הבעיות המאתגרות בהסתברות של Mosteller

מוטי בן-ארי

http://www.weizmann.ac.il/sci-tea/benari/

גרסה 0.7.1

16 בספטמבר 2022

תוכן העניינים

מבוא

Frederick Mosteller

Frederick Mosteller ‏(2006-1916) ייסד את המחלקה לסטטיסטיקה באוניברסיטת Harvard והיה ראש המחלקה מ-1957 ועד 1971. ל-Mosteller התעניין בחינוך בסטטיסקיה וחיבר ספרי לימוד חלוציים כולל [11] שהדגיש את הגישה ההסתברותי לסטטיסטיקה, ו-[10] שהיה אחת מספרי הלימוד הראשונים בניתוח מידע. בראיון תיאר Mosteller את ההתפתחות של גישתו להוראת הסטטיסטיקה [7].

מסמך זה

מסמך זה הוא "עיבוד" לספרו של Mosteller: **חמישים בעיות מאתגרות בהסתברות ופתרונותהן** [9]. הבעיות הפתרונות מוצגות ככל האפשר בצורה נגישה לקוראים עם ידע בסיסי בהסתברות, ובעיות רבות נגישות לתלמידי תיכון ולמורים. שכתבתי אתה הבעיות והפתרונות עם חישובים מפורטים, הסברים נוספים ואיורים. לעתים כללתי פתרונות נוספים.

הבעיות שונו כדי שיהיו נגישות: פישטתי את הבעיות, חילקתי אותן לתת-בעיות והוספתי רמזים. כהעדפה אישית ניסחתי אותן מחדש בצורה מופשטת יותר מ-Mosteller ולא השתמשתי ביחידות כגון אינצ'ים ומטבעות כגון דולרים. המספור והכותרות נשארו כדי להקל על השוואה עם ספרו של Mosteller.

מחשבונים מודרניים, כולל אפליקציות לסמארטפון, מסוגלים לבצע את כל החישובים ללא קושי, ובכל זאת התשמשתי בקירוב של Stirling

בסעיף האחרון נמצא חזרה על מושגים בסיסיים בהסתברות לפי [12]. בגלל שתלמידים עלולים לא להכיר מושגים כגון משתנה אקראי ותוחלת, המושגים הללו הוסברו יותר לעומק.

הבעיות המסומנות ב-D קשות יותר. אולם גם בעיות שאינן מסומנות ב-D יכולות להיותקשה ופתרון, ולכן אל נא להתייאש אם לא תוכלו לפתור אותן. בכל זאת שווה לנסות לפתור את כולן כי כל התקדמות לקראת פתרון מעודדת.

קוד מקור

רשיון זכויות יוצרים CC-BY-SA מאפשר לקוראים להפיץ את המסמך בחופשיות ולשנות אותו כפי שמתואר מרשיון. קוד מקור ב-LaTeX ו-Python ניתן למצוא ב:

`https://github.com/motib/probability-mosteller/`

הבעת תודה

אני אסיר תודה ל-Michael Woltermann עבור הצעותיו המקיפות. Aaron M. Montegomery הסביר לי בסבלות היבטים שונים של הילוך מקרי. David Fortus העיר הערות מועילות.

סימולציות

סימולציות *Monte Carlo* (על שם קזינו מופרסם במונקו) נכתבו בשפת התכנות Python 3. תכנית מחשב מבצעת ניסוי כגון "הטלת זוג קוביות" או "הטלת מטעה" מספר רב מאוד של פעמים, מחשב ממוצעים או יחס ההצלחות להפסדים ומציג אותם. השתמשתי במחוללי מספרים אקראיים הבנויים בתוך Python, `random.random()` ו-`random.randint()`, כדי לקבל תוצאות אקראיות לכל ניסוי.

כל תכנית מריצה סימולציה המורכת מ-10000 ניסויים והתוצאות מוצגות עם ארבע ספרות לאחר הנקודה העשרונית. כמעט תמיד התוצאה לא תהיה זהה לתוצאה שמתקבלת מחישוב ההסתברות או התוחלת.

שמות הקבצים הם `N-name.py` כאשר `N` הוא מספר הבעיה ו-`name` הוא שם הבעיה באנגלית.

שתי תוצאות מוצגות (באנגלית) עבור כל סימולציה:

- התוצאה התיאורטית שהיא **הסתברות** (Probability) או **תוחלת** (Expectation). בדרך כלל במקום להעתיק את הערכים המחושבים מהמסמך, התכנית מחשבת אותם מהנוסחאות.
- תוצאת הסימולציה שהיא **היחס בין מספר ההצלחות לבין מספר הניסויים** (Proportion) שמקביל להסתברות, או **ממוצע ההצלחות** (Average) שמקביל לתוחלת.

חשוב להבין שהסתברות ותוחלת הן מושגים תיאורטיים. **חוק המספרים הגדולים** מבטיח שהתוצאות של מספר רב של ניסויים תהינה קרובות לערכים התיאורטיים, אבל הם לא יהיו זהות. למשל, ההסתברות לקבל 6 בהטלת קוביה הוגנת היא $1/6 \approx 0.1667$. בהרצת סימולציה של 10000 הטלות קיבלתי טווח של ערכים: $0.1684, 0.1693, 0.1687, 0.1665, 0.1656$.

בעיות ופתרונות

1. מגרת הגרביים (The sock drawer)

במגרה נמצאות גרביים אדומות וגרביים שחורות. אם נשלוף שתי גרביים בצורה אקראית **ללא החזרה** ההסתברות ששתיהן אדומות היא $\frac{1}{2}$.

שאלה 1: מה המספר הקטן ביותר של גרביים שחורות שיכולות להיות במגרה? עבור מספר זה מה מספר הגרביים האדומות?

שאלה 2: מה המספר **הזוגי** הקטן ביותר של גרביים שחורות שיכולות להיות במגרה? עבור מספר זה מה מספר הגרביים האדומות?

פתרון 1

פתרון 1: יהי r מספר הגרביים האדומות במגירה ויהי b מספר הגרביים השחורות. $r \geq 2$ כי נתון שניתן לשלוף שתי גרביים אדומות, ו-$b \geq 1$ אחרת ההסתברות של שליפת שתי גרביים אדומות היתה 1. נכפיל את ההסתברויות של שתי השליפות:

$$P(\text{שניים אדומים}) = \frac{r}{r+b} \cdot \frac{(r-1)}{(r-1)+b} = \frac{1}{2}\,. \tag{1}$$

נפשט ונקבל משוואה ריבועית עבור המשתנה r:

$$r^2 - r(2b+1) - (b^2 - b) = 0\,. \tag{2}$$

r, b שנייהם מספרים שלמים חיוביים ולכן הדיסקרימיננט חייב להיות ריבוע של מספר שלם:

$$(2b+1)^2 + 4(b^2 - b) = 8b^2 + 1 \tag{3}$$

הדיסקרימיננט הוא ריבוע כאשר $b = 1$ (הערך הקטן ביותר). ממשוואה 2 מתקבל $r = 3$, כאשר אנו דוחים את הפתרון $r = 0$ כי $r \geq 2$. סכום מספר הגרביים הוא 4.

בדיקה: $\frac{1}{2} = \frac{2}{3} \cdot \frac{3}{4}$.

פתרון 2: נבדוק מספרים שלמים חיוביים זוגיים עבור b כדי למצוא את הקטן ביותר עבורו הדיסקרימיננט הוא ריבוע:

b	$8b^2+1$	$\sqrt{8b^2+1}$
2	33	5.74
4	129	11.36
6	**289**	**17**

עבור $b = 6$ הערך של r הוא 15 שמתקבל על ידי פתרון משוואה 2.

בדיקה: $\frac{1}{2} = \frac{14}{20} \cdot \frac{15}{21}$.

פתרון 2

פתרון 1: האם אי-שוויון זה נכון?

$$\frac{r}{r+b} \stackrel{?}{>} \frac{r-1}{(r-1)+b}\,. \tag{4}$$

$r \geq 2, b \geq 1$ ולכן שני המכנים חיוביים וניתן להכפיל את שני הצדדים:

$$\begin{aligned} r(r-1+b) &\stackrel{?}{>} (r-1)(r+b) \\ r^2 - r + rb &\stackrel{?}{>} r^2 - r + rb - b \\ b &\stackrel{?}{>} 0\,. \end{aligned}$$

$b > 1$ כך שמשווה 4 נכונה.

לפי משוואות 1, 4:

$$\left(\frac{r}{r+b}\right)^2 = \frac{r}{r+b} \cdot \frac{r}{r+b} > \frac{r}{r+b} \cdot \frac{r-1}{(r-1)+b} = \frac{1}{2}\,, \tag{5}$$

ובאופן דומה:

$$\left(\frac{r-1}{(r-1)+b}\right)^2 = \frac{r-1}{(r-1)+b} \cdot \frac{r-1}{(r-1)+b} < \frac{r}{r+b} \cdot \frac{r-1}{(r-1)+b} = \frac{1}{2}\,. \tag{6}$$

המכנה $r+b$ שונה מאפס ולכן ניתן לחשב שורש ריבועי ולפשט את משוואה 5:

$$\begin{aligned} \frac{r}{r+b} &> \sqrt{\frac{1}{2}} \\ r &> \frac{b}{\sqrt{2}-1} \\ r &> \frac{b}{\sqrt{2}-1} \cdot \frac{\sqrt{2}+1}{\sqrt{2}+1} \\ r &> b(\sqrt{2}+1)\,. \end{aligned}$$

באופן דומה עבור משוואה 6:

$$\begin{aligned} \frac{r-1}{(r-1)+b} &< \sqrt{\frac{1}{2}} \\ r-1 &< \frac{b}{\sqrt{2}-1} \\ r-1 &< b(\sqrt{2}+1)\,. \end{aligned}$$

משתי המשוואות נקבל:

$$r-1 < (\sqrt{2}+1)b < r\,. \tag{7}$$

עבור $b=1$ מתקבל $2.141 < r < 3.141$ ו-$r=3$, $b=1$ הוא פתרון.

פתרון 2: נבדוק מספרים זוגיים עבור b:

b	$(\sqrt{2}+1)b$	$< r <$	$(\sqrt{2}+1)b+1$	r	$P($שתי אדומות$)$
2	4.8	$< r <$	5.8	5	0.4762
4	9.7	$< r <$	10.7	10	0.4945
6	14.5	$< r <$	15.5	15	0.5000

Mosteller מעיר שקיים קשר בין בעיה זו לתורת המספרים ומביא פתרון נוסף: $b=35, r=85$.

סימולציה

```
Expectation of both red  = 0.5000
Average of both red for (red =  3, black =  1) = 0.5053
Average of both red for (red = 15, black =  6) = 0.5013
Average of both red for (red = 85, black = 35) = 0.4961
```

הערה

בשני הפתרונות אנחנו לא מוכיחים תנאי מספיק עבור הערכים של r, b. בפתרון 1 פיתחנו תנאי הכרחי: לפי משוואה 3 הדיסקרימיננט חייב להיות מספר שלם, ומחפשים ערכים של b שעומדים בדרישה זו. בפתרון 2 התנאי ההכרחי הוא ש-r, b מספקים את האי-שוויונות במשוואה 7 ואז חיפשנו ערכים שעומדים בדרישה זו. כתבתי תכנית קצרה לחפש פתרונות בטווח $[1, 50]$. התוצאות עבור ערכים מסביב ל-35 הן:

r	b	$\sqrt{8b^2+1}$	$P($שתי אדומות$)$
32	78	90.52	0.500917
33	80	93.34	0.499368
34	83	96.17	0.501474
35	85	99.00	0.500000
36	87	101.83	0.498601
37	90	104.66	0.500562

הנה הפתרונות עבור $b < 10^6$:

אדומות	שחורות
3	1
15	6
85	35
493	204
2871	1189
16731	6930
97513	40391
568345	235416

2. נצחונות עוקבים (Successive wins)

אתה משחק סדרה של שלושה משחקים נגד שני יריבים ואתה זוכה בסדרה אם אתה מנצח שני משחקים לפחות מתוך השלושה. ההסתברות שאתה מנצח במשחק נגד שחקן P_1 היא p_1 וההסתברות שאתה מנצח במשחק נגד שחקן P_2 היא p_2. נתון ש-$p_1 > p2$. באיזה המתסריטים שלהן יש סיכוי גדול יותר לזכות בסדרה?

- אתה משחק נגד P_1, P_2, P_1 בסדר זה.
- אתה משחק נגד P_2, P_1, P_2 בסדר זה.

פתרון 1

אתה זוכה אם: (א) אתה מנצח בשני השחקים הראשונים ומפסיד בשלישי, (ב) אתה מפסיד את המשחק הראשון ומנצח במשחק השני ובמשחק השלישי. (ג) אתה מנצח בשלושת המשחקים.

תהי p_{121} ו-p_{212} ההסתברויות שאתה זוכה בשני סדרי המשחק:

$$\begin{aligned} p_{121} &= p_1p_2(1-p_1) + (1-p_1)p_2p_1 + p_1p_2p_1 \\ p_{212} &= p_2p_1(1-p_2) + (1-p_2)p_1p_2 + p_2p_1p_2\,. \end{aligned}$$

קיים סיכוי גדול יותר לזכות בתסריט הראשון אם $p_{121} > p_{212}$, כלומר, אם:

$$\begin{aligned} p_1p_2(1-p_1) + (1-p_1)p_2p_1 + p_1p_2p_1 &\overset{?}{>} p_2p_1(1-p_2) + (1-p_2)p_1p_2 + p_2p_1p_2 \\ -p_1 &\overset{?}{>} -p_2 \\ p_2 &\overset{?}{>} p_1\,. \end{aligned}$$

נתון ש-$p_1 > p_2$ לכן כדאי לבחור את התסריט השני.

פתרון 2

הפתרון לא-איטואיטיבי. לפי האינטואיציה, כדאי לבחור לשחק שני משחקים נגד P_1 ואחד נגד P_2 כי יש סיכוי גבוה יותר לנצח משחק נגד P_1. אולם, הדרך היחידה לנצח את הסדרה היא בנצחון ב-**משחק האמצעי**, ולכן, כדאי לשחק את המשחק האמצעי נגד P_1, כי יש סיכוי גבוה יותר לנצח אותו.

סימולציה

```
For p1 = 0.6, p2 = 0.5
Proportion of P121 wins = 0.4166
Proportion of P212 wins = 0.4473
For p1 = 0.6, p2 = 0.4
Proportion of P121 wins = 0.3300
Proportion of P212 wins = 0.3869
For p1 = 0.6, p2 = 0.2
Proportion of P121 wins = 0.1625
Proportion of P212 wins = 0.2141
```

הסבר למה סכום היחסים אינו 1.

3. המושבע קל הדעת (The flippant juror)

קיימות שתי אפשרויות להגיע להכרעה: (א) פאנל של שלושה מושבעים המורכב משני מושבעים שמקבלים החלטה בלתי-תלויה עם הסתברות של p להגיע להחלטה הנכונה, ומושבע שלישי שמגיע להחלטה נכונה בהסתברות של $1/2$. ההכרעה הנכונה מתקבלת לפי הצבעת רוב. (ב) ההכרעה מתקבלת על ידי מושבע יחיד שיש לו הסתברות של p להגיע להחלטה נכונה. באיזו אפשרות הסתברות גבוהה יותר להגיע להכרעה נכונה?

פתרון

הפאנל מגיע להכרעה נכונה אם שלושת המושבעים מגיעים להחלטה נכונה או אם כל שני מושבעים מגיעים להחלטה נכונה. ההסתברות היא:

$$P(\text{נכונה החלטה}) = \overbrace{\left(p \cdot p \cdot \frac{1}{2}\right)}^{\text{שלושה נכונים}} + \overbrace{\left(p(1-p) \cdot \frac{1}{2} + (1-p)p \cdot \frac{1}{2} + p \cdot p \cdot \frac{1}{2}\right)}^{\text{שניים נכונים מתוך שלושה}} = p\,,$$

כך שאין הבדל בין שתי האפשרויות.

Simulation

```
Prediction: probabilities of (a) and (b) are equal
For p = 0.25, proportion correct of (a) = 0.5019, (b) = 0.5046
For p = 0.50, proportion correct of (a) = 0.5072, (b) = 0.4970
For p = 0.75, proportion correct of (a) = 0.5062, (b) = 0.5040
```

4. ניסוים עד להצלחה הראשונה (Trials until first success)

מה התוחלת של מספר ההטלות של קוביה עד שמופיע 6?

פתרון 1

ההסתברות שההטלה ה-i תהיה ההופעה הראשונה של 6 היא ההסתברות שב-$i-1$ הטלות יופיע אחד מחמשת המספרים האחרים כפול ההסתברות שבהטלה ה-i יופיע 6.

נפשט את הסימון: $p = 1/6$, $P = P(\text{הופעה ראשונה של 6 בהטלה } i)$, $E = E(\text{הטלה ראשונה של 6})$. אזי:

$$P = (1-p)^{i-1}p$$
$$E = 1p(1-p)^0 + 2p(1-p)^1 + 3p(1-p)^2 + 4p(1-p)^3 + \cdots = \sum_{i=1}^{\infty} ip(1-p)^{i-1}\,.$$

ללא ה- i הסכום הוא ההסתברות של הטלה של 6 בסופי של דבר:

$$P(6 \text{ הטלה בסופו של דבר של}) = \sum_{i=1}^{\infty} p(1-p)^{i-1} = p \cdot \frac{1}{1-(1-p)} = 1\,, \tag{8}$$

שהיא לא תוצאה מפתיעה.

ניתן לחשב את התוחלת כך:

$$\begin{array}{rcccccccc} E & = & p(1-p)^0 & + & p(1-p)^1 & + & p(1-p)^2 & + & p(1-p)^3 & +\cdots \\ & & & & p(1-p)^1 & + & p(1-p)^2 & + & p(1-p)^3 & +\cdots \\ & & & & & & p(1-p)^2 & + & p(1-p)^3 & +\cdots \\ & & & & & & & & p(1-p)^3 & +\cdots \end{array}$$

השורה הראשונה היא סכום הסדרה ההנדסית ממשוואה 8 שהוא 1. השורה השנייה היא אותה סדרה הנדסית עם איבר ראשון $p(1-p)$ ולכן הסכום הוא:

$$\frac{p(1-p)}{1-(1-p)} = 1-p\,.$$

באופן דומה, סכום השורה השלישית הוא $(1-p)^2$ וסכום השורה ה-i הוא $(1-p)^{i-1}$. לכן התוחלת היא סכום הסידרה ההנדסית:

$$E = 1 + (1-p) + (1-p)^2 + (1-p)^3 + \cdots = \frac{1}{1-(1-p)} = \frac{1}{p} = 6\,.$$

פתרון 2

הכפל את משוואה 8 ב-$1-p$ והחסר את תוצאה מאותה משוואה. התוצאה היא הסדרה ההנדסית במשוואה 8:

$$\begin{array}{rcll} E & = & p(1-p)^0 & + \quad 2p(1-p)^1 + 3p(1-p)^2 + 4p(1-p)^3 + \cdots \\ E\cdot(1-p) & = & & \quad\;\; p(1-p)^1 + 2p(1-p)^2 + 3p(1-p)^3 + \cdots \\ E\cdot(1-(1-p)) & = & p & + \quad p(1-p)^1 + p(1-p)^2 + p(1-p)^3 + \cdots \\ & = & 1 & \\ E & = & 1/p = 6\,. & \end{array}$$

פתרון 3

נתייחס להטלה הראשונה בנפרד משאר ההטלות. אם בהטלה הראשונה מופיע 6 (בהסתברות p) הטלה אחת מספיקה. אחרת, אם בהטלה לא מופיע 6 (הסתברות $1-p$) אזי ההטלות הבאות מרכיבות סדרה זהה לסדרה המקורית שהתוחלת שלה היא E. לכן התוחלת היא:

$$\begin{aligned} E &= 1\cdot p + (E+1)(1-p) \\ E &= \frac{1}{p} = 6\,. \end{aligned}$$

סימולציה

```
Expectation of first success = 6
Average of first success     = 6.0161
```

5. מטבע בריבוע (Coin in a square)

שאלה 1: מטילים מטבע על רשת (ללא גבולות) של ריבועים בגודל אחיד. המטבע נוחתת על הרשת בצורה אקראית כאשר למרכז המטבע מתפלגות אחידה בתוך ריבוע.

נתון ריבוע עם צלע באורך 8 ומטבע עם רדיוס 3, מה ההסתברות שהמטבע נוחתת כולה בתוך הריבוע?

שאלה 2: בכל הטלה אתה מרוויח 5 אם המטבע נוחתת בתוך ריבוע ומפסיד 1 אם היא נוגעת בצלע של ריבוע. מה תוחלת הרווח לכל הטלה?

שאלה 3: פתח נוסחה להסתברות שהמטבע נוחתת בתוך ריבוע אם אורך הצלע הוא a ורדיוס המטבע הוא r כאשר $r < a/4$.

פתרון

פתרון 1: איור 1(א) מראה ריבוע עם אורך צלע 8 וארבעה מעגלים בקוטר 3 חסומים על ידי פינות הריבוע. מרכזי המעגלים מרכיבים ריבוע פנימי שאורך הצלע שלו הוא 2. כל מטבע שמרכזה מחוץ לריבוע תחתוך צלע של הריבוע החיצוני. למיקום של מרכז המטבע התפלגות אחידה ולכן ההסתברות שהמטבע נוחתת בתוך הריבוע היא היחס בין שטח הריבוע הפנימי לשטח הריבוע החיצוני:

$$P(\text{המטבע נוחתת בתוך הריבוע}) = \frac{2\cdot 2}{8\cdot 8} = \frac{1}{16} = 0.0625\,.$$

פתרון 2:

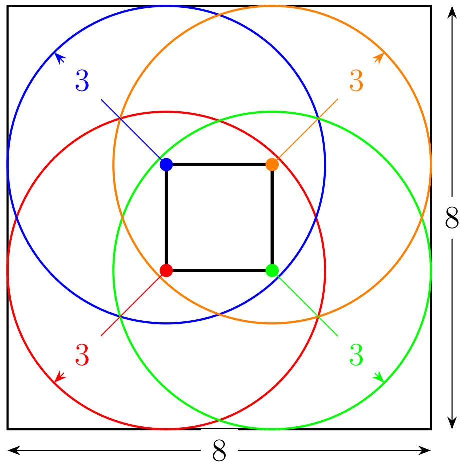
$$E(\text{הטלה לכל רווח}) = 5\cdot\frac{1}{16} + (-1)\cdot\frac{15}{16} = -\frac{10}{16} = -0.625\,.$$

פתרון 3: איור 1(ב) מראה ארבעה מעגלים חסומים על ידי פינות הריבוע. הצלע של הריבוע הפנימית הוא $a-2r$ ולכן:

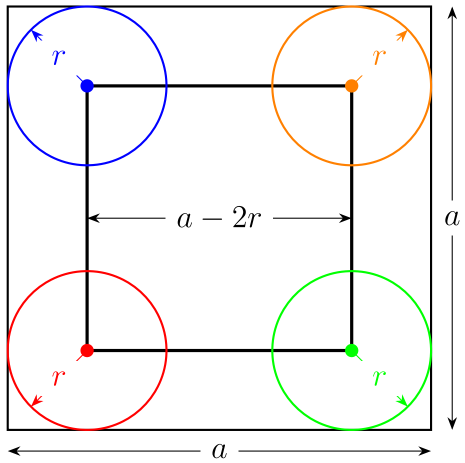
$$P(\text{המטבע נוחתת בתוך המעגל}) = \frac{(a-2r)^2}{a^2}\,.$$

סימולציה

```
For side = 8, radius = 1:
Probability of landing within the square = 0.5625
Proportion landing within the square     = 0.5704
For side = 8, radius = 2:
```



איור 1(א) מטבעות בתוך הריבוע



איור 1(ב) מטבעות בתוך ריבוע גדול

```
Probability of landing within the square = 0.2500
Proportion landing within the square     = 0.2481
For side = 8, radius = 3:
Probability of landing within the square = 0.0625
Proportion landing within the square     = 0.0639
For side = 8, radius = 4:
Probability of landing within the square = 0.0000
Proportion landing within the square     = 0.0000
```

6. הטלת מזל (Chuck-a-luck)

בחר מספר n בין 1 ל-6 והטל שלוש קוביות. אם לא מופיע n על אף קוביה אתה מפסיד 1; אם n מופיע על קוביה אחת אתה מרוויח 1; אם n מופיע על שתי קוביות אתה מרוויח 2; אם n מופיע על כל שלושת הקוביות אתה מרוויח 3. מה התוחלת של הרווח?

פתרון

תהי $P(k)$ ההסתברות ש-n מופיע על k קוביות. אזי:

$$E(\text{רווח לכל הטלה}) = -1P(0) + 1P(1) + 2P(2) + 3P(3)\,.$$

ההטלות של שלושת הקוביה הן בלתי-תלויות ולכן:

$$\begin{aligned} E(\text{רווח לכל הטלה}) = & -1\binom{3}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^3 + 1\binom{3}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^2 + \\ & 2\binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 + 3\binom{3}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^0 \end{aligned}$$

$$= \frac{1}{216}(-125 + 75 + 30 + 3)$$
$$= -\frac{17}{216} \approx -0.0787\,.$$

סימולציה

```
Expectation of winnings = -0.0787
Average winnings        = -0.0724
```

7. לרפא את המהמר הכפייתי (Curing the compulsive gambler)

רולט (roulette) הוא משחק מזל שמשחקים עם גלגל בעל 38 כיסים ממוספרים: 18 אדומים, 18 שחורים ו-2 ירוקים.[1] מסובבים את הגלגל, זורקים כדור לתוך הגלגל ומחכים שהכדור ינחת באחד הכיסים. הכדור נוחת בכיס אקראי עם התפלגות אחידה. אתה מהמר 1 שהכדור ינחת בכיס מסוים.[2] הקזינו זוכה אם הכדור נוחת בכיס ירוק, אחרת אתה מרוויח 36 אם המרת 1 על מספר הכיס בו נוחת הכדור. למעשה הרווח נטו הוא 35 כי התגמול של 36 כולל החזרה של דמי ההימור.

שאלה 1: מה תוחלת הרווח עבור 36 סבבים של משחק ברולט?

שאלה 2: חברך מציע להמר 20 שאחרי 36 סבבים אתה **תפסיד** כסף. מה תוחלת הרווח בהתחשב ברווח או הפסד גם של המשחק וגם של ההימור עם חברך?

פתרון

פתרון 1: ההסתברות של ניצחון בסבב אחד היא $1/38$ ולכן:

$$E(\text{רווח בסבב אחד}) = 35 \cdot \frac{1}{38} + (-1) \cdot \frac{37}{38} = -\frac{2}{38} \approx -0.0526$$
$$E(\text{רווח ב-36 סבבים}) = 36 \cdot -0.05266 = -1.8947\,.$$

פתרון 2: נבדוק את ארבעת התוצאות של 36 סבבים של רולט:

- אם אתה מפסיד בכל הסבבים ההפסד הוא 36.
- אם אתה זוכה בסבב אחד ומפסיד ב-36 סבבים אין רווח ואין הפסד.
- אם אתה זוכה בשני סבבים את מרוויח 70 ומפסיד 34 בשאר הסבבים כך שהרווח נטו הוא 36.
- אם אתה זוכה ב-k עבור $2 < k \leq 36$, הרווח נטו הוא $35k - (36-k) > 0$.

[1] ברולט אמריקאי נמצאים שני כיסים ירוקים וברולט אירופאי נמצא כיס ירוק אחד.
[2] זה סוג ההימור היחיד שמשתמשים בו בבעיות בספר זה.

לכן אתה מפסיד כסף רק אם אתה מפסיד את כל הסבבים:

$$\begin{aligned} P(\text{מפסיד ב-36 סבבים}) &= \left(\frac{37}{38}\right)^{36} \approx 0.3829 \\ E(\text{רווחים סה״כ}) &= \overbrace{-1.8947}^{\text{E של כל הסבבים}} + \overbrace{-20 \cdot 0.3829}^{\text{מפסיד בהימור}} + \overbrace{20 \cdot 0.6171}^{\text{זוכה בהימור}} \approx 2.7904\,. \end{aligned}$$

ברור שכדאי להסכים להימור המוצע!

סימולציה

```
Expectation of winning a round = -0.0526
Average winnings for a round   = -0.0593
```

בסימולציה היתה שונות גדולה שהוקטנה כל ידי הרצת מיליון ניסויים.

8. קלפים מושלמים בברידג׳ (Perfect bridge hand)

בחר באקראי 13 קלפים מחפיסה של 52 קלפים. מה ההסתברות שכולם מאותה סדרה?

פתרון 1

יש $\binom{52}{13}$ דרכים לבחור 13 קלפים מסדרה אחת. מתוכן יש ארבע דרכים שכולם מסדרה אחת, ולכן:

$$P(\text{בחירת 13 קלפים מאותה סדרה}) = 4 \cdot \frac{13! \cdot 39!}{52!} \approx 6.2991 \times 10^{-12}\,.$$

פתרון 2

יש 52 דרכים לבחור את הקלף הראשון. אח״כ יש 12 דרכים לבחור את הקלף השני מאותה סדרה מתוך 51 הקלפים שנשארו, 11 דרכים לבחור את הקלף השלישי, וכו׳. מכאן:

$$P(\text{בחירת 13 קלפים מאותה סדרה}) = \frac{52}{52} \cdot \frac{12}{51} \cdot \frac{11}{50} \cdots \frac{1}{40} = \frac{12!}{51!/39!} \approx 6.2991 \times 10^{-12}\,.$$

סימולציה

אין טעם להריץ סימולציה עם 52 קלפים כי ההסתברות תהיה אפס כמעט בוודאות. הרצתי סימולציה עם חפיסה של 16 קלפיםם ו-4 סדרות.

```
Probability of perfect hand = 0.0022
Proportion perfect hand     = 0.0020
```

9. משחק "קראפס" (Craps)^D

משחק ה-craps הוא משחק עם זוג קוביות. בהטלה הראשונה אתה זוכה אם סכום המספרים הוא 7 או 11, ואתה מפסיד אם הסכום הוא 2, 3 או 12. אם הסכום בהטלה הראשונה הוא $n = 4, 5, 6, 8, 9, 10$ (נקרא "נקודה" point), המשך להטיל את הקוביות עד שמופיעה הנקודה n (ניצחון) או 7 (הפסד).

שאלה 1: מה ההסתברות של המאורעות בהטלה הראשונה: ניצחון, הפסד, לא ניצחון ולא הפסד?

שאלה 2: מה ההסתברות לניצחון?

פתרון 1

פתרון 1: להסתברות של תוצאה בהטלה קוביה התפלגות אחידה וההטלות של שתי קוביות בלתי-תלויות, ולכן ההסתברות של כל תוצאה היא $1/36$. מספר הדרכים לקבל כל אחד מהמאורעות (הסכום של זוג הקוביות) הוא:

סכום	12	11	10	9	8	7	6	5	4	3	2
זוגות	1	2	3	4	5	6	5	4	3	2	1

בהטלה הראשונה יש 8 דרכים לקבל 7 או 11 וההסתברות לניצחון היא $8/36$. יש 4 דרכים לקבל $2, 3, 12$ וההסתברות להפסד היא $4/36$. ההסתברות לא לנצח ולא להפסיד בהטלה הראשונה היא:

$$1 - \frac{8}{36} - \frac{4}{36} = \frac{24}{36}\,.$$

פתרון 2: נעיין בשני מקרים:

- הנקודה היא 4. ההסתברות לנצח בהטלה השנייה (4) היא $3/36$ וההסתברות להפסיד (7) היא $6/36$. ההסתברות לא לנצח ולא להפסיד היא $1 - (3/36) - (6/36) = 27/36$.
- הנקודה היא 8. ההסתברות לנצח בהטלה השנייה (8) היא $5/36$ וההסתברות להפסיד (7) היא $6/36$. ההסתברות לא לנצח ולא להפסיד היא $1 - (5/36) - (6/36) = 25/36$.

אנו רואים שחייבים לחשב את ההסתברות לנצח בנפרד עבור כל אחת מהנקודות $4, 5, 6, 8, 9, 10$. נפתח נוסחה כללית להסתברות.

תהי P_n ההסתברות לנצח בהטלת הנקודה n בהטלה ו-Q_n ההסתברות לא לנצח ולא להפסיד בהטלה כלשהי. ניתן לחשב את W_n, ההסתברות לניצחון על ידי הטלת הנקודה n **לאחר ההטלה הראשונה**, על ידי חיבור:

- ההסתברות להופעת הנקודה בהטלה השנייה.
- ההסתברות לא לנצח ולא להפסיד בהטלה השנייה וההסתברות להופעת הנקודה בהטלה השלישית.
- ההסתברות לא לנצח ולא להפסיד בהטלה השנייה והשלישית וההסתברות להופעת הנקודה בהטלה הרביעית,
- . . .

$$
\begin{aligned}
W_n &= P_n + Q_n P_n + Q_n^2 P_n + Q_n^3 P_n + \cdots \\
&= P_n \left(1 + Q_n^1 + Q_n^2 + Q_n^3 + \cdots\right) \\
&= P_n \left(\frac{1}{1-Q_n}\right) .
\end{aligned}
$$

אתה מפסיד אם בהטלה כלשהי לאחר הראשונה מופיע 7 עם הסתברות $6/36$ ולכן:

$$
\begin{aligned}
Q_n &= (1-P_n) - (6/36) \\
W_n &= \frac{P_n}{P_n + (6/36)} .
\end{aligned}
$$

W_n עבור ששת הנקודות היא:

n	4	5	6	8	9	10
P_n	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$
W_n	$\frac{3}{9}$	$\frac{4}{10}$	$\frac{5}{11}$	$\frac{5}{11}$	$\frac{4}{10}$	$\frac{3}{9}$

נחשב את W, ההסתברות לנצח, על ידי חיבור ההסתברות לנצח בהטלה הראשונה לסכום ההסתברויות לנצח על ידי הטלת נקודה כפול ההסתברות להופעת **אותה נקודה** בהטלה הראשונה:

$$
W = \frac{8}{36} + \sum_{n \in \{4,5,6,8,9,10\}} P_n W_n \approx 0.4929 . \tag{9}
$$

שסיכוי שהקזינו יזכה במשחק אחד של craps הוא רק $0.5 - 0.4949 \approx 0.5\%$ אבל חוק המספרי הגדולים מבטיח שבסופו של דבר הם ינצחו ואתה תפסיד!

פתרון 2

פתרון 2: נעיין בסדרות ההטלות שלהן כאשר בכולן הנקודה היא 4:

$$
\begin{array}{ccccccccccc}
4 & 8 & 9 & 9 & 9 & 8 & 8 & 8 & 9 & 8 & 4 \\
4 & 8 & 9 & 9 & 9 & 8 & 8 & 8 & 9 & 8 & 7 \\
4 & 9 & 9 & 9 & 8 & 8 & 4 & & & &
\end{array}
$$

המשחק מסתיים רק אם מטילים 4 (ניצחון) או מטילים 7 (הפסד), כך שהופעות של 8 או 9 לא משפיעות על התוצאה. מכאן שהסתברות לנצח היא ההסתברות המותנית שתופיע 4 אם נתון שכבר הופיע 4 או 7. יהי f המאורע ש-4 מופיע ו-s המאורע ש-7 מופיע. אזי:

$$
P(f|f \cup s) = \frac{P(f) \cap P(f \cup s)}{P(f \cup s)} = \frac{P(f)}{P(f \cup s)} = \frac{3/36}{(3+6)/36} = \frac{3}{9} ,
$$

בדיוק הערך W_4 בטבלה לעיל. כעת ניתן להשתמש במשוואה 9 כדי לחשב את W.

השתמשנו בהסתברות בצורה סמויה בפתרון הראשון, כי W_n היא הסתברות שמותנית בהופעת הנקודה n בהטלה הראשונה.

סימולציה

```
Probability of winning = 0.4929
Proportion of wins     = 0.4948
```

13. דילמת האסיר (The prisoner's dilemma)

שלושה אסירים A, B, C מועמדים לשחרור מוקדם. וועדת השחרורים תשחרר שניים מהם כך שהאפשריות הן $\{A,B\}, \{A,C\}, \{B,C\}$ בהסתברות שווה של $1/3$ לכל זוג. לכן ההסתברות ש-A ישוחרר היא $2/3$. מפקד הכלא מוסר ל-A מידע נכון: את זהותו האסיר האחר שישוחרר B או C. אם ל-A נמסר ש-B ישוחרר, מה ההסתברות שהוא ישוחרר גם כן?

הערה: יש כאן בעיתיות מסויימת. מרגע שהוחלט על זהות האסירים שישוחררו אין כאן "הסתברות", כי הסתברות היא חישוב היחס הצפוי בין מספר הפעמים שמאורע מקיים לבין מספר הניסויים. מפקד הכלא יודע בוודאות את זהות האסירים. רצוי לדבר על ההסתברות מנקודת מבטו של A שעדיין לא יודע מה הוחלט.

פתרון 1

פתרון זה מראה חישוב נכון של ההסתברות שמבוססת על ספירה זהירה של המאורעות.

ארבעת המאורעות האפשריים הם:

e_1: ל-A נמסר ש-B ישוחרר ו-$\{A,B\}$ ישוחררו.

e_2: ל-A נמסר ש-C ישוחרר ו-$\{A,C\}$ ישוחררו.

e_3: ל-A נמסר ש-B ישוחרר ו-$\{B,C\}$ ישוחררו.

e_4: ל-A נמסר ש-C ישוחרר ו-$\{B,C\}$ ישוחררו.

ההסתברות של כל זוג להשתחרר שווה ולכן:

$$P(e_1) = P(e_2) = P(e_3 \cup e_4) = \frac{1}{3}\,.$$

אם $\{B,C\}$ ישוחררו, יימסר ל-A מידע נכון ש-B או C ישוחרר ובהסתברות שווה, ולכן $P(e_3) = P(e_4) = 1/6$. אנו מעוניינים לחשב את ההסתברות ש-A ישוחרר (מאורע e_1) בהינתן שמוסרים ל-A ש-B ישוחרר (מאורע $e_1 \cup e_3$):

$$P(e_1|e_1 \cup e_3) = \frac{P(e_1 \cap (e_1 \cup e_3))}{P(e_1 \cup e_3)} = \frac{P(e_1)}{P(e_1 \cup e_3)} = \frac{1/3}{1/3 + 1/6} = \frac{2}{3}\,.$$

פתרון 2

יהי $P(A) = P(B) = P(C) = 2/3$ ההסתברויות ש-A, B, C ישוחררו. A מעוניין בהסתברות המותנית $P(A|B)$ שהוא ישוחרר אם B ישוחרר:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{2/3} = \frac{1}{2}\,.$$

אבל זאת **לא** ההסתברות המותנית הנכונה. המידע החדש הוא לא ש-B ישוחרר אלא **שנמסר** ל-A ש-B ישוחרר. ההסתברות המותנת הנכונה היא $P(A|R_{AB})$ כאשר R_{AB} הוא המאורע של-A נמסר ש-B ישוחרר. כדי לחשב את $P(A|R_{AB})$:

$$P(A|R_{AB}) = \frac{P(A \cap R_{AB})}{P(R_{AB})}$$

עלינו לחשב את $P(A \cap R_{AB})$ ו-$P(R_{AB})$.

המידע על שחרורו של B הוא אמת ולכן:

$$P(A \cap R_{AB}) = P(\{A, B\}) = \frac{1}{3}\,.$$

אם $\{B, C\}$ ישוחררו אנו מניחים שההסתברות היא $1/2$ שמפקד יספר ל-A ש-B ישוחרר וגם ההסתברות היא $1/2$ שהמפקד יספר ש-C ישוחרר. לכן:

$$\begin{aligned}
P(R_{AB}) &= P(\{A, B\}) + \frac{1}{2} \cdot P(\{B, C\}) = \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2} \\
P(A|R_{AB}) &= \frac{P(A \cap R_{AB})}{P(R_{AB})} = \frac{1/3}{1/2} = \frac{2}{3}\,.
\end{aligned}$$

פתרון זה מדגיש את החשיבות של זיהוי נכון של המאורע שמביא את המידע החדש עבור חישוב ההסתברות המותנית.

פתרון אינטואיטיבי

בבעיית האסיר המידע ש-B ישוחרר נמסר ל-A **לאחר** שהוחלט על זהות המשוחררים. מרגע זה כלל לא משנה מה נמסר לו: אפשר לשקר לו או לא למסור לו כלום והמידע לא משנה את ההסתברות שהוא ישוחרר.

הבעיה של מונטי הול (Monty Hall Problem)

נציג כאן את הבעיה של מונטי הול (שלא מופיע בספר) כדי להראות את הקשר בינה לבין בעיית האסיר.

דיונים רבים בספרות נובעים מפרשנויות לניסוח הבעיה ולכן כדי למנוע בלבול אפרט לפרטי פרטים את כללי המשחק ואת ההנחות שלעתים נשארו סמויות:

1. בשעשועון בטלוויזיה המתחרה ניצב לפני שלוש דלתות סגורות. מאחרי כל אחת מהדלתות נמצא אחד הפרסים שהם מכונית אחת ושתי עזים.

2. המתחרה מעדיף לזכות במכונית ולא בעז![3]

[3] לא מצאתי הנחה סמויה זו כתובה בספרות אבל אם רוצים לדייק כדאי להציף אותה.

3. מיקום כל פרס הוא אקראי בהתפלגות אחידה.

4. המנחה יודע את המיקום של כל אחד מהפרסים.

5. המתחרה חייב לבחור דלת אחת.

6. לפני שמגלים למתחרה איזה פרס נמצא מאחורי הדלת שבחר, המנחה חייב לפתוח את אחת הדלתות שמאחוריהן נמצאות העזים. קיימות שתי אפשרויות:

- אם המתחרה בחר בדלת עם המכונית, הבחירה של המנחה בין שתי הדלתות האחרות היא אקראית בהתפלגות אחידה.
- אם המתחרה בחר בדלת עם עז, המנחה חייב לפתוח את הדלת עם העז השניה.

7. לאחר פתיחת הדלת המתחרה חייב להחליט אם להישאר עם הבחירתו המקורית או לשנות אותה ולבחור בדלת השניה שלא נפתחה.

8. המנחה פותח את הדלת שנבחרה והמתחרה זוכה בפרס שנמצא שם.

מה האסטרטגיה העדיפה עבור המתחרה: להישאר עם הבחירה המקורית או לשנות אותה?

לפי פתרון שכיח ההסתברות שהמכונית נמצאת מאחורי הדלת המקורית משתנה מ-$1/3$ ל-$1/2$ כי המכונית נמצאת מאחורי אחת משתי הדלתות שעדיין סגורות, ולכן לא משנה אם המתחרה בוחר להישאר עם החלטתו או לשנותה. פתרון זה **שגוי** כי אין כאן שני ניסויים בלתי-תלויים. לאחר פתיחת הדלת, המנחה לא מטיל מטבע כדי להחליט מחדש איפה לשים את המכונית ואיפה לשים את העז.

נניח ששינו את כללי המשחק כך שהמתחרה **לא** רשאי לשנות את החלטתו! ההסתברויות לא משתנות: ההסתברות לזכות במכונית נשארת $1/3$ והסתברות לזכות בעז נשארת $2/3$. מצב זה שקול לבעיית האסיר כי האסיר A לא רשאי לשנות את ההחלטה מי ישוחרר, ולכן לא משנה מה אומרים לו. עם הכללים החדשים לא משנה איזו דלת המנחה פותח, המתחרה לא יכול לעשות כלום.

בעיית מונטי הול שונה. הכלל שהמתחרה רשאי לשנות את החלטתו היא הסיבה היחידה שיש משמעות לפתיחת הדלת. שוב, ההסתברויות לא משתנות: ההסתברות שהמכונית נמצאת האחורי הדלת שנבחרה בהחלטה המקורית היא $1/3$, והסתברות שהיא נמצאת האחורי שתי הדלתות האחרות היא $2/3$. אבל כעת הסתברות $2/3$ "מרוכזת" בדלת אחת, זו שהמנחה לא פתח, ולכן, המתחרה יכול להכפיל את סיכוייו לזכות במכונית מ-$1/3$ ל-$2/3$ על ידי שינוי ההחלטה.

סימולציה

אין סימולציה לבעיית האסירים כי הבעיה שואלת אם ההסתברות משתנה לאור ידע חדש, אבל הראנו שאין שינוי. הנה תוצאות הסימולציה של בעיית מונטי הול עבור 10000 משחקים:

```
Wins when staying with original door = 0.3311
Wins when changing door              = 0.6689
```

14. איסוף תלושים (Collecting coupons)

נתון סדרה אינסופית של קופסאות ובתוך כל אחת נמצאים חמישה תלושים עם המספרים 1 עד 5. אתה שולף תלוש אחד מכל קופסה אחת אחרי השנייה.

שאלה 1: מה התוחלת של מספר התלושים שיש לשלוף עד שתקבל את כל חמשת המספרים?

שאלה 2: פתח נוסחה לתוחלת ל-n מספרים.

רמז: השתמש בפתרון לבעיה 4 (עמוד 10) והקירוב לסכום של מספרים הורמוניים (עמוד 85).

פתרון

פתרון 1: מה התוחלת של מספר השליפות עד שאתה מקבל מספר **שונה** מכל הקודמים? לפי בעייה 4 התוחלת היא $1/p$ כאשר p היא ההסתברות לשליפת מספר שונה. עבור השליפה הראשונה, ההסתברות היא 1 כך שהתוחלת היא $5/5 = 1$. עבור השליפה השנייה ההסתברות היא $4/5$ כך שהתוחלת היא $5/4$ וכך הלאה. לכן:

$$E(\text{כל חמשת המספרים}) = \frac{5}{5} + \frac{5}{4} + \frac{5}{3} + \frac{5}{2} + \frac{5}{1} = \; = \frac{1370}{120} \approx 11.4167 \,.$$

פתרון 2: נשמתמש באותה שיטה ובקירוב לסכום המספרים ההרומוניים:

$$E(\text{כל } n \text{ המספרים}) = n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots \frac{1}{2} + \frac{1}{1}\right) = nH_n \approx n\left(\ln n + \frac{1}{2n} + 0.5772\right) \,.$$

עבור $n = 5$ מתקבל:

$$E(\text{כל חמשת המספרים}) = 5H_5 \approx 5\left(\ln 5 + \frac{1}{10} + 0.5772\right) \approx 11.4332 \,.$$

סימולציה

```
For  5 coupons:
Expectation of draws = 11.9332
Average draws        = 11.4272
For 10 coupons:
Expectation of draws = 29.7979
Average draws        = 29.2929
For 20 coupons:
Expectation of draws = 72.4586
Average draws        = 72.2136
```

15. שורה בתיאטרון (The theater row)

סדר שמונה מספרים זוגיים ושבעה מספרים אי-זוגיים בשורה בצורה אקראית, למשל:

$$10 \quad 12 \quad 3 \quad 2 \quad 9 \quad 6 \quad 1 \quad 13 \quad 7 \quad 10 \quad 3 \quad 8 \quad 8 \quad 5 \quad 20\,,$$

שנוכל לכתוב כך:

$$E \quad E \quad O \quad E \quad O \quad E \quad O \quad O \quad O \quad E \quad O \quad E \quad E \quad O \quad E\,,$$

כי המספרים עצמם אינם חשובים.

מה התוחלת של מספר הזוגות השכנים שהם זוג/אי-זוגי או אי-זוגי/זוגי?

בדוגמה יש 10 זוגות שכנים שהם EO או OE.

רמז: התייחס לכל זוג שכנים בנפרד. מה ההסתברות שהם שונים?

פתרון

הטבלה שלהן מראה את עשרת הסידורים השונים עבור שלושה מספרים זוגיים ושני מספרים אי-זוגיים. מספר הזוגות השכנים השונים הוא 24 והממוצע הוא $24/10 = 2.4$.

זוגות	סידור	זוגות	סידור
3	$EEOEO$	1	$EEEOO$
4	$EOEOE$	2	$EEOOE$
2	$EOOEE$	3	$EOEEO$
2	$OEEEO$	3	$OEEOE$
1	$OOEEE$	3	$OEOEE$

נחזור לדוגמה עם 15 מספרים. תהי P_d ההסתברות שזוג נתון בסידור הוא EO או OE. אזי:

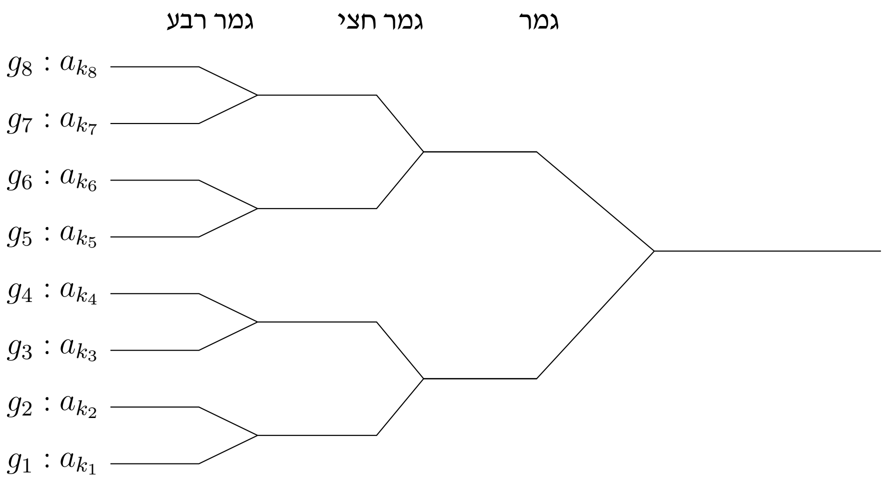
$$P_d = P(EO) + P(OE) = \frac{8}{15}\cdot\frac{7}{14} + \frac{7}{15}\cdot\frac{8}{14} = 2\cdot\frac{8}{15}\cdot\frac{7}{14} = \frac{8}{15}\,.$$

תהי E_d התוחלת של מספר הזוגות בסידור שהם EO או OE. זוג (EO, OE) תורם 1 למספר הזוגות השונים וזוג (EE, OO) תורם 0:

$$E_d = \sum_{\text{כל הזוגות}} 1\cdot P_d = 14\cdot\frac{8}{15} \approx 7.4667\,.$$

עבור עשרה מספרים:

$$\begin{aligned} P_d &= P(EO) + P(OE) = \frac{3}{5}\cdot\frac{2}{4} + \frac{2}{5}\cdot\frac{3}{4} = \frac{3}{5} \\ E_d &= 4\cdot\frac{3}{5} = \frac{12}{5} = 2.4\,. \end{aligned}$$



איור 2: טבלת משחקים לתחרות

סימולציה

```
For  5 places:
Expectation of different pairs = 2.4000
Average different pairs        = 2.3855
For 15 places:
Expectation of different pairs = 7.4667
Average different pairs        = 7.4566
For 27 places:
Expectation of different pairs = 13.4815
Average different pairs        = 13.4835
For 49 places:
Expectation of different pairs = 24.4898
Average different pairs        = 24.4725
```

16. האם השני בדירוג יזכה המקום שני? (Will the second-best be runner-up?)

לשמונה שחקים בתחרות $\{a_1, \ldots, a_8\}$ נקבעים משחקים $\{g_1, \ldots, g_8\}$ בצורה אקראית כך ששחקן a_{k_i} משחק את המשחק הראשון שלו במקום g_i (איור 2). השחקנים מדורגים כך שהטוב ביותר הוא a_1 והגרוע ביותר הוא a_8. השחקן הטוב יותר לעולם ינצח שחקן פחות טוב. ברור ששחקן a_1 ינצח בתחרות.

שאלה 1: מה ההסתברות שהשחקן a_2 יזכה במקום השני כאשר הוא משחק נגד a_1 בגמר ומפסיד?

שאלה 2: עבור 2^n שחקנים מה ההסתברות שהשחקן a_2 יזכה במקום השני כאשר הוא משחק נגד a_1 בגמר ומפסיד?

פתרון

פתרון 1: אם a_1 משחק באחד משחקים $\{g_1, g_2, g_3, g_4\}$ אף שחקן במשחקים הללו לא יגיע לגמר, לכן a_2 חייב לשחק באחד מהשחקים $\{g_5, g_6, g_7, g_8\}$. המסקנה המתבקשת היא שההסתברות ש-a_2 יזכה במקום השני היא $1/2$ כי הוא חייב לשחק באחד מארבעת המשחקים $\{g_1, g_2, g_3, g_4\}$. אולם, a_2 חייב לשחק באחד מארבעה המשחקים מתוך שבעת המשחקים ש-a_1 לא משחק בו וההסתברות היא $4/7$.

פתרון 2: באופן דומה, מתוך $2^n - 1$ המשחקים בהם a_1 לא משחק, a_2 חייב לשחק באחד מ-2^{n-1} המשחקים במחצית הטבלה שלא כוללת את a_1. מכאן:

$$P(\text{משחקים אחד נגד השני בגמר } a_1, a_2) = \frac{2^{n-1}}{2^n - 1}\,.$$

סימולציה

```
For   8 players:
Probability a2 is runner-up                = 0.5714
Proportion of games where a2 is runner-up  = 0.5707
For  32 players:
Probability a2 is runner-up                = 0.5161
Proportion of games where a2 is runner-up  = 0.5184
For 128 players:
Probability a2 is runner-up                = 0.5039
Proportion of games where a2 is runner-up  = 0.5060
```

17. אבירים תאומים (Twin knights) ^{D}

לשמונה שחקים בתחרות $\{a_1, \ldots, a_8\}$ נקבעים משחקים $\{g_1, \ldots, g_8\}$ בצורה אקראית כך ששחקן a_{k_i} משחק את המשחק הראשון שלו במקום g_i (איור 2). תהי $P(i, j)$ ההסתברות ש-a_i מנצח במשחק נגד a_j. לכל i, j, $P(i, j) = P(j, i) = 1/2$.

שאלה 1: מה ההסתברות שהשחקנים a_1, a_2 משחקים משחק אחד נגד השני.

שאלה 2: עבור 2^n שחקנים מה ההסתברות שהשחקנים a_1, a_2 משחקים משחק אחד נגד השני.

פתרון

פתרון 1: ללא הגבלת הכלליות נקבע ש-a_1 משחקת במשחק g_1. מה האפשרויות בהן a_1, a_2 ישחקו אחד נגד השני. בהסתברות $1/7$, a_2 משחק במשחק g_2. בהסתברות $2/7$, a_2 משחק במשחק g_3 או g_4, אבל הוא לא משחק נגד a_1 אלא אם **שניהם** מנצחים במשחק הראשון שלהם, כך שיש להכפיל את ההסתברות ב-$1/4$. בהסתברות $4/7$, a_2 משחק במשחק g_5, g_6, g_7, g_8, אבל הוא לא משחק נגד a_1 אלא אם **שניהם** מנצחים בשני המשחקים שלהם, כך שיש להכפיל את ההסתברות ב-$1/16$. מכאן:

$$P(\text{משחקים נגד השני } a_1, a_2) = \frac{1}{7} + \frac{1}{4} \cdot \frac{2}{7} + \frac{1}{16} \cdot \frac{4}{7} = \frac{1}{4}\,.$$

פתרון 2: תהי P_n ההסתברות שבתחרות עם 2^n שחקנים, a_1 ו-a_2 משחקים אחד נגד השני. ראינו ש-$P_3 = 1/4$. מה עם P_4? באותה שיטה:

$$\begin{aligned}P_4 &= \frac{1}{15} + \frac{1}{4}\cdot\frac{2}{15} + \frac{1}{16}\cdot\frac{4}{15} + \frac{1}{64}\cdot\frac{8}{15}\\ &= \frac{1}{15}\left(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}\right) = \frac{1}{8}\,.\end{aligned}$$

השערה סבירה היא ש-$P_n = 1/2^{n-1}$.

הוכחה 1: באותה שיטה תוך שימוש בנוסחה לסכום של סדרה הנדסית:

$$\begin{aligned}P_n &= \frac{1}{2^n-1}\sum_{i=0}^{n-1} 2^i \cdot \left(\frac{1}{2}\right)^{2i}\\ &= \frac{1}{2^n-1}\sum_{i=0}^{n-1} 2^{-i}\\ &= \frac{1}{2^n-1}\left(\frac{1-\left(\frac{1}{2}\right)^{(n-1)+1}}{1-\frac{1}{2}}\right) = \frac{1}{2^{n-1}}\,.\end{aligned}$$

הוכחה 2: באינדוקציה. טענת הבסיס היא: $P_3 = 1/4 = 1/2^{3-1}$.

יש שני צעדי אינדוקציה:

מקרה 1: נקבע ש-a_1 ו-a_2 משחקים בחצאים שונים של התחרות:

$$P(\text{משחקים בחצאים שונים } a_1, a_2) = \frac{2^{n-1}}{2^n-1}\,.$$

הם יכולים להיפגש רק במשחק הגמר ולכן כל אחד חייב לנצח בכל $n-1$ המשחקים שלו:

$$P(\text{נפגשים אם משחקים בחצאים שונים } a_1, a_2) = \frac{2^{n-1}}{2^n-1}\left(\frac{1}{2}\right)^{n-1}\left(\frac{1}{2}\right)^{n-1} = \frac{2^{-(n-1)}}{2^n-1}\,. \tag{10}$$

מקרה 2: נקבע ש-a_1 ו-a_2 משחקים באותו חצי תחרות:

$$P(a_1, a_2 \text{ משחקים בחצאים שונים}) = \frac{2^{n-1}-1}{2^n-1}\,.$$

שני השחקנים משחקים באותו חצי ולכן הבעיה היא למצוא P_{n-1}. לפי הנחת האינדוקציה:

$$P(\text{נפגשים אם משחקים באותו חצי } a_1, a_2) = \frac{2^{n-1}-1}{2^n-1}\cdot\frac{1}{2^{n-2}} = \frac{2^1-2^{-(n-2)}}{2^n-1}\,. \tag{11}$$

ממשוואות 10, 11 נקבל:

$$\begin{aligned}P_n &= \frac{2^{-(n-1)}}{2^n-1} + \frac{2^1-2^{-(n-2)}}{2^n-1}\\ &= \frac{2^{n-1}}{2^{n-1}}\cdot\frac{2^{-(n-1)}+2^1-2^{-(n-2)}}{2^n-1}\\ &= \frac{1}{2^{n-1}}\cdot\frac{2^0+2^n-2^1}{2^n-1} = \frac{1}{2^{n-1}}\,.\end{aligned}$$

סימולציה

```
For   8 players:
Probability that a1, a2 meet = 0.2500
Proportion a1, a2 meet       = 0.2475
For  32 players:
Probability that a1, a2 meet = 0.0625
Proportion a1, a2 meet       = 0.0644
For 128 players:
Probability that a1, a2 meet = 0.0156
Proportion a1, a2 meet       = 0.0137
```

18. תוצאה שווה בהטלת מטבע (An even split at coin tossing)

שאלה 1: אם אתה זורק מטבע הגון 20 פעמים מה ההסתברות לקבל "עץ" בדיוק 10 פעמים?

שאלה 2: אם אתה זורק מטבע הגון 40 פעמים מה ההסתברות לקבל "עץ" בדיוק 20 פעמים?

פתרון

פתרון 1: המטבע הוגן ולכן ההסתברות מתקבלת מהמקדם הבינומי:

$$P(\text{הטלות } 20\text{-ב "עץ" } 10) = \binom{20}{10}\left(\frac{1}{2}\right)^{10}\left(\frac{1}{2}\right)^{10} \approx 0.1762\,.$$

פתרון 2: אפשר לשער שהתוצאה תהיה זהה לשאלה הקודמת אבל:

$$P(\text{הטלות } 40\text{-ב "עץ" } 20) = \binom{40}{20}\left(\frac{1}{2}\right)^{20}\left(\frac{1}{2}\right)^{20} \approx 0.1254\,.$$

לפי חוק המספרים הגדולים מספר ההופעות של "עץ" ומספר ההופעות של "פלי" יהיו בערך שווים [12, Section 8.4], אבל הסיכוי נמוך מאוד שיהיו שווים בדיוק, והסיכוי למאורע זה הופך להיות קטן יותר ככל שמספר ההטלות גדל.

סימולציה

```
Probability of 10 heads for  20 tosses = 0.1762
Proportion  of 10 heads for  20 tosses = 0.1790
Probability of 20 heads for  40 tosses = 0.1254
Proportion  of 20 heads for  40 tosses = 0.1212
Probability of 50 heads for 100 tosses = 0.0796
Proportion  of 50 heads for 100 tosses = 0.0785
```

19. Isaac Newton **עוזר ל-** Samuel Pepys (Isaac Newton helps Samuel Pepys)

שאלה 1: מה ההסתברות לקבל **לפחות** 6 **אחד** כאשר מטילים 6 קוביות?

שאלה 2: מה ההסתברות לקבל **לפחות שני** 6 כאשר מטילים 12 קוביות?

שאלה 3: פתח נוסחה להסתברות לקבל לפחות n 6 כאשר מטילים $6n$ קוביות.

פתרון

פתרון 1: ההסתברות היא המשלים להסתברות לקבל אפס 6 ב-6 הטלות, שהיא המכפלה של לקבל מספר שונה מ-6 בכל ההטלות:

$$P(\text{לפחות 6 אחד}) = 1 - \left(\frac{5}{6}\right)^6 \approx 0.6651\,.$$

פתרון 2: ההסתברות היא המשלים להסתברות לקבל אפס או אחד 6 ב-12 הטלות:

$$P(6 \text{ לפחות שני}) = 1 - \left(\frac{5}{6}\right)^{12} - \binom{12}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{11} \approx 0.6187\,.$$

למאורע זה הסתברות קטנה יותר מהמאורע הקודם.

פתרון 3: ההסתברות היא המשלים להסתברות לקבל פחות מ-n 6 ב-$6n$ הטלות:

$$\begin{aligned} P(6\; n \text{ לפחות}) &= 1 - \binom{6n}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{6n-0} - \binom{6n}{1}\left(\frac{1}{6}\right)^1\left(\frac{5}{6}\right)^{6n-1} - \cdots \\ &= 1 - \sum_{i=0}^{n-1}\binom{6n}{i}\left(\frac{1}{6}\right)^i\left(\frac{5}{6}\right)^{6n-i}\,. \end{aligned}$$

סימולציה

```
For   6 dice to throw  1 sixes:
Probability = 0.6651
Proportion  = 0.6566
For  12 dice to throw  2 sixes:
Probability = 0.6187
Proportion  = 0.6220
For  18 dice to throw  3 sixes:
Probability = 0.5973
Proportion  = 0.5949
For  96 dice to throw 16 sixes:
Probability = 0.5424
Proportion  = 0.5425
For 360 dice to throw 60 sixes:
Probability = 0.5219
Proportion  = 0.5250
```

20. דו-קרב משולש (The three-cornered duel)^{D}

A, B, C מנהלים סדרה של דו-קרבות. לכל אחד מהם הסתברות קבועה לנצח ללא קשר לזהות היריב:

$$P(A) = 0.3, \quad P(B) = 1, \quad P(C) = 0.5\,.$$

מי שנפגע לא ממשיך להשתתף בקרבות. יורים את היריות בסדר קבוע A, B, C. אם שני יריבים עדיין עומדים, היורה יכול להחליט למי הוא יורה. אפשר להניח שכל אחד מקבל החלטה מיבטיבית נגד מי לירות.

שאלה 1: מה האסטרטגיה המיטבית של A?

שאלה 2: נניח ש-A יורה את היריה הראשונה באוויר. האם זו אסטרטגיה טובה יותר?

רמז: חשב את ההסתברויות המותנות של A לנצח בתלות בהחלטתה אל מי לירות קודם B או C.

פתרון

יהי I_X משתנה מסמן עבור הנצחות של X בסדרת קרבות:

$$I_X = \begin{cases} 1, & X \text{ מנצח בסדרה קרבות} \\ 0, & X \text{ מפסיד בסדרה קרבות}\,. \end{cases}$$

אסטרטגיה s_1 ל-X טובה יותר מאסטרטגיה s_2 אם התוחלת של I_X גדול יותר עבור s_1 מ-s_2.

סימון: $X \xrightarrow{H} Y$ מסמן ש-X יורה לעבר Y ופוגע. $X \xrightarrow{M} Y$ מסמן ש-X יורה לעבר Y ומחטיא.

פתרון 1: נחשב את ההסתברויות המותנות ש-A ינצח.

מקרה 1: A בוחר לירות את היריה הראשונה לעבר C.

אם $A \xrightarrow{M} C$ (הסתברות 0.7) אזי $C \xrightarrow{H} B$ כי C מסוכן יותר מ-A. כעת A יורה שוב לעבר B עם הסתברות 0.3 לפגוע, אבל אם A מחטיא, $A \xrightarrow{H} B$ בהסתברות 1 ו-A מפסיד.

אם $A \xrightarrow{H} C$ (הסתברות 0.3) אזי $A \xrightarrow{H} B$ 1 ו-A מפסידה.

חשב את התוחלת עם הערכים 1 כאשר A מנצחת ו-0 כאשר A מפסיד.

$$E(\text{מנצח } A | C \text{ בוחר לירות קודם ב- } A) =$$

$$\underbrace{1 \cdot (0.7 \cdot 1 \cdot 0.3)}^{A \xrightarrow{M} C, B \xrightarrow{H} C, A \xrightarrow{H} B} + \underbrace{0 \cdot (0.7 \cdot 1 \cdot 0.7 \cdot 1)}^{A \xrightarrow{M} C, B \xrightarrow{H} C, A \xrightarrow{M} B, B \xrightarrow{H} A} + \underbrace{0 \cdot (0.3 \cdot 1)}^{A \xrightarrow{M} C, B \xrightarrow{H} A} = 0.2100\,.$$

מקרה 2: A בוחר לירות את היירייה הראשונה לעבר B.

אם $B \xrightarrow{M} A$ (הסתברות 0.7), אזי כמו במקרה הקודם $C \xrightarrow{H} B$ ול-A הזדמנות אחת נוספת לפגוע ב-B (הסתברות 0.3), אחרת $A \xrightarrow{H} B$ בהסתברות 1 ו-A מפסיד.

אם $B \xrightarrow{H} A$ (הסתברות 0.3) אזי A, C יורים לסירוגין אחד לעבר השני עד שאחד נפגע. התסריטים האפשריים הם:

$$
\begin{array}{ll}
(1) & C \xrightarrow{H} A \\
(2) & C \xrightarrow{M} A \xrightarrow{H} C \\
(3) & C \xrightarrow{M} A \xrightarrow{M} C \xrightarrow{H} A \\
(4) & C \xrightarrow{M} A \xrightarrow{M} C \xrightarrow{M} A \xrightarrow{H} C \\
(5) & C \xrightarrow{M} A \xrightarrow{M} C \xrightarrow{M} A \xrightarrow{M} C \xrightarrow{H} A \\
(6) & C \xrightarrow{M} A \xrightarrow{M} C \xrightarrow{M} A \xrightarrow{M} C \xrightarrow{M} A \xrightarrow{H} C \\
 & \cdots
\end{array}
$$

ההסתברות ש-A מנצח כי הוא פוגע ב-C בסופו של דבר היא סכום ההסתברויות של התסריטים הזוגיים ברשימה:

$$
\begin{aligned}
P(A \text{ מנצח } | B \text{ ב- פוגע } A) &= (0.5 \cdot 0.3) + \\
&\quad (0.5 \cdot 0.7)(0.5 \cdot 0.3) + \\
&\quad (0.5 \cdot 0.7)(0.5 \cdot 0.7)(0.5 \cdot 0.3) + \cdots \\
&= 0.15 \sum_{i=0}^{\infty} 0.35^i = \frac{0.15}{1 - 0.35} = \frac{3}{13} \approx 0.2308 \,.
\end{aligned}
$$

באופן דומה, ההסתברות ש-C מנצח היא $\frac{0.5}{1-0.35} = \frac{1}{13} \approx 0.0760$.

התוחלת היא:

$$
E(A \text{ מנצח}) = E(A \text{ מנצח} \mid B \text{ מחטיא את } A) + E(A \text{ מנצח} \mid B \text{ פוגע ב- } A) =
$$

$$
\overbrace{1 \cdot (0.7 \cdot 1 \cdot 0.3)}^{A \xrightarrow{M} B, B \xrightarrow{H} C, A \xrightarrow{H} B} \;+\; \overbrace{0 \cdot (0.7 \cdot 1 \cdot 0.7 \cdot 1)}^{A \xrightarrow{M} B, B \xrightarrow{H} C, A \xrightarrow{M} B, B \xrightarrow{H} A} \;+\; \overbrace{1 \cdot 0.2308}^{A \xrightarrow{H} B, C \xleftrightarrow{H} *A, A \xrightarrow{H} C} \;+
$$

$$
\overbrace{0 \cdot 0.7692}^{A \xrightarrow{H} B, C \xleftrightarrow{H} *A, C \xrightarrow{H} A} \;\approx\; 0.2792 \,,
$$

שהיא גבוהה יותר מהתוחלת לנצח על ידי ירי תחילה לעבר C.

פתרון 2: אם A יורה לאוויר ולא פוגע באף יריב אזי $C \xrightarrow{H} B$ בהסתברות 1 ו-A יכול לנסות לפגוע ב-B בהסתברות 0.3. התוחלת היא:

$$
E(A \text{ מנצחת} \mid A \text{ יורה באוויר}) = 1 \cdot (0.3) + 0 \cdot (0.7) = 0.3 \,,
$$

שהיא גבוהה יותר מהתוחלת של שתי האסטרטגיות האחרות!

סימולציה

```
For A fires first at C:
Expectation of wins = 0.2100
Average wins        = 0.2138
```

```
For A fires first at B:
Expectation of wins = 0.2792
Average wins        = 0.2754
For A fires in the air:
Expectation of wins = 0.3000
Average wins        = 0.3000
```

21. לדגום עם או בלי החזרות? (Should you sample with or without replacement?) ^{D}

בכד A נמצאים 2 כדורים אדומים וכדור ירוק אחד, ובכד B נמצאים 101 כדורים אדומים ו-100 כדורים ירוקים. בחר כד אחד בצורה אקראי ושלוף שני כדורים באופן אקראי מהכד שנבחר. אתה מנצח אם אתה מזהה נכון אם כדורים נשלפו מכד A או כד B.

חשב את ההסתברויות לניצחון בכל אחד מהחוקים שלהן וקבע איזו שיטה נותן את ההסתברות הגבוהה ביותר לניצחון.

שאלה 1: אתה מחזיר את הכדור הראשון לפני השליפה השנייה.

שאלה 2: אתה לא מחזיר את הכדור הראשון לפני השליפה השנייה.

שאלה 3: לאחר שליפת הכדור הראשון אתה יכול להחליט אם להחזיר אותו או לא.

רמז: כאשר מחשבים הסתברויות:

$$\frac{101}{201} \approx \frac{100}{201} \approx \frac{100}{200} \approx \frac{1}{2}\,.$$

פתרון

יש ארבע תוצאות שנסמן RR, RG, GR, GG. לכל אחד מהחוקים נחשב את ההסתברויות המתנות של ארבעת התוצאות כתלות בזהות הכד A או B שנבחר תחילה. אחר כך נכפיל את ההסתברויות ב-$1/2$ כדי לקחת בחשבון את הבחירה האקראית של הכד.

פתרון 1: שליפה עם החזרה:

$$\begin{array}{lcccc}
P(RR|A) & = & \frac{2}{3}\cdot\frac{2}{3} & = & \frac{4}{9}\\
P(RR|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(RG|A) & = & \frac{2}{3}\cdot\frac{1}{3} & = & \frac{2}{9}\\
P(RG|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(GR|A) & = & \frac{1}{3}\cdot\frac{2}{3} & = & \frac{2}{9}\\
P(GR|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(GG|A) & = & \frac{1}{3}\cdot\frac{1}{3} & = & \frac{1}{9}\\
P(GG|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}
\end{array}$$

אם התוצאה היא RR ההסתברות שכד A נבחר (4/9) גבוהה מהסתברות שכד B נבחר $(1/4)$; אחרת, ההסתברות ש-B נבחר גבוהה יותר. לכן:

$$P(\text{ניצחון}) = \frac{1}{2}\left(\frac{4}{9}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right) = \frac{43}{72} \approx 0.5972\,.$$

פתרון 2: שליפה ללא החזרה:

$$\begin{array}{lclcl}
P(RR|A) & = & \frac{2}{3}\cdot\frac{1}{2} & = & \frac{1}{3}\\
P(RR|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(RG|A) & = & \frac{2}{3}\cdot\frac{1}{2} & = & \frac{1}{3}\\
P(RG|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(GR|A) & = & \frac{1}{3}\cdot 1 & = & \frac{1}{3}\\
P(GR|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(GG|A) & = & \frac{1}{3}\cdot 0 & = & 0\\
P(GG|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}
\end{array}$$

אם התוצאה היא GG ההסתברות שכד B נבחר גבוהה יותר (כמובן!) מההסתברות שכד A נבחר; אחרת, ההסתברות שכד A נבחר גבוהה יותר. לכן:

$$P(\text{ניצחון}) = \frac{1}{2}\left(\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\frac{1}{4}\right) = \frac{5}{8} = 0.6250\,,$$

שהיא גבוהה יותר מההסתברות לניצחון כאשר שליפה היא עם החזרה.

פתרון 3: ההחלטה נעשית על סמך התוצאות של השליפה הראשונה.

אם השליפה הראשונה היא מכד A ההסתברויות חייבות להיות מותנות גם בהחלטה להחזיר או לא. אם השליפה הראשונה היא מכד B היא לא משפיעה על ההסתברויות בגלל הקירוב ברמז.

$$\begin{array}{lclcl}
P(RR|A,w) & = & \frac{2}{3}\cdot\frac{2}{3} & = & \frac{4}{9}\\
P(RR|A,w/o) & = & \frac{2}{3}\cdot\frac{1}{2} & = & \frac{1}{3}\\
P(RR|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
P(RG|A,w) & = & \frac{2}{3}\cdot\frac{1}{3} & = & \frac{2}{9}\\
P(RG|A,w/o) & = & \frac{2}{3}\cdot\frac{1}{2} & = & \frac{1}{3}\\
P(RG|B) & = & \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4}\\
\hline
\end{array}$$

$$\begin{array}{llll} P(GR|A,w) & = \frac{1}{3}\cdot\frac{2}{3} & = & \frac{2}{9} \\ P(GR|A,w/o) & = \frac{1}{3}\cdot 1 & = & \frac{1}{3} \\ P(GR|B) & = \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4} \\ \hline P(GG|A,w) & = \frac{1}{3}\cdot\frac{1}{3} & = & \frac{1}{9} \\ P(GG|A,w/o) & = \frac{1}{3}\cdot 0 & = & 0 \\ P(GG|B) & = \frac{1}{2}\cdot\frac{1}{2} & = & \frac{1}{4} \end{array}$$

אם כדור אדום נשלף ראשונה אזי $\frac{1}{4} > \frac{4}{9}$ ו-$\frac{1}{4} < \frac{2}{9}$ עם החזרה, לעומת $\frac{1}{4} > \frac{1}{3}$ ו-$\frac{1}{4} > \frac{1}{3}$ ללא החזרה, לכן הכדור השני יכול לעזור בזיהוי הכד רק אם השליפה נעשית **עם** החזרה: אם אדום כד A ואם ירוק כד B. נבחר שליפה עם החזרה:

$$P(\text{ניצחון אם אדום ראשון}) = \frac{1}{2}\left(\frac{4}{9}+\frac{1}{4}\right) = \frac{25}{72}\,.$$

אם כדור ירוק נשלף ראשון אזי $\frac{1}{4} < \frac{2}{9}$ ו-$\frac{1}{4} < \frac{1}{9}$ עם החזרה, לעומת $\frac{1}{4} > \frac{1}{3}$ ו-$\frac{1}{4} < 0$ ללא החזרה, לכן הכדור השני יכול לעזור בזיהוי הכד רק אם השליפה נעשית **בלי** החזרה:

$$P(\text{ניצחון אם ירוק ראשון}) = \frac{1}{2}\left(\frac{1}{3}+\frac{1}{4}\right) = \frac{7}{24}\,.$$

ההסתברות לניצחון היא:

$$P(\text{ניצחון}) = \frac{25}{72}+\frac{7}{24} = \frac{23}{36} \approx 0.6389\,.$$

ההסתברות הגבוהה ביותר לניצחון מתקבלת כאשר ההחלטה לשלוף עם או בלי החזרה מתקבלת בהתאם לתוצאה של השליפה הראשונה.

סימולציה

```
With replacement:
Expectation of winning = 0.5972
Average wins           = 0.5976
Without replacement:
Expectation of winning = 0.6250
Average wins           = 0.6207
Decide after first draw:
Expectation of winning = 0.6389
Average wins           = 0.6379
```

22. הקלפי (The ballot box)

שני מועמדים A ו-B מתמודדים בבחירות. A קיבל a קולות ו-B קיבל b קולות, $a > b$. הקולות נספרים אחד-אחד וסכומי הביניים $(a_i, b_i), 1 \leq i \leq a+b$ מתעדכנים לאחר ספירת כל קול. מה ההסתברות שעבור לפחות i אחד, $a_i = b_i$?

שאלה 1: פתור עבור $a = 3, b = 2$ על ידי הכנת רשימה (a_i, b_i) עבור $1 \leq i \leq 5$.

שאלה 2: פתור לכל $a > b$.

רמז 1: מה ניתן להגיד על זהות המועמד המוביל עד לתיקו **הראשון**?

רמז 2: מה החשיבות של הקול הראשון שנספר.

פתרון

פתרון 1: מספר הדרכים לסדר את סכומי הביניים הוא $10 = \binom{5}{3} = \binom{5}{2}$ כי המיקום הקולות עבור מועמד אחד קובע את מיקום הקולות של המועמד השני. בטבלה שלהלן רשומים הסידורים האפשריים של הקולות וסכומי הביניים כאשר התיקו הראשון מודגש:

A	A	A	B	B	$(3,2)$	$(3,1)$	$(3,0)$	$(2,0)$	$(1,0)$
A	A	B	A	B	$(3,2)$	$(3,1)$	$(2,1)$	$(2,0)$	$(1,0)$
A	B	A	A	B	$(3,2)$	$(3,1)$	$(2,1)$	$(\mathbf{1},\mathbf{1})$	$(1,0)$
B	A	A	A	B	$(3,2)$	$(3,1)$	$(2,1)$	$(\mathbf{1},\mathbf{1})$	$(0,1)$
A	A	B	B	A	$(3,2)$	$(\mathbf{2},\mathbf{2})$	$(2,1)$	$(2,0)$	$(1,0)$
A	B	A	B	A	$(3,2)$	$(2,2)$	$(2,1)$	$(\mathbf{1},\mathbf{1})$	$(1,0)$
B	A	A	B	A	$(3,2)$	$(2,2)$	$(2,1)$	$(\mathbf{1},\mathbf{1})$	$(0,1)$
A	B	B	A	A	$(3,2)$	$(2,2)$	$(1,2)$	$(\mathbf{1},\mathbf{1})$	$(1,0)$
B	A	B	A	A	$(3,2)$	$(2,2)$	$(1,2)$	$(\mathbf{1},\mathbf{1})$	$(0,1)$
B	B	A	A	A	$(3,2)$	$(\mathbf{2},\mathbf{2})$	$(1,2)$	$(0,2)$	$(0,1)$

קיימים מצבי תיקו בכל הסידורים פרט לשני הראשונים ולכן:

$$P(\text{קיים תיקו עם } (3,2) \text{ קולות}) = \frac{8}{10} = \frac{4}{5}.$$

פתרון 2: נתחיל עם דיון איך לגשת לפתרון של השאלה השנייה. הנה מספר סידורים עבור $(a,b) = (3,2)$ עד לקבלת **התיקו הראשון**:

B מוביל עד תיקו	A מוביל עד תיקו
$B\ A$	$A\ B$
$B\ B\ A\ A$	$A\ A\ B\ B$

לכל סידור בו A מוביל עד לתיקו הראשון, קיים סידור שהוא תמונת ראי בו B מוביל עד לקבלת התיקו הראשון. השיקוף מתקבל על ידי החלפת כל A ב-B ולהפך. לפני התיקו הראשון אחד מהמועמדים חייב להוביל. אם הקול הראשון שנספר הוא עבור B חייב להיות תיקו בהמשך הספירה כי $a > b$.

ההסתברות שהקול הראשון היא עבור B היא:

$$P(\text{קול ראשון עבור } B) = \frac{b}{a+b}\,.$$

על ידי שיקוף המיקומים של הקולות, מספר הסידורים שמובילים לתיקו שמתחילים בקול עבור A שווה למספר הסידורים שמובילים לתיקו שמתחילים בקול עבור B. לכן:

$$P(\text{קיים תיקו}) = 2\cdot\frac{b}{a+b}\,.$$

בדיקה:

$$P(\text{קיים תיקו עם } (3,2) \text{ קולות}) = 2\cdot\frac{2}{2+3} = \frac{4}{5}\,.$$

סימולציה

```
For a =  3, b =  2:
Probability of a tie = 0.8000
Proportion of ties   = 0.8118
For a = 10, b =  8:
Probability of a tie = 0.8889
Proportion of ties   = 0.8977
For a = 20, b = 18:
Probability of a tie = 0.9474
Proportion of ties   = 0.9354
```

23. תיקו בהשוואת מטבעות (Ties in matching pennies)^D

הטל זוג מטבעות הוגנות N פעמים עבור N זוג, ורשום את מספר הפעמים שהזוגיות היא זוגי (עץ-עץ או פלי-פלי) ומספר הפעמים שהזוגיות היא אי-זוגי (עץ-פלי או פלי-עץ). מה ההסתברות לקבל תיקו (לא כולל התיקו $0-0$ בהתחלה)?

שאלה 1: פתור עבור $N=4$ על ידי רישום כל התוצאות האפשריות.

שאלה 2: פתור עבור $N=6$ על ידי פיתוח נוסחה להסתברות.

שאלה 3: הסבר מדוע ההסתברות למספר אי-זוגי $N+1$ שווה להסתברות של המספר הזוגי N.

רמז: השתמש בפתרון לבעיה 22.

פתרון

פתרון 1: נסמן את ההטלות עם זוגיות זוגי ב-E וההטלות עם זוגיות אי-זוגי ב-O. בעשרה מתוך ששה עשר סידורי ההטלות יש תיקו (מודגש):

EOOO	**EOOE**	**EOEO**	**EOEE**	**EEOO**	EEOE	EEEO	EEEE
OOOO	OOOE	OOEO	**OOEE**	**OEOO**	**OEOE**	**OEEO**	**OEEE**

פתרון 2: לפי בעיה 22:

$$P(i \text{ תיקו בהטלה}) = \begin{cases} 2i/N & i \le N/2 \text{ אם} \\ 2(N-i)/N & i \ge N/2 \text{ אם}, \end{cases} \tag{12}$$

כי בבעיית הקלפי ראינו שהערך הנמוך יותר קובע את ההסתברות.

ההסתברות ל-i זוגיים ניתן על ידי ההתפלגות הבינומית:

$$P(i \text{ זוגיים}) = \binom{N}{i}\left(\frac{1}{2}\right)^{i}\left(\frac{1}{2}\right)^{N-i} = \binom{N}{i}\left(\frac{1}{2}\right)^{N} = 2^{-N}\binom{N}{i}. \tag{13}$$

ההסתברות לתיקו היא הסכום מעל i של ההסתברות לקבל i זוגיים כפול ההסתברות לתיקו בהטלה ה-i (משוואה 12). עבור $N=6$ ותוך שימוש ב-$\binom{N}{i}=\binom{N}{N-i}$:

$$\begin{aligned} P(\text{תיקו}) &= \frac{2\cdot 2^{-6}}{6}\sum_{i=0}^{6} i\binom{i}{6} \\ &= \frac{1}{192}\left(0\cdot\binom{6}{0}+1\cdot\binom{6}{1}+2\cdot\binom{6}{2}+3\cdot\binom{6}{3}+4\cdot\binom{6}{2}+5\cdot\binom{6}{1}+6\cdot\binom{6}{0}\right) \\ &= \frac{1}{192}\left(2\cdot\binom{6}{1}+4\cdot\binom{6}{2}+3\cdot\binom{6}{3}\right) \\ &= \frac{132}{192} \approx 0.6875\,. \end{aligned}$$

פתרון 3: התיקו הראשון בהטלה ה-$N+1$ מתרחש רק עם הספירה כמעט זהות לאחר ההטלה ה-N:

$$\begin{gathered} ((N/2)-1,(N/2)+1) \\ ((N/2),(N/2)) \\ ((N/2)+1,(N/2)-1) \end{gathered}$$

אבל ללא תלות התוצאת ההטלה הסופית הערכים לא יהיו שווים.

סימולציה

```
For  4 tosses:
Probability of ties = 0.6250
Proportion of ties  = 0.6192
For  6 tosses:
Probability of ties = 0.6875
Proportion of ties  = 0.6900
For  7 tosses:
Probability of ties = 0.6875
```

```
Proportion of ties  = 0.6811
For 10 tosses:
Probability of ties = 0.7539
Proportion of ties  = 0.7559
For 20 tosses:
Probability of ties = 0.8238
Proportion of ties  = 0.8255
```

25. אורכים של מיתרים אקראיים (Lengths of random chords)

בחר מיתר אקראי במעגל היחידה. מה ההסתברות שאורכו של המיתר גדול מ-1?

כדי לפתור את הבעיה עליך להחליט איך לבחור מיתר אקראי. פתור את הבעיה עבור מהאפשרויות שלהלן:

שאלה 1: התפלגות אחידה של מרחק המיתר מהמרכז המעגל.

שאלה 2: התפלגות אחידה של הנקודה האמצעית של המיתר בתוך המעגל.

שאלה 3: התפלגות נקודות הקצה של המיתר על היקף המעגל.

פתרון

פתרון 1: מיתר ארוך יותר מהרדיוס אם הוא קרוב יותר למרכז ממיתר באורך 1. יהי $\overline{AB}$ מיתר באורך 1 ובנה גובה $h = \overline{OH}$ מהמרכז O אל המיתר (איור 3(א)). בגלל ש-$\triangle AOB$ שווה-צלעות $\triangle OHB$ הוא משולש ישר-זווית:

$$h = \sin \frac{\pi}{3} = \frac{\sqrt{3}}{2}\,.$$

יהי d המרחק של המיתר $\overline{DE}$ מהמרכז. לפי ההנחה ההתפלגות של d אחידה ב-$(0, 1)$ ולכן:

$$P(\overline{DE} > 1) = P(d < h) = \frac{h}{1} = \frac{\sqrt{3}}{2} \approx 0.8660\,.$$

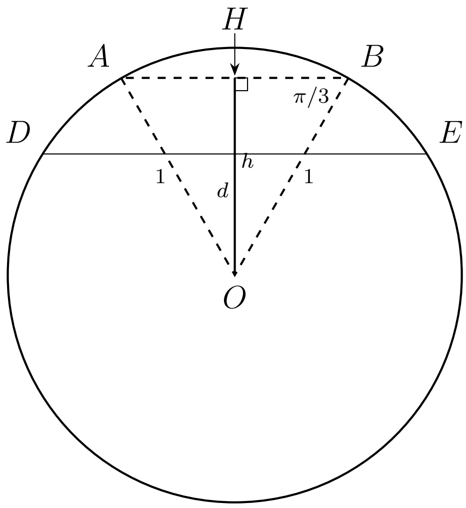
פתרון 2: בנה מעגל עם מרכז O ורדיוס h כאשר h הוא אורכו של גובה לגובה למיתר באורך 1. משיק לכל נקודה על המעגל יהיה מיתר $\overline{FE}$ שאורכו 1. אורכו של כל מיתר $\overline{EG}$ שנקודת האמצע שלו נמצאת בתוך המעגל יהיה גדול מ-1 (איור 3(ב)), ולכן ההסתברות שאורכו של מיתר גדול מ-1 היא היחס בין השטחים של שני המעגלים:

$$P(\overline{EG} > 1) = \frac{\pi \cdot h^2}{\pi \cdot 1^2} = h^2 = \frac{3}{4}\,.$$

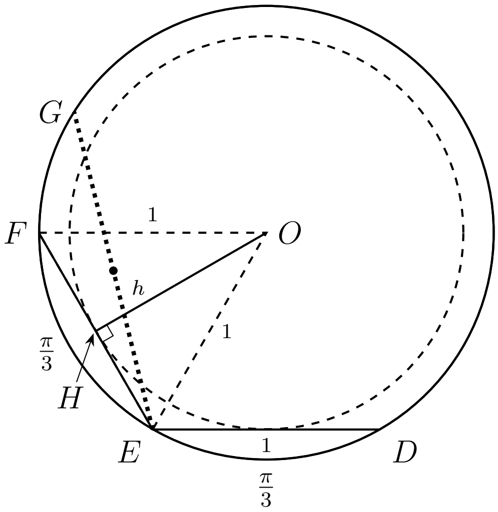
הסתברות זו היא הריבוע של ההסתברות שחישבנו בשאלה הקודמת.

פתרון 3: בנקודה שרירותית על ההיקף של מעגל היחידה (E באיור 3(ב)). כל נקודה אחרת על ההיקף (כגון G באיור) מגדירה מיתר שאורכו גדול מאחד אלא אם הנקודה שנבחרה נופל על הקשתות $\widehat{EF}$ או $\widehat{ED}$. לכן ההסתברות היא היחס בין הקשת $\widehat{FD}$ להיקף המעגל:

$$P(\overline{EG} > 1) = \frac{(2\pi - (2\pi/3)) \cdot 1}{2\pi \cdot 1} = \frac{2}{3}\,.$$



איור 3(א) מרחק של מיתר מהמרכז בהתפלגות אחידה ב-$(0,1)$



איור 3(ב) נקודת האמצע של מיתר בהתפלגות אחידה במעגל ונקודות הקצה של המיתר בהתפלגות אחידה בהיקף

סימולציה

הסימולציה היא עבור בחירת שתי נקודות על ההיקף.

```
Probability of long chords = 0.6667
Proportion of long chords  = 0.6627
```

26. הממהרים לדו-קרב (The hurried duelers)

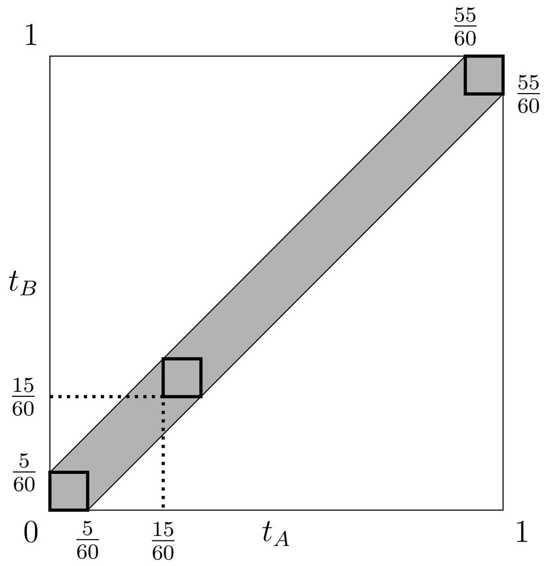
A ו-B מגיעים לנקודת מפגש בזמן אקראי עם התלפגות אחידה בתוך פרק זמן של שעה. אם A מגיע קודם ו-B לא מגיע במשך 5 דקות, A עוזב. באופן דומה אם B מגיע קודם ו-A לא מגיע במשך 5 דקות, B עוזב. מה ההסתברות שהם יפגשו?

בפרק הזמן של שעה הזמן הוא **רציף** בתחום $[0,1]$, כלומר, אי-אפשר לספור מספר בדיד של דקות או שניות כדי לחשב הסתברויות. כן ניתן לחשב הסתברויות של פרקי זמן.

רמז: צייר גרף כאשר ציר ה-x הוא זמן הגעתו של A וציר ה-y הוא זמן הגעתו של B.

פתרון

ללא הגבלת הכלליות הנח ש-A מגיע קודם. אם A מגיע ב-$t_A = 0$ ואם B מגיע לפני $t_B = 5/60$ הם נפגשים, אחרת הם לא נפגשים. מצב זה מוצג באיור 4 על ידי ריבוע קטן בראשית הצירים. אם A מגיע יותר מאוחר אזי גם B חייב להגיע באותו איחור. למשל, אם A מגיע ב-$t_A = 15$, B חייב להגיע בין $t_B = 15$ ו-$t_B = 20$. לכן הפגישה תתקיים בריבוע המתקבל על ידי הזזת הריבוע ב-15 מ-$(0,0)$ ל-$(15/60, 15/60)$.



איור 4: זמנים המבטיחים מפגש בין A ל-B

ההסתברות לפגישה היא היחס בין השטח האפור בגרף לשטחו של הריבוע הגדול. קל יותר לחשב את המשלים שהוא היחס בין שטח המשולשים הלבנים לשטחו של הריבוע הגדול:

$$\begin{aligned} P(\text{נפגשים } A, B) &= 1 - P(\text{לא נפגשים } A, B) \\ &= 1 - 2 \cdot \left(\frac{1}{2} \cdot \frac{55}{60} \cdot \frac{55}{60}\right) = \frac{23}{144} \approx 0.1597\,. \end{aligned}$$

סימולציה

```
Probability of meeting    = 0.1597
Proportion of meetings    = 0.1549
```

27. לתפוס את הזייפן הזהיר (Catching the cautious counterfeiter)

נתון n קופסאות ובכל אחת n מטבעות כאשר מטבע אחד בכל קופסה מזויף. שלוף מטבע אחת מכל קופסה ובדוק אם היא מזויפת או אמיתית. מה ההסתברות שכל המטבעות שנשלפות מזויפות?

שאלה 1: פתור עבור $n = 10$.

שאלה 2: פתור עבור $n = 100$.

שאלה 3: פתח נוסחה עבור ההסתברות כאשר n שרירותי.

שאלה 4: פתח נוסחה עבור ההסתברות כאשר n שואב לאיסוף.

פתרון

השליפות בלתי תלויות ולכן ההסתברות שכל המטבעות אמיתיות היא מכפלת ההסתברות עבור שליפה אחת.

פתרון 1:

$$P(\text{כל המטבעות אמיתיים}) = \left(\frac{9}{10}\right)^{10} \approx 0.3487\,.$$

פתרון 2:

$$P(\text{כל המטבעות אמיתיים}) = \left(\frac{99}{100}\right)^{100} \approx 0.3660\,.$$

פתרון 3:

$$P(\text{כל המטבעות אמיתיים}) = \left(\frac{n-1}{n}\right)^{n}\,.$$

פתרון 4:

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n} = \frac{1}{e}\,. \tag{14}$$

ניתן להוכיח את הגבול באמצעות חשבון דיפרנציאלי. תחילה ניתן לחשב את הגבול של הלוגריתם של הצד השמולי של משוואה 14:

$$\lim_{n\to\infty}\ln\left(1-\frac{1}{n}\right)^{n} = \lim_{n\to\infty} n\ln\left(1-\frac{1}{n}\right) = \lim_{n\to\infty}\frac{\ln\left(1-\frac{1}{n}\right)}{1/n}\,.$$

אם נחשב את הגבול נקבל $(\ln\ 1)/0 = 0/0$ אבל לפי חוק l'Hôpital נוכל להחליף את הביטוי בחילוק הנגזרות:

$$\lim_{n\to\infty}\ln\left(1-\frac{1}{n}\right)^{n} = \lim_{n\to\infty}\frac{\left(1-\frac{1}{n}\right)^{-1}(-(-n^{-2}))}{-n^{-2}} = -1$$
$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n} = e^{-1} \approx 0.3679\,.$$

סימולציה

```
For  10 boxes:
Probability of all real = 0.3487
Proportion all real     = 0.3480
For 100 boxes:
Probability of all real = 0.3660
Proportion all real     = 0.3730
For 200 boxes:
Probability of all real = 0.3670
Proportion all real     = 0.3690
```

28. לתפוס את הזייפן החמדן (Catching the greedy counterfeiter)

נתון n קופסאות ובכל אחת n מטבעות מהן m מזוייפות. שלוף מטבע אחת מכל קופסה ובדוק אם היא מזויפת או אמיתית. מה ההסתברות $P(n,m,r)$ ש-r מתוך המטבעות מזוייפות?

שאלה 1: פתח נוסחה עבור $P(n,m,r)$.

שאלה 2: חשב $P(20,10,2), P(20,10,8), P(20,5,2), P(20,5,4)$.

פתרון

פתרון 1: יש $\binom{n}{r}$ תת-קבוצות של קבוצת הקופסאות מהן המטבעות המזוייפות נשלפו. לפי ההתפלגות הבינומית:

$$P(n,m,r) = \binom{n}{r}\left(\frac{m}{n}\right)^r\left(\frac{n-m}{n}\right)^{n-r} .$$

פתרון 2:

$$\begin{aligned}
P(20,10,2) &= \binom{20}{2}\left(\frac{10}{20}\right)^2\left(\frac{10}{20}\right)^{18} \approx 0.0002\\
P(20,10,8) &= \binom{20}{8}\left(\frac{10}{20}\right)^8\left(\frac{10}{20}\right)^{12} \approx 0.1201\\
P(20,5,2) &= \binom{20}{2}\left(\frac{5}{20}\right)^2\left(\frac{15}{20}\right)^{18} \approx 0.0669\\
P(20,5,4) &= \binom{20}{4}\left(\frac{5}{20}\right)^4\left(\frac{15}{20}\right)^{12} \approx 0.1952\,.
\end{aligned}$$

סימולציה

```
For 10 bad coins,  2 draws:
Probability of counterfeit  = 0.0002
Proportion counterfeit      = 0.0002
For 10 bad coins,  8 draws:
Probability of counterfeit  = 0.1201
Proportion counterfeit      = 0.1181
For  5 bad coins,  2 draws:
Probability of counterfeit  = 0.0669
Proportion counterfeit      = 0.0688
For  5 bad coins,  4 draws:
Probability of counterfeit  = 0.1897
Proportion counterfeit      = 0.1905
```

Mosteller משתמש בגבול מבעיה 27 כדי להראות שעבור m, r נתון, כאשר n שואף לאינסוף:

$$\lim_{n\to\infty} P(n,m,r) = \frac{e^{-m}m^r}{r!}\,. \tag{15}$$

הנה השוואה של ההסתברויות עבור $m = 10, r = 8$ עבור ערכים עולים של n:

```
Limit = 0.11259903, binomial = 0.11482303, n = 100
Limit = 0.11259903, binomial = 0.11282407, n = 1000
Limit = 0.11259903, binomial = 0.11262155, n = 10000
Limit = 0.11259903, binomial = 0.11259926, n = 1000000
```

29. עובש בג'לטין (Moldy gelatin)

נתון לוח מלבני שמחולק ל-n משבצות ריבועיות קטנות. בכל משבצת יש r חיידקים בממוצע.

שאלה 1: פתח נוסחה להסתברות שיש בדיוק r חיידקים ב-n המשבצות.

שאלה 2: חשב את ההסתברות עבור $n = 100, r = 3$.

רמז: בעיה זו דומה לבעיה 28.

פתרון

פתרון 1: p, ההסתברות שבמשבצת אחת נמצא חידק (נתעלם מהאפשרות שחיידק אחד מוכל באופן חלקי בשתי משבצות או יותר), היא m/n. $P(n, m, r)$, ההסתברות שיש בדיוק r חיידקים ב-n משבצות' ניתנת על ידי ההתפלגות הבינומית:

$$P(n,m,r) = \binom{n}{r}\left(\frac{m}{n}\right)^r \left(\frac{n-m}{n}\right)^{n-r}\,.$$

פתרון 2:

$$P(10,3,3) = \binom{100}{3}\left(\frac{3}{100}\right)^3 \left(\frac{97}{100}\right)^{97} \approx 0.2275\,.$$

סימולציה

```
For  20 squares:
Probability of exactly  3 microbes  = 0.2428
Proportion of exactly   3 microbes  = 0.2436
Probability of exactly  5 microbes  = 0.2023
Proportion of exactly   5 microbes  = 0.1954
For 100 squares:
Probability of exactly  3 microbes  = 0.2275
Proportion of exactly   3 microbes  = 0.2247
Probability of exactly  5 microbes  = 0.1800
Proportion of exactly   5 microbes  = 0.1851
```

משוואה 15 מתאים גם כאן לחשב את הגבול כאשר n שואף לאינסוף:

$$\begin{aligned}\lim_{n\to\infty} P(n,m,r) &= \frac{e^{-m}m^r}{r!}\\ \lim_{n\to\infty} P(n,3,3) &= \frac{e^{-3}\cdot 3^3}{3!} \approx 0.2240\\ \lim_{n\to\infty} P(n,5,5) &= \frac{e^{-5}\cdot 5^5}{5!} \approx 0.1755\,.\end{aligned}$$

31. ימי הולדת זהים (Birthday pairings)

שאלה 1: עבור n אנשים מה ההסתברות $P(n)$ שלניים מהם או יותר יש יום הולדת זהה?

שאלה 2: מה מערך הקטן ביותר עבור n כך ש-$P(n) > 0.5$?

הנח שההתפלגות של ימי ההולדת היא אחידה בטווח $[1, 365]$.

רמז: חשב את המשלים להסתברות של-n אנשים ימי הולדת **שונים**.

פתרון

פתרון 1: בחר n אנשים אחד אחרי השני ובדוק אם יש להם אותו יום הולדת כמו הקודמים שנבחרו. עבור הראשון יש 365 ימים, עבור השני 364 ימים וכך הלאה:

$$1 - P(n) = \overbrace{\frac{365}{365}\cdot\frac{364}{365}\cdot\;\cdots\;\cdot\frac{365-(n-2)}{365}\cdot\frac{365-(n-1)}{365}}^{n} = \frac{365!/(365-n)!}{365^n}\,.$$

פתרון 2: חשב ערך זה עבור ערכים שונים של n או השתמש בתכנית מחשב כדי לחפש מ-1 ל-365 כדי למצוא את הערך הראשון עבורו $1 - P(n) < 0.5$. הערך הוא 23:

$$1 - P(23) = \frac{365!}{365^{23}\cdot 342!} \approx 0.4927\,.$$

רוב האנשים מופתעים שהתשובה היא ערך כל קטן.

מחשבון מודרני מסוגל לחשב $1 - P(n)$ אבל שווה לחשב אותו באמצעות הקירוב של Stirling שהוא $\ln n! \approx n\ln n - n$:

$$\begin{aligned}\ln(1-P(n)) &= \ln\left(\frac{365!}{342!\cdot 365^{23}}\right) = \ln 365! - \ln 342! - 23\ln 365\\ &\approx (365\ln 365 - 365) - (342\ln 342 - 342) - 23\;\ln 365\\ &\approx -0.7404\\ 1 - P(n) &\approx e^{-0.7404} = 0.4769\,.\end{aligned}$$

הקורא מוזמן לחשב את ההסתברות עם הקירוב המדוייק יותר:

$$\ln n! \approx n\ln n - n + \frac{1}{6}\left(8n^3 + 4n^2 + n + \frac{1}{30}\right) + \frac{1}{2}\ln\pi\,.$$

סימולציה

```
For 21 people:
Expectation of no pairs = 0.5563
Average no pairs        = 0.5497
For 22 people:
Expectation of no pairs = 0.5243
Average no pairs        = 0.5237
For 23 people:
Expectation of no pairs = 0.4927
Average no pairs        = 0.4933
For 24 people:
Expectation of no pairs = 0.4617
Average no pairs        = 0.4576
For 25 people:
Expectation of no pairs = 0.4313
Average no pairs        = 0.4345
```

32. למצוא עמית ליום הולדת (Finding your birthmate)

עמית ליום הולדת, בקיצור עמית, הוא אדם עם יום הולדת זהה לשלך.

מדוע מציאת עמית היא בעיה שונה ממציאת זוג עם ימי הולדת זהים?

שאלה 1: כמה אנשים עליך לשאול עד ש-$Q(n)$, ההסתברות למציאת עמית, גבוהה מ-0.5?

שאלה 2: פתור את הבעיה על ידי שימוש במשוואה 14.

פתרון

להרבה אנשים יכול להיות יום הולדת זהה שנחשב כהצלחה במציאת זוג, אבל לא הצלחה במציאת עמית אלא אם יום ההולדת של האדם האחר זהה לשלך.

פתרון 1: נמצא את n, המספר הקטן ביותר של אנשים, כך ש-$1 - Q(n)$, ההסתברות שלאף אחד מהם הוא לא עמית, פחות מ-0.5. ההסתברות שהאדם הראשון שאתה שואל אינו עמית היא $364/365$, אבל זאת גם ההסתברות שהשני, השלילי, ..., אינו עמית. הפתרון הוא ה-k הקטן ביותר כך ש:

$$1 - Q(n) = \left(\frac{364}{365}\right)^n < \frac{1}{2}\,.$$

בחיפוש עם מחשב נמצא ש-$n = 253$:

$$\left(\frac{364}{365}\right)^{253} \approx 0.4995\,.$$

פתרון 2: משוואה 14 היא:

$$\lim_{n\to\infty}\left(\frac{n-1}{n}\right)^n = \frac{1}{e}\,,$$

וניתן להשתמש בה לחשב את ההסתברות:

$$\begin{aligned}1-Q(n) &= \left(\frac{365-1}{365}\right)^n = \left[\left(\frac{364}{365}\right)^{365}\right]^{n/365}\\ &\approx e^{-n/365}\\ e^{-253/365} &\approx 0.5000\,.\end{aligned}$$

סימולציה

```
For 251 people:
Probability of no match = 0.5023
Proportion no match     = 0.5120
For 252 people:
Probability of no match = 0.5009
Proportion no match     = 0.5055
For 253 people:
Probability of no match = 0.4995
Proportion no match     = 0.4984
For 254 people:
Probability of no match = 0.4982
Proportion no match     = 0.4987
For 255 people:
Probability of no match = 0.4968
Proportion no match     = 0.5078
```

33. השוואת הבעיית יום הולדת זהה לבעיית עמית ליום הולדת

(Relating the birthday pairings and the birthmate problems)

סמן ב-$P(r)$ את ההסתברות למצוא זוג שלהם ימי הולדת זהים מתוך r אנשים (בעיה 31), וב-$Q(n)$ את ההסתברות שמתוך n אנשים לפחות אחד הוא עימית שלך (בעיה 32). נתון r עבור איזה n, $P(r) \approx Q(n)$?

פתרון 1

הפתרון מבוסס על [8].

מהפתרון לבעיית 31 מתקבל:

$$\begin{aligned}
P(r) &= \frac{365}{365}\cdot\frac{365-1}{365}\cdot\frac{365-2}{365}\cdot\ \cdots\ \cdot\frac{365-(r-1)}{365}\\
&= 1\left(1-\frac{1}{365}\right)\left(1-\frac{2}{365}\right)\cdot\ \cdots\ \cdot\left(1-\frac{r-1}{365}\right)\\
&\approx 1-\frac{1}{365}-\frac{2}{365}-\cdots-\frac{r-1}{365}\\
&= 1-\frac{1+2+3+\cdots+(r-1)}{365}\\
&= 1-\frac{r(r-1)/2}{365},
\end{aligned}$$

כאשר הקירוב במשוואה השלישית מתקבל מהשמטת חזקות של $1/365$ גדולות מאחת כי הן קטנות מדי להשפיע באופן מהותי על התוצאה.

באמצעות באותו קירוב, מהפתרון לבעיה 32 מתקבל:

$$\begin{aligned}
P_{\text{אין עמית}}(n) &= \overbrace{\left(1-\frac{1}{365}\right)\left(1-\frac{1}{365}\right)\cdots\left(1-\frac{1}{365}\right)}^{n}\\
&\approx 1-\overbrace{\frac{1}{365}-\frac{1}{365}\cdots-\frac{1}{365}}^{n}\\
&\approx 1-\frac{n}{365}
\end{aligned}$$

לכן $P(r)\approx Q(n)$ כאשר:

$$n=\frac{r(r-1)}{2}.$$

עבור $r=23$, $n=(23\cdot 22)/2=253$.

פתרון 2

Mosteller מביא את הפתרון האיטואיטיבי שלהלן:

> כאשר משווים בין בעיית הזוג והעמית, אנו שמים לב שעבור r אנשים בבעיית הזוג, קיימים $r(r-1)/2$ זוגות או **הזדמנויות** לידי הולדת זהים; לעומת זאת, אם שואלים n אנשים בבעיית העמית קיימות רק n הזדמנויות כדי שאמצא עמית אחד או יותר [8, p. 322].

מכאן הוא מסיק ש-$n\approx r(r-1)/2$.

ניתן להבין את הטיעון כך: בבעיית הזוג בחר תאריך שרירותי ושאל אם לשניים מתוך r **תאריך זה** הוא יום ההולדת שלהם. יש:

$$\binom{r}{2}=\frac{r!}{2!(r-2)!}=\frac{r(r-1)}{2}$$

דרכים שזה אפשרי. עבור בעיית העמית יום ההולדת שלך נתון, ולכל אחד מתוך n אנשים יש אותו סיכוי ליום הולדת זהה. על ידי השוואת שני הביטויים נקבל n עבורו $P(r) \approx Q(n)$.

סימולציה

תוכל להריץ את הסימולציות לבעיות $31, 32$ ולבדוק תוצאה זו.

34. חופש בימי הולדת (Birthday holidays)^{D}

בית חרושת נסגר בכל יום שהוא יום הולדת של אחד העובדים. אין חופשות נוספות.

שאלה 1: כמה עובדים כדאי להעסיק כדי למקסם את מספר ימי-העבודה בשנה אחת?

שאלה 2: מה התוחלת של היחס בין מספר ימי-העבודה המקסימליים לבין 365^2, מספר ימי-העבודה עם כל אחד מ-365 העובדים עובדים כל יום?

רמז: הוכח על ידי בדיקת מקרי הקצה שמקסימום חייב להתקיים. אחר כך פתח נוסחה של התוחלת של ימי-העבודה ביום אחד.

פתרון

פתרון 1: אם יש רק עובד אחד יהיו 364 ימי-עבודה. אם יש שני עובדים מספר ימי-העבודה הוא $363 + 363 = 726$ (כאשר נתעלם המאפשרות הזניחה שלשניהם אותו יום הולדת). בקצה השני אם יש מיליון עובדים מספר ימי-העבודה יהיה אפס כמעט בוודאות. אם מספר ימי-העבודה עולה ואחר כך חוזר לאפס, חייב להיות מקסימום בין נקודות הקצה.

כדי לפשט את הסימונים נסמן את המספר הימים בשנה ב-N ומספר העובדים ב-n.

יהי $p = \dfrac{N-1}{N} = 1 - \dfrac{1}{N}$. ההסתברות שיום נתון הוא יום-עבודה היא ההסתברות שלכל עובד יום הולדת בתאריך אחר:

$$\overbrace{\frac{N-1}{N} \cdot \;\cdots\; \cdot \frac{N-1}{N}}^{n} = \left(1 - \frac{1}{N}\right)^n = p^n \,.$$

ביום עבודה כל העובדים עובדים וביום חופש אף אחד לא עובד ולכן:

$$E(\text{ימי-עבודה ליום נתון}) = n \cdot p^n + 0 \cdot (1 - p^n) = np^n \,.$$

לכל ימי השנה אותה תוחלת ולכן:

$$E(\text{ימי-עבודה לשנה}) = Nnp^n \,. \qquad (16)$$

כדי למצוא את המקסימום נגזור את משוואה 16 ביחס ל-n ונשמתש ב-$(p^n)' = p^n \ln p$ ניתן להוכיח בעזרת חוק השרשרת:

$$(p^n)' = ((e^{\ln p})^n)' = (e^{n \ln p})' = e^{n \ln p}(n \ln p)' = (e^{\ln p})^n \ln p = p^n \ln p \,.$$

הנגזרת של משוואה 16 היא:

$$(Nnp^n)' = N(p^n + n(p^n)') = N(p^n + np^n \ln p)\,,$$

שהיא 0 כאשר:

$$n = -\frac{1}{\ln p}\,.$$

עבור $N = 365$ מתקבל $n = 364.5$ אבל n הוא מספר שלם ולכן המקסימום מתקבל ב-$n = 364$ או $n = 365$ שנותנים אותו תוחלת של ימי-עבודה:

$$\begin{aligned} E(\text{ימי-עבודה לשנה}) &= Nnp^n \\ &= 365 \cdot 364 \cdot \left(\frac{364}{365}\right)^{364} \\ &= 365 \cdot 364 \cdot \frac{365}{365}\left(\frac{364}{365}\right)^{364} \\ &= 365 \cdot 365 \cdot \left(\frac{364}{365}\right)^{365} \\ &= 48944\,. \end{aligned}$$

פתרון 2: התוחלת של היחס היא:

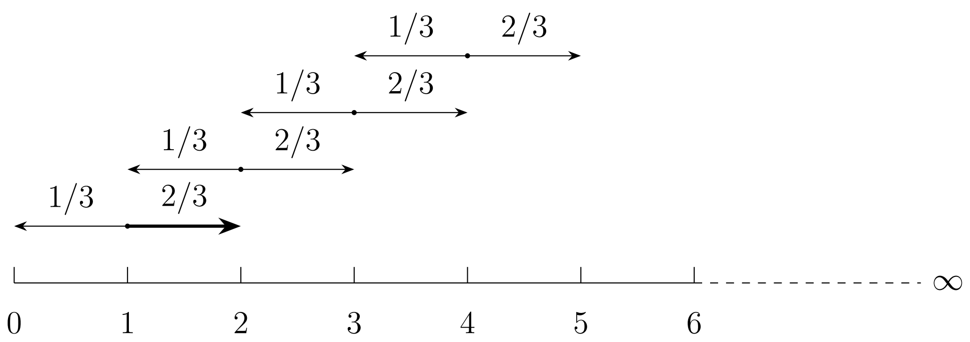
$$E(\text{ימי-עבודה אפשריים/ימי-עבודה מקסימליים}) = \frac{365 \cdot 365 \cdot \left(\frac{364}{365}\right)^{365}}{365 \cdot 365} = \left(\frac{364}{365}\right)^{365} \approx 0.3674\,.$$

לפי משוואה 14:

$$\lim_{n\to\infty} E(\text{ימי-עבודה אפשריים/ימי-עבודה מקסימליים}) = \lim_{N\to\infty}\left(1 - \frac{1}{N}\right)^N = \frac{1}{e}\,.$$

סימולציה

```
For 100 people
Expectation work-days    =  27742
Average work days        =  27743
Ratio work-days / 365**2 = 0.2082
For 250 people
Expectation work-days    =  45958
Average work days        =  45939
Ratio work-days / 365**2 = 0.3450
For 364 people
Expectation work-days    =  48944
Average work days        =  48936
Ratio work-days / 365**2 = 0.3674
```



איור 5: ?האם חלקיק יכול להגיע ל-0 (הציר אינסופי לימין)

```
For 365 people
Expectation work-days    =  48944
Average work days        =  48917
Ratio work-days / 365**2 = 0.3674
```

35. על שפת התהום (The cliff-hanger)

חלקיק מוצב על ציר ה-x במקום 1. בכל מקום על ציר ה-x הוא יכול לצעוד צעד ימינה עם הסתברות $2/3$ וצעד שמאלה עם הסתברות $1/3$ (איור 5).

שאלה 1: מה ההסתברות שהחלקיק יגיע למקום 0 בסופו של דבר?

שאלה 2: אם ההסתברות של צעד ימינה היא p וההסתברות של צעד שמאלה היא $1-p$, מה ההסתברות שהחלקיק יגיע למקום 0 בסופו של דבר? נתח את האפשרויות לערכים שונים של p.

רמז: השתמש בהסתרויות מותנות לאחר הצעד הראשון.

פתרון

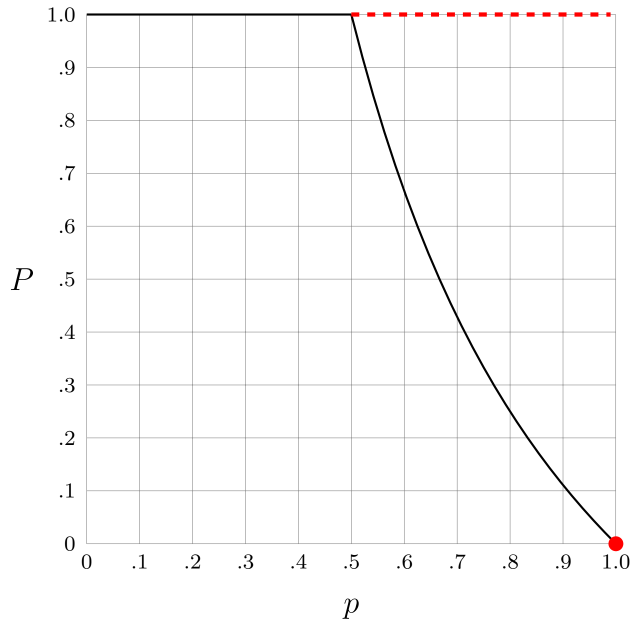
פתרון 1,2: נסמן צעד שמאלה ב-L וצעד ימינה ב-R. החלקיק יכול להגיע ל-0 ישירות על ידי צעד L עם הסתברות $\frac{1}{3}$, או על ידי צעד RLL עם הסתברות $\frac{2}{3}\cdot\frac{1}{3}\cdot\frac{1}{3}$, או על ידי צעד $RRLLL$ עם הסתברות $\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^3$, ... החישוב נראה כמו סידרה הנדסית פשוטה אבל הוא מתעלם מהאפשרויות כגון $RLRLL$.

נסמן ב-$P(i,j)$ את ההסתברות שהחלקיק מגיע ל-i מ-j. נחשב את ההתסתברות שהחלקיק מגיעה ל-0 מ-1 כתלות בצעד הראשון:

$$\begin{aligned} P(0,1) &= P(0,1|\text{צעד ראשון שמאלה}) + P(0,1|\text{צעד ראשון ימינה}) \\ &= (1-p)\cdot 1 + pP(1,2)P(0,1)\,. \end{aligned}$$

אבל $P(1,2)=P(0,1)$ ומתקבלת משוואה ריבועית ב-$P(0,1)$:

$$\begin{aligned} P(0,1) &= (1-p) + pP(0,1)^2 \\ pP^2(0,1) - P(0,1) + (1-p) &= 0 \\ P(0,1) &= 1,\ (1-p)/p\,. \end{aligned}$$



איור 6: הגרף של $P = \min(p/(1-p), 1)$ עבור $p \in [0, 1]$

אם $1/2 \geq p$ אזי $(1-p)/p \geq 1$, כך ש-$P = 1$ הוא הפתרון היחיד ובטוח שהחלקיק יגיע ל-0.

אם $p = 1$ אזי $P = 0$ כי אם החלקיק תמיד צועד ימינה הוא לא יכול להחזור ל-0.

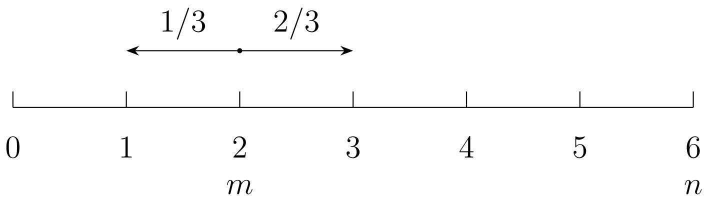
נניח ש-$P(0,1) = 1$ עבור $1/2 < p < 1$, כלומר, $P(0,1)$ **לא תלוי ב-**p. באיור 6 הקו האדום המקווקוו האדום מראה את עבור $P(1,0)$ כאשר p שואף ל-1 והנקודה האדומה מראה ש-$P(1,0) = 0$ כאשר p מגיע ל-1. $P(0,1)$ לא יכולה פתאום "לקפוץ" מ-1 ל-0 ולכן עבור $p > 1/2$ הפתרון היחיד הוא $P = (1-p)/p < 1$.[4]

עבור $P = 1/2, p = 2/3$ ועבור $P = 1, p = 1/2$. זה מפתיע כי לא היינו מצפים שהחלקיק יחזור תמיד ל-0 אם כיוון הצעד נקבע על ידי הטלת מטבע הוגן! אנו זקוקים למטבע ממש לא-הוגן (הסתברות של "עצי" שווה ל-2/3) כדי להשוות את הסיכויים לחזור ל-0 או לא.

סימולציה

```
For probability = 0.2500:
Probability of reaching 0 = 1.0000
Proportion reaching 0     = 1.0000
For probability = 0.5000:
Probability of reaching 0 = 1.0000
Proportion reaching 0     = 0.9612
For probability = 0.6667:
Probability of reaching 0 = 0.5000
```

[4] Mosteller כותב שזה נובע מרציפות אבל הוא לא נותן הוכחה.



איור 7: האם החלקיק יכול לחזור ל-0 (ציר סופי)?

```
Proportion reaching 0      = 0.5043
For probability = 0.7500:
Probability of reaching 0 = 0.3333
Proportion reaching 0      = 0.3316
For probability = 0.8000:
Probability of reaching 0 = 0.2500
Proportion reaching 0      = 0.2502
```

36. פשיטת הרגל של מהמר (Gambler's ruin) ^{D}

חלקיק מוצב על ציר ה-x במקום $m \geq 1$. בכל מקום על ציר ה-x הוא יכול לצעוד צעד ימינה עם הסתברות $p > 1/2$ וצעד שמאלה עם הסתברות $1-p$.

שאלה 1: מה ההסתברות שהחלקיק יגיע למקום 0 בסופו של דבר?

שאלה 2: יהי $m < n$. אם החלקיק מגיע למקום 0 או למקום n הוא מפסיק לצעוד (איור 7). מה ההסתברות שהחלקיק יגיע למקום 0? מה ההסתברות שהחלקיק יגיע למקום n?

הערה: בעיה 35 מייצגת מהמר המשחק עם כמות סופית של כסף נגד קזינו עם כמות בלתי מוגבלת של כסף. הבעיה מבקשת את ההסתברות שהמהמר יפסיד את כל כספו. בעיה 36(2) שואלת על מהמר אחד עם m שמשחק נגד מהמר שני עם $n-m$. הבעיה מבקשת את ההסתברויות ש**אחד** מהם מפסיד את כל כספו לשני.

פתרון

הפתרון מבוסס על [12, Chapter 2, Example 4m].

נסמן ב-$P(i,j)$ את ההסתברות שהחלקיק מגיע ל-i מ-j.

פתרון 1: ראינו בפתרון לבעיה 35 שעבור $p > 1/2$ (כאן נתון), אם חלקיק נמצא במקום 1 ההסתברות שלו להגיע ל-0 היא $r = (1-p)/p$. הסתברות זו לא תלויה במקום האבסולוטי של החלקיק ולכן:

$$P(i,j) = P(i+k, k+1) = P(i-k, j-k)\,,$$

ו-

$$P(0,m) = P(m-1,m)P(m-2,m-1)\cdots P(1,2)P(0,1) = r^m\,. \tag{17}$$

פתרון 2: סמן בקיצור $P_i = P(n,i)$ וחשב את P_i:

$$\begin{aligned} P_i &= pP_{i+1} + (1-p)P_{i-1} \\ pP_{i+1} &= (p+(1-p))P_i - (1-p)P_{i-1} \\ p(P_{i+1}-P_i) &= (1-p)(P_i - P_{i-1}) \\ P_{i+1} - P_i &= r(P_i - P_{i-1})\,. \end{aligned}$$

$P_0 = 0$ כי אם החלקיק נמצא ב-0 הוא מפסיק לצעוד. לכן:

$$\begin{aligned} P_2 - P_1 &= r(P_1 - P_0) = rP_1 \\ P_3 - P_2 &= r(P_2 - P_1) = r^2 P_1 \\ \cdots &= \cdots \\ P_i - P_{i-1} &= r(P_{i-1} - P_{i-2}) = r^{i-1}P_1\,. \end{aligned}$$

רוב הגורמים בצד השמאלי מצטמצמים כאשר מחברים את המשוואות:

$$\begin{aligned} P_i - P_1 &= P_1 \sum_{j=2}^{i} r^{j-1} \\ &= P_1 + P_1 \sum_{j=2}^{i} r^{j-1} - P_1 \\ P_i &= P_1 \sum_{j=0}^{i-1} r^j = P_1 \left(\frac{1-r^i}{1-r}\right)\,. \end{aligned} \tag{18}$$

אם חלקיק נמצא ב-n הוא כבר נמצא ב-n כך ש-$P_n = 1$:

$$\begin{aligned} 1 &= P_1 \left(\frac{1-r^n}{1-r}\right) \\ P_1 &= \left(\frac{1-r}{1-r^n}\right)\,. \end{aligned} \tag{19}$$

ממשוואות 18, 19:

$$P(n,i) = \left(\frac{1-r^i}{1-r^n}\right)\,. \tag{20}$$

בהוכחה סימטרית שמחליפה את p ו-$(1-p)$:

$$P(0,i) = \left(\frac{1-(1/r)^{n-i}}{1-(1/r)^n}\right)\,. \tag{21}$$

הקורא מוזמן להראות שהסכום של משוואות 20, 21 הוא 1 כלומר שמובטח שאחד המהמרים ינצח והשני יפסיד. עבור $m=1, n=3, p=2/3$:

$$\begin{aligned} P(0,1) &= \left(\frac{1-\left(\frac{1}{2}\right)^1}{1-\left(\frac{1}{2}\right)^3}\right) = \frac{4}{7} \\ P(3,1) &= \left(\frac{1-2^2}{1-2^3}\right) = \frac{3}{7}\,. \end{aligned}$$

סימולציה

```
For probability = 0.6667:
Probability of reaching (0,10) from 1 = (0.4995,0.5005)
Proportion reaching     (0,10) from 1 = (0.5056,0.4944)
Probability of reaching (0,10) from 4 = (0.0616,0.9384)
Proportion reaching     (0,10) from 4 = (0.0643,0.9357)
Probability of reaching (0,10) from 6 = (0.0147,0.9853)
Proportion reaching     (0,10) from 6 = (0.0123,0.9877)

For probability = 0.7500:
Probability of reaching (0,10) from 1 = (0.3333,0.6667)
Proportion reaching     (0,10) from 1 = (0.3395,0.6605)
Probability of reaching (0,10) from 4 = (0.0123,0.9877)
Proportion reaching     (0,10) from 4 = (0.0115,0.9885)
Probability of reaching (0,10) from 6 = (0.0014,0.9986)
Proportion reaching     (0,10) from 6 = (0.0015,0.9985)
```

ככל שלמהמר בצד שמאל יש יותר והסתברות גבוהה היותר לזכות בכל צעד, כך ההסתברות שלו לניצחון גדלה.

גרף של הצעדים

איור 8 נוצר בעזרת ספריית `matplotlib` של Python. קוד המקור מופיע בקובץ:

`36-gamblers-ruin-plot.py`

37. משחק נועז או משחק זהיר (Bold play vs. cautious play)

משחק הרולט מתואר בבעיה 7 (עמוד14).

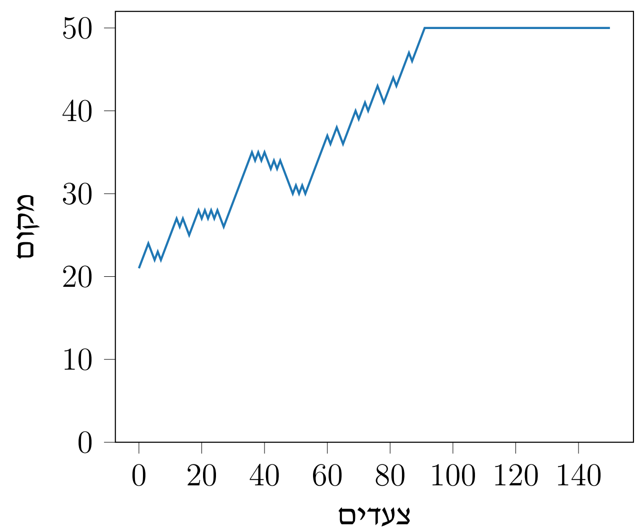
איזו מהאסטרגיות שלהלן עדיפה?

1. משחק נועז: להמר 20 בסיבוב אחד.
2. משחק זהיר: להמר 1 בכל סיבוב עד שאתה זוכה או מפסיד 20.

רמז: השתמש בתוצאות של בעיה 36.

פתרון

ההסתברות לנצח במשחק נועז היא $18/38 \approx 0.4737$.



איור 8: פשיטת רגל של מהמר $m=20, n=50, p=0.67$

משחק זהיר הוא בעיית "פשיטת רגל של מהמר" (בעיה 36): אתה מתחיל עם 20 ומשחק עד שאתה מגיע ל-40 (הרווחת 20) או אד שאתה מגיע ל-0 (הפסדת 20). ההסתברות לנצח במשחק זהיר נתונה על ידי משוואה 20 עם $p=18/38$ ו-$1-p=20/38$ כך ש-$r=20/18$:

$$
\begin{aligned}
r &= \frac{20}{38} \Big/ \frac{18}{38} = \frac{20}{18} \\
P(40,20) &= \frac{1-(20/18)^{20}}{1-(20/18)^{40}} \approx 0.1084\,.
\end{aligned}
$$

ברור שמשחק נועז עדיף על משחק זהיר.

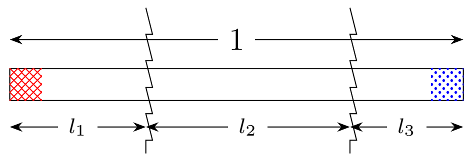
Mosteller מביא הסבר איטואטיבי: הימור בסיבובים רבים חושף את המהמר לאפשרות שהקזינו מצנח אם הכדור נוחת בכיס ירוק עם הסתברות של $2/38$.

סימולציה

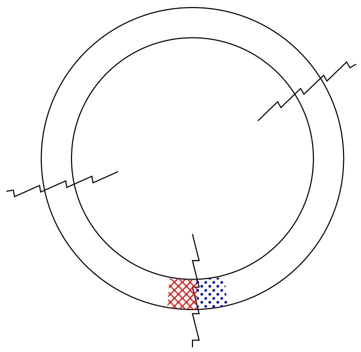
```
Probability of bold wins     = 0.4737
Proportion bold wins         = 0.4677
Probability of cautious wins = 0.1084
Proportion cautious wins     = 0.1094
```

39. הכימאי המגושם (The clumsy chemist)

נתון מספר רב של מקלות מזכוכית באורך 1. קצה אחד צבוע באדום (משובץ) ושני בכחול (מנוקד). כאשר זורקים אותם על הרצפה כל אחד נשבר לשלושה חלקים עם התפלגות אחידה של האורכים (איור 9(א)). מה התוחלת של אורכו של החלק בקצה הכחול?



איור 9(א) חלוקת מקל לשלושה חלקים



איור 9(ב) חלוקת טבעת לשלושה חלקים

רמז: במקום מקלות ישרים הנח שקבלת טבעות זכוכית (ללא סימנים) שגם הם נשברים לשלושה חלקים (איור 9(ב)).

פתרון 1

אין סימטריה במקלות כי הקצוות שונים מהחלק האמצעי. אולם הטבעת סימטרית ולכן ההתפלגויות של שלושת החלקים יהיו אחידות עם תוחלת $1/3$. על ידי צביעת אחת מנקודות השבירה כפי שמופיע באיור 9(ב), הבעיה הופכת להיות זהה לבעיית המקל כך שההתפלגויות זהות. לכן התוחלת של אורכי החלקים היא גם $1/3$.

פתרון 2

פתרון אלגנטי זה מבוסס על [5].

נניח שהמקל מייצג את קטע הקו $(0,1)$. המקל נשבר בשני מקומות שניתן לייצג על ידי שני משתנים אקראיים בלתי-תלויים עם התפלגות אחידה $X,Y \in (0,1)$. נחשב את ההסתברות $P(|X-Y|>a)$.

טבלה 1 מראה נקודות (x,y) כאשר $x,y \in \{0.0, 0.1, 0.2, \ldots, 0.9\}$ והנקודה העשרונית הושמטה. הערכים בטבלה הם $|X-Y|$. עבור $a=0.6$ הערכים למעלה משמאל (מודגשים) והערכים למטה מימין (מודגשים) הם התוצאות שמגדירות את $P(|X-Y|>a)$:

$$G(a) = P(|X-Y|>a) = 2 \cdot \frac{1}{2}(1-a)(1-a) = (1-a)^2 \,.$$

אנחנו מעוניינים המשלים:

$$F(a) = 1 - G(a) = P(|X-Y|<a) = 1-(1-a)^2 = 2a - a^2 \,.$$

זאת ההתפלגות ההסתברות המצטברת (CPD) עבור הקטע $(0,1)$. ניתן לקבל את פונקציית ההסתברות הצפיפות (PDF) על ידי גזירת ה-CDP:

$$f(a) = P(|X-Y|=a) = \frac{d}{da}F(a) = \frac{d}{da}(2a-a^2) = 2(1-a) \,.$$

a

	9	**9**	**8**	**7**	**6**	5	4	3	2	1	0	
	8	**8**	**7**	**6**	5	4	3	2	1	0	1	
	7	**7**	**6**	5	4	3	2	1	0	1	2	
a	6	**6**	5	4	3	2	1	0	1	2	3	
	5	5	4	3	2	1	0	1	2	3	4	
	4	4	3	2	1	0	1	2	3	4	5	
y	3	3	2	1	0	1	2	3	4	5	**6**	a
	2	2	1	0	1	2	3	4	5	**6**	**7**	
	1	1	0	1	2	3	4	5	**6**	**7**	**8**	
	0	0	1	2	3	4	5	**6**	**7**	**8**	**9**	
		0	1	2	3	4	5	6	7	8	9	

x a

טבלה 1: התפלגות נקודות השבירה ב-$(0,1) \times (0,1)$

התוחלת היא האינטגרל של ה-PDF כפול הערך:

$$E(|X-Y|) = \int_0^1 a \cdot 2(1-a)\,da = 2\left(\frac{a^2}{2} - \frac{a^3}{3}\right)\bigg|_0^1 = \frac{1}{3}\,.$$

סימולציה

```
Expectation of length of right piece = 0.3333
Average length of right piece        = 0.3359
```

40. האס הראשון (The first ace)

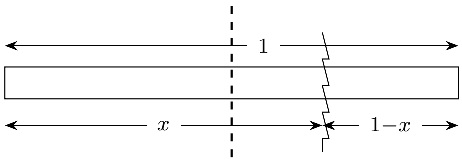
חלק קלפים מחפיסה מעורבת היטב עד שמופיע אס. מה התוחלת של מספר הקלפים שיש לחלק?

רמז: חשוב על חפיסת קלפים ללא האסים מסודרת בשורה.

פתרון הקלפים הם כמו "מקל" באורך 48 "שנשבר" על ידי 4 ל-5 חלקים. הפתרון של בעיה 39 מתאים גם כאן והתוחלת של חלק היא $48/5 = 9.6$.

סימולציה

```
Expectation of first ace = 9.6000
Average first ace        = 9.5805
```



איור 10 : שבירת מקל לשני חלקים

42. הקצה הקצר של המקל (The little end of the stick)

אתה שובר מספר גדול של מקלות זכוכית באורך 1 לשני חלקים. למקום השבירה התפלגות אחידה לאורך המקל.

שאלה 1: מה התוחלת של אורכו של החלק **הקטן** יותר?

שאלה 2: מה התוחלת של היחס בין אורכו של החלק הקטן לאורכו של החלק הגדול?

פתרון

פתרון 1: ההסתברות שנקודת השבירה היא בצד השמאלי של המקל היא $1/2$ שהיא גם ההסתברות שהנקודה בצד ימין. החלק הקטן יותר נמצא באותו צד שבו נמצאת נקודת השבירה. התוחלת של נקודת השבירה היא באמצע בין קצה המקל לבין אמצע המקל:

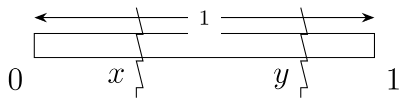
$$E(\text{אורך הקטן יותר}) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\,.$$

פתרון 2: ללא הגבלת הכלליות הנח שנקודת השבירה נמצאת בצד הימני של המקל (איור 10). היחס בין החלק הקטן והחלק הגדול הוא $(1-x)/x$ ויש להשתמש בקבוע נירמול (עמוד 86) כי התוחלת מחושבת מעל ל-$(1/2, 1)$, מחצית הטווח של x:

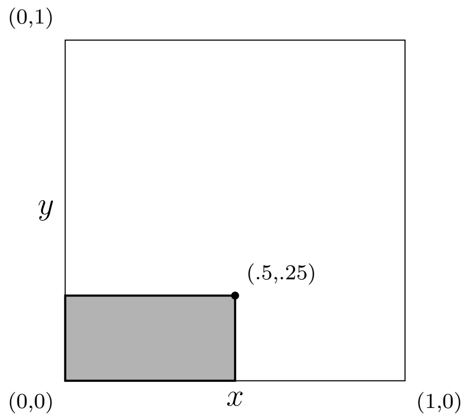
$$\begin{aligned}
E(\text{יחס גדול יותר / קטן יותר}) &= \left(\frac{1}{1-(1/2)}\right)\int_{\frac{1}{2}}^{1} \frac{1-x}{x}\,dx \\
&= 2\int_{\frac{1}{2}}^{1}\left(\frac{1}{x}-1\right)\,dx \\
&= 2(\ln|x|-x)\Big|_{\frac{1}{2}}^{1} = 2(0-1-\ln\tfrac{1}{2}+\tfrac{1}{2}) \approx 0.3863\,.
\end{aligned}$$

סימולציה

```
Expectation of length of smaller = 0.2500
Average length of smaller        = 0.2490
Expectation of smaller/larger    = 0.3863
Average smaller/larger           = 0.3845
```



איור 11(א) חלוקת מקל לשני חלקים



איור 11(ב) יצוג האורכים במעגל היחידה

43. המקל השבור **(The broken bar)**^{D}

אתה שובר מספר רב של מקלות זכוכית באורך 1 בשתי נדוקות שבירה (איור 11(א)).

שאלה 1: מה התוחלת של אורכו של החלק הקצר ביותר?

שאלה 2: מה התוחלת של אורכו של החלק הארוך ביותר?

רמז: x, y הם משתנים אקראים בלתי-תלויים בהתפלגות אחידה בתוך $(0, 1)$. ניתן להציג כל זוג (x, y) כנקודה בריבוע $(0, 1) \times (0, 1)$ (איור 11(ב)). מה ההסתברות ש-$(x, y) < (.5, .25)$?

רמז: עבור שאלה 1 הנח שהחלק השמאלי הוא הקצר ביותר ועבור שאלה 2 הנח שהחלק השמאלי הוא בארוך ביותר.

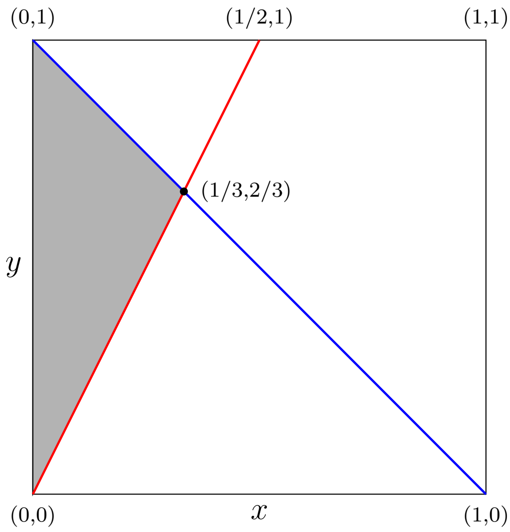
פתרון

פתרון 1: ללא הגבלת הכלליות הנח שהחלק השמאלי שאורכו x הוא החלק הקצר ביותר. מכאן ש-$x <$ $y - x$ ו-$y - x < 1 - y$ שניתן לפשט ולקבל $2x < y$ ו-$x + y < 1$.

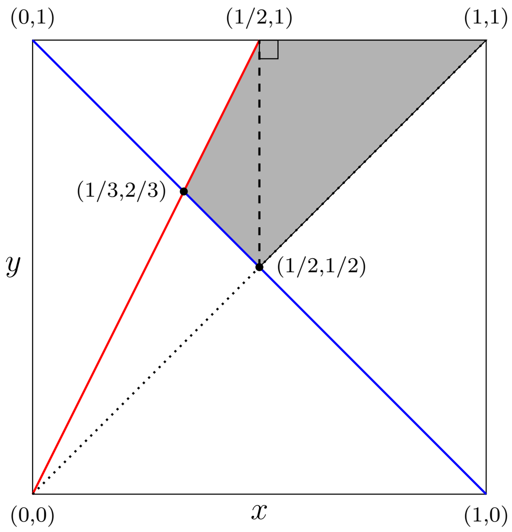
איור 12(א) מראה את הקווים $y = 2x$ (אדום) ו-$y = 1 - x$ (כחול). כדי לאמת את אי-השוויונות, (x, y) חייבת להיות באיזור באפור לשמאל לשני הקווים. ניתן לחשב את נקודת החיתוך $(1/2, 2/3)$ על ידי פתרון שתי המשוואות.

ניתן לחשב את התוחלת מעל חישוב האינטרגל של המכפלה של x וההפרש בין שני הקווים. קבוע הנירמול הוא השטח של הריבוע לחלק לשטח של האיזור האפור:

$$\begin{aligned} E(x) &= \frac{1}{1/6} \int_0^{1/3} x[(1-x) - 2x]\, dx \\ &= 6 \int_0^{1/3} (x - 3x^2)\, dx \\ &= 6 \left(\frac{x^2}{2} - x^3 \right) \Bigg|_0^{1/3} = \frac{2}{18} \approx 0.1111\,. \end{aligned}$$



איור 12(א) איזור אפור עבור המקל הקצר ביותר



איור 12(ב) איזור אפור עבור המקל הארוך ביותר

פתרון 2: כדי שהחלק השמאלי יהיה הארוך ביותר $x > y - x$ ו-$x > 1 - y$, ולכן (x, y) חייבת להיות לימינו של $y = 2x$ (אדום) ולימינו של $y = 1 - x$ (כחול) (איור 12(ב)). בנוסף, לפי ההנחה ש-x נמצא לשמאלו של y, (x, y) חייבת להיות לשמאלו של $y = x$ (מנוקד).

לפי הליניראיות של התוחלת ניתן לחלק את האיזור האפור לשני משולשים (קו מקווקו) ולחשב את התוחלת בנפרד. קבוע הנירמול הוא השטח של האיזור האפור שהוא $1/24 + 1/8 = 1/6$:

$$\begin{aligned}
E(x \text{ במשולש השמאלי}) &= 6\int_{1/3}^{1/2} x[2x-(1-x)]\,dx\\
&= 6\int_{1/3}^{1/2}\left(3x^2-x\right)\,dx\\
&= 6\left(x^3-\frac{x^2}{2}\right)\Bigg|_{1/3}^{1/2} = \frac{1}{9}\\
E(x \text{ במשולש הימני}) &= 6\int_{1/2}^{1} x(1-x)\,dx\\
&= 6\int_{1/2}^{1}(x-x^2)\,dx\\
&= 6\left(\frac{x^2}{2}-\frac{x^3}{3}\right)\Bigg|_{1/2}^{1} = \frac{1}{2}\\
E(x) &= \frac{1}{9}+\frac{1}{2} = \frac{11}{18} \approx 0.6111\,.
\end{aligned}$$

התוחלת של אורכו של החלק הבינוני היא $1 - \frac{2}{18} - \frac{11}{18} = \frac{5}{18} \approx 0.2778$.

סימולציה

```
Expectations: shortest = 0.1111, middle = 0.2778, longest = 0.6111
Averages:     shortest = 0.1115, middle = 0.2783, longest = 0.6102
```

44. לנצח במשחק לא הוגן (Winning an unfair game)^D

נתון מטבע לא הוגנת שההסתברות להופעת עץ היא $1/3 < p < 1/2$, הטל את המטבע מספר זוגי של פעמים $N = 2n$. אתה מנצח אם ורק אם **ביותר ממחצית** ההטלות מופיע עץ.

שאלה 1: עבור $p = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$, חשב את P_2, P_4, P_6. הסבר את הגבלת הבעיה ל-$1/3 < p < 1/2$.

שאלה 2: פתח נוסחה עבור P_N, ההסתברות לנצח ונוסחה עבור T_N, ההסתברות לתיקו.

שאלה 3: פתח נוסחה עבור ה-N עבורו יש את ההסתברות הגבוהה ביותר לנצח.

רמז: אם ההסתברות הגבוהה ביותר לנצח היא ב-N הטלות אזי $P_{N-2} \leq P_N$ ו-$P_N \geq P_{N+2}$.

פתרון

פתרון 1: ההטלות בלתי-תלויות ולכן נשמתמש בהתפלגות הבינומית:

$$\begin{aligned}
P_2 &= p^2 \\
P_4 &= 1 \cdot p^4 + \binom{4}{1} p^3 (1-p) \\
P_6 &= 1 \cdot p^6 + \binom{6}{1} p^5 (1-p) + \binom{6}{2} p^4 (1-p)^2 .
\end{aligned}$$

עבור $p = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$ התוצאות הן:

p	P_2	P_4	P_6
$1/4$	$1/16 = 0.0625$	$13/256 \approx 0.0501$	$154/4096 \approx 0.0376$
$1/3$	$1/9 \approx 0.1111$	$9/81 \approx 0.1111$	$73/729 \approx 0.1001$
$1/2$	$1/4 = 0.2500$	$5/16 = 0.3125$	$22/64 \approx 0.3435$

הגיוני לשער שכאשר N שואף לאינסוף הערכים של P_N יורדות עבור $p = \frac{1}{4}$ ו-$p = \frac{1}{3}$ (למרות שקצב הירידה קטן). הערכים של P_N עולות עד אחד עבור $p = \frac{1}{2}$. לפי רציפות ההסתברות הגדולה ביותר תהיה בטווח $1/3 < p < 1/2$.

פתרון 2: כדי לנצח, עץ חייב להופיע ב-$i \in \{n+1, n+2, \ldots, 2n-1, 2n = N\}$ הטלות. מההתפלגות הבינומית:

$$\begin{aligned}
P_N &= \sum_{i=n+1}^{2n} \binom{2n}{i} p^i (1-p)^{2n-i} \\
T_N &= \binom{2n}{n} p^n (1-p)^n .
\end{aligned}$$

פתרון 3: כדי שההסתברות עבור $N = 2n$ תהיה הגבוהה ביותר חייב להתקיים:

$$P_{2n-2} \leq P_{2n} \quad \text{ו-} \quad P_{2n} \geq P_{2n+2} .$$

מתי $P_{2n-2} \neq P_{2n}$?

מקרה 1: לאחר הטלה $2n-2$, עץ הופיע n פעמים ופלי $n-2$ פעמים (כך שהיית מנצח אם היית עוצר כאן), אבל פלי מופיע בשתי ההטלות הבאות. עכשיו יש n עץ ו-n פלי ולכן אתה מפסיד. ההסתברות היא:

$$\binom{2n-2}{n} p^n (1-p)^{n-2} (1-p)^2 \,.$$

מקרה 2**:** לאחר הטלה $2n-2$, עץ הופיע $n-1$ פעמים ופלי $n-1$ פעמים (כך שהיית מפסיד אם היית עוצר כאן), אבל עץ מופיע בשתי ההטלות הבאות. עכשיו יש $n+1$ עץ ו-$n-1$ פלי ולכן אתה מנצח. ההסתברות היא:

$$\binom{2n-2}{n-1} p^{n-1} (1-p)^{n-1} p^2 \,.$$

כדי לאמת את $P_{2n-2} \leq P_{2n}$, P_{2n-2} לא יכול לגדול כאשר P_{2n} נשאר ללא שינוי (מקרה 1), אבל P_{2n} יכול לגדול עד שהיא גבוהה מ-P_{2n-2} (מקרה 2). לכן:

$$\begin{aligned}
\binom{2n-2}{n} p^n (1-p)^{n-2} (1-p)^2 &\leq \binom{2n-2}{n-1} p^{n-1} (1-p)^{n-1} p^2 \\
\frac{1}{n}(1-p) &\leq \frac{1}{n-1} p \\
(n-1)(1-p) &\leq np \\
n &\leq \frac{1-p}{1-2p} \\
2n &\leq \frac{1}{1-2p} + 1 \,.
\end{aligned}$$

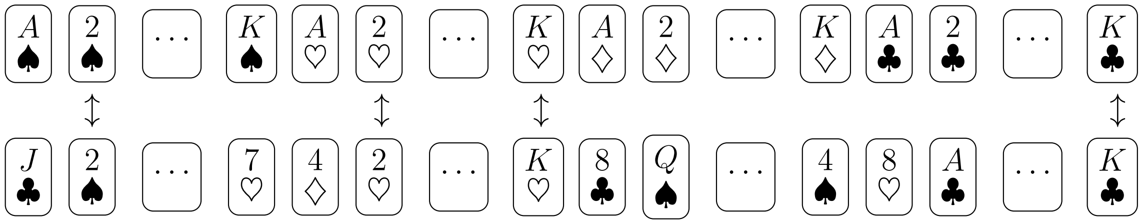
באופן דומה, כדי לאמת את $P_{2n} \geq P_{2n+2}$ חייב להיול ש:

$$\begin{aligned}
\binom{2n}{n+1} p^{n+1} (1-p)^{n-1} (1-p)^2 &\geq \binom{2n}{n} p^n (1-p)^n p^2 \\
\frac{1}{n+1}(1-p) &\geq \frac{1}{n} p \\
n(1-p) &\geq (n+1)p \\
n &\geq \frac{p}{1-2p} \\
2n &\geq \frac{1}{1-2p} - 1 \,.
\end{aligned}$$

לכן, ערך של $N = 2n$ עבורו מתקבלת ההסתברות הגבוהה ביותר הוא המספר השלם הזוגי הקרוב ביותר ל-$1/(1-2p)$.

סימולציה

```
For probability            = 0.3700
Optimal games to be played = 4
For  2 games, average won  = 0.1372
For  4 games, average won  = 0.1445
```



איור 13 : התאמת שתי חפיסות קלפים

```
For  6 games, average won   = 0.1431

For probability             = 0.4000
Optimal games to be played  = 6
For  4 games, average won   = 0.1820
For  6 games, average won   = 0.1845
For  8 games, average won   = 0.1680

For probability             = 0.4500
Optimal games to be played  = 10
For  8 games, average won   = 0.2671
For 10 games, average won   = 0.2646
For 12 games, average won   = 0.2640
```

45. ממוצע של מספר ההתאמות (Average number of matches)

סדר חפיסת קלפים בשורה בסדר הסטנדרטי ואז סדר חפיסה שניה בסדר אקראי מתחת לשורה הראשונה (איור 13). מה התוחלת של מספר ההתאמות של קלף בשורה הראשונה עם קלף בשורה מתחתיו?

פתרון

ההתפלגות אחידה וכלן לכל קלף בשורה השניה אותה המסתברות להתאים לקלף מעליו :

$$E(\text{מספר ההתאמות}) = 52 \cdot \frac{1}{52} = 1\,.$$

סימולציה

```
Expectation of matches = 1.00
Average of matches     = 1.01
```

46. הסתברויות של התאמות (Probabilities of matches)

סדר חפיסת קלפים בשורה בסדר הסטנדרטי ואז סדר חפיסה שניה בסדר אקראי מתחת לשורה הראשונה (איור 13). פתח נוסחה עבור $P(n,r)$, ההסתברות שיהיו בדיוק r התאמות של קלף בשורה הראשונה עם קלף בשורה מתחתיו! הנח ש-$P(k,0)$ נתון עבור $0 \leq k \leq n$.

פתרון

במבט ראשון נראה שבעיה זו דומה לבעיה 28 (לתפוס את הזייפן הזהיר) אבל קיים הבדל מהותי. השליפות מהקופסאות הן בלתי-תלויות אבל כאן ההתאמות תלויות אחת בשניה. למשל, אם יש התאמה בקלף הראשון (בהסתברות $1/n$), ההסתברות של התאמה בקלף השני היא $1/(n-1)$.

ההסתברות שקבוצה **נתונה** של r קלפים מתאימות היא:

$$\frac{1}{n} \cdot \frac{1}{n-1} \cdot \cdots \cdot \frac{1}{n+r-1} \,. \tag{22}$$

כדי לקבל בדיוק r התאמות, יש להכפיל משוואה 22 ב-$P(n-r,0)$, ההסתברות שאין בכלל התאמות בשאר $n-r$ הקלפים. לבסוף, יש $\binom{n}{r}$ דרכים לבחור r התאמות, ולכן:

$$\begin{aligned} P(n,r) &= \binom{n}{r} \frac{1}{n(n-1)(n+r-1)} P(n-r,0) \\ &= \frac{n!}{r!(n-r)!} \cdot \frac{1}{n!/(n-r)!} P(n-r,0) \\ &= \frac{1}{r!} P(n-r,0) \,. \end{aligned}$$

נוסחה זו פותרת את הבעיה כי $P(k,0)$ נתונה.

Mosteller מפתח נוסחה סגורה וגבול עבור $P(n,r)$:

$$P(n,k) = \frac{1}{k!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!} \tag{23}$$

$$\lim_{n-r\to\infty} P(n,k) \approx \frac{1}{k!} e^{-1} \,. \tag{24}$$

סימולציה

הרצתי את הסימולציה עבור $n=52$ קלפים וחישבתי את ההסתברות ממשוואה 24.

```
Probability of 1 matches = 0.3679
Proportion 1 matches     = 0.3710
Probability of 2 matches = 0.1839
Proportion 2 matches     = 0.1828
Probability of 3 matches = 0.0613
Proportion 3 matches     = 0.0569
Probability of 4 matches = 0.0153
Proportion 4 matches     = 0.0168
```

47. לבחור את הנדוניה הגדול ביותר (Choosing the largest dowry)^{D}

הנח סידרה של n קלפים עם הפנים למטה. על פניו של כל קלף נמצא מספר שלם חיובי אבל אין מידע על ההתלפגות שלהם. הפוך את הקלפים אחד-אחד ועיין במספרים. לאחר חשיפת כל אחד מהקלפים אתה יכול להכריז שמספר זה הוא הגדול ביותר בסידרה. אם אתה צודק אתה מנצח אחרת אתה מפסיד.

למשל, אם הסדרה היא $(47, 23, 55, 4)$, אתה מנצח רק אם אתה בוחר שת הקלף השלישי.

בחר קלף לפי אסטרטגיה זו: עבור r קבוע וותר על $r-1$ הקלפים הראשונים ובחר את הקלף הראשון שמספרו גדול מכל $r-1$ הקלפים.

שאלה 1: עבור $n=4$ ו-$r=3$ בדוק את כל התמורות ומצא בכמה מהן את מנצח.

שאלה 2: פתח נוסחה עבור ההסתברות לניצחון עבור n, r שרירותיים.

שאלה 3: מצא קירוב להסתברות כאשר $n, r \to \infty$.

רמז: נתון r באיזה מקומות יכול להופיע המספר הגדול ביותר m ובאיזה מקומות המספרים שהם פחות או שווים ל-m?

פתרון

פתרון 1: כדי לפשט את הסימון נשתמש בדירוג המספרים מנמוך לגבוה $1, 2, \ldots, n$ למרות שהערכים אמיתיים של המספרים לא ידועים.

יש 24 תמורות של ארבעה מספרים. לפי האסטרטגיה אתה מוותר על שני הקלפים הראשונים ובוחר או את הקלף השלישי או את הקלף הרביעי, כך שאתה מפסיד אם 4 נמצא בשני המקומות הראשונים. מה עם התמורה $(1, 2, 3, 4)$? אתה מוותר על $1, 2$ ובוחר 3 בגלל שהוא גובהה יותר מ-$1, 2$, אבל אתה מפסיד כי זה לא המספר הגדולה ביותר. מה עם התמורה $(1, 3, 2, 4)$? שוב, לפי האסטרטגיה אתה מוותר על $1, 3$, אבל מוותר גם על 2 כי הוא **לא** גדול מ-$1, 3$. כעת אתה בוחר 4 ומנצח. נסח טיעונים דומים לכל התמורות ובדוק שכל התמורות עם 4 במסגרת הן נצחונות:

$1\ 2 \mid 3\ 4$	$1\ 2 \mid \boxed{4}\ 3$	$1\ 3 \mid 2\ \boxed{4}$	$1\ 3 \mid \boxed{4}\ 2$	$1\ 4 \mid 2\ 3$	$1\ 4 \mid 3\ 2$
$2\ 1 \mid 3\ 4$	$2\ 1 \mid \boxed{4}\ 3$	$2\ 3 \mid 1\ \boxed{4}$	$2\ 3 \mid \boxed{4}\ 1$	$2\ 4 \mid 1\ 3$	$2\ 4 \mid 3\ 1$
$3\ 1 \mid 2\ \boxed{4}$	$3\ 1 \mid \boxed{4}\ 2$	$3\ 2 \mid 1\ \boxed{4}$	$3\ 2 \mid \boxed{4}\ 1$	$3\ 4 \mid 1\ 2$	$3\ 4 \mid 2\ 1$
$4\ 1 \mid 2\ 3$	$4\ 1 \mid 3\ 2$	$4\ 2 \mid 1\ 3$	$4\ 2 \mid 3\ 1$	$4\ 3 \mid 1\ 2$	$4\ 3 \mid 2\ 1$

ההסתברות לנצח היא $10/24$.

פתרון 2: אתה מפסיד אם המספר הגדול ביותר נמצא באחד המקומות $1, \ldots, r-1$. לכן כדי לנצח המספר הגדול ביותר חייב להיות במקום m כאשר $r \leq m \leq n$:

$$1 \quad 2 \quad \cdots \quad r-2 \quad r-1 \quad \overbrace{r \quad r+1 \quad \cdots \quad m-1 \quad m \quad m+1 \quad \cdots \quad n}^{\text{מספר גדול ביותר חייב להיות כאן}}\,.$$

לפי האסטרטגיה אתה מוותר על $r-1$ הקלפים הראשונים. אתה תבחר מקום m אם ורק אם **כל** במספרים ב-$(r, \ldots, m-1)$ קטנים **מכל** המספרים ב-$(1, \ldots, r)$. במילים אחרות, המספר הגדול ביותר בסידרה

$(1,\ldots,m-1)$ הוא **לא** בחלק השני של הסידרה $(r,\ldots m-1)$ אלא בחלק הראשון $(1,\ldots,r-1)$.
ההסתברות היא:

$$P(\text{המספר הגדול ביותר ב-}(1,\ldots,m-1)\text{ נמצא ב-}(1,\ldots,r-1)) = \frac{r-1}{m-1}\,.$$

נביא דוגמה כדי להקל על הבנת הטיעונים. נתון $1,\ldots,10$ ו-$r=5$:

$$\overbrace{8 \quad 10 \quad 9 \quad 4 \quad 1}^{\text{גדול ביותר נמצא כאן}} \quad 3 \quad 6 \quad 5 \quad 2$$

המספר הגדול ביותר נמצא במקום $m=9$. האולם המספר הגדול ביותר ב-$(1,\ldots,m-1=8)$ נמצא **בתוך** הסדרה $(r=5,\ldots,m-1=8)$ ולכן אתה לא תנצח. לפי האסטרטגיה תבחר 9, המספר הראשון שהוא גדול ביותר מכל המספרים ב-$(1,\ldots,r-1=4)$ ותפסיד כי $9>10$. לעומת זאת, אם הוחלפו המקומות של 9 ו-10 אזי המספר הגדול ביותר שהוא פחות מ-10 הוא 6 במקום $5=r<3$, ולכן לפי האסטרטגיה לא תבחר ב-$1,4$ ותנצח:

$$8 \quad 10 \quad \overbrace{\mathbf{9} \quad 4 \quad 1}^{\text{לא כאן}} \quad \overbrace{3 \quad 6 \quad 5 \quad 2}^{\text{כאן}}$$

$$8 \quad 9 \quad 10 \quad \overbrace{4 \quad 1}^{\text{לא כאן}} \quad \overbrace{3 \quad \mathbf{6} \quad 5 \quad 2}^{\text{כאן}}$$

ההסתברות שהמספר הגדול ביותר נמצא ב-m הוא $1/n$ ולכן:

$$P(\text{ניצחון}) = \sum_{m=r}^{n}\frac{1}{n}\cdot\frac{r-1}{m-1} = \frac{r-1}{n}\sum_{m=r}^{n}\frac{1}{m-1}\,. \tag{25}$$

עבור $r=3$, $n=4$, $P(\text{ניצחון})=5/12=10/24$, התוצאה שמצאנו על ידי בדיקת כל התמורות.

פתרון 3: נכתוב את משוואה 25 כך:

$$P(\text{ניצחון}) = \frac{r-1}{n}\left(\sum_{m=2}^{n}\frac{1}{m-1} - \sum_{m=2}^{r-1}\frac{1}{m-1}\right)\,. \tag{26}$$

עבור n,r גדולים ניתן למצוא קירוב לשתי הסדרות ההרמוניות במשוואה 26 כך:

$$P(\text{ניצחון}) = \frac{r}{n}(\ln n - \ln r) = \frac{r}{n}\ln\frac{n}{r} = -\frac{r}{n}\ln\frac{r}{n}\,. \tag{27}$$

נסמן $x=r/n$ ונמצא את המקסימום מהנגזרת:

$$\begin{aligned}(-x\ln x)' &= -x\cdot\frac{1}{x} + (-1)\ln x = 0\\ \ln x &= -1\\ x &= 1/e\,.\end{aligned}$$

לכן כדי למקסם את ההסתברות לנצח בחר $r\approx n/e$, וממשוואה 27:

$$P(\text{ניצחון}) \approx -\frac{1}{e}\ln\left(\frac{1}{e}\right) = \frac{1}{e} \approx \frac{1}{3}\,,$$

גבוהה הרבה יותר מהסתברות $1/n$ לנצח על ידי בחירת קלף אקראי.

סימולציה

הרצתי את הסימולציה עם 100 קלפים וערכי r קרובים ל-$100/e$:

```
Reject cards before r = 36:
Probability of wins   = 0.3674
Proportion wins       = 0.3641
Reject cards before r = 37:
Probability of wins   = 0.3678
Proportion wins       = 0.3759
Reject cards before r = 38:
Probability of wins   = 0.3679
Proportion wins       = 0.3548
Reject cards before r = 30:
Probability of wins   = 0.3590
Proportion wins       = 0.3601
```

48. בחירת המספר האקראי הגדול ביותר

(Choosing the largest random number) ^{D}

הנח סידרה של n קלפים עם הפנים למטה. על פניו של כל קלף נמצא מספר ממשי עם התפלגות אחידה ב-$[0.0, 1.0]$. הפוך את הקלפים אחד-אחד ועיין במספרים. לאחר חשיפת כל אחד מהקלפים, אתה יכול להכריז שמספר זה הוא הגדול ביותר בסידרה. אם אתה צודק אתה מנצח אחרת אתה מפסיד.

השתמש באסטרטגיה של בעיה 47: עבור r קבוע וותר על $r-1$ הקלפים הראשונים ובחר את הקלף הראשון שגדול מהמספר הגדול ביותר ב-$r-1$ קלפים הראשונים.

הגדרה: d, **ערך שווה-נפש**, הוא הערך שמתחתיו אתה מוותר על הקלף ומעליו אתה בחור את הקלף.

שאלה 1: חשב את d עבור $n=1$ וחשב את ההסתברות לנצח.

שאלה 2: חשב את d עבור $n=2$ וחשב את ההסתברות לנצח.

שאלה 3: חשב את d עבור $n=3$. אל תנסה לחשב את ההסתברות לנצח!

הערה: בבעיה 47 הערכים יכולים להיות $100, 200, 300$ או $100, 50, 20$ כך שחשיפת המספר הראשון לא מספק מידע על המספרים האחרים. בבעיה זו ההתלפגות אחידה ולכן אם המספר הראשון הוא 0.3 ההסתברות שהמספר השני יהיה גדול יותר היא 0.7.

פתרון

יהי v_1, v_2, v_3 המספרים על שלושת קלפים.

פתרון 1: אין ברירה אלא לבחור את הקלף הראשון כי אין קלפים אחרים. לכן אין ערך שווה-נפש. v_1 הוא המספר "הגדול ביותר" ו-$P(\text{ניצחון}) = 1$.

פתרון 2: אם אתה בוחר את הקלף הראשון $P(\text{ניצחון}) = v_1$ שהיא ההסתברות שהמספר על הקלף השני קטן יותר. אם אתה מוותר על הקלף הראשון, $P(\text{ניצחון}) = 1 - v_1$ שהיא ההסתברות ש-$v_2 > v_1$. לכן, אם $v_1 < 0.5$ בחר את הקלף השני כי $1 - v_1 > 0.5$ ואם $v_1 > 0.5$ בחר את הקלף הראשון. מכאן ש-$d = 0.5$.

נוסחה לחישוב ההסתברות לנצח:

$$P(\text{ניצחון}) = P(\text{ניצחון} \mid v_1 < 0.5)\, P(v_1 < 0.5) + P(\text{ניצחון} \mid v_1 > 0.5)\, P(v_1 > 0.5)\,. \tag{28}$$

$P(v_1 < 0.5) = 0.5$ נובע מההתפלגות האחידה. מה עם $P(\text{ניצחון} \mid v_1 < 0.5)$? לפי האסטרטגיה אתה מנצח אם $0.5 < v_2 < 1$ אבל גם אם $v_1 < v_2 < 0.5$. ההתפלגות של v_1 היא אחידה ב-$(0, 0.5)$ ולכן ההסתברות ש-$v_1 < v_2 < 0.5$ היא מחצית הטווח:

$$P(\text{ניצחון} \mid v_1 < 0.5) = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}\,.$$

מחישוב דומה עבור $v_1 > 0.5$ ומשוואה 28 מתקבל:

$$P(\text{ניצחון}) = \frac{3}{4} \cdot \frac{1}{2} + \frac{3}{4} \cdot \frac{1}{2} = \frac{3}{4}\,.$$

פתרון 3: אם אתה בוחר את הקלף הראשון, $P(\text{ניצחון}) = v_1^2$ כי הקלף השני והשלישי חייבים להיות קטנים מהראשון.

אם אתה מוותר על הקלף הראשון ובוחר את השני כי $v_1 < v_2$ אזי:

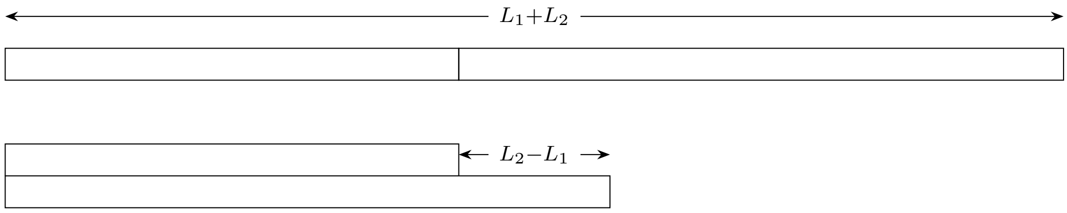
- $P(\text{ניצחון}) = v_1(1 - v_1)$ אם $v_1 < v_2$ (וותר על v_2) ו-$v_1 < v_3$ (בחר את v_3 ותנצח).
- $P(\text{ניצחון}) = (1 - v_1)v_1$ אם $v_1 > v_2$ (בחר את v_2) ו-$v_1 > v_3$ (תנצח כי $v_2 > v_3$).
- $P(\text{ניצחון}) = \frac{1}{2}(1 - v_1)^2$ אם $v_1 < v_2$ (בחר את v_2) ו-$v_1 < v_3$. הגורם $1/2$ לוקח בחשבון שניצחון תלוי ביחס: $v_2 > v_3$ (ניצחון) או $v_3 > v_2$ (הפסד).

הערך שווה-נפש d הוא ערך עבורו ההסתברות לנצח על ידי בחירת הקלף הראשון שווה להסתברות לנצח על ידי ויתור על הקלף הראשון:

$$\begin{aligned} d^2 &= 2d(1-d) + \frac{1}{2}(1-d)^2 \\ 5d^2 - 2d - 1 &= 0 \\ d &= \frac{1+\sqrt{6}}{5} \approx 0.6899\,. \end{aligned}$$

Gilbert&Mosteller [4, page 55] מראים שעבור $n = 3$:

$$P(\text{ניצחון}) = \frac{1}{3} + \frac{d}{2} + \frac{d^2}{1} - \frac{3d^3}{2} \approx 0.6617\,.$$



איור 14 : מדידת האורכים של שני מקלות

סימולציה

```
For  3 cards:
Indifference value = 0.6000
Probability of win = 0.6693
Proportion of wins = 0.6628
Indifference value = 0.6899
Probability of win = 0.6617
Proportion of wins = 0.6711
Indifference value = 0.7200
Probability of win = 0.6519
Proportion of wins = 0.6473
```

49. להכפיל את הדיוק (Doubling your accuracy)

נתון שני מקלות באורכים $L_1 < L_2$ ומכשיר למדידת אורך ששגיאת המדידה שלו ניתן על ידי התפלגות נורמלית [5] עם ממוצע 0 ושונות σ^2. ניתן למדוד את אורכי המקלות על ידי מדידת כל מקל בנפרד. האם יש דרך מדוייקת יותר?

פתרון

הנח את המקלות קצה לקצה ומדד $L_s = L_1 + L_2$ ואחר כך הנח אותם צד לצד ומדד $L_d = L_2 - L_1$ (איור 14). חשב L_1, L_2 :

$$\begin{aligned}\tfrac{1}{2}(L_s - L_d) &= \tfrac{1}{2}((L_1 + L_2) - (L_2 - L_1)) = L_1\\ \tfrac{1}{2}(L_s + L_d) &= \tfrac{1}{2}((L_1 + L_2) + (L_2 - L_1)) = L_2\,.\end{aligned}$$

השגיאות במדידות הן e_s, e_d כך שהשגיאות התוצאות הן :

$$\begin{aligned}\tfrac{1}{2}((L_s + e_s) - (L_d + e_d)) &= L_1 + \tfrac{1}{2}(e_s - e_d)\\ \tfrac{1}{2}((L_s + e_s) + (L_d + e_d)) &= L_2 + \tfrac{1}{2}(e_s + e_d)\,.\end{aligned}$$

[5] בעיה זו מניחה שהקורא מכיר התלפגויות נורמליות.

ממוצע של השגיאות במכשיר הוא 0 ולכן הממוצע של שתי המדידות גם כן 0. השונות יורדת למחצית מערכה הקודמת:[6]

$$\mathrm{Var}\left(\tfrac{1}{2}\left(e_s - e_d\right)\right) = \tfrac{1}{4}(\sigma^2 + (-1)^2\sigma^2) = \tfrac{1}{2}\sigma^2$$
$$\mathrm{Var}\left(\tfrac{1}{2}(e_s + e_d)\right) = \tfrac{1}{4}(\sigma^2 + \sigma^2) = \tfrac{1}{2}\sigma^2\,.$$

סימולציה

```
For L1 = 40, L2 = 16, variance = 0.50:
L1 mean = 39.9907, L1 variance = 0.2454
L2 mean = 16.0030, L2 variance = 0.2520

For L1 = 40, L2 = 16, variance = 1.00:
L1 mean = 39.9934, L1 variance = 0.4949
L2 mean = 15.9889, L2 variance = 0.4878

For L1 = 40, L2 = 16, variance = 2.00:
L1 mean = 39.9924, L1 variance = 0.9940
L2 mean = 16.0104, L2 variance = 1.0069
```

הממוצעים מדוייקים מאוד והשונויות יורדות למחצית מערכן הקודמת עבור $\sigma^2 = 0.5, \sigma^2 = 1.0$ אבל השונות לא מושפעות עבור $\sigma^2 = 2.0$.

50. משוואות ריבועיות אקראיות (Random quadratic equations)

תהי $x^2 + 2bx + c = 0$ משוואה ריבועית המוגדרת מעל $[-B, B] \times [-B, B]$ עבור $B \geq 1$.

שאלה 1: מה ההסתברות שהשורשים ממשיים?

שאלה 2: כאשר $B \to \infty$ מה ההסתברות שהשורשים ממשיים?

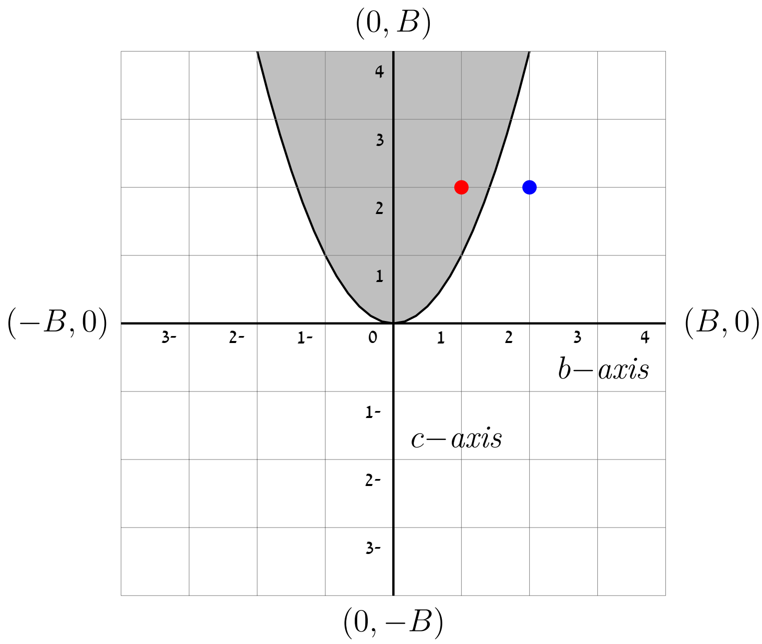
פתרון

פתרון 1: השורשים יהיו ממשיים אם הדיסקרימיננט $4b^2 - 4c \geq 0$ לא-שלילי. איור 15 מראה גרף של הפרבולה $c = b^2$ כאשר השורשים המרוכבים נמצא בשטח האפור. למשל, עבור $(b, c) = (1, 2)$, ל-$x^2 + 2x + 2$ שורשים מרוכבים (נקודה אדומה) ועבור $(b, c) = (2, 2)$ ל-$x^2 + 4x + 2$ שורשים ממשיים (נקודה כחולה).

נחשב את השטח האפור על ידי אינטגרציה:

$$\int_{-\sqrt{B}}^{\sqrt{B}} (B - b^2)\,db = Bb - \frac{b^3}{3}\bigg|_{-\sqrt{B}}^{\sqrt{B}} = \left(B^{3/2} - \frac{B^{3/2}}{3}\right) - \left(-B^{3/2} + \frac{B^{3/2}}{3}\right) = \frac{4}{3}B^{3/2}\,.$$

[6] המדידות בלתי-תלויות ולכן הקווריאנס הוא 0.



איור 15: עבור (b,c) בשטח האפור השורשים של $x^2+2bx+c$ מרוכבים

השטח הכולל של $[-B,B]\times[-B,B]$ הוא $4B^2$ ולכן:

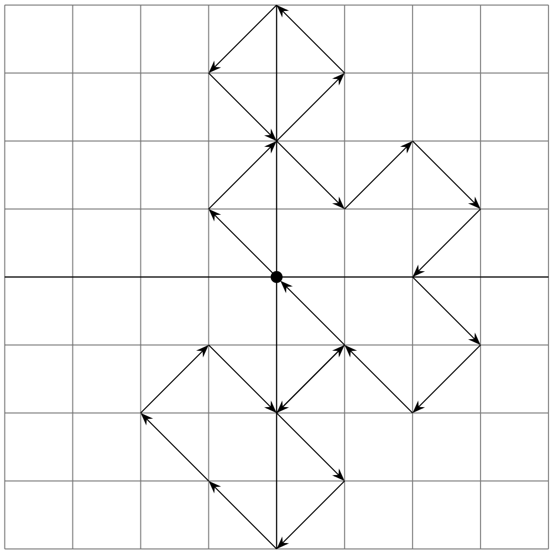
$$P(\text{שורשים מרוכבים}) = \frac{\frac{4}{3}B^{3/2}}{4B^2} = \frac{1}{3\sqrt{B}}$$
$$P(\text{שורשים ממשיים}) = 1-\frac{1}{3\sqrt{B}}\,.$$

פתרון 2:

$$\lim_{B\to\infty} P(\text{שורשים ממשיים}) = \lim_{B\to\infty}\left(1-\frac{1}{3\sqrt{B}}\right) = 1\,.$$

סימולציה

```
For B =  4:
Probability of real roots = 0.8333
Proportion real roots      = 0.8271
For B = 16:
Probability of real roots = 0.9167
Proportion real roots      = 0.9205
For B = 64:
Probability of real roots = 0.9583
Proportion real roots      = 0.9582
```



איור 16 : הילוך מקרי דו-ממדי

51. הילוך מקרי דו-ממדי (Two-dimensional random walk)

חלקיק נמצא במרכז של מערכת צירים דו-ממדית. החלקיק צועד ימינה או שמאלה על ציר ה-x עם הסתברות $1/2$ לכל כיוון **ובו-זמנית** צועד למעלה או למטה על ציר ה-y עם הסתברות $1/2$ לכל כיוון. איור 16 מראה הילוך מקרי של 22 צעדים שמתחיל ונגמר במרכז.

שאלה 1: מה ההסתברות שהחלקיק חוזר למרכז ב-2 צעדים?

שאלה 2: פתח נוסחה עבור התוחלת של מספר הביקורים של ההחלקיק במרכז.

שאלה 3: השתמש בקירוב של Stirling כדי לקבל הערכה של התוחלת של מספר הביקורים של ההחלקיק במרכז עבור n גדול.

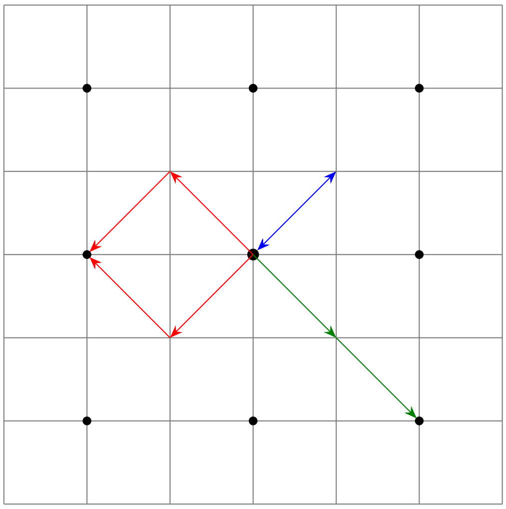
רמז: השתמש במשתנה מסמן כדי לחשב את התוחלת.

פתרון

פתרון 1: הנקודות באיור 17 מראות את המקומות האפשריים בהם החלקיק יכול להיות לאחר שני צעדים :

- המסלול הירוק מראה איך להגיע ל-$(\pm 2, \pm 2)$ על ידי שני צעדים באותו כיוון. ההסתברות היא $\left(\frac{1}{4}\right)^2 = \frac{1}{16}$.
- המסלול האדום מראה איך להגיע ל-$(\pm 2, 0)$ או ל-$(0, \pm 2)$. יש שני מסלולים אפשריים לכל נקודה ולכן ההסתברות היא $2 \cdot \left(\frac{1}{4}\right)^2 = \frac{2}{16}$.
- המסלול הכחול מראה איך להגיע ל-$(\pm 1, \pm 1)$ וחזרה למרכז. ההסבתרות היא $\frac{1}{16}$.

נסמן ב-$P_{2n}(x, y)$ ההסתברות שהחלקיק מגיע ל-(x, y) ב-$2n$ צעדים.



איור 17: שני צעדים בהילוך מקרי

ארבעת המסלולים הכחולים האפשריים הם היחידים שחוזרים למרכז ולכן:

$$P_2(0,0) = \frac{4}{16}\,.$$

פתרון 2: בחירת הכיוון בשני מצירים היא בלתי-תלויה:

$$P_{2n}(0,0) = P_{2n}(0,b) \cdot P_{2n}(a,0)\,, \tag{29}$$

כאשר a, b הם מספרים שלמים שרירותיים.

החלקיק יחזור למרכז אם ורק אם בשני הצירים מספר הצעדים $+1$ שווה המספר צעדים -1. יש $\binom{2n}{n}$ דרכים לסדר n צעדים של $+1$ ו-n צעדים של -1 ולכן:

$$P_{2n}(0,b) = P_{2n}(a,0) = \binom{2n}{n}\left(\frac{1}{2}\right)^n\left(\frac{1}{2}\right)^n$$

$$P_{2n}(0,0) = \left[\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right]^2\,. \tag{30}$$

הגדר משתנים מסמנים $I_{2n}(0,0)$ לחזרה למרכז ב-$2n$ צעדים ותהי $E(0,0)$ התוחלת של **מספר החזרות למרכז** במספר כלשהו של צעדים. ניתן לחשב:

$$E(0,0) = \sum_{n=1}^{\infty} E(I_{2n}(0,0))\,.$$

אפשר לשאול מה קורה אם החלקיק צעד שלושה צעדים וחזרה למרכז ושוב צועד שלושה צעדים וחזור למרכז. האם ערכו של $I_6(0,0)$ צריך להיות שניים ולא אחד? התשובה היא שהחזרה השניה מתרחשת ב-12 צעדים וייספר על ידי $I_{12}(0,0) = 1$.

ממשוואות 46, 47:

$$E(0,0) = \sum_{n=1}^{\infty} E(I_{2n}(0,0)) = E\left(\sum_{n=1}^{\infty} I_{2n}(0,0)\right) = \sum_{n=1}^{\infty} P_{2n}(0,0) = \sum_{n=1}^{\infty}\left[\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right]^2\,. \tag{31}$$

פתרון 3: לפי הקירוב של Stirling $n! \approx \sqrt{2\pi n}\,(n/e)^n$:

$$\begin{aligned} E_{2n}(0,0) &= \left[\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right]^2 \\ &= \left[\frac{(2n)!}{(n!)^2}\left(\frac{1}{2}\right)^{2n}\right]^2 \\ &\approx \left(\frac{1}{2}\right)^{4n}\frac{(\sqrt{2\pi\cdot 2n})^2\,(2n/e)^{4n}}{(\sqrt{2\pi n})^4\,(n/e)^{4n}} \\ &= \left(\frac{1}{2}\right)^{4n}\frac{4\pi n}{4\pi^2 n^2}\cdot\frac{(n/e)^{4n}\cdot 2^{4n}}{(n/e)^{4n}} \\ &= \frac{1}{\pi n} \\ E(0,0) &= \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\,, \end{aligned} \tag{32}$$

שהיא סידרה הרמונית שמתבדרת, כלומר, עם הסתברות 1 החלקיק חוזר למרכז!

בגלל ש-$E(0,0)=\infty$ מספר הפעמים שהחלקיק חוזר למרכז לא חסום. אולם, לפי האקסיומה הראשונה של הסתברות (עמוד 81), $P(0,0)$, ההסתברות שהחלקיק יחזור למרכז חייבת להיות $0 \leq P(0,0) \leq 1$ ולכן $P(0,0)=1$ שמשמעותה היא שבוודאות החלקיק חוזר למרכז. באופן כללי, התוחלת של משתנה אקראי היא אינסופית אם ורק אם ההסתברות היא אחד.

סימולציה

הרצתי את הסימולציה 100 עם מיליון צעדים בכל אחת.

```
Proportion returned to origin = 0.8700
```

ההסתברות שהחלקיק יחזור למרכז היא 1 ולכן התוצאה אמורה להיות קרוב ל-1.0000. המשמעות של התוצאה שקיבלתי היא שלמרות שהחלקיק יחזור למרכז, מספר הצעדים יכול להיות מאוד מאוד גדול.

52. הילוך מקרי תלת-ממדי (Three-dimensional random walk)^D

חלקיק נמצא במרכז של מערכת צירים תלת-ממדית. החלקיק צועד ימינה או שמאלה על ציר ה-x עם הסתברות $1/2$ לכל כיוון, **ובו-זמנית** צועד למעלה או למטה על ציר ה-y עם הסתברות $1/2$ לכל כיוון, **ובו-זמנית** צועד פנימה או החוצה על ציר ה-z עם הסתברות $1/2$ לכל כיוון.

שאלה 1: פתח נוסחה עבור התוחלת של מספר הפעמים שהחלקיק חוזר למרכז והשתמש בקירוב של Stirling כדי להעריך את ערכה.

רמז: פתח נוסחה להסתברות ואחר כך השתמש במשתנה מסמן.

שאלה 2: מה ההסתברות שהחלקיק יחזור למרכז **לפחות פעם אחת**?

רמז: השתמש בשיטה של בעיה 4.

פתרון

P_{2n}, ההסתברות לחזור למרכז לאחר $2n$ צעדים, נתון על ידי הכללת משוואה 29 לשלושה ממדים:

$$P_{2n} = P_{2n}(x=0\text{-חוזר ל}) \, P_{2n}(y=0\text{-חוזר ל}) \, P_{2n}(z=0\text{-חוזר ל}) \, . \tag{33}$$

$E(0,0)$, התוחלת של מספר הפעמים שהחלקיק חוזר למרכז, ניתנת על ידי הכללה של משוואה 31:

$$\begin{aligned}
E(0,0,0) &= \sum_{n=1}^{\infty} E(I_{2n}(0,0,0)) \\
&= E\left(\sum_{n=1}^{\infty} I_{2n}(0,0,0)\right) \\
&= \sum_{n=1}^{\infty} P_{2n}(0,0,0) \\
&= \sum_{n=1}^{\infty} \left[\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right]^3 .
\end{aligned}$$

מהקירוב של Stirling:

$$\begin{aligned}
P_{2n}(0,0,0) &= \left[\frac{(2n)!}{(n!)^2}\left(\frac{1}{2}\right)^{2n}\right]^3 \\
&\approx \left(\frac{1}{2}\right)^{6n} \frac{(\sqrt{2\pi\cdot 2n})^3\,(2n/e)^{6n}}{(\sqrt{2\pi n})^6\,(n/e)^{6n}} \\
&= \frac{1}{(\pi n)^{3/2}}
\end{aligned}$$

$$E(0,0,0) = \sum_{n=1}^{\infty} \frac{1}{(\pi n)^{3/2}} \approx 0.3772 \, . \tag{34}$$

Mosteller השתמש ב-18 איברים בחישוב שלו וקיבל 0.315. התכנית שלי השתמש ב-500 איברים וקיבלתי 0.3772.

שאלה 2: תהי P_1 ההסתברות שהחלקיק חוזר למרכז **לפחות פעם אחת**. מבעיה 4 אנו יודעים שהתוחלת של מספר הניסויים עד לראשון בו החלקיק **לא** חוזר למרכז היא $1/(1-P_1)$. לכן, התוחלת של מספר הניסויים עד שהחליק כן חוזר למרכז היא אחד פחות, כי החלקיק יכול לחזור למרכז מספר רב של פעמים עד שהוא לא חוזר [6]. מכאן ש:

$$\begin{aligned}
E(0,0,0) &= \frac{1}{1-P_1} - 1 \\
P_1 &= \frac{E(0,0,0)}{1+E(0,0,0)} \, .
\end{aligned}$$

בפתרון 1 חישבנו ש-$E(0,0,0) \approx 0.3772$ ולכן:

$$P_1 \approx 1 - \frac{1}{1+0.3772} \approx 0.2739 \, .$$

```
Expectation of reaching origin = 0.3772
Average times reached origin   = 0.3630
Probability of reaching origin = 0.2739
Proportion reached origin      = 0.2790
```

ממדים גדולים יותר: ניתן להכליל את משוואה 33 למספר כלשהו של ממדים וממשאוות 32, 34, סביר לשער ש-$E(0,0,0)$ **ביחס ישר** ל:

$$\sum_{n=1}^{\infty} \frac{1}{n^{d/2}}\,, \tag{35}$$

כאשר d הוא הממיד [1]. כעת נשתמש ב-*Cauchy condensation test* [14] על משוואה 35:

$$\sum_{n=1}^{\infty} \frac{1}{n^{d/2}}\,. \quad \text{מתכנסת אם ורק} \quad \sum_{n=1}^{\infty} \frac{2^n}{(2^n)^{d/2}} \quad \text{מתכנסת}$$

עבור $d=2$ התוצאה היא $\sum_{n=1}^{\infty} 1$ וברור שהיא מתבדרת.

עבור $d=3$, $E(0,0,0)$ מתכנסת כי:

$$\sum_{n=1}^{\infty} \frac{2^n}{(2^n)^{3/2}} = \sum_{n=1}^{\infty} \frac{2^n}{2^n \cdot 2^{n/2}} = \sum_{n=1}^{\infty} \frac{1}{(\sqrt{2})^n} = \frac{1}{\sqrt{2}-1} \approx 2.4\,.$$

עבור $d=4$, $E(0,0,0,0)$ מתכנסת כי $\sum_{n=1}^{\infty} \frac{1}{2^n} = 2$.

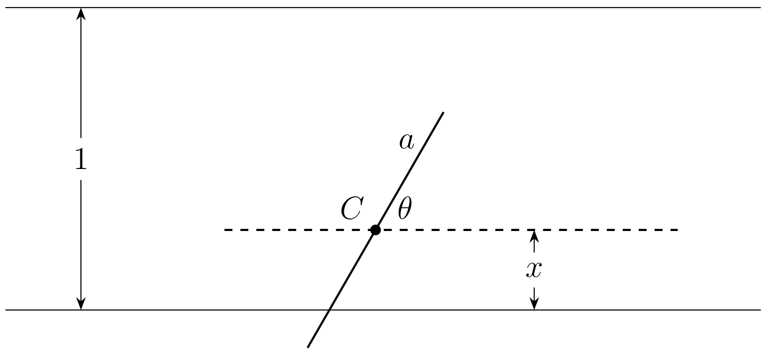
עבור ממדים גדולים יותר התוחלת של מספר החזרות למרכז הי סופית אבל ערכה יורדת, כך שיש פחות ופחות סיכוי שחלקיק יחזור למרכז בהילוך מקרי תלת-ממדי כלשהו.

53. המחט של Buffon (Buffon's needle) ^{D}

נתון משטח עם קווים מקבילים במרחק 1 אחד מהשני. קח מחט באורך $a \leq 1$ וזרוק אותו על המשטח. מה ההסתברות שהמחט חוצה קו?[7]

רמז: יש שני משתנים אקראיים (איור 18): x, המקום של מרכז המחט ביחס לקו הקרוב ביותר עם התפלגות אחידה בטווח $[0, 1/2]$, ו-θ, הזווית שבין המחט לבין הקווים המקבילים עם התפלגות אחידה בטווח $[0, \pi/2]$.

[7] Mosteller משתמש ב-l כאורך המחט וב-a כמחצית המרחק בין הקווים המקבילים. כדי להקל על החישובים אנו מניחים שהמרחק בין הקווים הוא 1. ניתן להתעלם מאפשרות שהמחט שוכב כולו לאורך אחד הקווים וכן את האפשרות שהוא נודע בשני קווים כי ההתסברות של המאורעות האלה היא אפס.



איור 18: המחט של Buffon

פתרון 1

תהי $p(a)$ ההסתברות שמחט באורך a חוצה קו והגדר משתנה מסמן:

$$I_{\text{חוצה קו}} = \begin{cases} 1, & \text{אם מחט באורך } a \text{ חוצה קו} \\ 0, & \text{אם מחט באורך } a \text{ חוצה לא קו}\,. \end{cases}$$

אזי:

$$E(I_{\text{חוצה קו}}) = 1 \cdot p(a) + 0 \cdot (1 - p(a)) = p(a)\,, \tag{36}$$

וניתן לחשב את ההסתברות על ידי חישוב התוחלת.

יהי m אנח לקווים המקביליים שעובר דרך מרכז המחט C ו-θ הזווית בין המחט לבין אחד מהקווים המקביליים. הטל את המחט על m כדי לקבל את הקטע הקו $\overline{CD}$ (איור 19). ההסתברות שהמחט חוצה קו היא:

$$P(\text{מחט באורך } a \text{ ובזווית } \theta \text{ חוצה קו}) = \frac{\overline{CE}}{1/2} = \frac{(a/2)\sin\theta}{1/2} = a\sin\theta\,. \tag{37}$$

התוחלת של מספר הקווים שהמחט חוצה מתקבלת על ידי אינטגרציה מעל לזוויות האפשריות:

$$E(\text{lines crossed}) = \frac{1}{(\pi/2) - 0}\int_0^{\pi/2} a\sin\theta\, d\theta = \frac{2}{\pi}\cdot a(-\cos\theta)\Big|_0^{\pi/2} = \frac{2a}{\pi}\,. \tag{38}$$

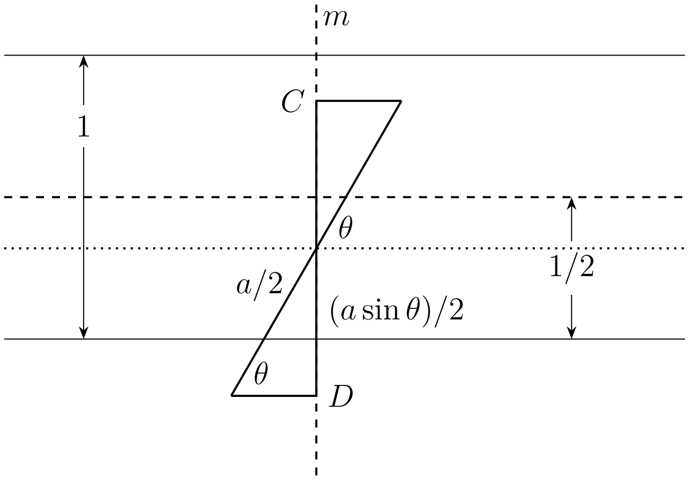
פתרון 2

הפתרון מבוסס על [2, Chapter 26].

תהי $E(x)$ התוחלת שמ מספר הקווים המקביליים שקו באורך x חוצה.

נתון מחט באורך a נשבור אותו למספר קטעים $\{a_1, \ldots, a_n\}$ ולפי הליניאריות של התוחלת:

$$E(a) = E\left(\sum_{i=1}^{n} a_i\right) = \sum_{i=1}^{n} E(a_i)\,,$$



איור 19 : משולש ישר-זווית לפתרון בעיית המחט של Buffon

ולכן לא משנה אם נחשב את התוחלת של כל קטע בנפרד. מכאן, שאם נכופף את המחט למעגל, התוחלת של מספר הקווים שהמעגל חוצה שווה למספר הקווים שהמחט חוצה.

נעיין בקו שמסובב למעגל C **בקוטר** 1 והיקף π. אם תזרוק את המעגל על המשטח הוא יחצה קו **בדיוק** פעמיים (איור 20), ולכן :

$$E(C) = 2\,. \tag{39}$$

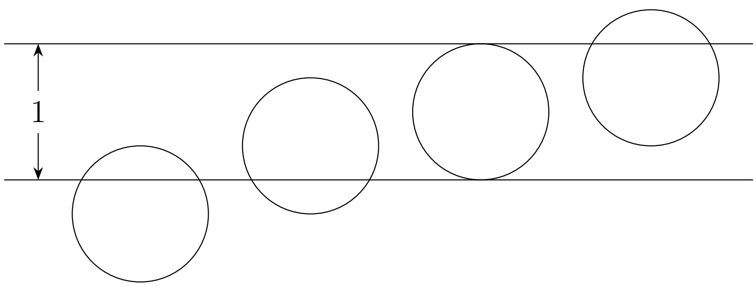
בנה מצולע משוכלל Q_n חסום על ידי c (אדום), ובנה מצולע משוכלל R_n שחוסם את c (כחול) (איור 21). כל קו ש-Q_n חוצה (אדום, מקווקו) חייב לחצות את המעגל וכל קו שחוצה את המעגל (כחול, מנוקד) חייב לחצות את R_n. לכן :

$$E(Q_n) \leq E(C) \leq E(R_n)\,. \tag{40}$$

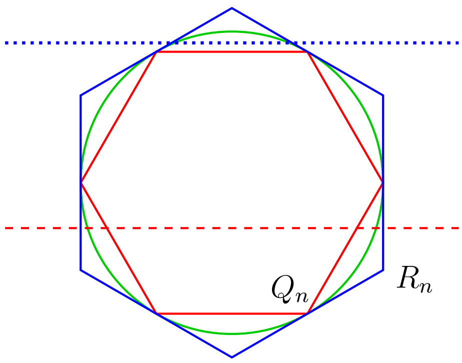
יהי a_Q, a_R סכומי האורכים של צלעות של Q_n, R_n, בהתאמה. לפי הליניאריות של התוחלת :

$$E(Q_n) = \sum_{i=1}^{n} E(a_Q \text{ צלעות של}) = a_Q E(1) \tag{41}$$

$$E(R_n) = \sum_{i=1}^{n} E(a_R \text{ צלעות של}) = a_R E(1)\,. \tag{42}$$



איור 20 : הפתרון של בעיית המחט של Buffon עם מעגלים



איור 21 : מצולעים כקירובים למעגל

כאשר $\infty \to n$ שני המצולעים הם קירובים למעגל ולכן :

$$\lim_{n\to\infty} a_Q = \lim_{n\to\infty} a_R = \pi\,, \tag{43}$$

ההיקף של המעגל. ממשוואות 41-43 מתקבל :

$$\begin{aligned}
&\lim_{n\to\infty} E(Q_n) = E(C) = \lim_{n\to\infty} E(R_n)\\
&E(C) = aE(1) = \pi E(1) = 2\\
&E(1) = \frac{2}{\pi}\\
&E(a) = aE(1) = \frac{2a}{\pi}\,.
\end{aligned}$$

סימולציה

$\pi = 2a/E$ ולכן ניתן לחשב קירוב לערכו על ידי הרצת סימולציה או זריקת מחטים על שולחן!

```
For length = 0.2:
Expectation of crossings = 0.1273
Average crossings        = 0.1308
Empirical value for pi   = 3.0581

For length = 0.5:
Expectation of crossings = 0.3183
Average crossings        = 0.3227
Empirical value for pi   = 3.0989

For length = 1.0:
Expectation of crossings = 0.6366
Average crossings        = 0.6333
Empirical value for pi   = 3.1581
```

54. המחט של Buffon עם רשת אופקי ואנכי
(Buffon's needle with horizontal and vertical rulings)

פתור את בעיית המחט של Buffon עבור משטח עם רשת אופקי ואנכי כאשר גודל המשבצות הוא 1×1. מחט יכול לחצות קו אנכי (כחול), קו אופקי (ירוק), שניהם (אדום) או אף אחד (כתום) (איור 22).

רמז: האם מספר הקווים האופקים והקווים האכנים שהמחט חוצה בלתי-תלויים?

פתרון

מספר הקווים האופקים והקווים האכנים שהמחט חוצה אכן בלתי-תלויים, ולפי הליניאריות של התוחלת:

$$\begin{aligned}
E(\text{קווים שמחט באורך } a \text{ חוצה}) &= E(\text{קווים אנכים שמחט באורך } a \text{ חוצה} + \\
&\quad \text{קווים אופקים שמחט באורך a חוצה}) \\
&\quad E(\text{קווים אנכים שמחט באורך } a \text{ חוצה}) + \\
&\quad E(\text{קווים אופקים שמחט באורך a חוצה}) \\
&= \frac{2a}{\pi} + \frac{2a}{\pi} = \frac{4a}{\pi}\,.
\end{aligned}$$

55. מחטים ארוכים (Long needles)[D]

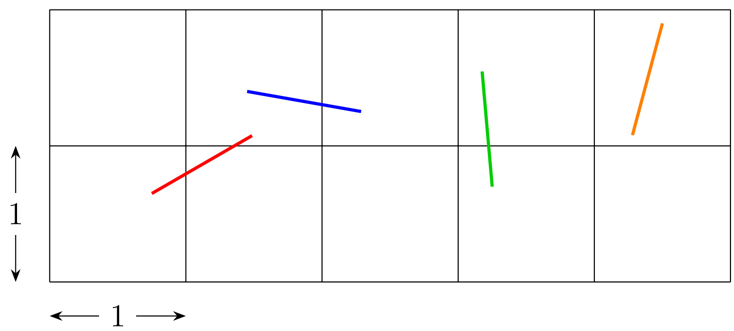
נתון מחט בבעיה של Buffon שאורכו a גדול מאחד.

שאלה 1: מה התוחלת של **מספר הקווים** שמחט חוצה?

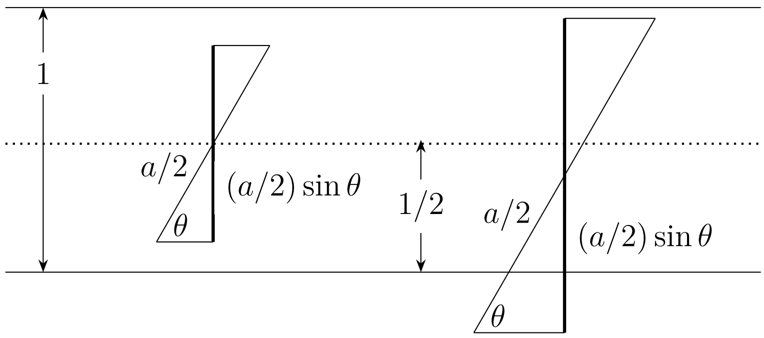
שאלה 2: פתח נוסחה עבור ההסתברות שהמחט חוצה **לפחות קו אחד**?

רמז: עבור איזו זוויות θ ההסתברות של חציית קו היא 1?

פתרון



איור 22: בעיית המחט של Buffon עבור משטח עם רשת אופקי ואנכי



איור 23 : מחטים ארוכים

פתרון 1: שבור את המחט לחלקים באורכים $\{a_1, a_2, \ldots, a_n\}$, $a_i < 1$, כך ש-$\sum_{i=1}^{n} a_i = a$. בפתרון של בעיה 53 ראינו ש :

$$E(a) = \sum_{i=1}^{n} E(a_i) = \frac{2a}{\pi}\,.$$

פתרון 2: הפתרון מבוסס על [13] ו-[2, Chapter 26].

לפי משוואה 37 ההסתברות שהמחט יחצה קו לפחות קו אחד היא $a \sin\theta$ **אם** $a\sin\theta \leq 1$, כלומר, אם $0 \leq \theta \leq \sin^{-1}(1/a)$. אולם אם $a\sin\theta > 1$ ההסתברות היא 1 (איור 23). נכליל את משוואה 38 עבור $a > 0$ שרירותי על ידי חלוקת האינטגרל לשני חלקים, אחד עבור $\theta < \sin^{-1}(1/a)$) ואחר עבור $\theta > \sin^{-1}(1/a)$):

$$\begin{aligned}
E(a) &= \frac{2}{\pi}\left(\int_0^{\sin^{-1}(1/a)} a\sin\theta\, d\theta + \int_{\sin^{-1}(1/a)}^{\pi/2} 1\, d\theta\right)\\
&= \frac{2}{\pi}\left(a(-\cos\theta)\Big|_0^{\sin^{-1}(1/a)} + \left(\frac{\pi}{2} - \sin^{-1}(1/a)\right)\right)\\
&= 1 + \frac{2}{\pi}\left(a\left(1 - \sqrt{1 - \frac{1}{a^2}}\right) - \sin^{-1}(1/a)\right)\,.
\end{aligned}$$

סימולציה

```
For length = 1.5:
Expectation of crossings = 0.7786
Average crossings        = 0.7780
For length = 2.0:
Expectation of crossings = 0.8372
Average crossings        = 0.8383
For length = 3.0:
Expectation of crossings = 0.8929
Average crossings        = 0.8897
```

56. הכד של Molina **(Molina's urns)**

שני כדים U_1, U_2 מכילים m כדורים כל אחד. ב-U_1 נמצאים w_1 כדורים לבנים ו-b_1 כדורים שחורים, וב-U_2 נמצאים w_2 כדורים לבנים ו-b_2 כדורים שחורים. מכל כד שלוף n כדורים **עם החזרה**. עבור ערכים שונים של $n > 1$ מצא w_1, b_1, w_2, b_2 כך ש:

$$P(\text{כל כדורים שנשלפו מ-}U_1\text{ לבנים}) = P(\text{כל כדורים שנשלפו מ-}U_2\text{ לבנים או שחורים}) .$$

פתרון

עבור $n = 2$ המשוואה שיש לפתור היא:

$$\left(\frac{w_1}{m}\right)^2 = \left(\frac{w_2}{m}\right)^2 + \left(\frac{b_2}{m}\right)^2$$
$$w_1^2 = w_2^2 + b_2^2 ,$$

וכל שלשת Pythagorean היא פתרון.

לפי המשפט האחרון של Fermat, שהוכח ב-1995 על ידי Andrew Wiles, אין פתרונות ל-$w_1^n = w_2^n + b_2^n$ עבור $n \geq 3$.

סימולציה

הרצתי את הסימולציה עבור $n = 2$ ומספר שלשות Pythagorean.

```
For w1 = 17, w2 = 8, b2 =15:
Proportion of two whites in urn 1          = 0.5523
Proportion of two whites or black in urn 2 = 0.5387
For w1 = 29, w2 = 20, b2 =21:
Proportion of two whites in urn 1          = 0.5003
Proportion of two whites or black in urn 2 = 0.5026
For w1 = 65, w2 = 33, b2 =56:
Proportion of two whites in urn 1          = 0.5381
Proportion of two whites or black in urn 2 = 0.5384
```

סקירה על הסתברות

סעיף זה סוקר מושגים בהתסבתרות. אביא דוגמה של כל מושג עבור הטלת קוביה הוגנת עם שש פאות.

ניסוי (trial) מושג לא מוגדר כאשר הכוונה היא לפעולה שיש לה תוצאות אפשריות.

תוצאה (outcome) התוצאה של ניסוי. אם אתה מטיל קוביה אחת תוצאה אפשרית היא 4.

מרחב מדגם (sample space) קבוצת כל התוצאות האפשריות של ניסוי. הקבוצה $S = \{1, 2, 3, 4, 5, 6\}$ היא מרחב המדגם של הטלת קוביה.

מאורע (event) תת-קבוצה של מרחב מדגם. תת-הקבוצה $e = \{2, 4, 6\} \subseteq S$ היא המאורע של הופעת מספר זוגי בהטלת קוביה.

משתנה אקראי (random variable) פונקציה ממרחב המדגם למספרים. יהי T מרחב המדגם של של הזוגות (הסדורות) הם התוצאות של הטלת זוג קוביות:

$$T = \{(a, b) | a, b \in \{1, 2, 3, 4, 5, 6\}\} .$$

הגדר משתנה אקראי X כפונקציה $X : T \mapsto \{2, 3, \ldots, 11, 12\}$ שממפה תוצאות של הטלת זוג קוביות לסכום המספרים על הקוביות:

$$X((a, b)) = a + b . \tag{44}$$

איחוד, חיתוך, משלים (union, intersection, complement) מאורעות הם קבוצות ולכן למושגים הללו יש את המשמעות הרגילה בתורת הקבוצות. יהי $e_1 = \{2, 4, 6\}$ ו-$e_2 = \{1, 2, 3\}$. אזי:

$$e_1 \cup e_2 = \{1, 2, 3, 4, 6\} \quad e_1 \cap e_2 = \{2\} \quad \overline{e_1} = S \setminus e_1 = \{1, 3, 5\} .$$

החיתוך הוא קבוצת המספרים הזוגיים מתוך שלושת האיברים הראשונים במרחב המדגם. המשלים הוא קבוצת המספרים האי-זוגיים מתוך מרחב המדגם.

זרים (mutually exclusive) שני מאורעות או יותר זרים זה לזה אם החיתוך שלהם הוא הקבוצה הריקה. שני המאורעות $e_1 = \{2, 4, 6\}$ ו-$e_2 = \{1, 3, 5\}$ זרים זה לזה כי $e_1 \cap e_2 = \emptyset$, כלומר, אין תוצאות שהן מספרים שהם גם זוגיים וגם אי-זוגיים.

הסתברות (probability) המשמעות האינטואיטיבית של הסתברות היא הגבול של התדירות היחסית של מאורע. יהי e מאורע ויהי n_e מספר הפעמים שהמאורע e מתרחש ב-n חזרות על הניסוי. $P(e)$, ההסתברות של המאורע e, היא:

$$P(e) = \lim_{n \to \infty} \frac{n_e}{n} .$$

הגדרה זו היא בעייתית כי אנחנו לא ממש יודעים אם הגבול קיים. ההגדרה תלויה על "חזרות על הניסוי" אולם אנו רוצים להגדיר הסתברות ללא קשר לסדרה מסוימת של ניסויים. **חוק המספרים הגדולים** מבטיח שהתפיסה האינטואיטיבית של הסתברות כי תדירות יחסית קרא מאוד למה שקורה כאשר חוזרים על ניסוי מספר רק של פעמים.

התיאוריה המודרנית של הסתברות מבוססת על שלוש אקסיומות שהן די אינטואיטיביות:

- עבור מאורע e, $P(e) \geq 0$.

- עבור כל התוצאות האפשריות במרחב S, $P(S) = 1$.
- עבור קבוצה של מאורעות זרים זה לזה $\{e_1, \ldots, e_n\}$:

$$P\left(\bigcup_{i=1}^{n} e_i\right) = \sum_{i=1}^{n} P(e_i)\,.$$

החזרה (replacement) בעיה שכיחה בהסתברות היא לשאול שאלות על שליפת כדור צבעוני מכד. חשוב שהבעיה תציין אם השליפה היא עם או בלי החזרה: לאחר שליפת הכדור האם מחזירים אותו לכד או לא לפני השליפה הבאה? נשלוף שני כדורים מכד עם שלושה כדורים אדומים ושלושה שחורים. למאורע שהשליפה של הכדור הראשונה שולף כדור אדום הסתברות של $\frac{1}{2} = \frac{3}{3+3}$. אם מחזירים את הכדור לפני השליפה השניה אזי ההסתברות שהשליפה של הכדור השני שולף כדור האדום נשארת $\frac{1}{2}$ ולכן ההסתברות ששני הכדורים הם אדומים היא $\frac{1}{4}$. אם לא מחזירים את הכדור ההסתברות שהכדור השני הוא אדום יורדת ל-$\frac{2}{5} = \frac{2}{2+3}$, ולכן ההסתברות ששני הכדורים הם אדומים היא $\frac{1}{4} = \frac{2}{5} \cdot \frac{1}{2}$.

התפלגות אחידה (uniformly distributed) אם הסתברויות של כל התוצאת במרחב שוות להסתברות התפלגות אחידה. אם S היא קבוצה סופית ולהסתברות שלה התפלגות אחידה אזי:

$$P(e) = \frac{|e|}{|S|}\,.$$

אם אתה מטיל קוביה **הוגנת** ההסתברות של התוצאות מתפלגת אחידה ולכן עבור $e = \{2,4,6\}$:

$$P(e) = \frac{|e|}{|S|} = \frac{|\{2,4,6\}|}{|\{1,2,3,4,5,6\}|} = \frac{1}{2}\,.$$

הסתברות מותנית (conditional probability) יהי e_1, e_2 מאורעות. $P(e_1|e_2)$, ההסתברות המותנית ש-e_1 מתרחש אם נתון ש-e_2 מתרחש, נתונה על ידי:

$$P(e_1|e_2) = \frac{P(e_1 \cap e_2)}{P(e_2)}\,.$$

יהי $e_1 = \{1,2,3\}$ המאורע שקוביה מראה מספר פחות או שווה ל-3 ויהי $e_2 = \{2,4,6\}$ המאורע שהקוביה מראה מספר זוגי. אזי:

$$P(e_2|e_1) = \frac{P(E_2 \cap E_1)}{P(e_1)} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{1/6}{1/2} = \frac{1}{3}\,.$$

אם אתה יודע שמספר הוא פחות או שווה ל-3, רק אחת משלושת התוצאה היא מספר זוגי.

בלתי-תלוי (independence) שני מאורעות בלתי-תלויים אם ההסתברות של החיתוך שלהם היא המכפלה של ההסתברויות הנפרדות:

$$P(e_1 \cap e_2) = P(e_1)\,P(e_2)\,.$$

במונחים של הסבתרות מותנית:

$$P(e_1|e_2) = \frac{P(e_1) \cap P(e_2)}{P(e_2)} = \frac{P(e_1)\,P(e_2)}{P(e_2)} = P(e_1)\,.$$

עבור מאורעות בלתי-תלויים e_1, e_2, ידיעה של ההסתברות של e_2 לא מספק מידע על ההסתברות של e_1. שלוש הטלות של קוביה הוגנת בלתי-תלויות ולכן ההסתברות שכולן מראות מספר זוגי היא $\frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$.

ממוצע (average) תהי $S = \{a_1, \ldots, a_n\}$ קבוצה של ערכים. אזי:

$$Average(S) = \frac{\sum_{i=1}^{n} a_i}{n} .$$

ממוצע מחושב מעל לקבוצה של ערכים אבל הממוצע לא חייב להיות איבר בקבוצה. אם יש 1000 משפחות בעיירה ולהן 3426 ילדים, הממוצע של מספר הילדים למשפחה היא 3.426 למרות שברור שאין משפחה עם 3.426 ילדים. אם אתה מטיל קוביה שש פעמים ומקבל את המספרים $\{2, 2, 4, 4, 5, 6\}$ הממוצע הוא:

$$\frac{2+2+4+4+5+6}{6} = \frac{23}{6} \approx 3.8 ,$$

שוב, לא איבר בקבוצה.

תוחלת (expectation) התוחלת של משתנה אקראי היא סכום ההסתברויות של כל תוצאה כפול הערך של משתנה האקראי עבור אותה תוצאה. עבור קוביה הוגנת לכל תוצאה יש הסתברות זהה ולכן:

$$E(\text{ערך קוביה}) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5 .$$

השתנה האקראי X ממשוואה 44 ממפה את המספרים המופיעים על זוג קוביות לסכום המספרים. ההסתברות של כל זוג היא $1/36$, אבל לזוגות $(2, 5)$ ו-$(5, 2)$ אותו סכום ולכן הם שייכים לאותה תוצאה. הערכים של המשתנה האקראי הם $\{2, \ldots, 12\}$ ומספר הדרכים לקבל כל אחד מהן הם:

סכום	2	3	4	5	6	7	8	9	10	11	12
זוגות	1	2	3	4	5	6	5	4	3	2	1

התוחלת היא הממוצע של ערכי המשתנה האקראי כפול **המשקל** שהוא ההסתברות של כל תוצאה. יהי E_s התוחלת של סכום הערכים כאשר מטילים זוג קוביות. אזי:

$$E_s = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + 4 \cdot \frac{3}{36} + \cdots + 10 \cdot \frac{3}{36} + 11 \cdot \frac{2}{36} + 12 \cdot \frac{1}{36} = 7 . \quad (45)$$

עבור קבוצה שרירותית של מאורעות $\{e_1, \ldots, e_n\}$ התוחלת היא:

$$E = \sum_{i=1}^{n} e_i P(e_i) .$$

ליניאריות של התוחלת (linearity of expectation)

נעיין שוב בתוחלת של הסכום של זוג קוביות (משוואה 45). תהי $E(e_6)$ התוחלת של המאורע שסכום הקוביות הוא 6. אזי:

$$E(e_6) = X(e_6)P(e_6) = 6 \cdot \frac{5}{36} ,$$

כי יש 5 זוגות מתוך 36 הזוגות האפשריים שסכומם 6: $(1,5), (2,4), (3,3), (4,2), (5,1)$. אבל ניתן לחשב את התוחלת כדלקמן כאשר $P(i, j)$ היא ההסתברות שהזוג (i, j) מופיע ו-$E(i, j)$ היא התוחלת

של $i+j$:

$$\begin{aligned}E(X(e_6)) &= 6\cdot P(1,5)+6\cdot P(2,4)+6\cdot P(3,3)+6\cdot P(4,2)+6\cdot P(5,1)\\ &= (1+5)\cdot\tfrac{1}{36}+(2+4)\cdot\tfrac{1}{36}+(3+3)\cdot\tfrac{1}{36}+(4+2)\cdot\tfrac{1}{36}+(5+1)\tfrac{1}{36}\\ &= E(1,5)+E(2,4)+E(3,3)+E(4,2)+E(5,1)\\ &= \sum_{i,j|i+j=6} E(i,j)\,.\end{aligned}$$

החישוב תלוי בעובדה שהמאורעות בלתי-תלויים כי ברור ש-$(2,4)$ ו-$(3,3)$ לא יכולים להופיע באותו ניסוי.

באותה שיטה ניתן להוכיח את ההכללה ([12, Section 4.9]):

$$E\left(\sum_{i=1}^{n} a_i e_i\right)=\sum_{i=1}^{n} a_i E(e_i)\,,$$

שנקראת הליניאריות של התוחלת. במקרה של שני משתנים אקראיים:

$$E(ae_1+be_2)=aE(e_1)+bE(e_2)\,.$$

משתנה מסמן (indicator variable) יהי e מאורע שההסתברות שלה היא $P(e)$. הגדר I_e, משתנה מסמן עבור e, כך [12, Chapter 4, Example 3b]:

$$I_e=\begin{cases}1, & \text{אם } e \text{ מתרחש}\\ 0, & \text{אם } e \text{ לא מתרחש}\,.\end{cases}$$

מכאן ש:

$$E(I_e)=1\cdot P(e)+0\cdot(1-P(e))=P(e)\,.$$

ניתן להכליל משוואה זו. נתונה קבוצה של מאורעות $\{e_1,\ldots\}$ והמשתנים המסמנים שלהם $\{I_1,\ldots\}$:

$$E\left(\sum_{i=1}^{\infty} I_i\right)=\sum_{i=1}^{\infty} I_i p(e_i)=\sum_{i=1}^{\infty}(1\cdot p(e_i)+0\cdot(1-p(e_i)))=\sum_{i=1}^{\infty} p(e_i)\,. \tag{46}$$

בנוסף:

$$\sum_{i=1}^{\infty} E(I_i)=E\left(\sum_{i=1}^{\infty} I_i\right)\,. \tag{47}$$

ההוכחה של נוסחה זו קשה והנוסחה תקיפה כאן בגלל שהמשתנים באקראיים לא שליליים.

משפט הבינום (binomial theorem) אם p היא ההסתברות של מאורע e אזי ההסתברות שהתוצאה של סדרה של n ניסויים בלתי-תלויים היא **בדיוק** k מאורעות e ניתנת על ידי **המקדם הבינומי (binomial coefficient)**:

$$\binom{n}{k}p^k(1-p)^{n-k}\,.$$

בהכללה, ההסתברות ש-e מתרחוש בין i ל-j פעמים היא:

$$\sum_{k=i}^{j}\binom{n}{k}p^k(1-p)^{n-k}\,.$$

לפי משפט הבינום:

$$\sum_{i=0}^{n}\binom{n}{i}x^i y^{n-i} = (x+y)^n$$
$$\sum_{i=0}^{n}\binom{n}{i}p^i (1-p)^{n-i} = (1+(1-p))^n = 1\,,$$

כפי שאפשר לצפות כי אחת התוצאות חייבת להתרחש.

סכום סדרה הרמונית (sum of a harmonic series) עבור n מספר שלם חיובי, הסדרה ההרמונית היא:

$$H_n = \sum_{k=1}^{n}\frac{1}{k} \approx \ln n + \frac{1}{2n} + \gamma\,,$$

כאשר $\gamma \approx 0.5772$ הוא **הקבוע של Euler (Euler's constant)**. כאשר n שואף לאינסוף הסדרה מתבדרת:

$$\sum_{k=1}^{\infty}\frac{1}{k} = \infty\,,$$

כי $\ln n$ אינו חסום.

הקירוב של Stirling (Stirling's approximation) קשה מאוד לחשב $n!$ עבור n גדול. נוח להשתמש באחת הנוסחאות של הקירוב של Stirling:

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$
$$\ln(n!) \approx n\ln n - n$$
$$\ln(n!) \approx n\ln n - n + \frac{1}{6}\left(8n^3+4n^2+n+\frac{1}{30}\right) + \frac{1}{2}\ln\pi\,.$$

התפלגות הסתברותית רציפה(Continuous probability distribution) התפלגויות הסתברות רציפות בדרך כלל לא מופיעות בספר אבל עבור קוראים עם הרקע המתאים אנו סורקים את המושגים הבסיסיים.

ניתן להגדיר הסתברויות מעל למשתנים אקראים רציפים. **פונקציית הסתברות צפיפות (probability density function (PDF))** $f(x) : \mathcal{R} \rightarrow \mathcal{R}$ ממפה תוצאה x לערך של הפונקציה וכך להגדיר:

$$P(x) = f(x)\,.$$

הסיבות למונח זו היא שההסתברות של ההופעה של כל מספר ממשי **בודד** היא אפס, ולכן הדרך הנכונה היא לתת הסתברויות לקטעים:

$$P(a < x < b) = \int_a^b f(x)\,dx\,.$$

האינטגרל הוא גם $P(a \le x \le b)$ כי ההסתברות של נקודות בודדות היא אפס.

כמו כל הגדרה של הסתברות, $P(x) \ge 0$ לכל x, וכן:

$$\int_{-\infty}^{\infty} P(x)\,dx = \int_{-\infty}^{\infty} f(x) = 1\,dx\,.$$

אם ערכו של האיטגרל אינו 1 חייבים להשתמש **בקבוע נירמול (normalization constant)**. אם ה-PDF מתפלגת אחידה בקטע $[a,b]$ אזי:

$$P(a \leq x \leq b) = \int_a^b 1\,dx = (b-a)\,,$$

ולכן חייבים להגדיר:

$$P(a \leq x \leq b) = \frac{1}{b-a}\int_a^b 1\,dx = \frac{1}{b-a}\cdot(b-a) = 1\,.$$

ניתן לחשב את התוחלת על ידי אינטגרציה של ה-PDF $f(x)$ כפול x:

$$E(x) = \int_{-\infty}^{\infty} x f(x)\,dx\,.$$

התפלגות הסתברות מצטברת (cumulative probability distribution (CPD)) עבור הקטע $[-\infty, a]$ מתקבלת על ידי אינטגרציה של ה-PDF:

$$P(x<a) = \int_{-\infty}^{a} f(x)\,dx\,.$$

ניתן לקבל את ה-PDF על ידי גזירה של ה-CPD:

$$P(x<a) = \frac{d}{da} CDP(x<a)\,.$$

מקורות

[1] Louigi Addario-Berry. When do 3D random walks return to their origin? MathOverflow. https://mathoverflow.net/q/45174.

[2] Martin Aigner and Günter M. Ziegler. Proofs from THE BOOK (Fifth Edition). Springer, .2014

[3] Matthew Carlton. Pedigrees, prizes, and prisoners: The misuse of conditional probability. Journal of Statistics Education, ,(2)13 .2005 https://doi.org/10.1080/10691898.2005.11910554.

[4] John P. Gilbert and Frederick Mosteller. Recognizing the maximum of a sequence. Journal of the American Statistical Association, ,73--35:(313)61 .1966

[5] Markus C. Mayer. Average distance between random points on a line segment. Mathematics Stack Exchange. https://math.stackexchange.com/q/1540015.

[6] Aaron M. Montgomery. Mosteller's solutions to random-walk problems. Mathematics Stack Exchange. URL: https://math.stackexchange.com/q/4460054.

[7] David S. Moore. A generation of statistics education: An interview with Frederick Mosteller. Journal of Statistics Education, ,(1)1 .1993 https://www.tandfonline.com/doi/pdf/10.1080/10691898.1993.11910453.

[8] Frederick Mosteller. Understanding the birthday problem. The Mathematics Teacher, ,325--322:(5)55 .1962

[9] Frederick Mosteller. Fifty Challenging Problems in Probability with Solutions. Dover, .1965

[10] Frederick Mosteller, Stephen E. Fienberg, and Robert E. K. Rourke. Beginning Statistics with Data Analysis. Addison-Wesley, .1983

[11] Frederick Mosteller, Robert E. K. Rourke, and George B. Thomas Jr. Probability With Statistical Applications. Addison-Wesley, .1961

[12] Sheldon Ross. A First Course in Probability (Tenth Edition). Pearson, .2019

[13] Wikipedia. Buffon's needle problem.

[14] Wikipedia. Cauchy condensation test.