הסתברות 2 מרתון –

אי שוויוניים:

מרקוב: אם X הוא משתנה מקרי אי שלילי, $t>0$ אז $P(X\geq t)\leq \frac{E(X)}{t}$.

- הערה: אם t קרוב לממוצע אז מרקוב לא יעיל, צריך שיהיה גדול כדי שיהיה לנו חסם.

צ'בישב : אם X מ"מ עם שונות סופית $t>0$,אז $P(|X-E(X)|\geq t)\leq \frac{Var(x)}{t^2}$.

- אשר שמים את הערך מוחלט אנחנו מגדילים את ההסתברות לכן אם רוצים ללא ערך מוחלט זה עדיין עובד רק שזה יותר קטן מהערך המוחלט.

צ'רנוף : אם $X_1,X_2\dots X_n$ מ"מ ב"ת המחזירים ערכים בטווח $[0,1]$ ואם $X=\Sigma X_i$ אז:

1. רוצה לדעת מה המרחק כמו צ'בישב: לכל $t>0$ מתקיים:
$P(X\geq E(X)+t)\leq e^{-\frac{2t^2}{n}}$ וגם $P(X\leq\ E(X)-t)\leq e^{-\frac{2t^2}{n}}$ או לחלופין $P(|X-E(X)|\geq t)\leq\ 2e^{-\frac{2t^2}{n}}$.
ככל שt גדול אנחנו יורדים אספונציאלית בניגוד לצ'בישב ולכן חזק יותר.
וגם בכיוון ההפוך .
2. ולכן $\epsilon>0$ מתקיים: $P\big(X\leq(1-\epsilon)E(X)\big)\leq e^{-\frac{\epsilon^2E(X)}{2}}$.
3. אם $\epsilon\leq\frac{3}{2}$ בנוסף למקודם, אז: $P\big(X\geq(1+\epsilon)E(X)\big)\leq e^{-\frac{\epsilon^2E(X)}{3}}$

צ'רנוף הופדינג: אם $X_1,X_2\dots X_n$ מ"מ ב"ת כך ש: $P(X_i=1)=P(X_i=-1)=\frac{1}{2}$, ואם $X=\Sigma X_i$ אז:

לכל $t>0$ מתקיים $P(X\geq t)\leq e^{-\frac{t^2}{2n}}$ וגם $P(X<-t)\leq e^{-\frac{t^2}{2n}}$.

תרגיל מבחן 2018 שאלה 1 א':

שאלה 1 (25 נקודות):

מטילים קוביה הוגנת 72000 פעמים, כאשר כל ההטלות בלתי תלויות. יהי X משתנה מקרי הסופר את מספר ההטלות שתוצאתן 1.

א. השתמשו באי שוויון צ'רנוף כדי להוכיח ש-$\Pr(10000\leq X\leq 14000)\geq 0.99$.

פתרון:

בדרך כלל שיש לנו גדול שווה אנחנו עושים את המשלים ז"א 1 פחות כדיל להביא את זה לצורה שאנחנו רוצים.

נגדיר : X_i = אינדיקטור האם יצא 1. מכאן: $P(X_i=1)=\frac{1}{6},P(X_i=0)=\frac{5}{6}$.

נשים לב ש: $X = \Sigma_{\mathrm{i}=1}^{7200}\mathrm{X_i}$

מכאן: $E(X_i) = P(X_i = 1) = \frac{1}{6}$.

ומכאן: $E(X) = E\left(\Sigma_{\mathrm{i}=1}^{7200}\mathrm{X_i}\right) = \Sigma_{\mathrm{i}=1}^{7200}E(\mathrm{X_i}) = 1200$

נשים לב ש: $P(10000 \le X \le 14000) = 1 - P(X \le 10000\ or\ X > 14000)$

נחשב:$P(X \le 10000\ or\ X > 14000)$ בעצם or זה + בתנאי שאין חיתוך מאורעות (וכאן זה הגיוני אני לא יכול להיות גם קטן מזה וגם גדול מזה) לכן:

$$P(X \le 10000\ or\ X > 14000) = P(X < 10000) + P(X > 14000) = P(X < E(X) - 2000) + P(X > E(X) + 2000)$$

יש לנו בעיה כי אין לנו $\le$ לכן פשוט נוסיף את השווה הזה וזה עוד מקרה שאנחנו מכניסים ומגדילים את ההסתברות.

$$P(X < E(X) - 2000) + P(X > E(X) + 2000) \le \ P(X \le E(X) - 2000) + P(X \ge E(X) + 2000)$$

$$P(X \le E(X) - 2000) + P(X \ge E(X) + 2000) \le 2e^{-\frac{\dot{2}2000^2}{72000}} = 2e^{-\frac{1000}{9}} \le 2e^{-\frac{100}{9}} \le 0.01$$

נציב ונקבל :

$$P(10000 \le X \le 14000) = 1 - P(X \le 10000\ or\ X > 14000) \ge 1 - 0.01 = 0.99$$

נעשה שימוש ב(2) וב(3)

דרך ב': צ'רנוף ס' 2.

ההתחלה אותו דבר.

חישוב:

$$P(X \le 10000\ or\ X > 14000)\ P(X < 10000) + P(X > 14000)$$
$$= P\left(X < \left(1 - \frac{1}{6}\right)E(X)\right) + P\left(X > \left(1 + \frac{1}{6}\right)E(X)\right) \le$$
$$P\left(X \le \left(1 - \frac{1}{6}\right)E(X)\right) + P\left(X \ge \left(1 + \frac{1}{6}\right)E(X)\right) \le e^{e^{-\frac{12000}{72}}} + e^{-\frac{12000}{108}} \le 0.01$$

ההמשך אותו דבר.

משפטי גבול:

החוק החלש של המספרים הגדולים:

נתונה לנו סדרה מ"מ $\{X_n\}$ ב"ת אחד בשני על אותו מרחב עם תוחלת μ לכל אחד. אזי לכל $\epsilon > 0$ מתקיים:

$$\lim_{\mathrm{n}\to\infty} P\left(\left|\frac{X_1 + \cdots X_n}{n} - \ \mu\right| \ge \epsilon\right) = 0$$

- אינטואיציה: ממוצע התוצאה של הטלת קוביה הוא 3.5 ככל שנטיל יותר קוביות (יותר מ"מ, בנוסחא כל X_i הוא תוצאת הטלת קוביה) הממוצע (חישוב הסכום וחלוקה במספר הקוביות) ייתקרב ל 3.5 יותר ויותר עד שההסתברות לחרוג מהממוצע אפילו בקצת שואפת ל0.

החוק החזק של המספרים הגדולים:

(אומר אותו דבר אבל מחמיר יותר)

נתונה לנו סדרה מ"מ $\{X_n\}$ ב"ת אחד בשני על אותו מרחב עם תוחלת μ לכל אחד ושונות סופית. אזי מתקיים:

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

- בהוכחה הגבלנו לשונות סופית כדי להשתמש בצ'בישב.

מושגים מכלילים את החוקים לעיל:

- התכנסות בהסתברות: סדרת משתנים מקריים $\{X_n\}$ מתכנסת בהסתברות למ"מ X. אם לכל $\epsilon>0$ מתקיים: $\lim_{n\to\infty}P(\{\omega\in\Omega:|X_n(\omega)-X(\omega)|\geq\epsilon\})=0$
החוק החלש זה מקרי פרטי של זה.
- התכנסות כמעט בוודאות: סדרת משתנים מקריים $\{X_n\}$ מתכנסת כמעט בוודאות למ"מ X. אם מתקיים: $P(\{\omega\in\Omega:\lim_{n\to\infty}X_n(\omega)=X(\omega)\})=1$

כלי חשוב בחסמים: חסם איחוד (*union bound*) = $P(\bigcup_{i=1}^{n}A_i)\leq\Sigma_{i=1}^{n}P(A_i)$ מוותרים על הכלה והדחה ולכן לא מוותרים על החיתוכים וזה יוצר גדול)

השיטה ההסתברותית:

שיטה לפתרון בעיות בקומבינטוריקה (שלא קשורות דווקא להסתברות) – בעיות קיום. (האם קיים או לא קיים)

ומשתמשים בהסתברות כדי להראות שמשהו קיים (אם יש לו הסתברות גדולה מ0) או לא קיים (הסתברות שווה ל0).

בעיית מספר רמזי: $R(k,l)$= מהו הגרף עם מס' הקוד' n המינימלי כך שיש בו בוודאות קליקה בגודל k או קבוצה ב"ת בגודל l. (וריאציה אחרת: אם נצבע את כל צלעות הגרף בשני צבעים בכל צביעה אפשרית שנרצה: אדום וכחול אז בהכרח נמצא קליקה בגודל k אדומה כולה או קליקה בגודל l כחולה כולה).
קליקה = תת גרף שלם (שכולם מחוברים לכולם)
קבוצה בלתי תלוה = תת גרף ריק (שאף אחד לא מחובר לאף אחד)

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

רעיון ההוכחה: אם נראה שקיים גרף עם n קוד' שאין בו קליקה בגודל t וגם אין בו קבוצה ב"ת בגודל t אז זה יוכיח שn קוד' לא מספיק וצריך יותר.

נבנה גרף עם n קוד' באופר אקראי, על כל צלע אופציונלית בגרף נטיל מטבע (הסתברות $\frac{1}{2}$) ואם יצא 1 נשים צלע וב0 לא נשים צלע.

נסמן את כל תתי הקבוצות של קוד' בגודל t ב $S_1,\ldots.S_{\binom{n}{t}}$. (למה $\binom{n}{t}$? כי זה הכמות של תתי הקבוצות שאני יכול ליצור בגרף בגודל n, כל פעם אני בוחר t קוד' מתוך הn)

נראה מה ההסתברות שG מכיל קליקה בגודל t או מכיל קבוצה ב"ת בגודל t.

(אם הינו עושים שזה לא מכיל ז"א $G\ is\ Bad$ היה לנו כפל וזה סתם באלגן, לכן כדי לעשות חיבור כי אז צריך רק מצב אחד שהוא טוב כדי להשפיע על היתר)

$$P(G\ is\ good)=P\left(\bigcup_{i=1}^{\binom{n}{2}}S_i\ is\ Good\right)\leq\sum_{i=1}^{\binom{n}{k}}P(S_i\ is\ Good)=\sum_{i=1}^{\binom{n}{t}}\frac{1}{2^{\binom{n}{t}}}\cdot 2=\binom{n}{t}2^{1-\binom{t}{2}}<1$$

שוויון ראשון = עוברים על כל הקבוצות ודורשים שלפחות אחת תהיה טובה (קליקה או ב"ת)

אי שוויון שני = חסם איחוד

שוויון שלישי = ההסתברות לכל צלע היא $\frac{1}{2}$ ויש $\binom{t}{2}$ צלעות אופציונליות בתת גרף עם t קוד' ולכן ההסתברות שכל הצלעות יהיו $\frac{1}{2}$ בחזקת כמות הצלעות ובאותו אופן ההסתברות שאף צלע לא תהיה ולכן זה כפול 2.

שוויון רביעי = חוקי חזקות.

אי שוויון חמישי = נתון בטענה.

קיבלנו שההסתברות שיהיה קליקה או קבוצה ב"ת בגרף קטנה ממש מ1 וזה אומר שקיימת הסתברות שלא יהיה קליקה וגם לא קב' ב"ת ולכן יש גרף עם n קוד' שלא מכיל קליקה או קבוצה ב"ת בגודל t ולכן רמזי צריך יותר מn.

<u>בשיטה ההסתברותית: יש את הכלים הבסיסיים בהתסברות (בדוגמה ברמזי) ויש את המומנט הראשון והשני:</u>

<u>**מומנט ראשון**</u> (חסם חלש וליניארי – מרקוב)

כלומר: נדבר בדרך כלל על מספרים שלמים ולכן: $P(X>0)=P(X\geq 1)$

ואז לפי מרקוב: $P(X>0)=P(X\geq 1)\leq \frac{E(X)}{1}=E(X)$

ולכן פשוט נוכל לחשב תוחלת ובאמצעותה לדעת האם $E(X)$ חיובי או לא.

בנוסף: נשתמש בשיטה זו כדי לחשוב גבולות. אם ככל שn גדל התוחלת שואפת ל0 זה אומר ש:

$\lim_{(n\to\infty)} P(X=0)=1$. כי הקיום תלוי בתחולת ואם התוחלת שואפת ל0 אז הקיום שואף ל0 ולכן ערכו של X יהיה 0 בהסתברות השואפת ל1.

<u>**מומנט שני**</u> (חסם חזק וריבועי – צ'בישב) כאשר התוחלת שואפת ל1 ומעלה ואז לא נותנת לנו שום חסם.

$$P(X=0)=P\big(X-E(X)=\ -E(X)\big)\leq P\big(|X-E(X)|=E(X)\big)\leq P\big(|X-E(X)|\geq E(X)\big)\leq \frac{Var(X)}{E(X)^2}$$

כאן המטרה הסופית היא להראות ש: $P(X=0)\leq \cdots \to 0$ ולכן $P(X=0)\to 0$.

<u>תרגיל:</u>

א. בונים גרף אקראי עם n קוד' בהסתברות $\frac{1}{2}$ לכל צלע.

הוכיחו שככל שn גדול, בהסתברות שואפת ל1 לא יהיה קוד' בודד. (קודקוד ללא צלעות).

- טיפ – תמיד לחשוב לעבור לאינדיקטורים.

פתרון: תגדיר את האינדיקטורים הבאים X_i: = האם i בודד 1 אחרת 0.

תוחלת של X_i היא $E(X_i)=P(X_i=1)=\left(\frac{1}{2}\right)^{n-1}$

$X=\Sigma_{i=1}^{n}X_i$

אם $X=0$ אז אין קוד' בודדים.

לפי מומנט ראשון:

$$P(X>0)=P(X\geq 1)\leq E(X)$$

נחשב את התוחלת: $P(X>0)\leq\ E(X)=E(\Sigma X_i)=\Sigma E(X_i)=n\cdot\left(\frac{1}{2}\right)^{n-1}=\frac{n}{2^{n-1}}\to 0$

לכן ההסתברות שיהיה לפחות קוד' אחד בודד שואפת ל0 ולכן ההסתברות שלא יהיו קוד' בודדים שואפת ל1.

ב. אותו תרגיל אבל ההסתברות לצלע $\frac{1}{n}$. הוכיחו שכעת בהסתברות שואפת ל1 כן יהיה קוד' בודד.

פתרון : נגדיר כמו מקודם.

התוחלת של X_i היא : $E(X_i)=P(X_i=1)=\left(1-\frac{1}{2}\right)^{n-1}$.

השונות של X_i היא: $Var(X_i)=\left(1-\frac{1}{n}\right)^{n-1}\cdot\left(1-\left(1-\frac{1}{n}\right)^{n-1}\right)$

מומנט ראשון:

נחשב את התוחלת : $P(X>0)\leq\ E(X)=E(\Sigma X_i)=\Sigma E(X_i)=n\cdot\left(1-\frac{1}{2}\right)^{n-1}\to\infty$

לא חסמנו לכן צריך מומנט שני:

$$P(X=0)\leq\frac{Var(X)}{E(X)^2}$$

נחשב: $E(X)=n\cdot\left(1-\frac{1}{2}\right)^{n-1}$

$$Var(X)=\sum_i Var(X_i)+2\sum_{i<j}Cov(X_i,X_j)=$$

$$Cov(X_i,X_j)=P(X_i=1,X_j=1)-P(X_i=1)P(X_j=1)=\left(1-\frac{1}{n}\right)^{2n-3}-\left(1-\frac{1}{n}\right)^{2n-2}$$

לכן:

$$Var(X)=n\cdot\left(1-\frac{1}{n}\right)^{n-1}\cdot\left(1-\left(1-\frac{1}{n}\right)^{n-1}\right)+2\binom{n}{2}\left(\left(1-\frac{1}{n}\right)^{2n-3}-\left(1-\frac{1}{n}\right)^{2n-2}\right)=O(n^2)$$

טעות חישוב לכתוב מחדש!

מרתון 2 –

גרפים אקראיים - מודל בשיטה ההסתברותית (הוכחות שמשהו קיים או לא קיי דרך הסתברות) כדי להוכיח תוכנות של גרפים.

2 מודלים של גרפים אקראיים:

- $G(n,m)$ – מרחב הסתברות של כל הגרפים עם n קוד' וm צלעות.
  לכל ההסתברות היא על "איפה להניח את הצלעות".
  מה ההסתברות לכל גרף ספציפי? בחר m מקומות מתוך ה $\binom{n}{2}$ הפוטנציאליים לצלע. לכן זה 1 חלקי כמות הגרפים השונים שיש עם n קודקודים וm צלעות $\binom{\binom{n}{2}}{m}$
  כלומר : $P\big(G \in G(n,m)\big) = \frac{1}{\binom{\binom{n}{2}}{m}}$. זו התפלגות אחידה.
- $G(n,p)$ – מרחב הסתברות של כל הגרפים על n קוד' כאשר כל צלע אפשרים בגרף היא נמצאת בהסתברות p. כאן זו לא התפלגות אחידה כי : $P\big(G \in G(n,p)\big) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$.

תוכנות של גרפים: תכונה Q של גרפים יכולה לתאר לדוגמה: גרף קשיר, יש קודקוד בודד בגרף וכ'ו..

כאשר גרף G מקיים את התכונה נסמן $G \in Q$, כלומר, תכונה היא קבוצת כל הגרפים שמקיימים אותה.

תכונה מונוטונית: (שייך ל $G(n,p)$)

- עולה: תכונה מונוטונית עולה אם הוספה של צלעות יכולה לקיים אותה או לשמר את הקיום אך לא לגרום לתכונה לא להתקיים. (ככל שמוסיפים צלעות יותר טוב לתכונה). פורמאלית:
  $P(G(n,p_1) \in Q) \le P(G(n,p_2) \in Q)$ כאשר $p_1 \le p_2$.
  או $P(G(n,m) \in Q) \le P(G(n,m+1) \in Q)$.
- יורדת באותו אופן, ככל שמוסיפים צלעות, יורד הסיכוי לקיים את התכונה.

דוגמה:

1. קשירות: תכונה עולה, יותר צלעות – סיכוי טוב לקשירות.
   0 צלעות – בטוח לא קשיר
   כל הצלעות – בטוח קשיר
   מהרגע שהגרף קשיר, הוספת צלעות לא יכולה לפגוע בקשירותו.

2. קובצה ב"ת בגודל k: תכונה יורדת, פחות צלעות סיכוי טוב יותר לקבוצה ב"ת גדולה.
3. קיום מעגל בגרף: תכונה עולה. יותר צלעות יותר סיכוי למעגל. ויותר מהותי הוספה לא תוריד את הסיכוי למעגל.
4. גרף דו צדדי קשירת: תכונה לא עולה ולא יורדהת.
   פחות צלעות – הסיכוי לקשירות יורד. יותר צלעות הסיכוי לדו צדדי יורדת.

סף תכונה ספציפית Q: פונקציה של הסתברויות, לכל גרף עם n קוד' יש הסתברות לצלע התלויה בn: $p(n)$.

$P(n)$: המקיימת: שאם ההסתברות לצלע קטנה בסדק גודל מ$p(n)$ התכונה לא תתקיים ואם ההסתברות לצלע גדולה בסדר גודל מ$p(n)$ אז התכונה כן תתקיים.

במילים אחרות, מהו הגבול בו עוברים מלא לקיים את התכונה לכן לקיים את התכונה.?

פה מדברים על משהו קיצוני יותר:

אם טיפה ירדנו מהגבול - ההסתברות שואפת ל0

אם טיפה עברנו את הגבול – ההסתברות שואפת ל1.

פורמאלית: פונקציה $p(n)$ היא סף אם:

$p=\omega\big(p(n)\big)$ אז $P(G(n,p)\in Q)\to 1$

$p=o(p(n))$ אז $P(G(n,p)\in Q)\to 0$

דוגמא:

מצאו את הסף עבור קיום קודקוד בודד בגרף $G(np)$.

פתרון: נסמן בX_i – אינדיקטור ש v_i הוא קודקוד בודד. $X=\Sigma X_i$

אם $X=0$ אז אין קודקוד בודדים.

ההסתברות שלא יהיה קודקוד בודד: $P(X=0)=1-P(X\geq 1)\geq 1-E(X)$

לפי מומנט ראשון: (מרקוב)$P(X\geq 1)\leq E(X)$ ולכן $-P(X\geq 1)\geq -E(X)$

נחרב את התוחלת: $E(X)=E(\Sigma X_i)=\Sigma E(X_i)=\Sigma P\big(v_i \text{ בודד}\big)=\Sigma(1-p)^{n-1}=n(1-p)^{n-1}$

סה"כ: $P(X=0)=1-P(X\geq 1)\geq 1-E(X)=1-n(1-p)^{n-1}$

קיבלנו שההסתברות שאין קודקוד בודד גדולה שווה מ1 פחות ביטוי.

המטרה: שההסתברות תשאף ל1 ולכן צריך לדאוג שהיא תהיה גדולה שווה מ1 ואז היא בהכרח תהיה שווה ל1. לכן צריך לדרוש שהביטוי ישאף ל0.

טענת עזר: $(1-x)^n\leq e^{-xn}$, $x=e^{\ln x}$.

מכאן: $n(1-p)^{n-1}\leq e^{\ln x}e^{-pn+p}=e^{\ln n-pn+p}$.

נדרוש: $e^{\ln n-pn+p}\ll 1$ במילים אחרות : $e^{\ln n-pn+p}=o(1)$.

ולכן $\ln n-pn+p\ll 0$. ומכאן: $\ln n\ll p(n-1)$. ולכן : $\frac{\ln n}{n-1}\ll p$

תשובה סופית: $p=\omega(\frac{\ln n}{n-1})$.

הראינו כיוון אחד עבור הסף. אם $p=\omega(\frac{\ln n}{n-1})$ אז ההסתברות שלא יהיה קודקוד בודד שואפת ל1.

ראינו שסף רגיל הוא בסדר גודל:

דוגמה: הסף עבור קיום משולש בגרף הוא $p(n)=\frac{1}{n}$

כלומר, אם ההסתברות לצלע היא יותר מ $\frac{1}{n}$ בסדר גודל: נגיד $\frac{1}{\sqrt{n}}$ אז הסיכוי שיהיה משולש שואף ל1. כלומר, אם ההסתברות לצלע היא יותר מ $\frac{1}{n}$ בסדר גודל: נגיד $\frac{1}{n^2}$ אז הסיכוי שיהיה משולש שואף ל0.

שימו לב שיש שטח אפור גדול: בגלל שמדברים בסדרי גודל אז אם שמים בהסתברות $c\cdot\frac{1}{n}$ אז לא ידוע אם יש או אין.

ספ חד: p_0 יקרא ספ חד של תכונה Q מונוטונית עולה אם :

$P(G(n,p) \in Q) \to 1$ אז $p \geq (1+\epsilon)p_0$

$P(G(n,p) \in Q) \to 0$ אז $p = (1+\epsilon)p_0$

בעוד שלכל תכונה מונוטונית עולה יש ספ לא בהכרח שיש לה ספ חד.

תרגיל בית 3:

(1)

פתרון:

א. $\lim P(\delta \geq \ln n) = 1$

כאשר רוצים להוכיח גבול השואף ל1 מראים חסם עליון על התנאי המשלים ששואף ל0.

נגדיר את $\deg_i v$ להיות הדרגה של v_i.

נמצא את התוחלת של $\deg_i v$: נשים לב ש $\deg_i v$ מתפלג: $Bin(n-1,p)$.

(כשמדברים על כמות צלעות לזכור שזה בינומי).

לכן: $E(\deg_i v) = (n-1)p$

נציג את p לפי הנתון: $E(\deg_{\mathrm{i}} v) = (n-1) \cdot \frac{10 \ln n}{n}$

נחשב: $P(\delta < \ln n) = P(\bigcup_{i=1}^{n} (\deg_i v < \ln n\ ) \leq \sum_{i=1}^{n} P(\deg_i v < \ln n)$

- שימוש ב*union boud*

כאשר $n \to \infty$ ניתן לומר

$$\Sigma_{\mathrm{i}=1}^{n} P(\deg_i v < \ln n) \leq \Sigma_{\mathrm{i}=1}^{n} P\left(\deg_i v < \frac{1}{10} \cdot E(\deg_i v)\right) \leq \Sigma_{\mathrm{i}=1}^{n} P\left(\deg_i v < \frac{1}{2} \cdot E(\deg_i v)\right)$$

לפי חסם צ'רנוף (2) עבור $\epsilon = \frac{1}{2}$ נקבל :

$$\Sigma_{\mathrm{i}=1}^{n} P\left(\deg_i v < \frac{1}{2} \cdot E(\deg_i v)\right) \leq \Sigma_{\mathrm{i}=1}^{n} e^{-\frac{1}{4} \cdot E\left(\frac{\deg v)}{2}\right)} = \Sigma_{\mathrm{i}=1}^{n} e^{-\frac{1}{4} \cdot (n-1) \cdot \frac{10 \ln n}{n}} = n e^{-\frac{1}{4} \cdot (n-1) \cdot \frac{10 \ln n}{n}} \to n e^{\frac{10 \ln n}{8}}$$

$$= n e^{\ln n^{-\frac{5}{4}}} = \frac{n}{n^{\frac{5}{4}}} \to 0$$

מכאן : $P(\delta \geq \ln n) = 1 - P(\delta < \ln n) \to 1$

ב. שוב נעשה את המשלים ושואף ל0.

נשים לב שכמו בס' א:

$$P(\Delta > 20 \ln n) = P\left(\bigcup_{i=1}^{n} (\deg_i v > 20 \ln n\right) \leq \sum_{i=1}^{n} P(\deg_i v > 20 \ln n) \leq \sum_{i=1}^{n} P(\deg_i v \geq 20 \ln n)$$

$$\leq \Sigma P\big(\deg_i v \geq 2E(\deg_i v)\big)$$

לפי חסם צ'רנוף (3) עבור $\epsilon = 1$ נקבל:

$$\Sigma P\big(\deg_i v \geq 2E(\deg_i v)\big) \leq e^{-\frac{1E(\deg_i v)}{3}} \leq \Sigma e^{-\frac{10(n-1)\ln n}{3n}} \to n e^{-\frac{10 \ln n}{3}} = n n^{-\frac{10}{3}} \to 0$$

מכאן: $P(\Delta \leq 2 - \ln n) = 1 - P(\Delta > 20 \ln n) \to 1$

שאלה 2:

פתרון:

כשפותרים ספ אחד מהם יהיה מומנט ראשון ויהיה אחד שהוא מומנט שני.

א. נגדיר: $S_1, \dots S_{\binom{n}{4}}$ את כל תתי הקבוצות בגודל 4 של קודקודים בתוך הגרף.

נגדיר: X_i אינדיקטור האומר האם S_i היא קליקה.

נחשב: $E(X_i) = P(X_i = 1) = p^6$.

נגדיר: $X = \Sigma_{i=1}^{\binom{n}{4}} X_i$ ומכאן: $E(X) = E(\Sigma X_i) = \Sigma E(X_i) = \binom{n}{4} p^6$

מכאן, לפי המומנט הראשון (מרקוב)

$$P(X \geq 1) \leq \frac{E(X)}{1} = \binom{n}{4} p^6 \leq n^4 p^6 \ll n^4 \left(n^{-\frac{2}{3}}\right)^6 = n^4 n^{-4} = 1$$

להיות קטן ממש בסדר גודל מקבוע זה לשאוף ל0.

לכן: $P(X \geq 1) \to 0$. כלומר ההסתברות שיש קליקה שואפת ל0.

ב. לפי המומנט השני:(צ'בישב)

$$P(X=0) = P\big(X - E(X) = -E(X)\big) \leq P\big(|X - E(X)| \leq E(X)\big) \leq \frac{Var(X)}{E(X)^2}$$

נותן לחשב את $Var(X)$:

נחשב :

$$Var(X) = Var\left(\Sigma_{i=1}^{\binom{n}{4}} X_i\right) = \Sigma_{i=1}^{\binom{n}{4}} \Sigma_{j=1}^{\binom{n}{4}} Cov(X_i, X_j)$$

$$= \Sigma_{i=1}^{\binom{n}{4}} Var(X_i) + 2\Sigma_{i<j}^{\binom{n}{4}} Cov(X_i, X_j)$$

נחשב כל מקרה בנפרד:

$$\Sigma_{i=1}^{\binom{n}{4}} Var(X_i) = \Sigma_{i=1}^{\binom{n}{4}} p^6(1-p^6) = \binom{n}{4} p^6(1-p^6)$$

אם אין חיתוך בקוד' בין $S_i \ and \ S_j$ או שיש רק קודקוד משותף איד – $X_i.X_j$ ב"ת שאין צלע משותפת ולכן $Cov(X_i, X_j) = 0$

אם יש חיתוך ב2 קודקודים אז יש צלע משותפת, במקרה כזה:

$$Cov(X_i, X_j) = E(X_i, X_j) - E(X_i)E(X_j) = P(X_i = 1, X_j = 1) - P(X_i = 1)P(X_j = 1)$$

$$= p^{11} - p^6 p^6 = p^{11} - p^{12}$$

מספר הקבוצות שיש להן חיתוך של 2 קוד':

$$\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}$$

נבחר 2 קודוקדים משותפים נשלים 2 ל S_i ואז נשלים 2 לS_j . מתקיים:

$$\binom{n}{2}\binom{n-2}{2}\binom{n-4}{2} = \frac{n!}{2!\,2!\,2!\,(n-6)!}$$

אם יש חיתוך ב3 קודקודים אז יש 3 צלע משותפת, במקרה כזה:

$$Cov(X_i, X_j) = E(X_i, X_j) - E(X_i)E(X_j) = P(X_i = 1, X_j = 1) - P(X_i = 1)P(X_j = 1)$$

$$= p^9 - p^6 p^6 = p^9 - p^{12}$$

מספר הקבוצות שיש להן חיתוך של 2 קוד':

$$\binom{n}{3}\binom{n-3}{1}\binom{n-4}{1}$$

נבחר 3 קודוקדים משותפים נשלים 1 ל S_i ואז נשלים 1 לS_j.

סה"כ נציב:

$$Var(X) = \Sigma_{i=1}^{\binom{n}{4}} Var(X_i) + 2\Big(\Sigma^{\binom{n}{4}}_{\text{אין חיתוך או יש 1}} Cov(X_i, X_j) + \Sigma_{\text{חיתוך של 2}} + \Sigma_{\text{חיתוך של 3}}$$

נציב:

$$Var(X) = \cdots \leq n^4 p^6 e^{-p^6} + n^6 p^{11} - n^6 p^{12} + n^5 p^9 - n^5 p^{12}$$

נציב בצ'בישב:

$$P(X=0) \leq \frac{Var(X)}{E(X)^2} = \frac{n^4 p^6 e^{-p^6} + n^6 p^{11} - n^6 p^{12} + n^5 p^9 - n^5 p^{12}}{\left(\binom{n}{4} p^6\right)^2}$$

נשים לב : $1 = \frac{1}{e^{n^{-4}}} = e^{-n^{-4}} \gg e^{(-p)^6}$ ולכן $e^{-p^6} \to 0$

מכאן:

$$\lim_{n\to\infty}\frac{n^4p^6e^{-p^6}+n^6p^{11}-n^6p^{12}+n^5p^9-n^5p^{12}}{\left(\binom{n}{4}p^6\right)^2}$$

$$=\lim_{n\to\infty}\frac{n^6p^{12}\left[\left(\frac{e^{-p^6}}{n^2p^6}\right)+p^{-1}-1+n^{-1}p^{-3}-n^{-1}\right]}{(n^4p^6)^2}\le\frac{n^2}{n^2}$$

טעות.. לתקן

טיפ: אם כתוב גדול בסדר גודל תציבו בp במונה דוגמה למשהו גדול יותר.

מרתון 3

אלגוריתמים אקראיים:

אלגוריתם המשתמש ברנדומליות :

יש שלושה אלגו' כאלה.

1. לאס וגאס – אלגו' שתמיד צודק בתשובה אבל זמן הריצה שלו תלוי במשתנה מקרי.
2. מונטה קרלו – אלגו' שטועה בהסתברות (בדרך כלל נמוכה) :
   a. טעות חד צדדית: אם צריך להחזיר *true* אז האלגוריתם תמיד צודק. אם צריך להחזיר *false* אז האלגוריתם יכול להחזיר *true* בהסתברות נמוכה.
   b. טעות דו צדדית: בכל מקרה יש טעות בהסתברות (בדרך כלל נמוכה).

בדרך כלל הם יהיו אלגו' פשוטים.

הניתוח ההסתברותי יהיה בדרך כלל יותר מסובך. כמעט תמיד נשתמש באינדיקטורים.

שיטות:

1. בדרך כלל מדובר בבחירה אקראית מתוך קבוצה של איברים.
2. חזרה על התהליך מספר פעמים מקטינה את ההסתברות לטעות.

דוגמא:

תרגיל 1 : מציאת איבר הגדול מהחציון ב$O(1)$.

אלגוריתם 1: בחירת איבר אקראי והחזרתו: טעות $\frac{1}{2}$

אלגוריתם 2: לקיחת 100 איברים ראשונים והחזרת הגדול מבניהם.

טעות: $\frac{1}{2^{100}}$.

תרגילים:

האם שני פולינומים שווים?

קלט: $F(x), G(x)$ שני פונקציות פולינומיות מסדר d. (החזקה הכי גבוהה היא d)

פלט: $true$ אם $F(x)=G(x)$ (זאת אותה פונקציה רק בייצוג אחר) $false$ אם זאת לא אותה פונקציה.

אלגוריתם:

1. בחר מספר : $k \in \{\}$
2. הצב בF, G וחשב.
3. אם $F(k) = G(k)$ החזר *true* אחרת החזר *false*.

סיבוכיות זמן: $O(d)$ עבור המכפלות, החיבורים והחזקות.

אם $F(x) = G(x)$ אין הסתברות לטעות.

אם $F(x) \neq G(x)$ ההסתברות לטעות היא שנבחר את אחת מלכל היותר *d* נקודות החיתוך. (כי יש *d* פתרונות). לכן הטעות היא לכל היותר $\frac{d}{|A|}$ ($|A|$ הקבוצה ש*k* יכולה להיות בה).

אם ניקח $|A| = 100d$, $k \in \{1,2,3 \dots 100d\}$ נקבל שהטעות היא $= \frac{d}{100d}$.

אלגוריתם עם חזרה:

1. בצע עבור *i* מ1 עד *k* :
   a. בחר מספר : $k \in \{1,2,3..2d\}$
   b. הצב בF, G וחשב.
   c. אם $F(k) \neq G(k)$ החזר *false*.
2. החזר *true*.

סיבוכיות זמן ריצה: $O(kd)$. (אם *k* קבוע אז זה $(O(d)$

אם $F(x) = G(x)$ אין הסתברות לטעות.

אם $F(x) \neq G(x)$ ההסתברות לטעות היא שנבחר את אחת מלכל היותר *d* נקודות החיתוך בכל איטרציה.

$$P(error) = P\left(\text{חיתוך נקודת בחרנו}\right)^k = \frac{1}{2^k}$$

תרגיל 2: מציאת חתך מינימלי בגרף.

חתך – קבוצה של צלעות שאם נוריד אותן מהגרף הוא יפסיק להיות קשיר.

כלומר מה מס' הצלעות המינימלי שיש להוריד כך שהגרף כבר לא יהיה קשיר.

אלגוריתם להחזרת חתך:

קלט: גרף *G* קשיר עם *n* קודקודים.

פלט: קבוצת צלעות שהיא חתך.

אלגוריתם:

1. קבע $G_0 = G$
2. עבור *i* מ1 עד *n-2* :
   a. בחר צלע אקראית $e_i \in E(G_{i-1})$ והורד אותה.
   b. שמור את הגרף שהתקבל בG_i .
3. החזר את הצלעות של $E(G_{n-2})$.

סיבוכיות: $O(n)$ איטרציות , בכל איטרציה נבחר קודקוד ואז נבחר צלע אחת שלו: $O(n)$.

סה"כ: $O(n^2)$.

האם האלגוריתם מחזיר חתך?

האלגוריתם מחזיר את כל צלעות הגרף פרט ל $n-2$ צלעות.

מכאן, החתך הוא כל צלעות הגרף פרט ל$n-2$ הללו ולכן הגרף שהיתקבל לאחר הורדת צלעות החתך יהיה בעל $n-2$ צלעות לכל הצלעות.

לפי משפט, גרף עם פחות מ$n-1$ צלעות הוא בטוח לא קשיר.

למטרת הניתוח ההסתברות, כל הורדת צלע תתבטא בכך שאנחנו ממזגים את שני הקודקודים ואת הצלעות שלהם אבל לא משלבים צלעות! אלה פשוט מוסיפים את הצלעות של שניהם לקודקוד חדש. במקרה כזה ויתכן ונקבל מולטי גרף.

מה ההסתברות שזה חתך מינימאלי?

יהא A קבוצה כלשהי של צלעות שהן חתך מינימלי של G בגודל k .

נגדיר משתנה מקרי E_i שייצג את המאורע : $e_i \notin A$. (באיטרציה ה i לא הורדנו צלע מA). $1 \leq i \leq n-2$

נחתח את ההסתברות:

כאילו ההסתברות להוציא חתך מינימלי כלשהו גדולה מההסתברות להוציא את A

$$P(the\ algo\ output\ \min cut) \geq P(the\ algo\ output\ A) \geq P(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_{n-2})$$

נבדוק ממה הביטוי $P(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_{n-2})$ גדול שווה:

לפי הסתברות מותנה –

$$P(E_1 \cap E_2 \cap E_3 \cap \ldots \cap E_{n-2}) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \cdot \ldots \cdot P(E_{n-2}|E_1 \cdot \ldots E_{n-3})$$

נחשב כל ביטוי בנפרד: נשים לב – שמכיוון שהחתך הוא בגודל k אז דרגת כל קודקוד היא לפחות k (אחרת הינו יכולים לבודד קודקוד על ידי הורדת כל הצלעות שמחוברת אליו ואז היה חתך בגודל קטן מk)

מכאן: $2|E(G)| = \Sigma_{v \in V} \deg(v) \geq k \cdot n$ ולכן אני יודע $|E(G)| \geq \frac{kn}{2}$

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

לאחר הורדת צלע 1 לא מA, נקבל גרף עם *n-1* קודקקודים. החתך יהיה באותו גודל לפחות כמו החתך המקורי שכן איחדנו בין 2 קצוות צלע שהורדנו ואז הקודקודים של 2 הצדדים עדיין מחוברים.

לכן מאותו חישבו כמו מקודם:

$$P(E_2|E_1) \geq 1 - \frac{k}{\frac{k(n-1)}{2}} = 1 - \frac{2}{n-1}$$

באותו אופן:

$$P(E_i| \ldots) \geq 1 - \frac{2}{n-i+1}$$

מכאן:

$$P(E_1 \cap E_2 \cap E_3 \cap \dots \cap E_{n-2}) \geq \left(1-\frac{2}{n}\right)\left(1-\frac{2}{n-1}\right)\dots\left(1-\frac{2}{3}\right) = \left(\frac{n-2}{n}\right)\left(\frac{n-3}{n-1}\right)\dots\left(\frac{1}{4}\right)\left(\frac{1}{3}\right)$$
$$= \frac{2}{n(n-1)} = \binom{n}{2}^{-1}$$

עם חזרות אם נחזור על האלגוריתם t פעמים ונחזיר את החתך הכי מינימלי שהתקבל.

(בגלל השילוב קודקודים יהיה לנו מספר אחר של קבוצה מינימלית – דוגמה במרתון)

ההסתברות לטעות לכל היותר : $\left(1-\frac{2}{n(n-1)}\right)^t \leq e^{-t\cdot\frac{2}{n(n-1)}}$

אנחנו רוצים שהe יצטמצם לכן ניקח $t = \ln\left(\sqrt{100}\right) \cdot n(n-1)$

נקבל : $e^{-t\cdot\frac{2}{n(n-1)}} = \frac{1}{50}$

ואם $t = \ln(n) \cdot n(n-1)$

נקבל : $\frac{1}{n^2}$

הסיבוכיות אבל תהיה מאוד גדולה.

תרגיל 3:

Let $n \geq k \geq 2$ and $1 \leq m \leq 2^{k-2}$ be integers. Let $\{A_1, \dots, A_m\}$ be a family of subsets of $\{1, \dots, n\}$, each of size k. Devise a randomized algorithm with the following properties:

(a) Its input is the number n and the family $\{A_1, \dots, A_m\}$.

(b) Its output is a red/blue colouring of the elements of the set $\{1, \dots, n\}$.

(c) With probability at least $1 - 2^{-100}$, the colouring produced by the algorithm is such that, for every $1 \leq i \leq m$, the set A_i contains at least one red element and at least one blue element.

(d) The running time of the algorithm is $O(n + km)$.

דוגמה: $n = 10$ $m = 5$ $k = 3$.

$$A_1 = \{2,3,4\}, A_2 = \{1,2,10\} \dots$$

m קבוצות בגודל k כל אחת שיש בה מספרים בין 1 לn .

a. מקבל את הנתונים
b. צביעה של כחול \ אדום אלמנטים מ1 עד n
c. כל שכל אחת מהקבוצות בקלט יהיה לפחות מספר אחד באדום ואחד בכחול.

פתרון:

אלגוריתם:

1. * *לאחר ניתוח ההסתברות ששם חסמנו ב*$\frac{1}{2}$ עבור i מ1 עד 100
   a. נבחר צביעה אקראית בכחול ואדם למספרים $\{1 \dots n\}$.
   b. עבור j מ1 עד m
      i. בדוק שיש בA_j איבר מכל צבע.
   c. אם בכל קבוצה היה לפחות אחד אדום ואחד כחול החזר את הצביעה.
2. החזר את הצביעה האחרונה.

סיבוכיות זמן:

בכל איטרציה בוחרים צבע לכל אחד מהמספרים 1 עד n: $O(n)$

ואז בודקים את תקינות הצביעה: $O(m\cdot k)$.

סה"כ $100(n+mk)=O(n+mk)$.

ניתוח הסתברות:

נגדיר את X_i האינדיקטור האם A_i צבועה באותו צבע. $X=\Sigma_{i=10}^{m}x_i$ (0 – שני צבעים שונים. 1 – אין הכל אותו צבע).

מכאן: $P(X_i=1)=\frac{2}{2^k}$

$$P(error)=\ P(X>0)=P\left(\bigcup_{i=1}^{m}X_i=1\right)\leq\sum_{i=1}^{m}P(X_i=1)=\frac{2m}{2^k}\leq\left(\frac{2^{k-1}}{2^k}\right)=\frac{1}{2}$$

- לאחר תיקון האלגוריתם-

מכאן ההסתברות לטעות לאחר 100 איטרציות היא לכל היותר $\frac{1}{2^{100}}$.

תרגיל 4:

Let $\mathcal{F}\subseteq\mathbb{Z}_2^n$ be a family of binary vectors of length n. For any two binary vectors $\bar{x}=(x_1,\ldots,x_n)$ and $\bar{y}=(y_1,\ldots,y_n)$, define the distance between them to be

$$dist(\bar{x},\bar{y})=|\{1\leq i\leq n:x_i\neq y_i\}|.$$

Assume that the distance between any two vectors in $\mathcal{F}$ is at least $n/10$. Devise a randomized algorithm with the following properties:

(a) Its input are vectors $\bar{x},\bar{y}\in\mathcal{F}$.

(b) Its output is either $\bar{x}=\bar{y}$ or $\bar{x}\neq\bar{y}$.

(c) If $\bar{x}=\bar{y}$, then the algorithm will output $\bar{x}=\bar{y}$.

(d) If $\bar{x}\neq\bar{y}$, then the algorithm will output $\bar{x}\neq\bar{y}$ with probability at least $1-2^{-100}$.

(e) The running time of the algorithm is constant (i.e., independent of n).

דוגמה:

$$x=01001\ y=11100\ dist(x,y)=|\{1,3,5\}|$$

פתרון:

אלגוריתם:

1. עבור i מ1 עד 100
   a. הגרל מספר j בין 1 לn.
   b. בדוק האם $x[j]=y[j]$. אם לא – החזר ש$x\neq y$.
2. החזר שx=y.

זמן ריצה: $O(1)$.

ניתוח הסתברות לטעות:

אם נותנים לי $x=y$ אז לכל הגרלת אינדקס j נקבל: *x[j]=y[j]* ולכן יוחזר בהסתברות 1. כלומר אין טעות.

אם $x \neq y$ אז : נגדיר את E להיות קבוצת כל האינדקסים כך ש: $x[i] \neq y[i]$ כלומר:

$E = \{1 \leq i \leq n : x[i] \neq y[i]\}$ לפי הנתונים $|E| \geq \frac{n}{10}$.

נחשב:

$$P(int\ iteration\ i\ the\ choosen\ j \notin E) = 1 - P(int\ iteration\ i\ the\ choosen\ j \in E)$$

$$= 1 - \frac{|E|}{n} \leq 1 - \frac{\frac{1}{n}}{n} = \frac{9}{10}$$

$$P(error) = P\left(\bigcap_{i=1}^{100} int\ iteration\ i\ the\ choosen\ j \notin E\right) = \prod_{i=1}^{100} int\ iteration\ i\ the\ choosen\ j \notin E$$

$$\leq \left(\frac{9}{10}\right)^{100} \nless \frac{1}{2^{100}}$$

בעיה עם 100 לכן צריך לתקן את החזקה כדי לגרום למספר להיות קטן יותר מ$\frac{1}{2^{100}}$.

נתקן ל*1000* איטרציות:

אלגוריתם:

1. עבור *i* מ1 עד 1000
   a. הגרל מספר *j* בין 1 ל*n*.
   b. בדוק האם $x[j] = y[j]$. אם לא – החזר ש$x \neq y$.
2. החזר ש*x=y*.

$$\left(\frac{9}{10}\right)^{1000} < 0.4^{100} < \frac{1}{2^{100}}$$

תרגיל 4:

מבחן 2019

שאלה 2 (25 נקודות):

כתבו אלגוריתם המקיים את כל התכונות הבאות:

1. **קלט:** מספרים ממשיים $0 \leq a_1, a_2, \dots, a_n \leq 1$, $\varepsilon > 0$ ו-$\delta > 0$.
2. **פלט:** מספר ממשי x.
3. לכל קלט, הפלט של האלגוריתם צריך לקיים $\left|\frac{a_1+\dots+a_n}{n} - x\right| < \varepsilon$ בהסתברות $1-\delta$ לפחות.
4. זמן הריצה של האלגוריתם קבוע (כלומר לא תלוי ב-n אבל יכול להיות תלוי ב-ε וב-δ).

פתרון:

אלגוריתם:

1. הגדר *sum =0*
2. עבור *i* מ1 עד *t*
   a. בחר איבר אקראי בין 1 לn : $sum += a_j$
3. החזר $x = \frac{sum}{t}$

זמן ריצה: $O(1)$ (t לא תלוי בn אלה רק ב$\varepsilon\&\delta$)

הסתברות לטעות:

נגדיר את X_i להיות האיבר (הערך) שנבחר בשלב הi . $X = \Sigma_{i=1}^{t} X_i$

- בגלל שכל a_j הוא בין 0 ל1 ניתן לעשות שימוש בצ'רנוף כי צריך סכום של משנים בין 0 ל1.

מכאן: $E(X_i) = \frac{a_1 + .. + a_n}{n}$ ולכן: $E(X) = t(\frac{a_1 + .. + a_n}{n})$

$$P(error) = P\left(\left|\frac{a_1 + .. + a_n}{n} - \frac{X}{t}\right| \geq \varepsilon\right) = P\left(\left|t\frac{a_1 + .. + a_n}{n} - X\right| \geq t\varepsilon\right) = P(|X - E(X)| \geq \varepsilon t$$

לפי צ'רנוף:

$$\leq 2e^{-2\cdot\frac{(t\varepsilon)^2}{t}} = 2e^{-2\cdot\varepsilon^2 t}$$

נדרוש : $2e^{-2\cdot\varepsilon^2 t} \leq \delta$ ומכאן (נוציא לאן על מספקים קטנים לכן הופך סימן):

$$-2\varepsilon^2 t \leq \ln\frac{\delta}{2}$$

$$2\varepsilon^2 t \geq \ln\frac{2}{\delta}$$

ומכאן:

$$t \geq \frac{\ln\frac{2}{\delta}}{2\varepsilon^2}$$

ולכן נבצע לפחות $\left\lceil\frac{\ln\frac{2}{\delta}}{2\varepsilon^2}\right\rceil$ איטרציות.

מרתון 4 –

מרחב הסתברות רציף:

לא ניתן להכליל באופן טבעי את המחשב הדיסקרטי (הבדיד) (או בן מניה) למרחב הרציף.

במרחב הבדיד אמרנו שלכל מאורע יש הסתברות: $P:\Omega \rightarrow [0,1]$ ואמרנו שהתכונות הן:

1. $0 \leq P(\omega) \leq 1$
2. $\Sigma_{\omega \in \Omega} P(\omega) = 1$

במאורעות: $P(A) = \Sigma_{\omega \in \mathrm{A}} P(\omega)$

אם נעליל למרחב לא בן מניה: לדוגמה $\Omega = [0,1]$ (כל הממשיים בין 0 ל1)

ננסה לתת התפלגות אחידה: אותה הסתברות לכל הדגימות:

נניח ש: $P(\omega) = p$. לפי הצפיפות, קיים $n > \frac{1}{p}$ ומכאן : $P\left(\left\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}\right\}\right)$.

מצד שני אם ההסתברות לכל איבר או אפילו לכל קבוצה בת מניה הוא 0 אז:

$$0 = \Sigma_{\omega \in [0,1]} P(\omega) = P([0,1]) = 1$$

לכן במרחבים כאלו, כל דגימה לא תקבל הסתברות (הסתברות 0) אלא כל תת קבוצה לא בת מניה תקבל הסתברות.

במחרב ההסתברות כזה: $P: P(\Omega) \rightarrow [0,1]$

1. $P(\omega) = 0$
2. $P(A) = 0$ לכל A בת מניה.
3. $0 \leq P(X) \leq 1$.
4. $P(\Omega) = 1$

לא חייבים לדבר על כל תתי הקבוצות, לעיתים ניתן לדבר על חלק מהם. במקרה כזה תת הקבוצה:

$F \subseteq P(\Omega)$ תקרא סיגמא אלגברה: הקבוצה שיש בה את הקבוצה הריקה, את הכל והיא סגורה לפעולות על קבוצות כאשר מספר הפעולות הוא בן מניה.

משתנים מקריים רציפים:

משתנה X על מרחב הסתברות (R, F, P) יקרא משתנה מקרי רציף אם קיימת פונקציה: $f_x: R \rightarrow R$

כך ש: $P(X \in A) = \int_A f_x(x)dx$.

לפונקציה f_X קוראים פונקציית צפיפות.

תכונות:

1. $P(X = a) = \int_a^a f_X(x)dx = 0$
2. $P(X \in R) = P(R) = \int_{-\infty}^{\infty} f_X(x)dx = 1$
3. $P(a \leq X \leq b) = P(a < X < b) = \int_b^a f_X(x)dx$

מסקנה – כדי להוכיח שמדובר במרחב הסתברותי צריך להוכיח שהאינטגרל על הכל שווה ל1.

פונקציית צפיפות מצטברת: $F_X(a) = P(X \le a) = \int_{-\infty}^{a} f_X(x)dx$

$\int_a^b f_X(x)dx = F_X(b) - F_X(a)$ כי אם רוצים את השטח בין a ל-b אז לוקחים את כל השטח ממינוס אינסוף עד b ומורידים את השטח ממינוס אינסוף עד a.

לכן פונקציית הצפיפות המצטברת היא פונקציה הקדומה של פונקציית הצפיפות.

תרגיל: (תרגילי בית)

2. For what values of the parameter C are the following functions probability density functions?

(a) $f : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by

$$f(x) = \begin{cases} C(4x - 2x^2) & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(b) $g : \mathbb{R} \longrightarrow \mathbb{R}$ is defined by

$$g(x) = \begin{cases} Ce^{-x/100} & \text{if } x \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

פתרון:

א. נבדוק: $\int_{-\infty}^{\infty} f(x)dx = \int_0^2 C(4x - 2x^2)dx = C\left(2x^2 - \frac{2}{3}x^3\right)|_0^2 = C\left(8 - \frac{16}{3}\right) = 1$

ומכאן: $C = \frac{3}{8}$.

ב. $\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} Ce^{-\frac{x}{100}}dx = -100Ce^{-\frac{x}{100}}|_0^{\infty} = 100C = 1$

לכן $C = \frac{1}{100}$.

תוחלת של משתנה מקרי רציף:

במקרה הבדיד ראינו ש: $E(X) = \Sigma_X P(X = x)$

עבור משתנה רציף נקבל ש: $E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)dx$

תכונות של תוחלת:

1. ליניאריות התוחלת נשמרת: $E(aX + b) = aE(X) + b$.
2. $E\big(g(X)\big) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x)dx$.

שונות:

$$Var(X) = E(X^2) - E(X)^2$$

תכונות: $Var(aX + b) = a^2 Var(X)$

(כמו מה שהיה במרחב בדיד.)

תרגילים:

שאלה 4 (25 נקודות):

תהי

$$f(x)=\begin{cases}4x & 0\le x<1/2\\ 4-4x & 1/2\le x\le 1\end{cases}$$

א. הוכיחו ש-f היא פונקציית צפיפות.

ב. יהי X משתנה מקרי עם פונקציית צפיפות f. חשב את פונקציית ההסתברות המצטברת של X, את התוחלת שלו ואת השונות שלו.

א. $\int_{-\infty}^{\infty} f(x)dx=\int_0^{\frac{1}{2}}4xdx+\int_{\frac{1}{2}}^{1}(4-4x)dx=2x^2|_0^{\frac{1}{2}}+(4x-2x^2)|_{\frac{1}{2}}^{1}=\frac{1}{2}+(2-1.5)=1$

ב. $F_X(a)=P(X\le a)=\int_{-\infty}^{a} f(x)dx$

אם $a<0$ נקבל: $F_X(a)=P(X\le a)=\int_{-\infty}^{a} f(x)dx=0$

אם $0\le a<\frac{1}{2}$ נקבל: $F_X(a)=P(X\le a)=\int_0^{\frac{1}{2}} f(x)dx=\int_0^{\frac{1}{2}}4xdx=2a^2$

אם $\frac{1}{2}\le a\le 1$ נקבל:

$$F_X(a)=P(X\le a)=\int_0^{\frac{1}{2}} f(x)dx+\int_{\frac{1}{2}}^{a} f(x)dx=\frac{1}{2}+\int_{\frac{1}{2}}^{a}(4-4x)dx=\frac{1}{2}+(4x-2x^2)|_{\frac{1}{2}}^{a}=4a-2a^2-1$$

אם $a>1$ אז: $F_X(a)=P(X\le a)=\int_{-\infty}^{a} f(x)dx=\int_{-\infty}^{\infty} f(x)dx=1$

תוחלת:

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_0^{\frac{1}{2}}4x^2dx+\int_{\frac{1}{2}}^{1}(4x-4x^2)dx=\frac{4}{3}x^3|_0^{\frac{1}{2}}+2x^2-\frac{4}{3}x^3|_{\frac{1}{2}}^{1}=\frac{1}{6}+\frac{2}{3}-\frac{1}{3}=\frac{1}{2}$$

שונות:

תחילה נחשב:

$$E(X^2)=\int_{-\infty}^{\infty} x^2f(x)dx=\int_0^{1/2}4x^3dx+\int_{1/2}^{1}(4x^2-4x^3)dx=x^4|_0^{1/2}+\tfrac{4}{3}x^3-x^4|_{1/2}^{1}=\tfrac{1}{16}+\tfrac{1}{3}-\tfrac{1}{6}+\tfrac{1}{16}=\tfrac{1}{8}+\tfrac{1}{6}=\tfrac{7}{24}$$

ומכאן: $Var(X)=E(X^2)-(E(X))^2=\frac{7}{24}-(\frac{1}{2})^2=\frac{1}{24}$.

התפלגויות מיוחדות:

1. **התפלגות אחידה רציפה בקטע $[a,b]$:**

פונקציית הצפיפות היא: $f_X(x)=\frac{1}{b-a}\ if\ a\le x\le b$

פונ' הצפיפות המצטברת: $F_X(x)=\frac{x-a}{b-a}\ if a\le x\le b, 1\ if\ x>b, 0\ if\ x<a$

תוחלת: $E(X)=\frac{a+b}{2}$

שונות: $Var(X)=\frac{(b-a)^2}{12}$

2. **התפלגות אקספוננציאלית\מעריכית : פרמטר λ.**

פונקציית הצפיפות היא: $f_X(x)=\lambda\cdot e^{-\lambda x}\ if\ x\ge 0, 0\ if\ x<0$

פונ' הצפיפות המצטברת: $F_X(x)=P(X\le x)=1-e^{-\lambda x}\ if\ x\ge 0, 0\ if\ x<0$

תוחלת: $E(X)=\frac{1}{\lambda}$

שונות: $Var(X)=\frac{1}{\lambda^2}$

תכונה שהתפלגות זו מקיימת: חוסר זיכרון.

$P(X > s + t \mid X > t) = P(X > s)$ ז"א ניתן ממש לשכוח מהתנאי שיש לנו, זה יתקיים לכל $s, t \in \mathbb{R}^+$.

במילים אחרות: $P(X > s + t) = P(X > s) \cdot P(X > t)$.

משפט: משתנה רציף אי-שלילי יהיה חסר זיכרון **אם ורק אם** הוא מתפלג מעריכית.

הוכחה:

כיוון 1: אם X מתפלג מעריכית אז:

$$P(X > s)P(X > t) = \big(1 - F_X(s)\big) \cdot \big(1 - F_X(t)\big) = e^{-\lambda s} \cdot e^{-\lambda t} = e^{-\lambda(s+t)}$$
$$= 1 - F_X(s + t) = P(X > s + t)$$

כיוון 2: אם X מקיים חוסר זיכרון: $P(X > s + t) = P(X > s) \cdot P(X > t)$ לכל $s, t \in \mathbb{R}^+$.

תהי $F_X(x)$ פונקציית הצפיפות המצטברת של X. נסמן : $g(x) = 1 - F_X(x)$

המטרה: להוכיח ש: $g(x) = e^{-\lambda x}$ עבור $\lambda > 0$ כלשהי.

לפי הנתון: $g(s + t) = g(s)g(t)$.

טענת עזר: אם $m, n \in \mathbb{N}^+$ אז $g\left(\frac{m}{n}\right) = g^m\left(\frac{1}{n}\right)$

הוכחה: $g\left(\frac{m}{n}\right) = g\left(\frac{m-1}{n} + \frac{1}{n}\right) = g\left(\frac{(m-1)}{n}\right)g\left(\frac{1}{n}\right) = \cdots = g^m\left(\frac{1}{n}\right)$ באינדוקציה על m.

נשים לב ש: $g(1) = g\left(\frac{n}{n}\right) = g^n\left(\frac{1}{n}\right)$ (לפי טענת העזר) ומכאן: $g\left(\frac{1}{n}\right) = g(1)^{\frac{1}{n}}$

באותו אופן: $g\left(\frac{m}{n}\right) = \left(g\left(\frac{1}{n}\right)\right)^m = \left(g(1)^{\frac{1}{n}}\right)^m = g(1)^{\frac{m}{n}}$.

מכיוון שהמשתנה רציף אז $F_X(x)$ פונקציה רציפה. לכם $g(x)$ רציפה.

כלל: אם עבור פונקציה רציפה: $f(q) = g(q)$ לכל $q \in \mathbb{Q}$ אז $g(x) = f(x)$ לכל $x \in \mathbb{R}$.
(בגלל הצפיפות של הממשיים עם הרציונליים).

מסקנה: $g(x) = g(1)^x = e^{\ln g(1)^x} = e^{\ln g(1) x}$ לכל $x \in \mathbb{R}^+$.

נבחר: $\lambda = -\ln g(1)$.

אם $g(1) = 1$ אז: $g(x) = 1$ לכל x ואז: $F_X(x) = 0$ וזו סתירה כי $F_X(\infty) = 1$.

אם $g(1) = 0$ אז: $g(x) = 0$ לכל x ואז: $F_X(x) = 1$ לכל $x \geq 0$ וגם $F_X(x) = 0$ לכל $x < 0$.

סתירה כי הפונקציה רציפה לכן לא ייתכן שב-0 יש לה קפיצה.

ולכן הוכחנו שניתן לבחור ככה ולכן המשתנה מתפלג מעריכית.

3. התפלגות נורמלית: פרמטרים – ממוצע μ ושונות σ^2 כאשר $\sigma > 0$ היא סטיית תקן.

פונקציית הצפיפות היא: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.

פונ' הצפיפות המצטברת: לא קיימת פונקציה סגורה נורמאלית. משתמשים בדרך כלל בהזזה להתפלגות נורמאלית סטנדרטית.

התפלגות נורמאלית סטנדרטית:

התפלגות נורמאלית כך ש: $\mu = 0$ ו $\sigma = 1$.

בדרך כלל טבלת הערכים היא על ההתפלגות הסטנדרטית.

~~אם רוצים לעבור ללא סטנדרטי אז עושים הזזה.~~

פונ' הצפיפות המצטברת של ההתפלגות הסטנדרטית: $\Phi(x) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$. הסימון הוא עבור התפלגות נורמאלית הסטנדרטית.

לפונקציה הזו יש טבלת ערכים.

תוחלת: $E(X) = \mu$. בסטנדרטית =0.

שונות: $Var(X) = \sigma^2$. בסטנדרטית =1.

<u>תכונות:</u>

א. אם X מתפלג נורמאלית עם ממוצע μ ושונות σ^2 וגם $Y = aX + b$.
אז Y מתפלג נורמאלית עם ממוצע $a\mu + b$ ושונות: $a^2\sigma^2$.
לכן כדי להמיר משתנה שמתפלג נורמאלית למשתנה מתפלג נורמאלית סטנדרטית:

$$Z = \frac{X - \mu}{\sigma}$$

ב. לכל $x \in R$. $\Phi(x) + \Phi(-x) = 1$. במילים אחרות: $\Phi(-x) = 1 - \Phi(x)$.

משפט הגבול המרכזי:

אם $\{X_i\}_{i=1}^{\infty}$ סדרה של אינסוף משתנים מקריים:

1. ב"ת אחד בשני
2. בעלי התפלגות זהה עם תוחלת μ ושונות σ^2 (גם כן זהות) כאשר $\sigma > 0$.

אם נסמן לכל $n \in \mathbb{N}$: $Y_n = \frac{X_1 + \cdots X_n - n\mu}{\sigma\sqrt{n}}$. כך ש$F_n$ היא פונקציית הצפיפות המצטברת של Y_n. אז:

$$\lim_{n \to \infty} F_n(x) = \Phi(x)$$

המשפט הופך כל סדרת משנים (עם התנאים לעיל) למשתנה נורמאלי סטנדרטי.

המשפט עוזר כאשר לא ידוע ההתפלגות אלא רק הממוצע וסטיית הקן או שידוע ההתפלגות אך היא קשה לחישוב ואנו רוצים חסמים על הסתברות מסוימת.

אם אין אינסוף משתנים אז המשפט נותן הערכה ולא משהו מדויק אבל צריך לפחות כמות גדולה של משתנים וניתן יהיה להתשמש בY_n עבור n ספציפי יחסית גדול.

<u>משפט:</u> (*Berry – Esseen*)

אם $\{X_i\}_{i=1}^{\infty}$ סדרה של אינסוף משתנים מקריים:

1. ב"ת אחד בשני .
2. בעלי התפלגות זהה עם תוחלת 0 ושונות σ^2 (גם כן זהות) כאשר $\sigma > 0$.
3. בנוסף: $E(|X_i|^3) = \rho$

אם נסמן לכל $n \in \mathbb{N}$: $Y_n = \frac{X_1 + \cdots X_n - n\mu}{\sigma\sqrt{n}}$. כך ש$F_n$ היא פונקציית הצפיפות המצטברת הסטנדרטית של Y_n.
אז:

$$|F_n(x) - \Phi(x)| \leq \frac{\rho}{\sigma^3\sqrt{n}}$$

המשפט מבטיח בהיתן התנאים הנוספים את הקצב שבו מתקרבים להתפלגות נורמאלית.

תרגילים:

שאלה 4 (30 נקודות):

א. (10 נקודות) יהי X משתנה מקרי עם תוחלת $E(X) = 3/5$ ופונקציית צפיפות

$$f_X(x) = \begin{cases} ax^2 + b & if\ 0 \le x \le 1 \\ 0 & otherwise \end{cases}$$

מצאו את a ו-b.

ב. (10 נקודות) יהי Y משתנה מקרי רציף המתפלג אחיד על הקטע $(0,5)$. מה ההסתברות שלמשוואה $4z^2 + 4zY + Y + 2 = 0$ יש שני פתרונות ממשיים שונים?

ג. (10 נקודות) יהי Z משתנה מקרי נורמלי סטנדרטי ויהי $a > 0$ מספר ממשי. הוכיחו ש- $\Pr(|Z| > a) = 2\Pr(Z > a)$.

א. משוואה 1: $\int_0^1 (ax^2 + b)dx = \frac{a}{3}x^3 + bx\,|_0^1 = \frac{a}{3} + b = 1$

משוואה 2: $E(X) = \int_0^1 (ax^3 + b)dx = \frac{a}{4}x^4 + \frac{b}{2}x^2|_0^1 = \frac{a}{4} + \frac{b}{2} = \frac{3}{5}$

נבודד: $b = 1 - \frac{a}{3}$ ונציב : $\frac{a}{4} + \frac{1}{2} - \frac{a}{6} = \frac{3}{5}$.

ומכאן: $\frac{1}{10} = \frac{a}{12}$ ולכן $a = \frac{6}{5}$.

ומכאן: $b = \frac{3}{5}$.

ב. עבור $y \in (0,5)$ נקבל את המשוואה: $4z^2 + 4zy + y + 2 = 0$.

לפי נוסחת השורשים יש 2 פתרונות ממשיים שונים רק אם $z_{1,2} = \frac{b \pm \sqrt{b^2 - 4ac}}{2a}$

כאשר: $b^2 - 4ac > 0$.

ומכאן: $16y^2 - 16(y+2) > 0$

ומכאן: $y^2 - y - 2 > 0$ חילוק ב16

הפתרונות הם $y = 2, y = -1$ ולכן: $y > 2$ (עבור 1- הוא לא בתחום *(0,5)*).

לכן: $P(two\ solution) = P(Y > 2) = 1 - P(Y \le 2) = 1 - \frac{(2-0)}{5-0} = \frac{3}{5}$

ג. נשים לב שמתקיים: $P(|Z| > a) = P(Z > a) + P(Z < -a) = P(Z > a) + \Phi(-a) =$
$P(Z > a) + 1 - \Phi(a) = P(Z > a) + 1 - P(Z < a) = P(Z > a) + 1 - 1 + P(Z > a) =$
$P(Z > a) + P(Z > a) = 2P(Z > a)$

1. Let $X \sim U([1,2])$ be a random variable with the continuous uniform distribution. Calculate $\mathbb{E}(X^n)$ for every positive integer n.

2. Let $X_1, \ldots, X_n$ be independent exponentially distributed random variables with parameters $\lambda_1, \ldots, \lambda_n$, respectively. Let $X = \min\{X_i : 1 \le i \le n\}$ and let $\lambda = \sum_{i=1}^{n} \lambda_i$. Prove that X is exponentially distributed with parameter λ.

3. A fair 6-sided die is rolled 420 times, all dice rolls being mutually independent. Let S denote the sum of the resulting numbers.
   (a) Use Chebyshev's inequality to find a lower bound on $Pr(1400 \le S \le 1540)$.
   (b) Use the central limit theorem (CLT) to estimate $Pr(1400 \le S \le 1540)$.

פתרון:

1. $E(X^n) = \int_{-\infty}^{\infty} x^n f_X(x) = \int_1^2 \frac{x^n 1}{1} dx = \frac{1}{n+1} x^{n+1}|_1^2 = 2^{n+1} - \frac{1}{n+1}$

2. צ"ל: $P(X \le x) = 1 - e^{-\lambda x}$. זו המצטברת בדרך כלל יותר קלה..

לפי הגדרת X, $P(X \le x) = P(X_1 \le x \; or \; X_2 \le or \ldots X_n \le x)$
נשים לב ש: $P(X_1 \le or \, .. X_n \le x) = 1 - P(X_1 > x \; and \ldots X_n > x)$ ולכן:
$P(X \le x) = 1 - P(X_1 > x \; and \ldots X_n > x)$
כל המשתנים ב"ת ולכן : $P(X \le x) = 1 - P(X_1 > x) \cdot \ldots \cdot P(X_n > x)$
כל X_i מתפלג מעריכית ולכן : $P(X_i > x) = e^{-\lambda_i x}$ ומכאן:
$$P(X \le x) = 1 - e^{\lambda_1 x} \cdot \ldots \cdot e^{-\lambda_n x} = 1 - e^{-x(\lambda_1 + \cdots + \lambda_n)} = 1 - e^{-x\lambda}$$
משל.

3. סעיפים:

א. אי שוויון צ'ביצ'ב: $P(|X - E(X)| \ge a) \le \frac{Var(X)}{a^2}$
נגדיר את S_i להיות התוצאה של ההטלה ה i . עבור $1 \le i \le 420$.
$S = \Sigma S_i$ נחשב תוחלת ושונות של כל אחד מהם.
S_i מתפלג אחיד על 1-6 ולכן: $E(S_i) = \frac{1+6}{2} = \frac{7}{2}$. $Var(S_i) = \frac{6^2-1}{12} = \frac{35}{12}$.
מכיוון שכל המשתנים ב"ת:
$$E(S) = E(\Sigma S_i) = \Sigma^{420}{}_{i=1} E(S_i) = 420 \cdot \frac{7}{2} = 1470$$
$$Var(S) = Var\left(\Sigma_{i=1}^{420} S_i\right) = \Sigma_{i=1}^{420} Var(S_i) = 420 \cdot \frac{35}{12} = 1225$$
פתרון: $P(1400 \le S \le 1540) = P(-70 \le S - E(S) \le 70) = P(|S - E(X)| \le 70)$
$$= 1 - P(|S - E(X)| > 70) = 1 - P(|S - E(X)| \ge 71) \ge 1 - \frac{Var(S)}{72^2} = 1 - \frac{1225}{72^2}$$

ב. נפעיל את המשפט למרות שאין לנו אינסוף משתנים כיוון עבור 420 זה כבר יחסית קרוב.
נשים לב שלכל S_i התוחלת היא אותו דבר $\frac{7}{2}$ והשונות גם כן $\frac{35}{12}$ והם ב"ת ועם אותה התפלגות (אחידה).
$$P(1400 \le S \le 1540) = P\left(1400 \le \Sigma_{i=1}^{420} S_i \le 1540\right)$$

נחסר את התוחלת:

$$P\left(-70 \leq \Sigma_{i=1}^{420} S_i - 420\frac{7}{2} \leq 70\right)$$

צריך לחלק עכשיו את הביטוי:

$$P\left(-\frac{70}{\sqrt{420 \cdot \frac{35}{12}}} \leq \frac{\left(\Sigma_{i=1}^{420} S_i - 420\frac{7}{2}\right)}{\sqrt{420 \cdot \frac{35}{12}}} \leq \frac{70}{\sqrt{420 \cdot \frac{35}{12}}}\right)$$

$$= P(-2 \leq Y_{420} \leq 2) = \Phi(2) - \Phi(-2) = \Phi(2) - \left(1 - \Phi(2)\right) = 2\Phi(2) - 1$$

המשתנה Y_n מתפלג נורמלית סטנדרטי .

$$2\Phi(2) - 1 \approx 2 \cdot 0.9772 - 1 = 0.9544$$

מרתון 4 –

פתרון תרגילים ומבחנים:

מבחן 2019 סמסטר א'

שאלה 1 (30 נקודות):

תהי $\{X_i\}_{i=1}^{\infty}$ סדרה של משתנים מקריים בלתי תלויים כך ש-$X_i \sim U(\{-1,0,1\})$ לכל $i \in \mathbb{N}$. לכל $n \in \mathbb{N}$ יהי $S_n = \sum_{i=1}^{n} X_i$.

א. (15 נקודות) השתמשו באי שוויון צ'רנוף (כפי שנלמד בהרצאות) כדי למצוא חסם עליון טוב ככל האפשר על $\Pr(|S_n| > 2\sqrt{n})$.

ב. (15 נקודות) השתמשו במשפט הגבול המרכזי כדי למצוא הערכה טובה ככל האפשר ל-$\Pr(|S_n| > 2\sqrt{n})$.

א. לפי צ'רנוף צריך שתטווח הערכים של כל משתנה מקרי בסכום יהיה בין 0 ל1.

הערה: כדי להזיז משתנה X_i שהוא בטווח $[a,b]$ במקרה שלנו $[-1,1]$ אז צריך להחסיר את הערך התחתון ביותר ואז את כל התוצאה לחלק בערך העליון ביותר. כלומר $Y_i = \frac{(X_i - a)}{b-a}$

לכן נגדיר: $Y_i = \frac{(X_i+1)}{2}$ (שיטה: להחסיר את המספר הכי קטן שאפשר להוסיף כדי שאפס יהיה החסם התחתון, ואז נחלק במספר הגדול ביותר).

כעת Y_i מתפלג: $U(\{0, \frac{1}{2}, 1\})$. ונגדיר : $Y = \Sigma Y_i$

השלב השני בחסמים הוא לחשב את התוחלת של כל משתנה מקרי ואז את התוחלת של הסכום.

$E(Y_i) = \frac{0+\frac{1}{2}+1}{3} = \frac{1}{2}$. ומכאן : $E(Y) = E(\Sigma Y_i) = \Sigma E(Y_i)$.

השלב הבא הוא למצוא את היחס בין המ"מ בשאלה למ"מ החדש שהגדרנו:

$$S_n = \Sigma_{i=1}^{n} X_i = \Sigma(2Y_i - 1) = 2\Sigma Y_i - n = 2Y - n$$

נציב בהסתברות:

$$P|S_n| > 2\sqrt{n} = P(|2Y - n| > 2\sqrt{n}) = P\left(\left|Y - \frac{n}{2}\right| > \sqrt{n}\right) = P(|Y - E(Y)| > \sqrt{n}) \leq P(|Y - E(Y)| \geq \sqrt{n})$$

לפי צ'רנוף הראשון :

$$P(|Y - E(Y)| \geq \sqrt{n}) \leq 2e^{-\frac{2n}{n}} = 2e^{-2}$$

ב. כדי להשתמש במשפט צריך לעשות הזזה: להוריד את הסכום את התוחלת כפול מספר המ"מ לחלק לשורש של השונות כפול השורש של מספר המשתנים המקריים.

כלומר: $Y_n = \left(\frac{S_n - nE(X_i)}{\sqrt{nVar(X_i)}}\right)$.

נותר לחשב את הVAR של X_i. ומכאן: $Var(X_i) = E(X_i^2) - E(X_i)^2$.

נחשב:

$E(X_i) = \frac{(-1)+0+1}{3} = 0, E(X_i^2) = \frac{1+0+1}{3} = \frac{2}{3}$ סה"כ : $Var(X_i) = \frac{2}{3}$.

נציב בהסתברות:

$$P(|S_n| > 2\sqrt{n}) = P(S_n > 2\sqrt{n}\ or\ S_n < \ -2\sqrt{n})$$

אבל ננסה כיוון שני של וגם:

$$P(|S_n| > 2\sqrt{n}) = 1 - P(S_n \le 2\sqrt{n}\ and\ S_n \ge -2\sqrt{n})$$

נבצע הזזה לפי המשפט:

$$1 - P(-2\sqrt{n} \le S_n \le 2\sqrt{n}) = 1 - P\left(-\frac{2\sqrt{n}}{\sqrt{n}\sqrt{\frac{2}{3}}} \le \frac{S_n}{\sqrt{n}\sqrt{\frac{2}{3}}} \le \frac{2\sqrt{n}}{\sqrt{n}\sqrt{\frac{2}{3}}}\right) =$$

$$1 - P(-\sqrt{6} \le Y_n \le \sqrt{6}) \approx 1 - (\Phi(\sqrt{6}) - (\Phi(-\sqrt{6}) \approx 1 - (\Phi(\sqrt{6}) - 1 + (\Phi(\sqrt{6})$$

$$= 2 - 2\Phi(\sqrt{6}) = 2 - 2\Phi(2.44) \approx 2 - 2 \cdot 0.9927 = 0.0146$$

שאלה 2 (25 נקודות):

כתבו אלגוריתם המקיים את כל התכונות הבאות:

1. **קלט:** מספרים ממשיים $0 \le a_1, a_2, \dots, a_n \le 1$, $\varepsilon > 0$ ו-$\delta > 0$.
2. **פלט:** מספר ממשי x.
3. לכל קלט, הפלט של האלגוריתם צריך לקיים $\left|\frac{a_1+\dots+a_n}{n} - x\right| < \varepsilon$ בהסתברות $1-\delta$ לפחות.
4. זמן הריצה של האלגוריתם קבוע (כלומר לא תלוי ב-n אבל יכול להיות תלוי ב-ε וב-δ).

אלגוריתם: קלט $0 \le a_1, \dots a_n \le 1$, $\varepsilon, \delta > 0$.

- (לאחר חישוב) $k = \left\lceil \frac{\ln\frac{2}{\delta}}{2\varepsilon^2} \right\rceil$
- עבור i מ1 עד k בצע:
  - בחר איבר מתוך: $\{a_1, \dots, a_n\}$ באופן אקראי עם החזרות וללא תלות בבחירות הקודמות.
- החזר את הממוצע של האיברים שנבחרו:
  - אם האיברים שנבחרו הם $b_1, .. b_k$ החזר $x = \frac{(b_1+\dots+b_k)}{k}$.

זמן הריצה: קבוע ולא תלוי בn .

ניתוח ההסתברות: נגדיר: $X_i =$ המספר הi שנבחר מתוך הקבוצה $\{a_1, \dots, a_n\}$.

ונגדיר: $X = \Sigma_{i=1}^{k} X_i$.

נחשב את התוחלת נשים לב שX_i מתפלג: $U\{a_1, \dots, a_n\}$) ולכן: $E(X_i) = \frac{(a_1+\dots+a_n)}{n}$

ומכאן: $E(X) = E(\Sigma^k X_i) = \Sigma^k E(X_i) = k \cdot \frac{(a_1+\dots+a_n)}{n}$.

מכאן: $P(\ \ |\frac{(a_1+\dots+a_n)}{n} - x| \ge \varepsilon) = P(|\frac{E(X)}{k} - \frac{X}{k}\ \ | \ge k\varepsilon) = P(|X - E(X)| \ge k\varepsilon)$

מכיוון ש $0 \le a_1, \dots, a_n \le 1$ אז $0 \le X_i \le 1$ ולכן נשתמש בצ'רנוף ונקבל:

$$P(|X - E(X)| \ge k\varepsilon) \le 2e^{-\frac{2(k\varepsilon)^2}{k}} \le \delta$$

חישוב צד לאי שיווין האחרון: $2e^{-\frac{2(k\varepsilon)^2}{k}} \le \delta$

מכאן: $-2\frac{(k\varepsilon)^2}{k} \le \ln\frac{\delta}{2}$

ולכן: $\varepsilon^2 k \ge \frac{\ln\frac{\delta}{2}}{-2}$

ומכאן : $k \ge \frac{\ln\frac{2}{\delta}}{2\varepsilon^2}$

לכן: $P(\quad |\frac{(a_1+\cdots+a_n)}{n} - x| < \varepsilon) = 1 - P(\quad |\frac{(a_1+\cdots+a_n)}{n} - x| \ge \varepsilon) \ge 1-\delta$

שאלה 3 (25 נקודות):

יהיו $2 \le k \le n$ וכן $m < 2^{k-1}$ מספרים טבעיים. תהי $\{A_1,\dots,A_m\}$ משפחה של תת קבוצות של $\{1,\dots,n\}$ כך ש- $|A_i| = k$ לכל $1 \le i \le m$. הוכיחו שקיימת צביעה של איברי $\{1,\dots,n\}$ בשלושה צבעים כך ש- A_i תכיל איברים משני צבעים שונים לפחות לכל $1 \le i \le m$.

פתרון:

דוגמא : $n = 5$ $\{1,2,3,4,5\}$. $k = 3$. $\{A_1 = \{1,2,3\}, A_2 = \{1,2,4\}, A_3 = \{2,4,5\}\}$.

דוגמת צביעה: 1,2 כחול, 3,4 אדום, 5 צהוב.

נשתמש בשיטה ההסתברותית, נצבע את איברי $\{1,2,\dots,n\}$ אקראית בשלישה צבעים באופן אחיד ובלתי תלוי.

נגדיר: X_i = אינדיקטור הקבוצה A_i צבועה כולה בצבע אחיד.

נגדיר: $X = \Sigma^m X_i$.

נחשב תוחלת : $E(X_i) = P(X_i = 1) = 3 \cdot \left(\frac{1}{3}\right)^k$

ומכאן: $E(X) = E(\Sigma X_i) = \Sigma^m E(X_i) = 3m \cdot \left(\frac{1}{3}\right)^k$

סה"כ: $P(Error) = P(X > 0) = P(\cup_{i=1}^m (X_i = 1) \le \Sigma^m P(X_i = 1) = \Sigma^m E(X_i) = 3m \cdot \left(\frac{1}{3}\right)^k$

$$< 3 \cdot 2^{k-1}\left(\frac{1}{3}\right)^k = 2^{k-1}\left(\frac{1}{3}\right)^{k-1} < 1$$

ולכן ההסתברות לצביעה לא טובה קטנה ממש מ-1 ומכאן ההסתברות לצביעה טובה גדולה ממש מ0 ולכן קיימת צביעה טובה.

שאלה 4 (30 נקודות):

א. (10 נקודות) יהי X משתנה מקרי עם תוחלת $E(X)=3/5$ ופונקציית צפיפות

$$f_X(x)=\begin{cases}ax^2+b & if\ 0\le x\le 1\\ 0 & otherwise\end{cases}$$

מצאו את a ו-b.

ב. (10 נקודות) יהי Y משתנה מקרי רציף המתפלג אחיד על הקטע $(0,5)$. מה ההסתברות שלמשוואה $4z^2+4zY+Y+2=0$ יש שני פתרונות ממשיים שונים?

ג. (10 נקודות) יהי Z משתנה מקרי נורמלי סטנדרטי ויהי $a>0$ מספר ממשי. הוכיחו ש- $\Pr(|Z|>a)=2\Pr(Z>a)$.

א. מכיוון שזו פונקציית צפיפות מתקיים: $\int_{-\infty}^{\infty}f_x(x)dx=1$.

ומכאן: $\int_{-\infty}^{\infty}f_x(x)dx=\int_0^1(ax^2+b)dx=\frac{ax^3}{3}+bx\mid_0^1=\frac{a}{3}+b=1$

לפי הגדרת תוחלת: $E(X)=\int_{-\infty}^{\infty}xf_x(x)dx=\frac{3}{5}$

ומכאן: $\int_{-\infty}^{\infty}xf_x(x)dx=\int_0^1x(ax^2+b)dx=\frac{ax^4}{4}+\frac{bx^2}{2}\mid_0^1=\frac{a}{4}+\frac{b}{2}=\frac{3}{5}$

נפתור מערכת 2 המשוואות: נבודד : $b=1-\frac{a}{3}$ נציב: $\frac{a}{4}+\frac{1}{2}-\frac{a}{6}=\frac{3}{5}$.

ומכאן: $\frac{1}{10}=\frac{a}{12}$ ולכן : $a=\frac{2}{10}=\frac{6}{5}$ ומכאן: $b=1-\frac{2}{5}=\frac{2}{5}$.

ב. לפי נוסחת השורשים נדרוש: $(4Y)^2-4\cdot4\cdot(Y+2)>0$.

ומכאן:

$Y^2-Y-2>0$ ומכאן: $Y<-1\ or\ Y>2$.

המשתנה Y מקבל ערכים בטווח $(0,5)$ ולכן נדרוש ש$Y>2$.

לכן: $P(Y>2)=1-P(Y\le2)=1-\frac{2-0}{5-0}=\frac{3}{5}$.

ג. נחשב: $P(|Z|>a)=1-P(|Z|\le a)=1-P(-a\le Z\le a)=1-\big(\Phi(a)-\Phi(-a)\big)=$

$2-2\Phi(a)=2\big(1-\Phi(a)\big)=2P(Z>a)$

מבחן 2018 א'

שאלה 1 (25 נקודות):

מטילים קוביה הוגנת 72000 פעמים, כאשר כל ההטלות בלתי תלויות. יהי X משתנה מקרי הסופר את מספר ההטלות שתוצאתן 1.

א. השתמשו באי שוויון צ'רנוף כדי להוכיח ש- $\Pr(10000\le X\le14000)\ge0.99$.

ב. השתמשו במשפט הגבול המרכזי כדי להוכיח ש- $\Pr(11900\le X\le12100)\le0.7$.

א. נגדיר: X_i = אינדיקטור שבהטלה i יצא 1. ולכן: $X=\Sigma^{72000}X_i$.

נחשב את התוחלת: $E(X_i)=P(X_i=1)=\frac{1}{6}$. ומכאן: $E(X)=\Sigma^{72000}E(X_i)=12000$

ומכאן לפי צ'רנוף:

$P(10000\le X\le14000)=P(-2000\le X-E(X)\le2000)=P(|X-E(X)|>2000)$

$$= 1 - P(|X - E(X)| > 2000) \geq 1 - \ P(|X - E(X)| \geq 2000) \geq 1 - 2e^{2 \cdot \frac{2000^2}{72000}}$$
$$= 1 - 2e^{-\frac{2000}{18}} = 1 - 2e^{-\frac{1000}{9}} \geq \ 1 - 2e^{-\frac{100}{9}} \geq 1 - 0.001 = 0.99$$

ב. נשתמש במשפט הגבול המרכזי:

נחשב את השונות: $Var(X_i) = P(X_i = 1) \cdot P(X_i = 0) = \frac{1}{6} \cdot \frac{5}{6} = \frac{5}{36}$

מכאן:

$$P(11900 \leq X \leq 12100) = P\left(\frac{11900 - 12000}{\sqrt{72000 \cdot \frac{5}{36}}} \leq \frac{X - 12000}{\sqrt{72000 \cdot \frac{5}{36}}} \leq \frac{12100 - 12000}{\sqrt{72000 \cdot \frac{5}{36}}}\right) =$$

$$P\left(-1 \leq \frac{X - 12000}{\sqrt{72000 \cdot \frac{5}{36}}} \leq 1\right) = \Phi(1) - \Phi(-1) = 2\Phi(1) - 1 \ \approx 2 \cdot 0.8413 - 1 = 0.6826$$
$$< 0.7$$

שאלה 2 (25 נקודות):

כתבו אלגוריתם המקיים את כל התכונות הבאות:

1. **קלט:** קבוצה כלשהי A המכילה n מספרים ממשיים.
2. **פלט:** מספר $x \in A$.
3. לכל קלט, הפלט של האלגוריתם צריך לקיים $|\{y \in A : y < x\}| \geq n/3$ וגם $|\{y \in A : y > x\}| \geq n/3$ בהסתברות $1 - 2^{-100}$ לפחות.
4. זמן הריצה של האלגוריתם קבוע (כלומר לא תלוי ב-n).

האלגוריתם: קלט קבוצה A של n מספרים ממשיים.

- חשב: (אחרי חישוב) $k = 1000$.
- עבור i מ1 עד k בצע:
  - בחר איבר מתוך הקבוצה A באופן אקראי אחיד ובלתי תלוי בבחירות הקודמות.
- החזר את החציון מבין k האיברים שנבחרו.

זמן ריצה קבוע – לא תלוי בn .

נוכיח את ההסתברות:

נגדיר: X_i = אינדיקטור האומר שx הוא בשליש העליונים.

Y_i = אינד' שאומר ש x הוא בשליש התחתונים.

נשים לב שכדי שהחציון של האיברים שנבחרו יהיה בשליש האמצעי, נדרוש בהכרח פחות חצי מהאיברים שנבחרו יהיו בשליש התחתון ופחות מחצי יהיו בשליש העליון.

נגדיר: $X = \Sigma^k X_i, Y = \Sigma^k Y_i$

נחשב תוחלת: $E(X_i) = E(Y_i) \leq \frac{1}{3}$. כיוון שאם גודל הקבוצה מתחלק ב-3 אז כל החלקים שווים ולכן יש $\frac{1}{3}$ סיכוי להיות בכל חלק, אם זה לא מתחלק אז לפי התרגיל צריך לפחות $\frac{n}{3}$ ולכן מעגלים למעלה ולכן החלקים העליונים והתחתונים יהיו גדול יותר מ $\frac{1}{3}$ מהקבוצה.

מכאן: , $E(X) = \Sigma^k E(X_i) \leq \frac{1}{3}k$ אותו דבר $E(Y)$.

$$P(mean\ in\ middle) = P\left(X < \frac{k}{2}, Y < \frac{k}{2}\right) = 1 - P\left(X \geq \frac{k}{2} \cup Y \geq \frac{k}{2}\right)$$
$$\geq 1 - \left(P\left(X \geq \frac{k}{2}\right) + P\left(Y \geq \frac{k}{2}\right)\right)$$

חישוב צד:

$$P\left(X \geq \frac{k}{2}\right) = P\left(X \geq \frac{3}{2} \cdot \frac{k}{3}\right) \leq P\left(X \geq \left(1 + \frac{1}{2}\right) \cdot E(X)\right) \leq e^{-\frac{E(X)}{12}} \rightarrow not\ good\ \ldots$$

$$P\left(X \geq \frac{k}{2}\right) = P\left(X \geq \frac{k}{3} + \frac{k}{6}\right) \leq P\left(X \geq E(X) + \frac{k}{6}\right) \leq e^{-\frac{2k^2}{36k}} = e^{-\frac{k}{18}}$$

$$P\left(Y \geq \frac{k}{2}\right) = P\left(Y \geq \frac{k}{3} + \frac{k}{6}\right) \leq P\left(X \geq E(Y) + \frac{k}{6}\right) \leq e^{-\frac{2k^2}{36k}} = e^{-\frac{k}{18}}$$

סה"כ: $P(mean\ in\ middle) \geq 1 - \left(P\left(X \geq \frac{k}{2}\right) + P\left(Y \geq \frac{k}{2}\right)\right) \geq 1 - 2e^{-\frac{k}{18}} \geq 1 - 2^{-100}$

$e^{-\frac{k}{18}} \geq 4^{-\frac{k}{18}} = 2^{-\frac{k}{9}} \geq 2^{-101}$ מתקיים.

מרתון פתירת מבחנים:

מבחן 2017 א' ב'

שאלה 1 (25 נקודות):

דוגמים עשרה מספרים מתוך $\{1,...,100\}$ באופן אחיד, בלתי תלוי **ובלי החזרה**. השתמש באי-שוויון צ'בישב (הרגיל או החד צדדי, לפי ראות עיניכם) על מנת להוכיח שההסתברות שלפחות 4 מן המספרים שנדגמו הם בין 1 ל-10 היא לכל היותר $1/9$.

פתרון:

נגדיר X = למספר המספרים שיצאו בין 1ל 10ל ולכן הוא מתפלג היפרגיאומטרית עם פרמטרים: $N = 100$ $D = 10, n = 10$ ולכן: $E(X) = \frac{nD}{N} = 1$.

$Var(X) = n \cdot \frac{\left(\frac{D}{N}\right)\left(\frac{1-D}{N}\right)(N-n)}{N-1}$.

$$P(X \geq 4) \leq P(|X - E(X)| \geq 3) \leq \frac{\frac{81}{99}}{9} < \frac{1}{9}$$

שאלה 2 (25 נקודות):

נתון אלגוריתם A המקיים את כל התכונות הבאות:

1. **קלט:** מספר טבעי $n > 1$.
2. **פלט:** " n ראשוני" או " n פריק".
3. לכל קלט, הפלט של האלגוריתם נכון בהסתברות $2/3$ לפחות.
4. זמן הריצה של האלגוריתם הוא $f(n)$.

השתמש ב- A כדי לכתוב אלגוריתם B המקיים את כל התכונות הבאות:

1. **קלט:** שלושה מספרים טבעיים $n,m,k > 1$.
2. **פלט:** "כן" אם בדיוק אחד מבין n,m ו- k הוא ראשוני ו"לא" אחרת.
3. לכל קלט, הפלט של האלגוריתם נכון בהסתברות 0.99 לפחות.
4. זמן הריצה של האלגוריתם הוא $\theta(f(n))$.

פתרון:

קודם להסתכל על זמן הריצה – מה מרשים לי? בדרך כלל זה קבוע.

בדרך כלל ננסה לחפש את התשובה הנכונה דרך דגימה.

אלגוריתם B:

- עבור i מ1 עד t:
  - הרץ את A על n,m,k.
- אם קיבלנו לפי A שבדיוק אחד הוא ראשוני – החזר כן.

- החזר – לא.

זמן ריצה: הרצה של אלגוריתם A על הקלטים מספר קבוע של פעמים ולכן: $\theta(f(n))$

הסתברות לטעות:

נניח שבקלט יש בדיוק אחד ראשוני. בה"כ n.

ההסתברות לתשובה נכונה היא: או שהאלגוריתם A יענה נכון על כולם.

או שA יענה שגוי על n ואז יענה כן על בדיוק מהשניים האחרים..

$$P(Correct) = P(A(n) = 1, A(m) = 0, A(k) = 0) + P(A(n) = 0, A(m) = 1, A(k) = 0) + P(A(n) = 0, A(m) = 0, A(k) = 1) \geq \left(\frac{2}{3}\right)^3 + 0 + 0 = \frac{8}{27}$$

$$P(Correct) = 3 \cdot P(1,0,0) + 2 \cdot P(0,1,0) \geq \left(\frac{2}{3}\right)^3 + 0 + 0 = \frac{8}{27}$$

(לקחנו רק מקרה שהראשון צודק אבל היתר אמרנו שההסתברות שלהם היא 0)

נניח שעכשיו בקלט אין בדיוק אחד ראשוני.

$$P(Correct) = 1 - [P(A(n) = 1, A(m) = 0, A(k) = 0) + P(A(n) = 0, A(m) = 1, A(k) = 0) + P(A(n) = 0, A(m) = 0, A(k) = 1)] \geq 1 -$$

שאלה 4 (25 נקודות):

הוכיחו שההסתברות שקיים קודקוד ב-$G(n,1/2)$ שדרגתו גדולה מ-$0.51n$ או קטנה מ-$0.49n$ שואפת לאפס כאשר n שואף לאינסוף.

פתרון:

נסמן משתנה מקרי X_i את הדרגה של i. מכאן: $E(X_i) = \left(\frac{1}{2}\right) \cdot (n-1) < \frac{n}{2}$. בממוצע אני אהיה מחובר לחצי מכלל הקודקודים שאני יכול להתחבר אליהם.

$$P\left(\bigcup_{i=1}^{n} (X_i > 0.51n \text{ or } X_i < 0.49n)\right) \leq \sum_{i=1}^{n} P(X_i > 0.51n \text{ or } X_i < 0.49n)$$

חישוב צד:

$$P(X_i > 0.51n \text{ or } X_i < 0.49n) = P(X_i - 0.5n > 0.01n \text{ or } X_i - 0.5n < -0.01n) \leq P(|X_i - E(X)| > 0.01n) \leq P(|X_i - E(X)| > 0.01n) \leq 2e^{\frac{-2(0.01n)^2}{n-1}}$$

הערה: כל משתנה בינומי מורכב מאינדיקטורים וניתן לעשות עליו שימוש של צ'רנוף!

נחזור ונציב:

$$\sum_{i=1}^{n} P(X_i > 0.51n \text{ or } X_i < 0.49n) \leq \sum_{i=1}^{n} 2e^{\frac{-2(0.01n)^2}{n-1}} = n \cdot 2e^{\frac{-2(0.01n)^2}{n-1}}$$

ומכאן: $= \frac{2n}{e^{0.0002n}} \to 0$

מבחן 2020

שאלה 1 (25 נקודות):

תהי $\{X_i\}_{i=1}^{\infty}$ סדרה של משתנים מקריים בלתי תלויים כך שלכל $i \in \mathbb{N}$ מתקיים

$$\Pr(X_i = -9) = 1/10, \quad \Pr(X_i = 1) = 9/10$$

לכל $n \in \mathbb{N}$ יהי $S_n = \sum_{i=1}^{n} X_i$.

א. (13 נקודות) השתמשו באי שוויון צ'רנוף (כפי שנלמד בהרצאות) כדי למצוא חסם עליון טוב ככל האפשר על $\Pr(|S_n| > 10\sqrt{n})$.

ב. (12 נקודות) השתמשו במשפט הגבול המרכזי כדי למצוא הערכה טובה ככל האפשר ל-$\Pr(|S_n| > 10\sqrt{n})$.

א. נגדיר $Y_i = \frac{X_i - 9}{10}$ ומכאן, Y_i הוא אינדיקטור כך ש: $P(Y_i = 0) = P(X_i = -9) = \frac{1}{10}$

$P(Y_i = 1) = P(X_i = 1) = \frac{9}{10}$.

נגדיר: $Y = \Sigma Y_i$.

$Y = \frac{S_n + 9n}{10}$ ומכאן: $S_n = 10Y - 9n$ ולכן: $E(Y) = \Sigma E(Y_i) = \frac{\Sigma 9}{10} = \frac{9n}{10}$ לפי ליניאריות התוחלת.

לכן: $P(|S_n| > 10\sqrt{n}) \leq P(|10Y - 9n| \geq 10\sqrt{n}) = P\left(\left|Y - \frac{9n}{10}\right| \geq \sqrt{n}\right) \leq 2e^{-2\cdot\frac{(\sqrt{n})^n}{n}} =$

$2e^{-2}$.

לפי צ'רנוף.

ב. לפי משפט הגבול המרכזי: $E(X_i) = 0$. $Var(X_i) = E(X_i^2) - E(X_i)^2 = 9$

$$P(|S_n| > 10\sqrt{n}) = 1 - P(|S| \leq 10\sqrt{n}) = 1 - P(-10\sqrt{n} \leq S \leq 10\sqrt{n}) =$$

$$1 - P\left(-\frac{(10\sqrt{n})}{3\sqrt{n}} \leq \frac{S_n}{3\sqrt{n}} \leq \frac{10\sqrt{n}}{3\sqrt{n}}\right) = 1 - \left[\Phi\left(\frac{10}{3}\right) - \Phi\left(-\frac{10}{3}\right)\right] = 2 - 2\Phi\left(\frac{10}{3}\right)$$

שאלה 2 (25 נקודות):

כתבו אלגוריתם המקיים את כל התכונות הבאות:

1. **קלט:** מספרים טבעיים $n \geq k \geq 2$ ומשפחה $\{A_1, \ldots, A_m\}$ של תת קבוצות של $\{1, \ldots, n\}$ כך שמתקיים $|A_i| = k$ לכל $1 \leq i \leq m$ וכן $m \leq 2^{k-2}$.
2. **פלט:** צביעה של איברי $\{1, \ldots, n\}$ בשני צבעים (נניח, אדום וכחול).
3. בהסתברות $1 - 2^{-100}$ לפחות, הצביעה של איברי $\{1, \ldots, n\}$ תהיה כזו ש-A_i תכיל איברים משני הצבעים (כלומר, לפחות איבר אחד אדום ולפחות איבר אחד כחול) לכל $1 \leq i \leq m$.
4. זמן הריצה של האלגוריתם הוא $O(n + mk)$.

אלגוריתם:

- עבור k מ1 עד t:
- צבע אקראית את האיברים בכחול בהסתברות $\frac{1}{2}$ ואותו דבר באדום.
  - עבור כל הקבוצות. עבור $1 \leq i < m$:
  - $b = 0\,, r = 0$
    - עבור $1 \leq j \leq k$:
      - בדוק האם האיבר j בתוך A_i צבוע בכחול.
      - אם כן: $b++$.
      - אחרת: $r++$.
    - אם $b = 0\ or\ r = 0$ עבור לk הבא.
  - אם עברת על כל הקבוצות ובכל קבוצה יש לפחות אחד אדור ואחד כחול – החזר את הצביעה.
- החזר את הצביעה.

סיבוכיות זמן ריצה:$O\big(t \cdot (n + mk)\big)$ כי מבצעים t סיבובים ובכל סיבוב עוברים על האיברים $n - 1$ ובוחרים עבורם צבע לאחר מכן עוברים על כל תתי הקבוצות של כל אחת מהן בגודל k ומוודאים את הצביעה.

הפרמטר t קבוע ולכן זמן הריצה הוא $O(n + mk)$.

הסתברות להצלחה – היא 1 פחות ההסתברות שלא צבענו טוב בכל איטרציה של t.

$$P(Good\ Coloring) = 1 - (P(Bad\ coloring))^t$$

חישוב צד:

$$P(Bad\ coloring) = P\left(\bigcup_{i=1}^{m} A_i\ is\ bad\right) \leq \sum_{i=1}^{m} A_i is\ bad = \sum_{i=1}^{m} 2 \cdot \frac{1}{2^k} = \frac{m}{2^{k-1}} \leq \frac{2^{k-2}}{2^{k-1}} = \frac{1}{2}$$

נציב:

$$P(Good\ Coloring) = 1 - (P(Bad\ coloring))^t \geq 1 - 2^{-t}$$

$t = 100$ וככה נקבל את התוצאה כנדרש.

שאלה 3 (25 נקודות):

הוכיחו שההסתברות שקיימת ב-$G(n,1/2)$ קבוצה שלטת (כלומר קבוצת קודקודים S כך שכל קודקוד של $G(n,1/2)$ שאינו ב-S מחובר בקשת לקודקוד של S) מגודל 100 שואפת לאפס כאשר n שואף לאינסוף.

X_s – להיות אינדיקטור האם S היא קבוצה שלטת ב$G\left(n, \frac{1}{2}\right)$.

$$E(X_s) = P(X_s = 1) = 1 - P(X_s = 0) = 1 - \frac{n - 100}{2^{100}}$$

מכאן: $X = \Sigma_{S \subseteq V} X_S$. $E(X) = \binom{n}{100}\left[1 - \frac{n-100}{2^{200}}\right]$לכן:

$$P(exists\ DomSet) = P(X > 0) = P(X \geq 1) = P\left(X - E(X) \geq 1 - \binom{n}{100}\left[1 - \frac{n - 100}{2^{200}}\right]\right)$$

$\leq e$^(

חישוב צד:

$$1-\binom{n}{100}\left[1-\frac{n-100}{2^{200}}\right]=1+2O(n^{100}-n$$

** לשלים מהמרתון.

$$P(S\ is\ good)=P(\cup X_i=1)\leq \Sigma 1-\frac{n-100}{2^{100}}\leq \binom{n}{100}\left(1-\frac{n-100}{2^{100}}\right)\leq n^{100}\cdot e^{-\frac{(n-100)}{2^{100}}}\to 0$$

שאלה 4 (25 נקודות):

יהי X משתנה מקרי עם פונקציית צפיפות

$$f_X(x)=ae^{-|x|}$$

המוגדרת לכל x ממשי.

א. (7 נקודות) מצאו את a.
ב. (9 נקודות) חשבו את התוחלת של X.
ג. (9 נקודות) חשבו את השונות של X.

א. f_X היא פונקציית צפיפות ולכן האינטגרל על כל הערכים =1.

מכאן נדרוש: $\int_{-\infty}^{\infty} ae^{-|x|}=1$.

נחשב: $\int_{-\infty}^{\infty} ae^{-|x|}=\int_{-\infty}^{0} ae^{x}+\int_{0}^{\infty} ae^{-x}=\lim_{b\to-\infty} ae^0-ae^b+\lim_{b\to\infty} -ae^{-b}+ae^0$

$$=a+a\overset{?}{=}1$$

מכאן: $a=\frac{1}{2}$.

ב. $E(X)=\int_{-\infty}^{\infty} xae^{-|x|}=\frac{1}{2}\int_{-\infty}^{\infty} xe^{-|x|}=\frac{1}{2}\left(\int_{-\infty}^{0} xe^{x}+\int_{0}^{\infty} xe^{-x}\right)$

לפי אינטגרציה בחלקים:

$$\int_{-\infty}^{0} xe^{x}=xe^{x}|_{-\infty}^{0}-\int_{-\infty}^{0} e^{x}=-1$$

$$\int_{0}^{\infty} xe^{-x}=-xe^{-x}|_{0}^{\infty}-\int_{0}^{\infty} -e^{-x}=1$$

סה"כ: $E(X)=\frac{1}{2}(-1+1)=0$.

ג. $Var(X)=E(X^2)-E(X)^2$

$$E(X^2)=\frac{1}{2}\left(\int_{-\infty}^{0} x^2e^{x}+\int_{0}^{\infty} x^2e^{-x}\right)$$

<u>**הערה:**</u> שאני שם x^2 זה רק של המקדם של התוחלת ולא של הפונקציה.

$$\int_{-\infty}^{0} x^2e^{x}=x^2e^{x}|_{-\infty}^{0}-\int_{-\infty}^{0} 2xe^{x}=-2(-1)=2$$

(כי אני צריך שוב לעשות אינטגרציה בחלקים אבל אם מוציאים את ה-2 החוצה זה כמו למעלה)

$$\int_0^\infty x^2e^{-x} = -x^2e^{-x}|_0^\infty - \int_0^\infty -2e^{-x} = 2$$

סה"כ: $E(X^2) = 2$

לכן: $Var(X) = 2 - 0 = 2$.

מבחן 2017 א'

שאלה 1 (25 נקודות):

מטילים קוביה תקנית 360000 פעם. יהי X משתנה מקרי הסופר את מספר ההטלות שתוצאתן 6.

מצאו מספרים a ו- b עבורם מתקיים $\Pr(54000 \le X \le 63000) \approx \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx$.

פתרון:

נשתמש במשפט הגבול המרכזי:

נגדיר $X_i = 1$ אם יצא 6 בהטלה הi, 0 אחרת.

$X = \Sigma X_i$ מתפלג בינומית עם $n = 36000, p = \frac{1}{6}$.

לכן : $E(X) = np = 60000$ וגם $Var(X) = np(1-p) = 50000$.

$$E(X_i) = \frac{1}{6}\ Var(X_i) = \frac{5}{36}$$

מכאן:

$$P(54000 \le X \le 63000) = P\left(\frac{(54000-60000)}{\sqrt{\frac{5}{36}\cdot 360000}} \le \frac{(X-60000)}{\sqrt{\frac{5}{36}\cdot 360000}} \le \frac{(63000-60000)}{\sqrt{\frac{5}{36}\cdot 360000}}\right)$$

$$P\left(\frac{-4000}{\sqrt{\frac{5}{36}\cdot 360000}} \le \frac{(X-60000)}{\sqrt{\frac{5}{36}\cdot 360000}} \le \frac{4000}{\sqrt{\frac{5}{36}\cdot 360000}}\right)$$

ולפי המשפט: $a = \frac{-4000}{\sqrt{\frac{5}{36}\cdot 360000}}, b = \frac{4000}{\sqrt{\frac{5}{36}\cdot 360000}}$

שאלה 2 (25 נקודות):

כתבו אלגוריתם המקיים את כל התכונות הבאות:

1. **קלט:** קבוצה כלשהי A המכילה n מספרים ממשיים.
2. **פלט:** מספר $x \in A$.
3. לכל קלט, הפלט של האלגוריתם צריך לקיים $|\{y \in A : y < x\}| \ge 2n/3$ בהסתברות $1-2^{-100}$ לפחות.
4. זמן הריצה של האלגוריתם קבוע (כלומר לא תלוי ב- n).

פתרון:

אלגוריתם על קלט A = קבוצה של n מספרים ממשיים.

- עבור i מ1 עד t:
  - בחר איבר מA באופן אקראי ללא תלות ועם החזרות.
- החזר את המקסימום מבין t האיברים שנבחרו.

זמן ריצה: $O(t)$ כפול זמן דגימה שהוא קבוע ולכן הזמן הוא $O(1)$.

הסתברות:

נגדיר: X_i = אינדיקטור האם האיבר הi שנבחר הוא מהשליש העליון של A.

מכאן: מכיוון שמחזירים את המקסימום אז האלגוריתם יחזיר נכון אם לפחות 1 מהאיברים הוא מהשליש העליון.

מה ההסתברות שאיבר אחד יהיה מתוך ה$\frac{2}{3}$ התחתונים: $\frac{2}{3} = \frac{\frac{2n}{3}}{n}$.

לכן:

$$P(error) = P\left(all\ a_i is\ from\ lowest\frac{2}{3}\right) = \left(\frac{2}{3}\right)^{t=200} = \left(\frac{4}{9}\right)^{\frac{t=200}{2}} < \left(\frac{1}{2}\right)^{100}$$

שאלה 4 (25 נקודות):

יהיו $n \geq k \geq 4$ מספרים טבעיים ותהי $\{A_1, \dots, A_m\}$ משפחה של תת קבוצות של $\{1, \dots, n\}$ כך שמתקיים $|A_i| = k$ לכל $1 \leq i \leq m$ וכן $m < 4^{k-1}/3^k$. הוכיחו שקיימת צביעה של איברי $\{1, \dots, n\}$ בארבעה צבעים כך ש- A_i תכיל איברים מכל ארבעת הצבעים לכל $1 \leq i \leq m$.

נשתמש בשיטה ההסתברותית:

נצבע אקראית את איברי *1-n* באופן אחיד ב4 צבעים אחיד ובלתי תלוי.

נגדיר: X_i = להיות אינדיקטור שאומר שA_i לא צבועה ב4 צבעים.

מכאן:

$$P(bad\ coloring) = P\left(\bigcup_{i=1}^{m} X_i = 1\right) \leq \Sigma_{i=1}^{m} P(X_i = 1) = \Sigma_{i=1}^{m} 4 \cdot \left(\frac{3}{4}\right)^k = m \cdot 4 \cdot \left(\frac{3}{4}\right)^k$$
$$< \frac{4^{k-1}}{3^k} \cdot \frac{3^k}{4^{k-1}} = 1$$

קיבלנו שההסתברות לצביעה רעה קטנה ממש מ-1 ולכן קיימת צביעה טובה.

מבחן לדוגמה 2017:

שאלה 1 (25 נקודות):

חלקיק נמצא בנקודה 0 על ציר המספרים. בכל שלב הוא נע צעד אחד ימינה (כלומר מ- t ל- $t+1$) בהסתברות $1/2$ או שמאלה (כלומר מ- t ל- $t-1$) בהסתברות $1/2$ באופן בלתי תלוי בשלבים הקודמים. הוכיחו שההסתברות שאחרי 100 השלבים הראשונים החלקיק נמצא בנקודה שהיא מימין ל-20 היא לכל היותר e^{-2}.

פתרון:

נגדיר את X_i = הצעד בשלב הi כאשר 1 צעד ימינה ו1- הוא צעד שמאלה. כל אחד בהסתברות $\frac{1}{2}$.

נגדיר $X = \Sigma^{100} X_i$. מכאן X הוא הנקודה הסופית לאחר 100 צעדים.

מכאן לפי צ'רנוף הופדינג : $P(X > 20) \leq \ P(X \geq 20) \leq e^{-\frac{400}{200}} = e^{-2}$.

שאלה 2 (25 נקודות):

תהי F משפחה של וקטורים בינאריים באורך n. כל וקטור $(x_1,\ldots,x_n) \in F$ מקיים $\sum_{i=1}^{n} x_i \geq 2n/3$ (במקרה זה נאמר שהוקטור $(x_1,\ldots,x_n)$ הוא גדול) או $\sum_{i=1}^{n} x_i \leq n/3$ (במקרה זה נאמר שהוקטור $(x_1,\ldots,x_n)$ הוא קטן). כתבו אלגוריתם המקיים את כל התכונות הבאות:

1. **קלט:** וקטור $(x_1,\ldots,x_n) \in F$ כלשהו.
2. **פלט:** "$(x_1,\ldots,x_n)$ קטן" או "$(x_1,\ldots,x_n)$ גדול".
3. לכל קלט, הפלט של האלגוריתם צריך להיות נכון בהסתברות $1-2^{-100}$ לפחות.
4. זמן הריצה של האלגוריתם קבוע (כלומר לא תלוי ב- n).

פתרון:

אלגוריתם: קלט ווקטור מתוך F:

- עבור i מ1 עד t:
  - בחר ביט בוקטור באופן אקראי ובלתי תלוי והוסף לסכום.
- אם הסכום גדול מ $\frac{t}{2}$ אז החזר "גדול"
- אחרת החזר "קטן"

(ניקח t אי זוגי כדי לא להיות באמצע בדיוק)

מכיוון שt הוא הקבוע זמן הריצה קבוע.

הסתברות:

נגדיר X_i = הביט שנבחר בשלב הi.

$X = \Sigma^{t} X_i$.

נניח שהקלט גדול.

מכאן: $P(X_i = 1) \geq \frac{2}{3}$ – זה כי אנחנו מניחים שלפחות $\frac{2}{3}$ הם אחדות!

.$E(X) \geq \frac{2t}{3}$

נשתמש באי שיוויון צ'רנוף: $P(X \leq E(X) - a) \leq e^{-\frac{2a^2}{t}}$

ההסתברות לטעות במקרה זה היא: $\frac{t}{2} = \frac{2t}{3} - a$ וככה נגיע לa.

$$P(Error) = P\left(X < \frac{t}{2}\right) \leq P\left(X < \frac{2t}{3} - \frac{t}{6}\right) \leq P\left(X < E(X) - \frac{t}{6}\right) \leq P\left(X \leq E(X) - \frac{t}{6}\right)$$

$$\leq e^{-\frac{2\left(\frac{t}{6}\right)^2}{t}} = e^{-\frac{t}{18}} < e^{-100} < 2^{-100}$$

נניח שהקלט גדול.

מכאן: $P(X_i = 1) \geq \frac{1}{3}$ – זה כי אנחנו מניחים לכל היותר $\frac{1}{3}$ הם אחדות!

.$E(X) \leq \frac{t}{3}$

נשתמש באי שיוויון צ'רנוף: $P(X \geq E(X) + a) \leq e^{-\frac{2a^2}{t}}$

ההסתברות לטעות במקרה זה היא: $\frac{t}{2} = \frac{2t}{3} - a$ וככה נגיע לa.

$$P(Error) = P\left(X > \frac{t}{2}\right) \leq P\left(X > \frac{t}{3} + \frac{t}{6}\right) \leq P\left(X \geq E(X) + \frac{t}{6}\right) \leq P\left(X \leq E(X) + \frac{t}{6}\right)$$

$$\leq e^{-\frac{2\left(\frac{t}{6}\right)^2}{t}} = e^{-\frac{t}{18}} < e^{-100} < 2^{-100}$$

שאלה 4 (25 נקודות):

הוכיחו כי $\lim_{n\to\infty} \Pr(B(n,n,1/2)$ קשיר$)=1$.

הוכיחו שאם לוקחים גרף דו צדדים עם n קודקודים בכל צד ושמים צלע בין שני הצדדים בהסתברות $\frac{1}{2}$ אז הגרף כמעט בוודאות קשיר ככל ש $n \to \infty$.

נוכיח משהו חזק יותר: שבהסתברות שואפת ל1, לכל 2 קודקודים בצד אחד יש שכן משותף ושאין קודקוד בודד.

הוכחה: נראה שבין כל 2 קודקודים בגרף יש מסלול (אז זה אומר שהוא קשיר). יהיו $u, v \in V$ אם הם באותו צד אז קיים לפי ההנחה קודקוד w שהוא שכן משותף לכן קיים המסלול $u \to w \to v$ (ולהיפך).

אם הם לא באותו צד אז מכיוון שאף קודקוד הוא לא בודד, קיים שכן w לu בצד השני. כעת, באותו אופן כמו במקרה הראשון, בין w לv יש מסלול (כי הם באותו צד) ולכן קיים מסלול מu לv.

ההסתברות שיש קודקוד בודד: $P(isolated\ vetex) \le 2n \cdot \left(\frac{1}{2}\right)^n = n \cdot 2^{1-n} \to 0$

ההסתברות שיש שני קודקודים ללא שכן משותף:

$$P(u, v\ without\ commin\ neighbor) \le 2 \cdot \binom{n}{2} \cdot \left(1 - \left(\frac{1}{2}\right)^2\right)^n \to 0$$

מה ההסתברות שהגרף קשיר:

$$P\left(B\left(n, n, \frac{1}{2}\right) is\ connected\right) \ge 1 - P(isolated\ vetex) - P(u, v\ without\ commin\ neighbor) \to 1$$

ולכן ההסתברות לגרף קשיר שואפת ל1.

מבחן 2018:

שאלה 3 (25 נקודות):

יהי $G \sim G(n, 1/n)$ גרף מקרי. הוכיחו שההסתברות שקיימת קבוצה של קודקודים של G מגודל $1 \le t \le n/10$ המכילה לפחות $3t$ קשתות שואפת ל-0 כאשר n שואף לאינסוף.

חסם הדוק ל$\binom{n}{k}$ הוא: $\binom{n}{k} \le \left(\frac{en}{k}\right)^k$ עדיף להשתמש בזה כאשר k תלוי בn.

$X_s =$ אינדיקטור שלקבוצה S בגודל t יש לפחות $3t$ צלעות, אז:

$$E(X_s) = P(X_s = 1) \le \binom{\binom{t}{2}}{3t} \cdot \left(\frac{1}{n}\right)^{3t} \le \left(\frac{et^2}{3t}\right)^{3t} \left(\frac{1}{n}\right)^{3t} = \left(\frac{et}{3n}\right)^{3t}$$

מכאן: ההסתברות שלא קיימת אף קבוצה כזו מאף גודל $1 \le t \le \frac{n}{10}$ הוא:

$$P(Good) \le \sum_{t=1}^{\frac{n}{10}} \sum_{S \subseteq V, |S| = t} P(X_s = 1) \le \sum_{t=1}^{\frac{n}{10}} \sum_{S \subseteq V, |S| = t} \left(\frac{et}{3n}\right)^{3t} = \sum_{t=1}^{\frac{n}{10}} \binom{n}{t} \left(\frac{et}{3n}\right)^{3t} \le \sum_{t=1}^{\frac{n}{10}} \left(\frac{en}{t}\right)^t \left(\frac{e^3 t^3}{3^3 n^3}\right)^t$$
$$= \sum_{t=1}^{\frac{n}{10}} \left(\frac{e^4 t^2}{3^3 n^2}\right)^t \le \sum_{t=1}^{\frac{n}{10}} \left(\frac{3t}{n}\right)^{2t}$$

טריק זה לעשות פיצול סיגמא:

$$u, \sum_{t=1}^{\sqrt{n}} \left(\frac{3t}{n}\right)^{2t} + \sum_{t=\sqrt{n}}^{\frac{n}{10}} \left(\frac{3t}{n}\right)^{2t} \le \sum_{t=1}^{\sqrt{n}} \left(\frac{3\sqrt{n}}{n}\right)^2 + \sum_{t=\sqrt{n}}^{\frac{n}{10}} \left(\frac{3\left(\frac{n}{10}\right)}{n}\right)^{2\sqrt{n}}$$

$$= \sqrt{n} \cdot \left(\frac{9}{n}\right) + \left(\frac{n}{10} - \sqrt{n}\right) \cdot \left(\frac{3}{10}\right)^{2\sqrt{n}} \to 0$$

שאלות מתרגולים:

תרגול 5:

Exercise 1 For two vertices u and v in a graph G, we denote by $dist_G(u,v)$ the length of a shortest path connecting them. The diameter of a graph is defined to be $\operatorname{diam}(G) = \max_{u,v} dist_G(u,v)$.

Let $G \sim G(n,m)$, where $m = \left\lceil \frac{1}{2}\binom{n}{2} \right\rceil$. Prove that

$$\lim_{n\to\infty} \Pr(\operatorname{diam}(G(n,m)) > 2) = 0.$$

נגדיר במקום: $G\left(n,\frac{1}{2}\right)$.

נגדיר: X_{uv} אינדיקטור שהמרחק (מסלול קצר ביותר) בין $u,v \geq 3$.

מכאן: $P(X_{uv}=1)=\frac{1}{2}\cdot$

$$P(Diam(G)>2)=P(\bigcup_{u,v}(X_{uv}=2) \leq \sum_{u.v} P(X_{uv}=1)$$

(?????)

1. Suppose that A is a Monte-Carlo randomized algorithm for some problem P, which, given any input, outputs a correct solution with probability at least $1/2$, and whose running time on inputs of size n is $f(n)$. Suppose also that, for any input x of A, the size of the output of $A(x)$ is not larger than the size of x. Finally, suppose that B is a deterministic algorithm which, given a potential solution for P of size m, verifies its correctness in time $g(m)$, where g is a non-decreasing function. Devise a Las-Vegas algorithm C for P whose expected running time is $O(f(n)+g(n))$.

פתרון:

אלגוריתם A עם טעות על כל קלט של $\frac{1}{2}$ לכל היותר בזמן $f(n)$.

$A(x)=y$ כך ש: $|y| \leq |x|$.

נתון: אלגוריתם B לא הסתברותי שמקבל את y ומוודא שy הוא הפתרון לא בזמן: $g(n)$.

פתחו אלגוריתם לאס וגאס לאותה בעיה בזמן ריצה של $O(f(n)+g(n))$.

אלגוריתם של C: על קלט x:

- הרץ את $A(x)=y$.
- $B(y)$ אם ענה כן – החזר את y. אם לא – חזור להתחלה.

ההסתברות שC עונה נכון היא 1.

זמן הריצה של C:

נסמן בX את מספר האטרקציות עד שנעצור.

בכל איטרציה יש סיכוי להצלחה (כי אלגוריתם A מחזיר תשובה נכונה בהסתברות $\frac{1}{2}$)

והאיטרציות הן בלתי תלויות. וממשיכים לבצע עד ההצלחה הראשונה.

לכן X מתפלג גיאומטרית עם הסתברות: $\frac{1}{2}$ לפחות.

מכאן: $T_c = X \cdot \left(f(n) + g(n)\right)$ זה זמן הריצה.

ולכן: $E(X) \leq 2$.

מכאן : $E(T_c) \leq 2\left(f(n) + g(n)\right)$

סה"כ זמן הריצה הממוצע הוא לכל היותר $O\left(\left(f(n) + g(n)\right)\right)$.

תרגיל 4:

4. Let $X_1, \ldots, X_{1200}$ be mutually independent random variables such that, for every $1 \leq i \leq 1200$, X_i is uniformly distributed over the segment $[0,1)$. For every $1 \leq i \leq 1200$, Let Y_i be a rounding of X_i to the nearest integer, i.e.

$$Y_i = \begin{cases} 0 & \text{if } X_i < 1/2 \\ 1 & \text{if } X_i \geq 1/2 \end{cases}$$

Use the Central Limit Theorem to estimate the following probabilities:

(a) $Pr\left(\left|\sum_{i=1}^{1200} X_i - \sum_{i=1}^{1200} Y_i\right| \leq 10\right)$.

(b) $Pr\left(\sum_{i=1}^{1200} |X_i - Y_i| > 310\right)$.

א. נשים לב ש: $\Sigma^{1200} X_i - \Sigma^{1200} Y_i = \Sigma(X_i - Y_i)$

נגדיר $Z_i = X_i - Y_i$.

Z_i מתפלג אחיד על $[-\frac{1}{2}, \frac{1}{2})$

$$E(X_i) = \int_0^1 x dx = \frac{x^2}{2}|_0^1 = \frac{1}{2}$$

$$E(Y_i) = P(Y_i = 1) = P\left(X_i \geq \frac{1}{2}\right) = \int_{\frac{1}{2}}^{1} 1 dx = x|_{\frac{1}{2}}^{1} = \frac{1}{2}$$

מכאן: $E(Z_i) = E(X_i) - E(Y_i) = \frac{1}{2} - \frac{1}{2} = 0$.

$$Var(Z_i) = Var(X_i - Y_i) = Cov(X_i - Y_i, X_i - Y_i)$$
$$= Cov(X_i, X_i) + Cov(Y_i, Y_i) - 2Cov(X_i, -Y_i)$$

חישוב צד:

$$Cov(X_i, Y_i) = E(X_i Y_i) - E(X_i)E(Y_i) = \int_{\frac{1}{2}}^{1} x dx - \frac{1}{4}$$
$$= -\frac{1}{2} + \frac{1}{8} + \frac{1}{4} = -\frac{1}{8}$$
$$Cov(X_i, X_i) = Var(X_i) = \frac{((1-0)^2)}{12} = \frac{1}{12}$$
$$Cov(Y_i, Y_i) = Var(Y_i) = \frac{1}{4}$$

סה"כ: $Var(Z_i) = \frac{1}{12}$.

נגדיר: $Z = \Sigma Z_i$

מכאן: $P(|\Sigma^{1200}X_i - \Sigma^{1200}Y_i| \leq 10) = P(|Z| \leq 10) = P(-10 \leq Z \leq 10) =$

$$P\left(-\frac{10}{\sqrt{1200 \cdot \frac{1}{12}}} \leq \frac{Z}{\sqrt{1200 \cdot \frac{1}{12}}} \leq \frac{10}{\sqrt{1200 \cdot \frac{1}{12}}}\right) = P\left(-1 \leq \frac{Z}{\sqrt{1200 \cdot \frac{1}{12}}} \leq 1\right)$$

ב.

נגדיר $Z_i = |X_i - Y_i|$.

פונקציית הצפיפות היא:

$$f_{Z_i}(x) = \begin{cases} x & ,0 \leq x < \frac{1}{2} \\ 1-x & ,\frac{1}{2} \leq x < 1 \end{cases}$$

$$E(Z_i) = \int_0^{\frac{1}{2}} x^2 dx + \int_{\frac{1}{2}}^1 x - x^2 dx = \cdots = \frac{1}{8}$$

$$E\left(Z_i^2\right) = \int_0^{\frac{1}{2}} x^3 dx + \int_{\frac{1}{2}}^1 x^2 - x^3 dx = \cdots = \frac{7}{96}$$

$$Var(Z_i) = E\left(Z_i^2\right) - E(Z_i)^2 = \frac{11}{192}$$

נגדיר: $Z = \Sigma Z_i$

אלגוריתם:

- הגדר: K, M, N
- למשך: $1 \leq i \leq t$ סיבובים הרץ :
  - הרצץ את A על k, m, n בכל פעם שA מחזיר "פריק" – הוסף 1 לK, N, M בהתאמה.
- אם **רק** אחד מהמשתנים גדול מ$\frac{2t}{3}$ החזר "כן".
- אחרת, "לא".

הגדרת K_i, M_i, N_i – אינדיקטורים המקבלים 1: כאשר A מחזיר "ראשוני" ומקבל 0 אשר A מחזיר "פריק".

נניח כי הקלט: בה"כ k הוא ראשוני והיתר לא.

לכן: $E(K_i) = P(X_i = 1) \geq \frac{2}{3}, E(M_i) = P(M_i = 1) < \frac{1}{3},$

נגדיר: $K = \sum_{i=1}^{t} K_i$. (וכך גם לM_i, N_i) לכן כדי לבדוק מה ההסתברות שאנחנו נקבל "כן":

$P\left(K \geq \frac{2t}{3}\right) + P\left(M < \frac{2t}{3}\right) + P\left(N < \frac{2t}{3}\right)$.

או ללכת בכיוון אחר:

אלגוריתם:

- הגדר: X

- למשך: $1 \leq i \leq t$ סיבובים הרץ :
  - הרץ את A על k, m, n בכל פעם שA מחזיר "פריק" רק על 1 הוסף 1 לX.
- אם $|X|$ גדול מ$\frac{2t}{3}$ החזר "כן".
- אחרת, "לא".

נגדיר X_i – אינדיקטור המקבל 1 אם מההרצה הi קיבלנו שרק 1 פריק.

ונגדיר $X = \Sigma_{i=1}^{t} X_i$.

נניח כי הקלט מכיל ראשוני, לכן: $P(X_i = 1) = \left(\frac{2}{3}\right)^3 + 2 \cdot \left[\left(\frac{1}{3}\right) \cdot \left(\frac{1}{3}\right) \cdot \left(\frac{2}{3}\right)\right]$
השבר הראשון הוא כאשר האלגוריתם צודק בכולם ואחרי ה+ זה כאשר הוא טועה לגבי ההבחנה של הראשוני אבל גם טועה לגבי ההבחנה של האחד מהפריקים.

לכן על מנת לקבל תשובה נכונה: $P\left(X \geq \frac{2t}{3}\right)$