{algebra}$, (סיגמא-אלגברה).
3) $\mathbb{P}:\mathcal{F}\to[0,1]$ היא פונקציית הסתברות סיגמא-אדטיבית שמקיימת $\mathbb{P}(\emptyset)=0$ ו- $\mathbb{P}(\Omega)=1$

כאשר Ω סופית או בת מניה, אפשר לקחת $\mathcal{F}=2^{\Omega}$ ואז אפשר לרשום $(\Omega,\mathbb{P})$.

1 משתנים מקריים רציפים:

יהי $X:\mathbb{R}\to\mathbb{R}$ משתנה מקרי במרחב הסתברות $(\mathbb{R},\mathcal{F},\mathbb{P})$,
(ℝ היא ה- Ω, ו- $\mathcal{F}\subseteq 2^{\Omega}$. 2^{Ω} זה המאורעות (תת הקבוצות של מרחב המדגם), $\mathcal{F}$ זה המאורעות שעבורם אנחנו יודעים להגדיר את ℙ.)
כאשר $\mathcal{F}$ היא סיגמא-אלגברה שכוללת את כל המקטעים, כלומר $(a,b),(a,b],[a,b]\in\mathcal{F}$ וכו'. אנחנו רוצים ש$\mathcal{F}$ תהיה סיגמא-אלגברה כדי שאם נדע את ההסתברות של מאורע, נוכל לדעת את ההסתברות של המשלים שלו, ואם נדע את ההסתברות של קבוצת מאורעות נוכל לדעת את ההסתברות של האיחוד והחיתוך.
נאמר ש X **רציף** (או רציף בהחלט) אם יש דרך מיוחדת לחשב את ההסתברות של המאורעות שתלויים ב X.
פורמלית: אם קיימת פונקציה $f_X:\mathbb{R}\to\mathbb{R}$ כך ש (1):

$$\mathbb{P}(X\in B)=\int_B f_X(x)dx \qquad \text{לכל } B\in\mathcal{F}.$$

בפרט, צריך שיתקיים: $\mathbb{P}(\mathbb{R})=\int_{-\infty}^{\infty}f_X(x)dx=1$.

הפונקציה f_X נקראת **פונקציית הצפיפות** (PDF – Probability Density Function) של X.

מ- (1) נובע ש: $\mathbb{P}(a\le X\le b)=\int_a^b f_X(x)dx$ לכל שני $a\le b$ ממשיים.

בפרט, $\mathbb{P}(X=c)=\int_c^c f_X(x)dx=0$ לכל $c\in\mathbb{R}$. ולכן: לכל $a\le b$ ממשיים:

$$\mathbb{P}(a\le X\le b)=\mathbb{P}(a<X\le b)=\mathbb{P}(a\le X<b)=\mathbb{P}(a<X<b)=\int_a^b f_X(x)dx$$

לכל $a\in\mathbb{R}$ נגדיר $F_X(a)=\mathbb{P}(X\le a)=\int_{-\infty}^{a}f_X(x)dx$. בהמשך נראה למה זה ככה.

הפונקציה F_X נקראת **פונקציית ההתפלגות המצטברת** (CDF – Cumulative Distribution Function) של X.

זה נותן לנו גם ש: $\mathbb{P}(a\le X\le b)=\mathbb{P}(X\le b)-\mathbb{P}(X<a)=\mathbb{P}(X\le b)-\mathbb{P}(X\le a)=F_X(b)-F_X(a)$.

דוגמה 1: יהי X משתנה מקרי רציף, ותהי פונקציית הצפיפות שלו: (ה-PDF)

$$f_X(x)=\begin{cases}c(1-x^2), & if -1<x<1\\ 0, & otherwise\end{cases} \qquad \text{עבור קבוע } c.$$

כדי למצוא את הערך של c נשתמש במשוואה $\int_{-\infty}^{\infty}f_X(x)dx=1$. מתקיים:

$$1=\int_{-\infty}^{\infty} f_X(x)dx=\int_{-\infty}^{-1} 0dx+\int_{-1}^{1} c(1-x^2)dx+\int_{1}^{\infty} 0dx=\int_{-1}^{1} c(1-x^2)dx=c\left[x-\frac{x^3}{3}\right]|_{x=-1}^{x=1}=$$
$$=c\left[\left(1-\frac{1}{3}\right)-\left(-1+\frac{1}{3}\right)\right]=\frac{4c}{3}$$

כלומר $c=3/4$. ועכשיו נוכל לחשב הסתברויות עם X, לדוגמה:

$$\mathbb{P}(X\le 0)=\int_{-\infty}^{0} f_X(x)dx=\int_{-1}^{0}\frac{3(1-x^2)}{4}dx=\left[\frac{3x}{4}-\frac{x^3}{4}\right]|_{x=-1}^{x=0}=0-\left(-\frac{3}{4}+\frac{1}{4}\right)=\frac{1}{2}$$

דוגמה 2: יהי X משתנה מקרי רציף עם פונקציית צפיפות f_X ופונקציית התפלגות מצטברת F_X.
נרצה למצוא את ה- PDF ו- CDF של המשתנה המקרי $Y=2X$. לכל $a\in\mathbb{R}$ מתקיים:

$$F_Y(a)=\mathbb{P}(Y\le a)=\mathbb{P}(2X\le a)=\mathbb{P}\left(X\le\frac{a}{2}\right)=F_X\left(\frac{a}{2}\right)$$

נגזור את F_Y לפי a ונקבל: $f_Y(a)=\frac{1}{2}f_X\left(\frac{a}{2}\right)$

1.1 תוחלת ושונות של משתנה מקרי רציף:

ניזכר שאם X הוא בדיד, התוחלת היא סכום ההסתברויות על התומך: $\mathbb{E}(X)=\sum_x x\mathbb{P}(X=x)$.

באופן דומה, במ"מ רציף, התוחלת היא: $\mathbb{E}(X)=\int_{-\infty}^{\infty} xf_X(x)dx$.

דוגמה 3: יהי X מ"מ רציף עם PDF המוגדרת: $f_X(x)=\begin{cases}2x, & if\ 0\le x\le 1\\ 0, & otherwise\end{cases}$

נשים לב שזה באמת PDF, כי: $\int_{-\infty}^{\infty} f_X(x)dx=\int_0^1 2xdx=x^2|_{x=0}^{x=1}=1$, אז:

$$\mathbb{E}(X)=\int_{-\infty}^{\infty} xf_X(x)dx=\int_0^1 2x^2dx=2\frac{x^3}{3}|_{x=0}^{x=1}=\frac{2}{3}$$

באופן דומה למקרה של מ"מ בדיד, אפשר להשתמש בPDF כי לחשב את התוחלת של פונקציה כלשהי של X.

טענה 1.1: יהי X מ"מ רציף עם PDF f_X. אזי, לכל פונקציה $g:\mathbb{R}\to\mathbb{R}$ מתקיים:

$$\mathbb{E}\big(g(X)\big)=\int_{-\infty}^{\infty} g(x)f_X(x)dx$$

כלומר נוכל להשתמש בפונקציה של x, וההבדל היחיד הוא שכופלים ב- $g(x)$.

למה 1.2: הוכחה עבור פונקציות אי-שליליות:

יהי Y מ"מ רציף אי"ש עם PDF f_Y. אזי:

$$\mathbb{E}(Y)=\int_0^{\infty}\mathbb{P}(Y>y)dy$$

הוכחה: מכיוון שמתקיים $\mathbb{P}(Y>y)=\int_y^{\infty} f_Y(x)dx$ לכל $y\in\mathbb{R}$, נובע ש:

$$\int_0^{\infty}\mathbb{P}(Y>y)dy=^{א}\int_0^{\infty}\left(\int_y^{\infty} f_Y(x)dx\right)dy=^{ב}\int_0^{\infty}\left(\int_0^{x} dy\right)f_Y(x)dx=^{ג}\int_0^{\infty} xf_Y(x)dx=^{ד}\mathbb{E}(Y)$$

א – נציב את הנוסחה.

ב – החלפת סדר אינטגרציה: בשלב הראשון, $y \in [0,\infty); x \in [y,\infty)$. כלומר תמיד $0 \le y \le x$. האינטגרל הפנימי הוא לפי x והחיצוני הוא לפי y. אנחנו רוצים להחליף את הסדר, כדי שהאינטגרל החיצוני יהיה לפי x והפנימי לפי y. במקום לסכום לפי x בתחום $[y,\infty)$, (כאשר y בתחום $[0,\infty)$) נסכום לפי y בתחום $[0,x)$. (כאשר x בתחום $[0,\infty)$).
ואם קבענו בחוץ את x, אז $f_Y(x)$ הוא קבוע בזמן שעושים את האינטגרל לפי y. אז נוציא אותו החוצה.

ג – האינטגרל שהוא רק dy זה בעצם האינטגרל של הפונקציה 1, שזה פשוט y. ואז בחישוב מציבים $x - 0$.

ד – לפי הגדרה, ובגלל ש-Y אי שלילי אז אפשר להתעלם מהתחום השלילי.

הוכחת טענה 1.1:

א – בהנחה ש g אי"ש, מלמה 1.2 נובע:

$$\mathbb{E}\big(g(X)\big) =^{א} \int_0^\infty \mathbb{P}(g(X) > y)\, dy =^{ב} \int_0^\infty \int_{x:g(x)>y} f_X(x)\, dx\, dy =^{ג} \int_{x:g(x)>0} \left(\int_0^{g(x)} dy\right) f_X(x) dx =^{ד}$$
$$= \int_{x:g(x)>0} g(x) f_X(x)\, dx =^{ה} \int_{-\infty}^\infty g(x) f_X(x)\, dx$$

ב – לפי (1). האינטגרל על הקבוצה המתאימה.

ג – החלפת סדר אינטגרציה. החיצוני הוא y, בתחום $[0,\infty)$. הפנימי x הוא כך ש $y < g(x)$.

ד – האינטגרל הפנימי שווה $g(x)$, ו- $f_X(x)$ הוא קבוע מבחינתו.

ה – כי g אי"ש ואם היא שווה 0, הכל יוצא אפס. אז בכל מקרה נסכום רק על הקבוצה שבה $g(x) > 0$.

טענה 1.3: לינאריות התוחלת

יהיו $a, b \in \mathbb{R}$ ויהי X מ"מ רציף. אז מתקיים: $\mathbb{E}(aX + b) = a \cdot \mathbb{E}(X) + b$

הוכחה: תהי f_X ה-PDF של X. אז מטענה 1.1 נובע:

$$\mathbb{E}(aX + b) = \int_{-\infty}^\infty (ax + b) f_X(x)\, dx = a \underbrace{\int_{-\infty}^\infty x f_X(x)\, dx}_{\mathbb{E}(X)} + b \underbrace{\int_{-\infty}^\infty f_X(x)\, dx}_{=1} = a\mathbb{E}(X) + b$$

כמו במקרה של מ"מ בדיד, **השונות** של X מוגדרת: $Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$,
וחישוב ישיר נותן: $Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2$

דוגמה 4: יהי X מ"מ רציף עם PDF המוגדרת: $f_X(x) = \begin{cases} 2x, & if\ 0 \le x \le 1 \\ 0, & otherwise \end{cases}$

ראינו כבר שזו אכן PDF, ושמתקיים: $\mathbb{E}(X) = 2/3$. לפי טענה 1.1 נקבל:

$$\mathbb{E}(X^2) = \int_{-\infty}^\infty x^2 f_X(x)\, dx = \int_0^1 2x^3\, dx = \frac{x^4}{2}\Big|_{x=0}^{x=1} = \frac{1}{2}$$

ולכן: $Var(X) = \mathbb{E}(X^2) - \big(\mathbb{E}(X)\big)^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}$

טענה 1.4: יהיו $a, b \in \mathbb{R}$ ויהי X מ"מ רציף, אז: $Var(aX + b) = a^2 Var(X)$.

הזזה בסקלר לא משפיעה, כפל בסקלר יוצא בריבוע.

ההוכחה זהה להוכחה מהסתברות 1 על מ"מ בדיד.

הרצאה 11

1 משתנים מקריים רציפים נפוצים:

נראה מספר מ"מ רציפים עם התפלגויות נפוצות.

1.1 התפלגות אחידה:

עבור $a<b$ ממשיים, נאמר שמ"מ רציף X הוא בעל התפלגות **אחידה** על הקטע [a, b] אם פונקציית הצפיפות שלה היא:

$$f_X(x)=\begin{cases}\frac{1}{b-a}, & if\ a\le x\le b\\ 0, & else\end{cases}$$

נשים לב שזו אכן PDF כי:

$$\int_{-\infty}^{\infty} f_X(x)\,dx=\frac{1}{b-a}\int_a^b dx=1$$

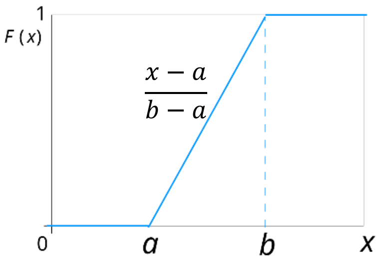
נרצה למצוא את ה-CDF של X. נשים לב שלכל $x<a$ מתקיים $F_X(x)=0$, ולכל $x>b$ מתקיים $F_X(x)=1$.
לכל $a\le x\le b$ מתקיים:

$$F_X(x)=\int_{-\infty}^{x} f_X(t)\,dt=\int_{-\infty}^{x}\frac{1}{b-a}dt \stackrel{א}{=} \frac{1}{b-a}\int_a^x dt=\left(\frac{x}{b-a}\right)-\left(\frac{a}{b-a}\right)=\frac{x-a}{b-a}$$

א – כי לכל ערך קטן מ-a נקבל 0. השבר הוא קבוע ויוצא מהאינטגרל.

כלומר:

$$F_X(x)=\begin{cases}0 & if\ x<a\\ \left(\frac{x-a}{b-a}\right) & if\ a\le x\le b\\ 1 & if\ x>b\end{cases}$$



דוגמה 1: יהי X מ"מ שמתפלג אחיד על הקטע [0,1], ויהי $Y=aX+b$ עבור $a>0, b\in\mathbb{R}$ כלשהם.
נרצה למצוא את ההתפלגות של Y: נשים לב שמתקיים (לפי הצבת ערכי הקטע בנוסחה):

$$F_X(x)=\begin{cases}0, & x<0\\ x, & 0\le x\le 1\\ 1, & x>1\end{cases}$$

אז לכל $y\in\mathbb{R}$ מתקיים (בשלב הרביעי משתמשים בנתון ש $a>0$):

$$F_Y(y)=\mathbb{P}(Y\le y)=\mathbb{P}(aX+b\le y)=\mathbb{P}\left(X\le\frac{y-b}{a}\right)=\begin{cases}0, & if\ \frac{y-b}{a}<0\\ \frac{y-b}{a}, & if\ 0\le\frac{y-b}{a}\le 1\\ 1, & if\ \frac{(y-b)}{a}>1\end{cases}\stackrel{א}{=}$$

$$=\begin{cases}0, & if\ \ y<b\\ \frac{y-b}{(a+b)-b}, & if\ \ b\le y\le a+b\\ 1, & if\ \ y>a+b\end{cases}$$

א – נכפיל את כל האגפים ב- a, ונוסיף את b לכל האגפים. בשבר הוספנו b והחסרנו b.

כלומר Y מתפלג אחיד על הקטע $[b,(a+b)]$.

אינטואיטיבית זה הגיוני, כי עשינו בסה"כ פעולה לינארית על X – "מתחנו" לפי a ו"הזזנו" לפי b. העתקה לינארית חח"ע ועל.

נחשב את התוחלת והשונות של X:

$$\mathbb{E}(X)=\int_{-\infty}^{\infty}xf_X(x)\,dx=\frac{1}{b-a}\int_a^b x\,dx=\frac{1}{b-a}\cdot\frac{x^2}{2}\bigg|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{(b-a)(b+a)}{2(b-a)}=\frac{b+a}{2}$$

כדי לחשב שונות נחשב את $\mathbb{E}(X^2)$:

$$\mathbb{E}(X^2)=\int_{-\infty}^{\infty}x^2f_X(x)\,dx=\frac{1}{b-a}\int_a^b x^2\,dx=\frac{1}{b-a}\cdot\frac{x^3}{3}\bigg|_a^b=\frac{b^3-a^3}{3(b-a)}=\frac{a^2+ab+b^2}{3}$$

ולכן:

$$\begin{aligned}Var(X)=\mathbb{E}(X)-\left(\mathbb{E}(X)\right)^2&=\frac{a^2+ab+b^2}{3}-\left(\frac{b+a}{2}\right)^2=\frac{4(a^2+ab+b^2)}{12}-\frac{3(b^2+2ab+a^2)}{12}\\&=\frac{4a^2+4ab+4b^2}{12}-\frac{3b^2+3ab+3a^2}{12}=\frac{a^2+ab+b^2}{12}=\frac{(b-a)^2}{12}\end{aligned}$$

1.2 התפלגות מעריכית (אקספוננציאלית)

אפשר לחשוב עליה בתור גרסה רציפה של ההתפלגות הגיאומטרית.

עבור $\lambda>0\in\mathbb{R}$ כלשהו, מ"מ רציף X הוא בעל התפלגות **מעריכית** עם פרמטר λ אם הPDF שלו מוגדר:

$$f_X(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases}$$

נשים לב שזאת אכן פונקציית צפיפות כי:

$$\int_{-\infty}^{\infty}f_X(x)\,dx=\int_0^{\infty}\lambda e^{-\lambda x}\,dx\overset{\text{א}}{=}-e^{-\lambda x}\Big|_0^{\infty}=\underbrace{\Big(-\underbrace{e^{-\lambda\infty}}_{e^{-\infty}=0}\Big)}_{0}-\underbrace{\Big(-\underbrace{e^{-\lambda\cdot 0}}_{e^{0}=1}\Big)}_{-1}=0-(-1)=1$$

א – ניזכר שהנגזרת של $e^{f(x)}$ היא $f'(x)e^{f(x)}$, כלומר כשנגזור את $e^{-\lambda x}$ נקבל $-\lambda e^{-\lambda x}$. צריך בלי המינוס, אז נתקן: $\left(-e^{-\lambda x}\right)'=(-\lambda)\left(-e^{-\lambda x}\right)=\lambda e^{-\lambda x}$.

נרצה למצוא את ה-CDF של X: נשים לב שאם $x<0$ אז $F_X(x)=0$ ומצד שני, אם $x\geq 0$ אז

$$F_X(x)=\int_{-\infty}^{x}f_X(t)\,dt=\int_0^x\lambda e^{-\lambda t}\,dt=-e^{-\lambda t}\Big|_0^x=-e^{-\lambda x}-(-1)=1-e^{-\lambda x}$$

כלומר

$$F_X(x)=\begin{cases}0, & x<0\\ 1-e^{-\lambda x}, & x\geq 0\end{cases}$$

תזכורת מהסתברות 1:

יהי $X\sim Geom(p)$, x שלם חיובי:

$$\begin{aligned}\mathbb{P}(X>x)&=\sum_{k=x+1}^{\infty}\mathbb{P}(X=k)=\sum_{k=x+1}^{\infty}p(1-p)^{k-1}=p\sum_{k=x+1}^{\infty}(1-p)^{k-1}=p\cdot\frac{(1-p)^x}{1-(1-p)}\\&=(1-p)^x\overset{\text{א}}{\approx}e^{-px}\end{aligned}$$

א – עבור p קטן

אם $X\sim Exp(\lambda)$ אז: (x יכול להיות ממשי, λ לא חייב להיות קטן)

$$\mathbb{P}(X>x)=1-\underbrace{F_X(x)}_{\mathbb{P}(X\leq x)}=1-\left(1-e^{-\lambda x}\right)=e^{-\lambda x}$$

ניתן לראות את הדמיון ביניהם. בשניהם מודדים "עד להצלחה". בבדיד מודדים ניסיונות, ברציף מודדים זמן.

דוגמה 2: יהי X מ"מ מעריכי עם פרמטר $\lambda>0$. נגדיר $Y=cX$ עבור $c>0$ ממשי כלשהו. נרצה למצוא את ההתפלגות של Y. מכיוון ש $c>0$, לכל $x\in\mathbb{R}$ מתקיים ש $x<0$ אמ"מ $cx<0$. לכן:

$$F_Y(y)=\mathbb{P}(Y\leq y)=\mathbb{P}(cX\leq y)=\mathbb{P}\left(X\leq\frac{y}{c}\right)=\begin{cases}0, & y<0\\ 1-e^{-\lambda y\backslash c}, & y\geq 0\end{cases}$$

כלומר גם Y מתפלג מעריכית, עם פרמטר λ/c.

נחשב את התוחלת והשונות של X: נעשה אינטגרציה בחלקים עם $u=x,v'=\lambda e^{-\lambda x}$ ונקבל את משוואה (1):

$$\mathbb{E}(X)=\int_{-\infty}^{\infty}xf_X(x)dx=\underbrace{\int_0^{\infty}x\cdot\lambda e^{-\lambda x}dx}_{\int uv'}=\underbrace{x\cdot\int_0^{\infty}-e^{-\lambda x}dx}_{uv}-\underbrace{\int_0^{\infty}1\cdot -e^{-\lambda x}dx}_{\int u'v}=$$
$$\overset{\text{א}}{=}-xe^{-\lambda x}|_0^{\infty}-\int_0^{\infty}-e^{-\lambda x}dx\overset{\text{ב}}{=}(0-0)+\frac{1}{\lambda}\underbrace{\int_0^{\infty}\lambda e^{-\lambda x}dx}_{PDF}=\frac{1}{\lambda}$$

תזכורת לאינטגרציה בחלקים: $(uv)'=u'v+uv'\Rightarrow uv=\int u'v+\int uv'\Rightarrow\int uv'=uv-\int u'v$

א – כשמציבים ∞, מקבלים $-\infty e^{-\infty}=-\infty\cdot\frac{1}{e^{\infty}}$, שנראה כאילו זה $-\frac{\infty}{\infty}$, לא מוגדר. אבל פונקציה עם אינסוף במעריך שואפת לאינסוף הרבה יותר מהר מ"סתם" אינסוף, אז זה שואף לאפס. לכן בשלב הבא זה $(0-0)$.

ב – נשים לב שזה דומה למשהו שכבר חישבנו לו את האינטגרל, רק בלי ה-λ. אז נחלק ונכפיל את האינטגרל בקבוע.

התוחלת בהתפלגות גיאומטרית הייתה $\frac{1}{p}$, ושוב רואים את הדמיון ביניהם.

כדי לחשב את השונות של X, נחשב את $\mathbb{E}(X^2)$. נעשה אינטגרציה בחלקים עם $u=x^2,v'=\lambda e^{-\lambda x}$ ועם (1) נקבל:

$$\mathbb{E}(X^2)\overset{\text{א}}{=}\int_{-\infty}^{\infty}x^2f_X(x)dx=\int_0^{\infty}x^2\lambda e^{-\lambda x}dx=-x^2e^{-\lambda x}|_0^{\infty}-\int_0^{\infty}-2xe^{-\lambda x}dx$$
$$\overset{\text{ב}}{=}(0-0)+2\int_0^{\infty}xe^{-\lambda x}dx\overset{\text{ג}}{=}\frac{2}{\lambda}\underbrace{\int_0^{\infty}x\cdot\lambda e^{-\lambda x}dx}_{\mathbb{E}(X)=1/\lambda}=\frac{2}{\lambda^2}$$

א – כמו שכבר ראינו, חישוב תוחלת של פונקציה של משתנה מקרי.

ב – כמו א בשלב הקודם.

ג – נשים לב שזה דומה לתוחלת. נחלק ונכפיל בקבוע כדי להביא את האינטגרל למצב הרצוי.

ולכן:

$$Var(X)=\mathbb{E}(X^2)-\left(\mathbb{E}(X)\right)^2=\frac{2}{\lambda^2}-\left(\frac{1}{\lambda}\right)^2=\frac{2}{\lambda^2}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}$$

תכונות חוסר זיכרון

משתנה מקרי אי שלילי X ייקרא **חסר זיכרון** אם לכל $s,t \geq 0$ ממשיים מתקיים: $\mathbb{P}(X > s + t \mid X > t) = \mathbb{P}(X > s)$

כלומר, ההסתברות ש: $X > s + t$ בהינתן ש: $X > t$, שווה להסתברות ש: $X > s$.

לדוגמה, אם נטיל מטבע באופן בלתי תלוי, ההסתברות שנקבל עץ ב-4 פעמים הבאות שווה להסתברות שנקבל עץ ב-4 פעמים הבאות גם בהינתן זה שיצא עץ כבר 5 פעמים ברצף. זה משתנה מקרי ש"לא זוכר את ההיסטוריה".

נשים לב שמתקיים באופן כללי: (נוסחת הסתברות מותנה)

$$\mathbb{P}(X > s + t \mid X > t) = \frac{\mathbb{P}(X > s + t, X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > t)}$$

ולכן, משתנה מקרי אי שלילי X יהיה חסר זיכרון אמ"מ, לכל $s,t \geq 0$ ממשיים מתקיים: (2)

$$\mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t)$$

משפט 1.1: יהי X משתנה מקרי רציף אי-שלילי. אזי X חסר זיכרון אמ"מ הוא מתפלג מעריכית.

הוכחה: בכיוון הראשון, נניח ש X מתפלג מעריכית עם פרמטר $\lambda > 0$.

אז, לכל $s,t \geq 0$ מתקיים:

$$\begin{aligned}
\mathbb{P}(X > s)\mathbb{P}(X > t) &= \big(1 - \mathbb{P}(X \leq s)\big)\big(1 - \mathbb{P}(X \leq t)\big) = \big(1 - F_X(s)\big)\big(1 - F_X(t)\big) = \\
&= \left(1 - \left(1 - e^{-\lambda s}\right)\right)\left(1 - \left(1 - e^{-\lambda t}\right)\right) = e^{-\lambda s}e^{-\lambda t} = e^{-\lambda(s+t)} = 1 - \left(1 - e^{-\lambda(s+t)}\right) = \\
&= 1 - F_X(s + t) = 1 - \mathbb{P}(X \leq s + t) = \mathbb{P}(X > s + t)
\end{aligned}$$

בכיוון השני, נניח ש X משתנה מקרי רציף אי-שלילי חסר זיכרון. נגדיר את F_X ה- CDF של X.

אם נצליח להראות ש $F_X(x) = 1 - e^{-\lambda x}$ לכל $x \geq 0$ (ושווה 0 אחרת), זה יראה ש-X מתפלג מעריכית.

כדי לא לגרור את ה- "-1", נגדיר את הפונקציה: $g(x) = 1 - F_X(x)$ לכל $x \geq 0$.

ועכשיו מספיק להוכיח ש: $g(x) = e^{-\lambda x}$. (עבור $x \geq 0$)

מכיוון ש-X חסר זיכרון מתקיים לפי (2): (3)

$$\begin{aligned}
g(s + t) &= 1 - F_X(s + t) = 1 - \mathbb{P}(X \leq s + t) = \mathbb{P}(X > s + t) = \mathbb{P}(X > s)\mathbb{P}(X > t) \\
&= \big(1 - \mathbb{P}(X \leq s)\big)\big(1 - \mathbb{P}(X \leq t)\big) = \big(1 - F_X(s)\big)\big(1 - F_X(t)\big) = g(s)g(t)
\end{aligned}$$

אנחנו נטען ש: $g\left(\frac{m}{n}\right) = g^m\left(\frac{1}{n}\right)$ לכל $m,n > 0$ שלמים.

(כאשר $g^m\left(\frac{1}{n}\right)$ הכוונה היא ל- $\left(g\left(\frac{1}{n}\right)\right)^m$. הסימון הוא לצורך הנוחות. לא להתבלבל עם הרכבה).

נקבע n כלשהו ונוכיח באינדוקציה על m. הטענה מתקיימת באופן טריוויאלי עבור $m = 1$. זה בסיס האינדוקציה.

נניח שהטענה מתקיימת עבור $m \geq 1$ כלשהו ונוכיח עבור $(m + 1)$. מתקיים:

$$g\left(\frac{m + 1}{n}\right) = g\left(\frac{m}{n} + \frac{1}{n}\right) \overset{\text{א}}{=} g\left(\frac{m}{n}\right)g\left(\frac{1}{n}\right) \overset{\text{ב}}{=} g^m\left(\frac{1}{n}\right)g\left(\frac{1}{n}\right) = g^{m+1}\left(\frac{1}{n}\right)$$

א – לפי (3). ב – לפי הנ"א.

זה נתן לנו מידע על g ברציונליים חיוביים. זה יאפשר לנו לקבוע את g: נשים לב ש:

$$g(1) = g\left(\frac{n}{n}\right) = g^n\left(\frac{1}{n}\right) \overset{\text{א}}{\Longrightarrow} g\left(\frac{1}{n}\right) = \big(g(1)\big)^{1/n}$$

א – נוציא שורש בדרגה n.

כלומר אם אנחנו יודעים את $g(1)$, אנחנו יודעים את $g\left(\frac{1}{n}\right)$ לכל n טבעי. מכאן נובע ש:

$$g\left(\frac{m}{n}\right)=\left(g\left(\frac{1}{n}\right)\right)^m=\left(g(1)^{1/n}\right)^m=\left(g(1)\right)^{\frac{m}{n}}$$

אז אנחנו יודעים לחשב את g לכל רציונלי (כתלות ב- $g(1)$).

ניזכר שX רציפה ולכן גם g רציפה. ומכיוון שהיא רציפה, אנחנו יודעים אותה לכל ממשי. (קצת אינפי).

כלומר מתקיים לכל $x\geq 0$ ממשי: $g(x)=\left(g(1)\right)^x$.

אז עכשיו אנחנו יודעים ש- g אקספוננציאלית, אז זה כבר בכיוון הנכון. עכשיו צריך להביא אותה לצורה של $e^{-\lambda x}$:

נקבע $\lambda=-\ln g(1)$. אנחנו צריכים שיתקיים: $0<g(1)<1$ (כדי ש-λ יהיה מוגדר וחיובי).

מכיוון ש $g(1)$ היא הסתברות של מאורע, מתקיים $0\leq g(1)\leq 1$.

נב"ש ש $g(1)=1$, אזי לפי מה שמצאנו מתקיים $g\equiv 1$ ואז $F_X\equiv 0$.
זאת סתירה לדרישה ש $F_X(\infty)=1$ (דרישה עבור כל CDF).

נב"ש ש $g(1)=0$, אז $F_X(x)=1$ לכל $x>0$, ו- $F_X(x)=0$ לכל $x\leq 0$.
סתירה לכך ש- F_X רציפה.

ולכן, מתקיים:

$$g(x)=\left(g(1)\right)^x=e^{\ln\left(g(1)\right)^x}=e^{x\ln g(1)}=e^{-\lambda x}$$

כנדרש.

הרצאה 12

1 משתנים מקריים נורמליים

משתנה מקרי רציף X ייקרא בעל התפלגות **נורמלית** עם פרמטרים μ,σ^2 (כאשר $\sigma>0$) אם הPDF שלו מוגדרת לכל $x\in\mathbb{R}$:

$$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$$

נוכיח שזו אכן PDF: ראשית, נציב $y=(x-\mu)/\sigma$ כך שמתקיים: $\frac{dx}{dy}=\frac{1}{\sigma}$.

אנחנו רוצים להראות: $\int_{-\infty}^{\infty}f_X(x)dx=1$. קודם נציב את הנוסחה מלמעלה, ואת החלק הקבוע (השבר) נוציא החוצה. ואז נציב.
בגלל איך שקבענו את y, מתקיים:

$$\frac{y^2}{2}=\frac{\left(\frac{x-\mu}{\sigma}\right)^2}{2}=\frac{(x-\mu)^2}{2\sigma^2}$$

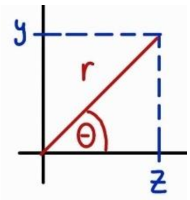
בנוסף, גבולות האינטגרל נשארים כמו שהם כי גם כשנציב $x=-\infty,\infty$ בנוסחה של y נקבל אינסוף.

במקום dx צריך לרשום dy, כלומר צריך לגזור את y לפי x. ולכן ה-σ נעלמת מהמכנה. ולכן מתקיים:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-(x-\mu)^2/(2\sigma^2)}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-y^2/2}dy$$

ולכן, מספיק להוכיח ש $\int_{-\infty}^{\infty}e^{-y^2/2}dy=\sqrt{2\pi}$. נגדיר $I=\int_{-\infty}^{\infty}e^{-y^2/2}dy$ ונוכיח ש $I^2=2\pi$:

$$I^2=I\cdot I=\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy\overset{\text{א}}{=}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy\int_{-\infty}^{\infty}e^{-\frac{z^2}{2}}dz=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{y^2+z^2}{2}}dydz$$



א – נחליף את שם המשתנה בשביל הנוחות, ונכתוב את מכפלת האינטגרלים כאינטגרל כפול.

כדי לחשב את האינטגרל הכפול נעשה הצבה של z,y לייצוג קוטבי.
כלומר נתייחס ל-(y,z) בתור נקודה על מערכת צירים.
נציב $y=r\sin\theta\,,z=r\cos\theta$ וזה ייתן לנו ש $dzdy=rd\theta dr$.[1] ולכן:
$r\in[0,\infty)$ כי הוא מייצג אורך. $\theta\in[0,2\pi]$ – זווית של העיגול.
$y^2+z^2=r^2\sin^2\theta+r^2\cos^2\theta=r^2(\sin^2\theta+\cos^2\theta)=r^2$ ע"פ זהות טריגו ולכן:

$$I^2=\int_0^{\infty}\left(\underbrace{\int_0^{2\pi}re^{-\frac{r^2}{2}}d\theta}_{2\pi}\right)dr=2\pi\int_0^{\infty}re^{-r^2/2}dr=-2\pi e^{-r^2/2}\Big|_0^{\infty}=2\pi$$

הפונקציה לא תלויה בθ, (מבחינת θ זה קבוע) אז האינטגרל הפנימי לפי $d\theta$ זה פשוט θ. בשלב השני, זה בדיוק מה שרצינו – e בחזקת פונקציה כפול הנגזרת הפנימית.

כלומר, $I=\sqrt{2\pi}$, כנדרש.

טענה 1.1: יהי X מ"מ נורמלי עם פרמטרים μ,σ^2. יהי $Y=aX+b$ עבור ממשיים b, $a>0$ כלשהם.
אז Y מתפלג נורמלית עם פרמטרים $a\mu+b,\ a^2\sigma^2$.

הוכחה: יהי F_Y ה-CDF של Y. אז, לכל $x\in\mathbb{R}$ מתקיים:

$$F_Y(x)=\mathbb{P}(Y\le x)=\mathbb{P}(aX+b\le x)=\mathbb{P}\left(X\le\frac{x-b}{a}\right)=F_X\left(\frac{x-b}{a}\right)$$

נגזור לפי y ונקבל:

---

[1] שקר מתמטי כלשהו. אין הוכחה לזה פה, לא למדנו באינפי. פשוט תזרמו עם זה

$$f_Y(x) = \frac{1}{a} f_X\left(\frac{x-b}{a}\right) = \frac{1}{a\sigma\sqrt{2\pi}} e^{-\left(\frac{x-b}{a}-\mu\right)^2/(2\sigma^2)} = \frac{1}{a\sigma\sqrt{2\pi}} e^{-(x-b-a\mu)^2/(2a^2\sigma^2)}$$

כלומר Y מתפלג נורמלית עם פרמטרים $a\mu + b,\ a^2\sigma^2$. $\qquad Y \sim N(a\mu + b, a^2\sigma^2)$.

מסקנה 1.2: יהי $X \sim N(\mu, \sigma^2)$. אז אם ניקח $Y = aX + b$ עם $a = \frac{1}{\sigma}, b = -\frac{\mu}{\sigma}$, כלומר: $Y \coloneqq (X-\mu)/\sigma$, זה מקיים: $Y \sim N(0,1)$

התפלגות $N(0,1)$ נקראת **התפלגות נורמלית סטנדרטית** (Standard Normal Distribution). נשים לב שה-PDF שלה היא:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

נהוג לסמן את ה-CDF של התפלגות סטנדרטית ב"פי" $\Phi(x)$, כלומר:

$$F_X(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

בפרט, $\Phi(\infty) = 1$. תכונה חשובה של Φ, היא סימטריות ביחס לאפס:

טענה 1.3: לכל $x \in \mathbb{R}$ מתקיים $\Phi(-x) = 1 - \Phi(x)$.

הוכחה: לכל $x \in \mathbb{R}$ מתקיים:

$$\underbrace{\Phi(x)}_{\mathbb{P}(X \le x)} + \Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt$$

$$=^{\text{א}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt + \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-y^2/2} dy =^{\text{ב}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt = 1$$

א – נציב $y = -t$. אז $\frac{dy}{dt} = -1$. ולכן "הופכים" את האינטגרל.

ב – אינטגרל על PDF.

עכשיו, אם $X \sim N(\mu, \sigma^2)$, אז $Y \coloneqq (X-\mu)/\sigma$ הוא סטנדרטי, ולכן:

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

כלומר, מספיק לעבוד עם ה-PDF הסטנדרטי.

דוגמה 1: יהי $X \sim N(\mu, \sigma^2)$. נרצה לחשב ערך של a שעבורו $\mathbb{P}(X \le a) \approx 0.99$. מתקיים:

$$0.99 \approx \mathbb{P}(X \le a) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

לפי טבלת ההתפלגות הסטנדרטית, נראה ש $\Phi(2.33) \approx 0.9901$. פתרון מקורב יהיה $\frac{a-\mu}{\sigma} = 2.33$ או $\frac{a-\mu}{\sigma} = 2.33\sigma + \mu$.

נחשב את התוחלת ושונות של $X \sim N(\mu, \sigma^2)$. יהי $Y = (X-\mu)/\sigma$, אז:

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} x f_Y(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = -\frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} = 0 - 0 = 0$$

לפי לינאריות התוחלת, נקבל:

$$\mathbb{E}(X) =^{\text{א}} \mathbb{E}(\sigma Y + \mu) = \sigma \mathbb{E}(Y) + \mu = \mu$$

א – כי $Y = \frac{X-\mu}{\sigma}$ אז $X = \sigma y + \mu$.

כדי לחשב את השונות של X צריך לחשב את $\mathbb{E}(Y^2)$. נעשה אינטגרציה בחלקים עם $u = x, v' = xe^{-x^2/2}$ ונקבל:

$$\mathbb{E}(Y^2)=\int_{-\infty}^{\infty}x^2f_Y(x)dx=\frac{1}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty}x^2e^{-x^2/2}dx}_{\int uv\prime}=\frac{1}{\sqrt{2\pi}}\left[\underbrace{-xe^{-x^2/2}\Big|_{-\infty}^{\infty}}_{uv}-\underbrace{\int_{-\infty}^{\infty}-e^{-x^2/2}dx}_{\int u\prime v}\right]$$

$$=\frac{1}{\sqrt{2\pi}}\left[(0-0)-\underbrace{\int_{-\infty}^{\infty}-e^{-x^2/2}dx}_{\int u\prime v}\right]=\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-x^2/2}dx}_{\text{PDF}}=1$$

ולכן, $Var(Y)=\mathbb{E}(Y^2)-\big(\mathbb{E}(Y)\big)^2=1-0=1$.

ונקבל: $Var(X)=Var(\sigma Y+\mu)=\sigma^2Var(Y)=\sigma^2$.

דוגמה 2: יהי $X\sim N(0,1)$, ותהי $g:\mathbb{R}\to\mathbb{R}$ פונקציה גזירה שמקיימת: (1)

$$\lim_{x\to\infty}g(x)\cdot e^{-\frac{x^2}{2}}=\lim_{x\to-\infty}g(x)\cdot e^{-\frac{x^2}{2}}=0$$

א) נוכיח ש $\mathbb{E}\big(g'(X)\big)=\mathbb{E}\big(X\cdot g(X)\big)$:

נעשה אינטגרציה בחלקים עם $u=g(x),v'=xe^{-x^2/2}$, ועם (1) נקבל:

$$\mathbb{E}\big(X\cdot g(X)\big)=\frac{1}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty}x\cdot g(x)\cdot e^{\frac{-x^2}{2}}dx}_{\int uv\prime}=-\frac{1}{\sqrt{2\pi}}\cdot\underbrace{g(x)\cdot e^{\frac{-x^2}{2}}}_{uv}\Bigg|_{-\infty}^{\infty}-\frac{1}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty}-g'^{(x)}\cdot e^{\frac{-x^2}{2}}dx}_{\int u\prime v}$$

$$=0-0+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}g'^{(x)}\cdot e^{\frac{-x^2}{2}}dx=\mathbb{E}\big(g'(X)\big)$$

ב) נוכיח ש $\mathbb{E}(X^{n+1})=n\cdot\mathbb{E}(X^{n-1})$

נקבע שלם חיובי n כלשהו, ונגדיר את g: $g(x)=x^n$. אז לפי (א) נקבל:

$$\mathbb{E}(X^{n+1})=\mathbb{E}(X\cdot X^n)=\mathbb{E}\big(X\cdot g(x)\big)=\mathbb{E}\big(g'(x)\big)=\mathbb{E}(n\cdot X^{n-1})=n\mathbb{E}(X^{n-1})$$

ג) נחשב את $\mathbb{E}(X^4)$:

בגלל ש $X\sim N(0,1)$, נקבל לפי (ב):

$$\mathbb{E}(X^4)=3\cdot\mathbb{E}(X^2)=3\cdot\mathbb{E}(X^0)=3\cdot\mathbb{E}(1)=3\cdot1=3$$

הרצאה 13

1 משפט הגבול המרכזי – Central Limit Theorem

(זה משפט "הגבול המרכזי". כלומר זה משפט על הגבול המרכזי, ולא "משפט גבול" מרכזי. הגבול הוא מרכזי, לא המשפט.)
למרות שהמשפט לא "מרכזי" (מסתבר), הוא תוצאה משמעותית בתורת ההסתברות.

משפט 1.1 (משפט הגבול המרכזי)**:** תהי $\{X_i\}_{i=1}^{\infty}$ סדרה של משתנים מקריים שמקיימת:

א. בלתי תלויים
ב. כולם בעלי התפלגות זהה
ג. בעלי תוחלת סופית μ
ד. שונות סופית $\sigma^2 > 0$

לכל שלם חיובי n, יהי $Y_n = \frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$ ותהי F_n ה-CDF של Y_n. כלומר $F_n(a) = \mathbb{P}(Y_n \le a)$ לכל $a \in \mathbb{R}$.
אזי מתקיים לכל $a \in \mathbb{R}$: $\lim_{n\to\infty} F_n(a) = \Phi(a)$.

למה הגדרנו ככה את Y_n? ניזכר קודם במ"מ נורמלי סטנדרטי. התוחלת שלו היא 0, והשונות היא 1.
אז כדי שנטען ש-Y_n מתפלג בצורה דומה למ"מ נורמלי סטנדרטי, אז נצטרך שהתוחלת תהיה 0, והשונות 1 (לפחות בקירוב).
אבל התחלנו עם מ"מ שיש לו תוחלת שונה מ-0, וסכמנו n כאלה. אז נוריד את התוחלת של הסכום – $n\mu$.
באופן דומה, בגלל שהם בת"ל אז השונות של הסכום זה סכום השונויות - $n \cdot \sigma^2$. אז כדי לקבל 1, נחלק באותו דבר.
אבל כשכופלים (או מחלקים) מ"מ בקבוע, הקבוע יוצא בריבוע. אז נחלק בשורש של זה.

דוגמה 1: נטיל מטבע הוגן 1000 פעמים, באופן בת"ל. יהי X המספר הכולל של הטלות שיצאו עץ.
בהרצאה 1 השתמשנו באי-שוויונות צ'בישב וצ'רנוף כדי לתת חסם תחתון על $\mathbb{P}(450 < X < 550)$.
על ידי צ'בישב קיבלנו 0.9, ועל ידי צ'רנוף קיבלנו $1 - 2e^{-5} \approx 0.98652$.
עכשיו נשתמש במשפט הגבול המרכזי.
לכל הטלה $1 \le i \le 1000$, יהי X_i האינדיקטור לעץ.
נשים לב שכל ה- X_i מקיימים את 4 התנאים, ובפרט $\mathbb{E}(X_i) = 1/2$, $Var(X_i) = \mathbb{E}(X_i^2) - \left(\mathbb{E}(X_i)\right)^2 = 1/4$.
יהי $Y = \frac{X-500}{\sqrt{250}}$, אז נקבל לפי 1.1: (1)

$$\begin{aligned}
\mathbb{P}(450 < X < 550) &= \mathbb{P}\left(\frac{450 - 1000 \cdot \frac{1}{2}}{\sqrt{\frac{1}{4}} \cdot \sqrt{1000}} < \frac{X - 1000 \cdot \frac{1}{2}}{\sqrt{\frac{1}{4}} \cdot \sqrt{1000}} < \frac{550 - 1000 \cdot \frac{1}{2}}{\sqrt{\frac{1}{4}} \cdot \sqrt{1000}}\right) \\
&= \mathbb{P}\left(-\frac{50}{\sqrt{250}} < Y < \frac{50}{\sqrt{250}}\right) = \mathbb{P}\left(-\frac{2500}{250} < Y < \frac{50}{250}\right) = \mathbb{P}\left(-\sqrt{10} < Y < \sqrt{10}\right) \\
&= \mathbb{P}\left(Y < \sqrt{10}\right) - \mathbb{P}\left(Y \le -\sqrt{10}\right) \approx^{\text{א}} \Phi\left(\sqrt{10}\right) - \Phi\left(-\sqrt{10}\right) = \Phi\left(\sqrt{10}\right) - \left(1 - \Phi\left(\sqrt{10}\right)\right) \\
&= 2\Phi\left(\sqrt{10}\right) - 1 \approx^{\text{ב}} 0.9984
\end{aligned}$$

א – משפט 1.1. ב – לפי הטבלה.

קיבלנו חסם הדוק יותר. בנוסף, הוא חסם תחתון וגם עליון.
מצד שני, זו תוצאה איכותית ולא כמותית. זה אומר לנו מה קורה בגבול אבל לא ספציפית עבור ערך כלשהו.
הבעיה נמצאת ב"שוויון" א. אמרנו שזה *בערך* התוצאה, אבל התוצאה מדברת על אינסוף הטלות ולא ספציפית על 1000.
אנחנו יודעים שיש פער כלשהו ואנחנו לא יודעים בדיוק כמה.
נגיד אם הפער הוא 1/2, אז החסם גרוע. אם הפער הוא יותר מ1, אז בכלל חרגנו מהטווח שהגיוני להסתברות.
יכול להיות שהחסם האמיתי רחוק ממה שמצאנו.
תוצאה כמותית הייתה אולי נותנת חסם פחות טוב אבל היינו יודעים שהוא נכון.

לכאורה נשמע שאי אפשר להשתמש ב CLT לחישובים כאלה, ואכן יש גרסאות כמותיות של CLT. נראה אחת מהן.
(בפועל בקורס הזה, נעבוד עם ה CLT כאילו זה קירוב תקין).

משפט 1.2 (משפט ברי-אסן)**:** תהי $\{X_i\}_{i=1}^{\infty}$ סדרה של מ"מ בת"ל בעלי אותה התפלגות.
נניח שמתקיים: $\mathbb{E}(X_i) = 0$, $Var(X_i) = \sigma^2 > 0$, $\mathbb{E}(|X_i|^3) = \rho$ עבור כל $i \in \mathbb{N}$, כאשר $\sigma, \rho \in \mathbb{R}$.
לכל שלם חיובי n, יהי $Y_n = \frac{X_1+\cdots+X_n}{\sigma\sqrt{n}}$ ויהי F_n ה CDF של Y_n. כלומר, $F_n(a) = \mathbb{P}(Y_n \le a)$ לכל $a \in \mathbb{R}$.
אזי מתקיים לכל $a \in \mathbb{R}$ ולכל n שלם חיובי: $|F_n(a) - \Phi(a)| \le \frac{\rho}{\sigma^3\sqrt{n}}$.

נשים לב שזה מזכיר את הגדרת הגבול, ולא סתם: הטענה של CLT היא לפי גבול, קרי: לכל אפסילון, מתישהו ה CDF תהיה קרובה עד כדי אפסילון ל-Φ. אז הטענה פה קובעת את "אפסילון" לפי n. זו תוצאה כמותית. עבור n קטן החסם אולי לא יהיה הדוק, אבל הוא מדויק.

נשתמש במשפט 1.2 כדי לחשב שוב את החסם מדוגמה 1.
לכל $1 \le i \le 1000$, יהי $Z_i = X_i - 1/2$.
נשים לב ש $\mathbb{E}(Z_i) = 0$. השונות לא משתנה, אז $Var(Z_i) = Var(X_i) = 1/4$.
כלומר $\sigma := \sqrt{Var(Z_i)} = 1/2$. לבסוף, $\mathbb{P}(|Z_i| = 1/2) = 1$.
ולכן $\mathbb{E}(|Z_i|^3) = (1/2)^3 = 1/8$.
יהי $Y = \frac{Z_1+\cdots+Z_{1000}}{\sqrt{1/4}\cdot\sqrt{1000}} = \frac{X-500}{\sqrt{250}}$. ממשפט 1.2, נקבל: (2)

$$\mathbb{P}\left(Y \le \sqrt{10}\right) \ge \Phi\left(\sqrt{10}\right) - \frac{\frac{1}{8}}{\left(\frac{1}{2}\right)^3 \cdot \sqrt{1000}} \ge \Phi\left(\sqrt{10}\right) - 0.03163$$

וגם (3):

$$\mathbb{P}\left(Y \le -\sqrt{10}\right) \le \Phi\left(-\sqrt{10}\right) + \frac{\frac{1}{8}}{\left(\frac{1}{2}\right)^3 \cdot \sqrt{1000}} \le \Phi\left(-\sqrt{10}\right) + 0.03163$$

בדומה ל(1), נקבל עם (2),(3):

$$\begin{aligned}
\mathbb{P}(450 < X < 550) &= \mathbb{P}\left(\frac{450 - 1000\cdot\frac{1}{2}}{\sqrt{\frac{1}{4}}\cdot\sqrt{1000}} < \frac{X - 1000\cdot\frac{1}{2}}{\sqrt{\frac{1}{4}}\cdot\sqrt{1000}} < \frac{550 - 1000\cdot\frac{1}{2}}{\sqrt{\frac{1}{4}}\cdot\sqrt{1000}}\right) \\
&= \mathbb{P}\left(-\frac{50}{\sqrt{250}} < Y < \frac{50}{\sqrt{250}}\right) = \mathbb{P}\left(-\frac{2500}{250} < Y < \frac{50}{250}\right) = \mathbb{P}\left(-\sqrt{10} < Y < \sqrt{10}\right) \\
&= \mathbb{P}\left(Y < \sqrt{10}\right) - \mathbb{P}\left(Y \le -\sqrt{10}\right) \ge \left[\Phi\left(\sqrt{10}\right) - 0.03163\right] - \left[\Phi\left(-\sqrt{10}\right) + 0.03163\right] \\
&= 2\Phi\left(\sqrt{10}\right) - 1 - 0.06326 \ge 0.935
\end{aligned}$$

זה יותר טוב מהחסם שמצאנו עם צ'בישב, אבל פחות טוב מהחסם שמצאנו עם צ'רנוף.

דוגמה 2: נזרוק קובייה הוגנת 360,000 פעמים, בת"ל. יהי X מספר הזריקות שיצאו 6. נרצה למצוא $a, b \in \mathbb{R}$ שמקיימים:

$$\mathbb{P}(54000 \le X \le 63000) \approx \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx$$

(שאלה ממבחן, 2017. הטריק שאמור לעזור בשאלה הוא שהצד הימני זה בעצם $\Phi(b) - \Phi(a)$).
הרעיון הוא להשתמש ב CLT, ואפילו לא צריך להשתמש בטבלה.

לכל $1 \le i \le 360000$, יהי X_i האינדיקטור לכך שיצא 6. אזי $x = \sum_{i=1}^{360000} X_i$ הוא מספר הפעמים שיצא 6. ולכל i בתחום מתקיים: $\mathbb{E}(X_i) = 1/6$, $Var(X_i) = 5/36$. אז לפי CLT מתקיים:

$$\mathbb{P}(54000 \le X \le 63000) = \mathbb{P}\left(\frac{54000 - 360000 \cdot \frac{1}{6}}{\sqrt{\frac{5}{36}} \cdot \sqrt{360000}} \le \frac{X - 360000 \cdot \frac{1}{6}}{\sqrt{\frac{5}{36}} \cdot \sqrt{360000}} \le \frac{63000 - 360000 \cdot \frac{1}{6}}{\sqrt{\frac{5}{36}} \cdot \sqrt{360000}}\right)$$

$$= \mathbb{P}\left(-12\sqrt{5} \le \frac{X - 360000 \cdot \frac{1}{6}}{\sqrt{\frac{5}{36}} \cdot \sqrt{360000}} \le 6\sqrt{5}\right)$$

$$= \mathbb{P}\left(\frac{X - 360000 \cdot \frac{1}{6}}{\sqrt{\frac{5}{36}} \cdot \sqrt{360000}} \le 6\sqrt{5}\right) - \mathbb{P}\left(\frac{X - 360000 \cdot \frac{1}{6}}{\sqrt{\frac{5}{36}} \cdot \sqrt{360000}} < -12\sqrt{5}\right) \approx$$

$$\approx \Phi(6\sqrt{5}) - \Phi(-12\sqrt{5}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{6\sqrt{5}} e^{-\frac{x^2}{2}}dx - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-12\sqrt{5}} e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-12\sqrt{5}}^{6\sqrt{5}} e^{-x^2/2}dx$$

נסיק שהתנאי הרצוי מתקיים עם: $a = -12\sqrt{5}, b = 6\sqrt{5}$.

דוגמה 3: נשתמש ב CLT כדי להוכיח ש:

$$\lim_{n\to\infty} e^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\right) = \frac{1}{2}$$

תהי $\{X_i\}_{i=1}^{\infty}$ סדרה בת"ל של מ"מ כאשר $X_i \sim Poi(1)$ לכל $i \in \mathbb{N}$. לכל n שלם חיובי, יהי $Y_n = \sum_{i=1}^{n} X_i$. אזי, ראינו בהסתברות 1 ש: $Y_n \sim Poi(n)$ לכל $n \in \mathbb{N}$. (משתנה מקרי שהוא סכום של שני משתנים מקריים שמתפלגים פואסון, בעצמו מתפלג פואסון עם פרמטר שהוא סכום הפרמטרים. באינדוקציה אפשר להרחיב את זה לכל סכום של n מ"מ) בפרט, (4):

$$\mathbb{P}(Y_n \le n) = \sum_{j=0}^{n} \mathbb{P}(Y_n = j) =^{\text{א}} \sum_{j=0}^{n} e^{-n} \cdot \frac{\lambda^j}{j!} = e^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\right)$$

א – אם $X \sim Poi(\lambda)$, אז $\mathbb{P}(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$

וגם, לפי CLT מתקיים (5): ננרמל את Y_n לפי התנאים של CLT:

מתקיים: $\mathbb{E}(Y_n) = n, Var(Y_n) = n$. אז נחסיר n כדי שהתוחלת תהיה 0, ונחלק בשורש השונות כדי שנקבל שונות 1.

$$\lim_{n\to\infty} \mathbb{P}(Y_n \le n) = \lim_{n\to\infty} \mathbb{P}(Y_n - n \le 0) = \lim_{n\to\infty} \mathbb{P}\left(\frac{Y_n - n}{\sqrt{n}} \le 0\right) = \Phi(0) = \frac{1}{2}$$

אם נחבר את (4), (5) נקבל:

$$\lim_{n\to\infty} e^{-n}\left(1 + n + \frac{n^2}{2!} + \cdots + \frac{n^n}{n!}\right) = \lim_{n\to\infty} \mathbb{P}(Y_n \le n) = \frac{1}{2}$$

כנדרש. ■

עד כאן הסתברות 2.