תרגול 13

משתנים מקריים:

משתנה מקרי ממשי: בהינתן (Ω, Pr) מרחב הסתברות בדיד, אזי f פונקציה ממשית $f: \Omega \to R$ נקראת משתנה מקרי ממשי על מרחב ההסתברות.

תוחלת של משתנה מקרי ממשי f מסומנת ב- E(f) והיא הממוצע המשוקלל - $E[f] = \sum_{x\in\Omega} Pr(x)f(x)$.

- בהינתן (Ω, Pr) מרחב הסתברות בדיד, ו- f מ"מ שלו, $\Pr(f = a) = \sum_{\substack{x\in\Omega \\ f(x)=a}} \Pr(x)$.
- תוחלת של מ"מ ע"פ התפלגות הערכים שלו: $E(f) = \sum_{a\in R} a * \Pr(f = a)$.

ליניאריות התוחלת – יהי $(\Omega, \Pr)$ מרחב הסתברות בדיד, $f_1,\ldots,f_n$ משתנים מקריים ממשיים המוגדרים עבורו, ו- $a_0, a_1, \ldots, a_n$ קבועים, אזי $f = a_0 + \sum_{i=1}^n a_i f_i$ הוא מ"מ ממשי, ו- $E(f) = a_0 + \sum_{i=1}^n a_i E(f_i)$.

אינדיקאטור (משתנה מקרי מציין) – משתנה מקרי ממשי f שהטווח שלו כולל רק את הערכים {0,1} ייקרא משתנה אינדיקאטור. משתנה אינדיקאטור f למאורע A מקיים ש- E(f) = Pr(A).

- f,g מ"מ על אותו מרחב הסתברות ייקראו **בלתי-תלויים** אם Pr(f=a,g=b) = Pr(f=a)Pr(g=b) לכל a,b.
  - עבור 2 מ"מ בת"ל, E(fg) = E(f)E(g).

<u>שאלה 1</u>

נגדיר: $M = \{s_1, s_2, \ldots, s_k \colon\ s_i \in \{1, \ldots, i\}\} \subseteq \{1, \ldots, k\}^k$

בוחרים בהתפלגות אחידה סדרה מ- M. מהי התוחלת של סכום ספרות הסדרה הנבחרת?

<u>פתרון:</u>

נגדיר את f כמ"מ המתאים לכל סדרה במרחב המדגם את סכום הספרות שלה. מליניאריות התוחלת, $E[f]$ הינו סכום תוחלות המשתנים המקריים $f_1, \ldots, f_k$, כאשר f_i מתאים לכל סדרה את ערך האיבר ה - i שלה.

$$Pr(f_i = t) = \frac{1 \cdot 2 \cdot 3 \cdots (i-1)(i+1) \cdots k}{1 \cdot 2 \cdot 3 \cdots (i-1) \cdot (i) \cdot (i+1) \cdots k} = \frac{1}{i}$$

למעשה, בהינתן סדרה אקראית s ואינדקס i, ניתן למצוא $i-1$ סדרות אחרות ב - M שזהות ל - s בכל האיברים למעט s_i, וכל אחת מקבלת ערך אחר במקום ה - i. לכן, משיקולי סימטריה ההסתברות של הספרה ה- i לקבל ערך $1 \le t \le i$ היא $\frac{1}{i}$ לכל t.

מכאן ש- $E[f_i] = \sum_{t=1}^{i} Pr(f_i = t) \cdot t = \sum_{t=1}^{i} \frac{1}{i} t = \frac{(i+1)i}{2\cdot i} = \frac{i+1}{2}$.

לכן, $E[f] = \sum_{i=1}^{k} E[f_i] = \sum_{i=1}^{k} \frac{i+1}{2} = \frac{(2+k+1)k}{4} = \frac{k(k+3)}{4}$.

<u>הגדרה</u>: **גרף מקרי** G הינו גרף מעל קבוצת הקדקודים $\{1,2,\dots,n\}$, שההסתברות להגריל אותו מבין אוסף הגרפים מעל הקדקודים $\{1,2,\dots,n\}$ מפולגת אחיד, כלומר $Pr(G)=2^{-\binom{n}{2}}$.

<u>שאלה 2</u>

יהא G גרף מקרי עם 10 קודקודים.

א. מה ההסתברות שיש ב-G בדיוק 15 צלעות?
ב. כמה משולשים בממוצע יש ב-G ?

<u>פתרון:</u>

א. גודל מרחב המדגם הוא 2^{45}, ההתפלגות אחידה. כלומר, יש לנו 2^{45} גרפים אפשריים אותם אנו יכולים להגריל, כל אחד בהסתברות $1/2^{45}$. נגדיר את המאורע A – בגרף המקרי 15 צלעות. מספר הגרפים עם 15 צלעות הוא כמספר האפשרויות לבחור 15 צלעות מתוך 45 צלעות אפשריות – $\binom{45}{15}$.
לכן ההסתברות שבגרף שהגרלנו יש בדיוק 15 צלעות היא $\frac{\binom{45}{15}}{2^{45}}$.

ב. בגרף ישנן $\binom{10}{3}=120$ שלשות של קודקודים. נמספר אותן בין 1 ל-120. עבור כל שלשת קודקודים $1\le i\le 120$, נסמן ב-A_i, את המאורע שהשלשה ה-i יוצרת משולש. נסמן ב-f_i משתנה אינדיקטור עבור המאורע A_i.
נבחין שהמשתנה $f=f_1+f_2+\cdots+f_{120}$ מציין לנו את מספר המשולשים בגרף.
ההסתברות ששלשת קודקודים מסוימת יוצרת משולש, כלומר כל 3 הצלעות ביניהן מופיעות, היא $\frac{1}{2^3}$.
לכן, לכל i,
$$E(f_i)=\Pr(A_i)=\frac{1}{2^3}$$
נחשב את התוחלת של f:

$$E(f)=E(f_1+f_2+\cdots+f_{120})=E(f_1)+E(f_2)+\cdots+E(f_{120})=\frac{120}{2^3}=15$$

<u>שאלה 3</u>

מבצעים את הניסוי הבא:
<u>בשלב א'</u> מטילים קובייה הוגנת בה 2 פאות המסומנות ב – 1, שתי פאות מסומנות ב - 2 ושתיים ב - 3.
נסמן ב – f מ"מ המתאר את תוצאת ההטלה.
<u>בשלב ב'</u> מטילים מטבע הוגן מספר פעמים השווה לתוצאת ההטלה בשלב א'.
נסמן ב - g מ"מ המתאר את מספר הפעמים שיצא בהם עץ בהטלה.
א. מהי תוחלת $f+g$?
ב. מהי תוחלת $f\cdot g$?

<u>פתרון:</u>

א. נבחין כי f ו-g בבירור מ"מ תלויים, שכן תוצאת הראשון קובעת את מספר הטלות המטבע בשני. אולם מלינאריות התוחלת די לחשב בנפרד את התוחלת של כל אחד מהמ"מים כדי לקבל את תוחלת הסכום.
$E[f]=\frac{1}{3}\cdot 1+\frac{1}{3}\cdot 2+\frac{1}{3}\cdot 3=2$.
נותר לחשב את $E[g]$: נסמן לכל $1\le i\le 3$ את המאורע A_i - בהטלה ה - i יצא עץ. למען שלמות ההגדרה, נאמר ש - A_i לא מתקיים אם לא הוטלה מטבע בפעם ה - i. נסמן ב - g_i משתנה אינדיקאטור עבור A_i. נבחין כי תוחלת g שווה לסכום תוחלות ה - g_i-ים, לכן נשאר לחשב את $E[g_i]=Pr(A_i)$ עבור כל i. כדי שיוטל המטבע בפעם ה-i צריך תוצאה של לפחות i בקובייה, ובהנחה שזה התקיים, בהסתברות $\frac{1}{2}$ ייצא עץ. מכאן ש - $Pr(A_i)=\frac{1+(3-i)}{3}\cdot\frac{1}{2}$. לכן $E[g]\ =\ E[g_1]+E[g_2]+E[g_3]=\frac{3}{6}\ +\ \frac{2}{6}\ +\ \frac{1}{6}=1$, ובסה"כ נקבל $E[f+g]=3$.

ב. <u>פתרון ישיר</u>: נתבונן במרחב הסתברות (Ω,Pr) הבא:

$$\Omega=\{(1,\text{עץ}),(1,\text{פלי}),(2,\text{פלי},\text{עץ}),(2,\text{פלי}\ \text{פלי})(2,\text{עץ},\text{פלי}),(2,\text{עץ},\text{עץ}),$$

$$(3,\text{פלי},\text{פלי},\text{עץ}),(3,\text{פלי},\text{עץ},\text{עץ}),(3,\text{פלי},\text{עץ},\text{פלי}),(3,\text{עץ},\text{פלי},\text{עץ})$$

$$(3,\text{עץ},\text{עץ},\text{פלי}),(3,\text{עץ},\text{פלי},\text{פלי}),(3,\text{עץ},\text{עץ},\text{עץ}),(3,\text{פלי},\text{פלי},\text{פלי})\}$$

נשים לי כי ההסתברות לאיברים במרחב המדגם אינה אחידה: ההסתברות לכ"א מהשניים הראשונים היא $\frac{1}{6}$, ההסתברות לכ"א מהארבעה הבאים – $\frac{1}{12}$, ההסתברות לכ"א משמונת הבאים – $\frac{1}{24}$.
נחשב למשל את $Pr(A_2)$. ע"פ ההסתברויות למאורעות הבסיסיים שבמאורע A_2,
$Pr(A_2)=2\cdot\frac{1}{12}+4\cdot\frac{1}{24}=\frac{1}{6}+\frac{1}{6}=\frac{1}{3}$.
במקרה של המכפלה מצבנו פחות טוב – עלינו לחשב את פונקציית ההתפלגות המשותפת של שני האיברים. כדי לחשב את $E[fg]=\sum_{a,b\in R}(a*b)pr(f=a,g=b)$ ניצור את הטבלה הבאה:

	$f=1$	$f=2$	$f=3$
$g=0$	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{24}$
$g=1$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{3}{24}$
$g=2$	0	$\frac{1}{12}$	$\frac{3}{24}$
$g=3$	0	0	$\frac{1}{24}$

אם כן,

$$E[fg]=\sum_{a,b\in R}(a*b)pr(f=a,g=b)=\sum_{a\in R}\sum_{b\in R}(a*b)pr(f=a,g=b)=$$
$$\sum_{b\in R}(1*b)pr(f=1,g=b)+\sum_{b\in R}(2*b)pr(f=2,g=b)+\sum_{b\in R}(3*b)pr(f=3,g=b)=$$
$$1\cdot pr(f=1,g=1)+2\cdot pr(f=2,g=1)+4\cdot pr(f=2,g=2)+3\cdot pr(f=3,g=1)+$$
$$+6\cdot pr(f=3,g=2)+9\cdot pr(f=3,g=3)=$$
$$\frac{1}{6}+2\cdot\frac{1}{6}+4\cdot\frac{1}{12}+3\cdot\frac{3}{24}+6\cdot\frac{3}{24}+9\cdot\frac{1}{24}=\frac{4+8+8+9+18+9}{24}=\frac{56}{24}=\frac{7}{3}$$

<u>שונות ואי-שיוויונים מרקוב וצ'בישב:</u>

שונות: של משתנה מקרי ממשי f מסומנת בתור $V(f)$ והיא: $V(f)=E((f-E(f))^2)=E(f^2)-E^2(f)$

- עבור 2 מ"מ בת"ל f,g מתקיים $V(f+g)=V(f)+V(g)$.

אי שיוויון מרקוב: עבור מ"מ <u>אי-שלילי</u> f וכל $C>0$ מתקיים: $Pr(f\geq C)\leq\frac{E(f)}{C}$.

אי שיוויון צ'בישב: עבור כל מ"מ f וכל $C>0$ מתקיים: $Pr(|f-E(f)|\geq C)\leq\frac{V(f)}{C^2}$.

<u>שאלה 4</u>

נתונים לנו 5 כדורים שונים במשקלים 50 , 40 ,30 , 20 , 10 (גרם). בוחרים כדור אחד בהתפלגות אחידה. נסמן ב-f את המ"מ שמתאים עבור כל כדור את המשקל שלו.

חשבו את התוחלת ואת השונות של f.

<u>פתרון:</u>

נבחין כי בכל מקרה על מנת לחשב את השונות של מ"מ, עלינו לחשב תחילה את התוחלת. מרחב המדגם שלנו הוא 5 הכדורים, ומפני שבוחרים כל כדור בהתפלגות אחידה ההסתברות לבחירת כדור כלשהו היא $\frac{1}{5}$.

נמצא את התוחלת של f :

$$E(f) = \sum_{x\in\Omega} Pr(x)f(x) = \frac{1}{5}\cdot 10 + \frac{1}{5}\cdot 20 + \frac{1}{5}\cdot 30 + \frac{1}{5}\cdot 40 + \frac{1}{5}\cdot 50 =$$

$$= \frac{(10+50)\,5}{5\cdot 2} = 30$$

כעת נחשב את השונות של f , לפי הגדרה 1: $V(f) = E((f-E(f))^2)$

$$V(f) = E((f-E(f))^2) = E((f-30)^2) = \sum_{x\in\Omega} Pr(x)(f(x)-30)^2 = \frac{1}{5}\cdot(10-30)^2 + \frac{1}{5}\cdot(20-30)^2 + \frac{1}{5}\cdot(30-30)^2 + \frac{1}{5}\cdot(40-30)^2 + \frac{1}{5}\cdot(50-30)^2 =$$

$$\frac{1}{5}(400+100+0+100+400) = \frac{1000}{5} = 200$$

ניתן לחשב גם לפי הגדרה 2: $V(f) = E(f^2) - E^2(f)$

$$E(f^2) = \sum_{x\in\Omega} Pr(x)f^2(x) = \frac{1}{5}\cdot 10^2 + \frac{1}{5}\cdot 20^2 + \frac{1}{5}\cdot 30^2 + \frac{1}{5}\cdot 40^2 + \frac{1}{5}\cdot 50^2$$

$$= \frac{1}{5}(100+400+900+1600+2500) = \frac{5500}{5} = 1100$$

$$E^2(f) = 30^2 = 900$$

$$\Rightarrow \quad V(f) = E(f^2) - E^2(f) = 1100 - 900 = 200$$

<u>שאלה 5</u>

יהיו f_1, f_2 2 מ"מ בלתי תלויים אי-שליליים, אשר לכל אחד מהם יש תוחלת 100 ושונות 100, ונסמן: $f = f_1 + f_2$.

תנו חסם תחתון הדוק ככל האפשר להסתברות שמתקיים $f < 240$.

<u>פתרון:</u>

ראשית, נעזר באי-שוויון מרקוב:

מלינאריות התוחלת $E(f) = E(f_1+f_2) = E(f_1) + E(f_2) = 100 + 100 = 200$.

לפי אי-שוויון מרקוב $Pr(f \geq 240) \leq \frac{200}{240} = \frac{5}{6}$

ולכן $Pr(f < 240) = 1 - Pr(f \geq 240) \geq \frac{1}{6}$.

כעת ננסה לשפר את החסם בעזרת אי-שוויון צ'בישב.
מאחר ש- f_1, f_2 בת"ל מתקיים כי $V(f) = V(f_1 + f_2) = V(f_1) + V(f_2) = 100 + 100 = 200$

כעת,

$$Pr(f < 240) > Pr(160 < f < 240) = Pr(160 - 200 < f - 200 < 240 - 200) = Pr(-40 < f - E(f) < 40) = Pr(|f - E(f)| < 40) = 1 - Pr(|f - E(f)| \geq 40)$$

לפי אי שיוויון צ'ביבשב $Pr(|f - E(f)| \geq 40) \leq \frac{\mathrm{V(f)}}{40^2} = \frac{200}{1600} = 0.125$

לכן, $Pr(f < 240) > 1 - Pr(|f - E(f)| \geq 40) \geq 1 - 0.125 = 0.875$

שאלה 6

הוכיחו שלכל מספר טבעי n קיים גרף תחרות עם n קודקודים ולפחות $n!\,2^{-(n-1)}$ מסלולי המילטון.

פתרון:

ישנם $n!$ מסלולי המילטון אפשריים, כמספר האפשרויות לסדר את הקודקודים במסלול, נמספר מסלולים אלו. נגריל גרף תחרות G בהתפלגות אחידה. נבחין כי עבור 2 קוד' x, y, ההסתברות שקיימת הצלע $< x, y >$ היא חצי (בדומה עבור הצלע $< y, x >$). נגדיר את f_i להיות מ"מ שבהנתן גרף תחרות G עם n מחזיר 1 אם מסלול ההמילטון ה-i נמצא בגרף G ו- 0 אחרת, וזאת לכל $1 \leq \mathrm{i} \leq n!$.

מתקיים כי $E(f_i) = \Pr(f_i = 1) = 2^{-(n-1)}$, מאחר שבמסלול ההמילטון ה-i כל צלע תופיע בהסתברות חצי.
בנוסף, נגדיר $f = f_1 + \cdots + f_{n!}$, ונבחין כי $\mathrm{f}(G)$ הוא מספר מסלולי ההמילטון ב- G.

מלינאריות התוחלת נקבל כי מספר מסלולי ההמילטון הממוצע בגרף תחרות שהגרלנו G הוא

$$E(f) = E(f_1 + \cdots + f_{n!}) = E(f_1) + \cdots + E(f_{n!}) = n!\,2^{-(n-1)}$$

לכן בהכרח קיים גרף תחרות עם לפחות $n!\,2^{-(n-1)}$ מסלולי המילטון.

שאלה 7

הוכח כי בהינתן גרף מקרי עם הרבה קדקודים, ככל הנראה הוא לא דו חלקי.
כלומר: אם נגדיר B = המאורע בו הגרף המוגרל הוא דו חלקי, אז $lim_{n\to\infty} Pr(B) = 0$.
פתרון:
כזכור, גרף דו חלקי G מעל קבוצת קדקודים V, מגדיר חלוקה של V לשתי קבוצות: $U, W = V \backslash U$, כך שכל הצלעות ב - G מחברות בין קדקודים מ - U לקדקודים מ - W.
עבור $U \subseteq V$ נתון, נסמן ב - B_U את המאורע בו הגרף המקרי G דו חלקי, ו - U מהווה את צד שמאל (בה"כ). אם נסמן $|U| = k$, אז יש $k(n-k)$ צלעות אפשריות כאלה בגרף כזה, ולכן המאורע B_U מכיל $2^{k(n-k)}$ גרפים המפולגים אחיד. לכן $Pr(B_U) = \frac{2^{k(n-k)}}{2^{\binom{n}{2}}} = 2^{k(n-k)-\binom{n}{2}}$.
כעת נתבונן ב- $k(n-k)$ כפונקציה של k. המקסימום של פונקציה זו מתקבל בנקודה בה $k = \frac{n}{2}$, והערך בנקודה זו הוא $\frac{n^2}{4}$. קיבלנו, אם כן: $Pr(B_U) = 2^{k(n-k)-\binom{n}{2}} \leq 2^{\frac{n^2}{4}-\binom{n}{2}} = 2^{\frac{-n(n-2)}{4}}$.
לכן: $Pr(B) = Pr(\bigcup_{U\subseteq V} B_U) \leq \sum_{U\subseteq V} Pr\,(B_U) \leq 2^n \cdot 2^{\frac{-n(n-2)}{4}} = 2^{\frac{-n(n-6)}{4}} \xrightarrow[n\to\infty]{} 0$ כנדרש.