בעית יום ההולדת והתאמתה לתופעות נוספות

מאת: אורי גרינבלט, קבוץ מעברות

1. בעית יום ההולדת ופתרונה

בחדר נמצאים x אנשים. מה ההסתברות שיהיו ביניהם לפחות שניים בעלי אותו יום הולדת?

כדי לפתור את הבעיה, נחשוב על הבעיה המשלימה: מה ההסתברות שבקבוצה של x אנשים לא יהיו שניים בעלי אותו יום הולדת? אם ההסתברות במקרה זה היא Q הרי שההסתברות שימצאו לפחות שניים בעלי אותו יום הולדת היא $P = 1 - Q$.

נפתור תחילה את הבעיה השנייה. לאדם הראשון - יש 365 ימים "לבחור" מתוכם יום הולדת. לאדם השני יש 364 ימים מתוך 365. ההסתברות שלא יהיה לשניהם אותו יום הולדת היא $\frac{364}{365}$. לאדם השלישי, נותרו 363 ימים מתוך 365, כדי שלא יהיה לו יום הולדת יחד עם אחד משני קודמיו, וההסתברות לכך היא $\frac{363}{365} \cdot \frac{364}{365}$. לאדם הרביעי, נותרו 362 ימים, וההסתברות היא $\frac{362}{365} \cdot \frac{363}{365} \cdot \frac{364}{365}$, וכן הלאה.

באופן כללי: לאדם ה-x (x טבעי, קטן מ- 365) נותרו $365-(x-1)$ ימים, כך שלא יהיה לו יום הולדת משותף עם אחד מקודמיו. מכאן שההסתברות שבקבוצה של x אנשים לא יימצאו שניים בעלי אותו יום הולדת היא:

$$Q_{365,x} = \frac{364\cdot 363\cdot 362\cdot \ldots \cdot[365-(x-1)]}{365\cdot 365\cdot 365\cdot \ldots \cdot 365} = \frac{365!}{(365-x)!\cdot 365^x}$$

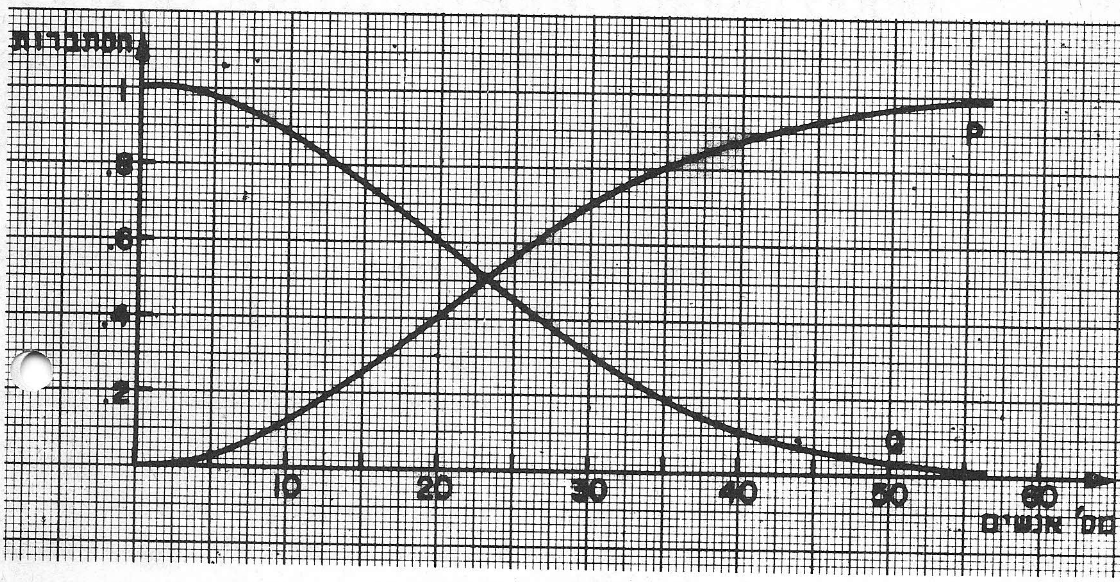
כאמור, ההסתברות $P_{365,x}$ שבקבוצה של x אנשים ימצאו לפחות שניים בעלי אותו יום הולדת היא:

$$P_{365,x} = 1 - Q_{365,x}$$

נביא כאן טבלה הנותנת את ערכי Q, P עבור ערכים שונים של x (הטבלה לקוחה מתוך ספרו של E. Parzen [1]):

x	4	12	16	20	22	23	24	28	32	40	48	56
Q	0.984	0.833	0.716	0.589	0.524	0.493	0.462	0.346	0.247	0.109	0.039	0.012
P	0.016	0.167	0.284	0.411	0.476	0.507	0.538	0.654	0.753	0.891	0.961	0.988

נצייר גם גרף הממחיש את אופי ההשתנות של P ו- Q כאשר x משתנה:



המפתיע בתוצאות אלו הוא המספר הקטן יחסית של אנשים שצריכים להיות בקבוצה על מנת שתהיה הסתברות מעל 0.5 שיהיו ביניהם לפחות שניים בעלי אותו יום הולדת. מתוך הטבלה אנו רואים שמספר זה הוא רק 23.

אם מספר האנשים בקבוצה עולה על 50, נוכל לומר בבטחון כמעט מוחלט כי ישנם לפחות שניים ביניהם בעלי אותו יום הולדת.

מכיוון שכיתת בי"ס מונה בדרך כלל מעל 25 תלמידים, מאלף הדבר לנסות ולבדוק תופעה זו בבית הספר. סביר מאד שנקבל כי מעל מחצית הכיתות שיש בהן 25 תלמידים ומעלה, כוללות בתוכן ל פ ח ו ת שני ילדים בעלי אותו יום הולדת (כולל תאומים!).

אנו מניחים כאן שיום הולדת היא תופעה אקראית ולכל יום בשנה יש אותה הסתברות לשמש כיום הולדת. ברור שאם היינו שואלים על יום מסויים בשנה, למשל: מה ההסתברות שימצאו בקבוצה שני אנשים אשר יום הולדתם חל ב- 1 באפריל? התשובה היתה שונה לגמרי (מהי?).

2. בעיות דומות

ננסה למצוא בעיות אחרות הנפתרות בשיטה דומה. נרשום לדוגמה את מספריהן של x מכוניות הנוסעות בכביש, ונשאל: מהי ההסתברות שתהיינה ביניהן שתיים כאלה, שמספריהן יהיו זהים בשלוש הספרות האחרונות? במילים אחרות, בהנחה שמספרי הרכב הם אקראיים, אנו שואלים מה ההסתברות לכך שבין x מספרים אקראיים בני שלוש ספרות יהיו שניים שווים.

חישוב ההסתברות $P_{1000,x}$ נעשה באופן דומה, ומתברר שכאשר נרשום את המספרים של 38 מכוניות, יש הסתברות של מעל 0.5 לכך שתהיינה לפחות שתיים בעלות התכונה הנ"ל.

כדוגמה שנייה לבעיה דומה לבעיית יום ההולדת, נוכל לקחת שבר עשרוני אין-סופי לא-מחזורי (כדוגמת $\sqrt{2}$ או π) ולשאול: אם נבחר באקראי x ספרות רצופות בשבר זה, מה ההסתברות שתהיינה ביניהן שתי ספרות שוות? כאן אנו שואלים מהי ההסתברות שבין x ספרות אקראיות (מתוך 10) תהיינה שתיים שוות. מתברר שמספיקות 4 ספרות על מנת להגיע להסתברות של 0.5.

נרשום למשל את המספר π עד 20 ספרות: 3 . 1 4 1 5 9 2 6 5 3 5 8 9 7 9 3 2 3 8 4 . אם ניצור את כל 17 הקבוצות של 4 ספרות שכנות, נקבל שב- 8 מתוכן (3141, 1415, 6535, 5358, 8979, 9793, 9323, 3238) יש שתי ספרות שוות. אם נשאל אותה שאלה ביחס לקבוצות בנות 5 ספרות נקבל שההסתברות היא 0.7. ואמנם, אם ניצור את 16 הקבוצות של ספרות שכנות, נמצא שב- 12 מתוכן יש שתי ספרות שוות. מובן שאפשר לבחור קבוצות ספרות גם בדרך אקראית אחרת, לאו דווקא ספרות רצופות.

דוגמה משטח אחר היא: מה ההסתברות שבקבוצה של x אנשים יהיו שניים ששמם מתחיל באותה אות? או: נגמר באותה אות, מכיל אותה אות, מכיל אותו זוג אותיות, מכיל אותו מספר אותיות, נותן אותו סכום בגימטריא (כנראה, ששמו של מלך פרס - ארתחשסתא - הוא בעל הסכום הגבוה ביותר - 1370), וכו'. בכל המקרים הללו, הנחת האקראיות בוודאי אינה בדוקה, אך חקירתם מביאה לתוצאות מעניינות.

באופן כללי: נניח שברשותנו x כדורים שיש לחלקם באקראי ל-n תאים, וקיים $n > x$ (מדוע יש לדרוש זאת?). נוכל להגיד שההסתברות $P_{n,x}$ שיהיו לפחות שני כדורים באותו תא היא:

$$P_{n,x} = 1 - Q_{n,x} = 1 - \frac{n!}{(n-x)!} \cdot \frac{1}{n^x}$$

3. נוסחות קירוב

בשאלה המקורית דנו במיוחד בבעיה של הקשר בין x ל- n כאשר הערך של P (או Q) הוא 0.5. על מנת לחשב הלכה למעשה את ההסתברות המבוקשת בעזרת הנוסחה הנ"ל, מתברר כי יש לבצע חישובים מייגעים למדי, בגלל חישוב העצרת. אף בעזרת מחשב כיס פשוט העבודה אינה קלה ביותר (אם אין בו פונקציה זו). נוכל על כן להשתמש כאן בנוסחת קירוב שהוכיח המתימטיקאי סטירלינג(*):

$$n! \simeq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

(את הוכחתה אפשר למצוא בספרי מתמטיקה מתקדמים, למשל בספרו של R. Courant, [2]).
בעזרת נוסחת סטירלינג נקבל כי:

$$Q_{n,x} \simeq \frac{1}{e^x}\left(\frac{n}{n-x}\right)^{n-x+\frac{1}{2}}$$

(*) ג'יימס סטירלינג היה מתמטיקאי סקוטי נודע שחי בשנים 1692-1770.

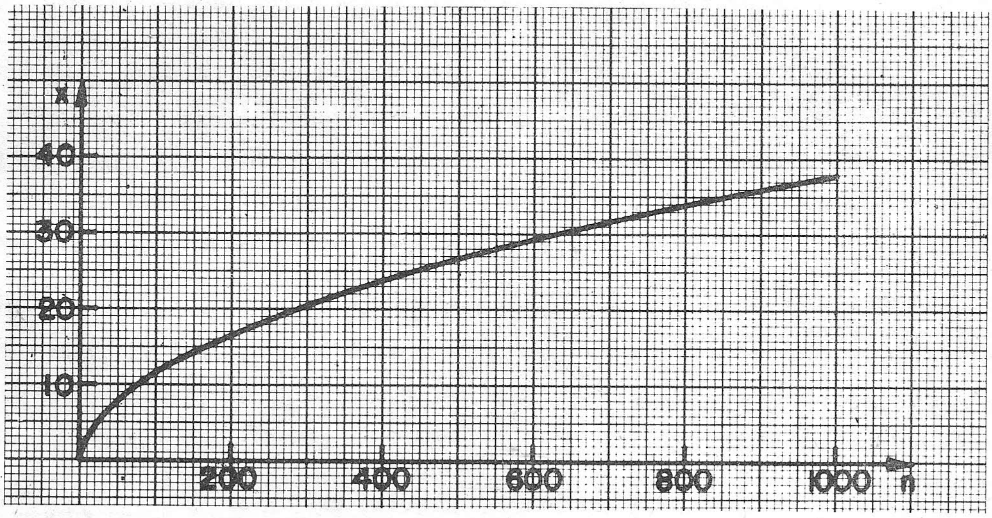
נוסחה זו נוחה למדי לחישוב בעזרת לוגריתמים, סרגל חישוב (בעל סקלה לוג-לוג) או מחשב כיס.

יתרון נוסף לנוסחה זו, הוא שכאן x יכול להיות גם מספר לא שלם. למשל בשביל בעיית יום ההולדת מקבלים שערך x הנותן $P = 0.5$ הוא 22.768 (*)

כאשר מחשבים בעזרת נוסחה זו את ערכי x עבור n שונים, הנותנים $Q = 0.5$ מקבלים את הטבלה:

n	10	20	50	100	200	500	1000
x	4.03	5.56	8.61	12.05	16.9	26.6	37.5

כאשר מציירים את הגרף מתקבלת התמונה הבאה:



גרף זה דומה מאד לפרבולה. ננסה לצייר גרף של x^2 כפונקציה של n ואמנם מתקבל קו ישר ששיפועו הנמדד הוא 1.41 (ראה בגרף הבא) כלומר $x^2 = 1.41n$ (**).
ומכאן נקבל את נוסחת הקירוב: $x = 1.19\sqrt{n}$.

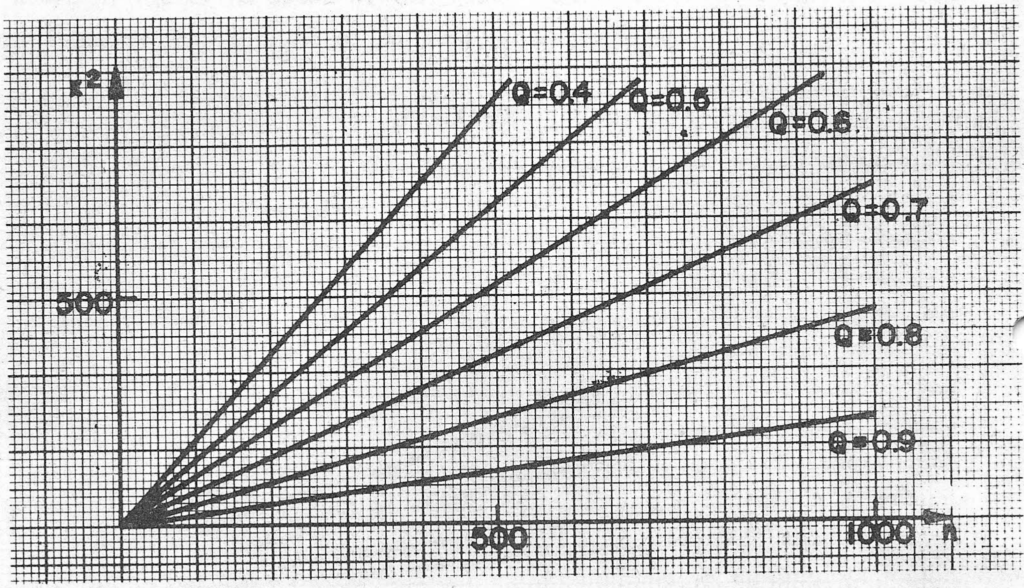
לדוגמה, בשביל $n = 365$ קבלנו בבעית יום ההולדת הסתברות $P = 0.5$ כאשר $x = 22.7$ ואמנם $1.19\sqrt{365} \simeq 22.7$. באותו אופן בשאלת מספרי המכוניות קבלנו $P = 0.5$ כאשר $x = 37.5$. כאן $n = 1000$ ואמנם $1.19\cdot\sqrt{1000} \simeq 37.6$.

(*) המשמעות של x לא שלם היא שאם ניקח מספר רב של קבוצות אנשים, אשר המספר הממוצע של האנשים בהן הוא x , אזי ההסתברות שימצאו בכל קבוצה לפחות שני אנשים בעלי אותו יום הולדת היא 0.5.

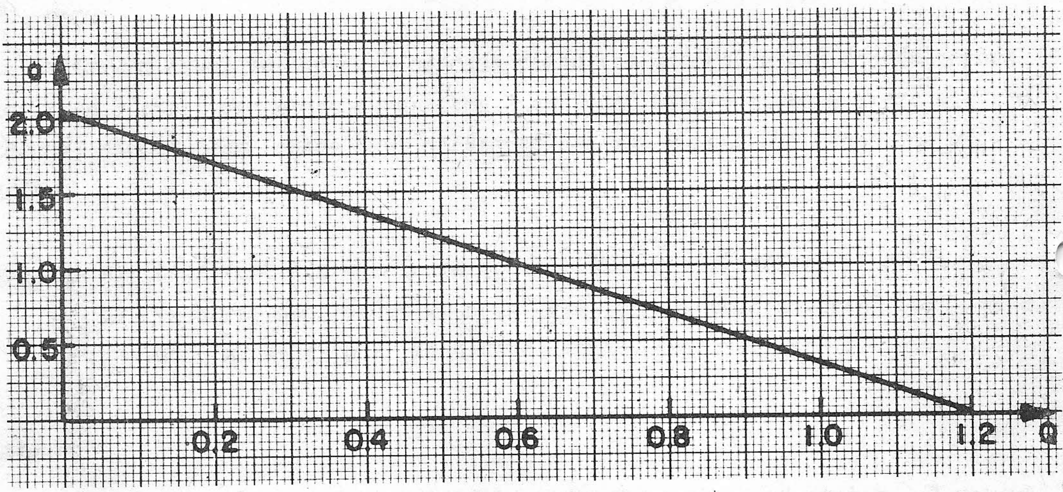
(**) בחישוב בעזרת שיטת הריבועים המינימליים התקבל כי הפונקציה היא מהצורה: $x \propto n^{0.485}$, ואילו השיפוע כאשר מציירים את x^2 כפונקציה של n הוא 1.405.

4. פתרון גרפי של הקשר בין x, q ו-n

אם במקום $Q = 0.5$ ניקח ערכים אחרים, נקבל בכל המקרים קווים ישרים, ששיפועיהם כמובן תלויים בערך של Q:



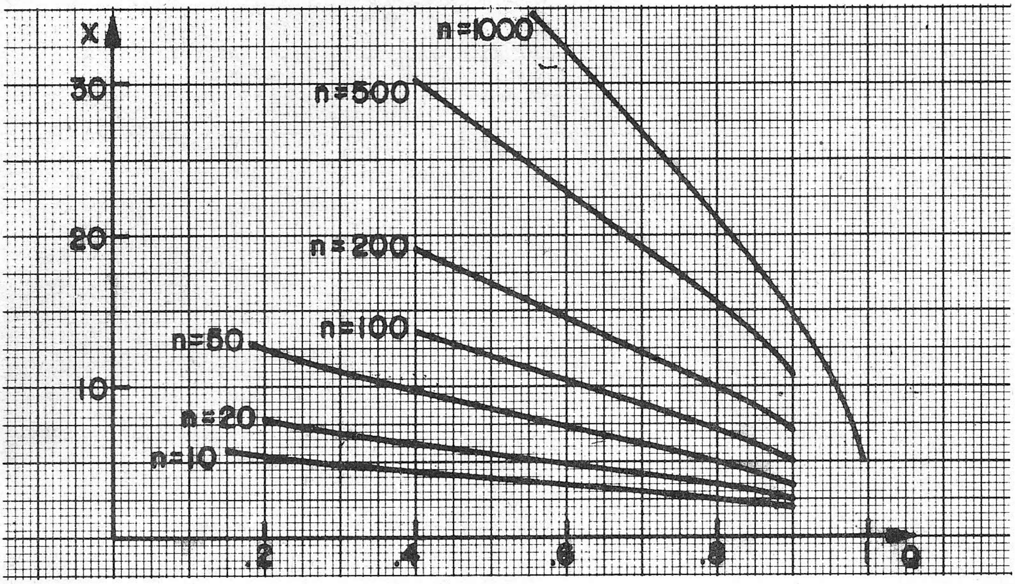
נצייר גרף נוסף, של השיפועים (המחושבים עבור $x = a\sqrt{n}$) כפי שנמדדו מתוך גרף זה, כפונקציה של Q ונקבל את השרטוט הבא:



מתוך שרטוט זה ניתן לקבל את נוסחת השיפוע: $a = 2.04 - 1.70\ Q$ (*) מכאן נוכל לרשום את x כפונקציה של n ו- Q ע"י נוסחת הקירוב הבאה: $x = (2.04 - 1.70\ Q)\sqrt{n}$.

(*) בחישוב בעזרת שיטת הריבועים המינימליים התקבלה הנוסחה $a = 2.036 - 1.69\ Q$.

כאשר נסינו לחקור מעט את משמעות התופעה, ציירנו גרף של x כפונקציה של Q, בשביל ערכי n שונים, וקבלנו את משפחת העקומות הבאה: מתברר שעבור Q "לא גדול מדי ולא קטן מדי" (בתחום שבין 0.8 - 0.4) הקו הוא בעל שיפוע קבוע.



5. סיכום

נסינו להראות במאמר זה, כי בעיה מפורסמת כבעית יום ההולדת, ניתנת להכללה והרחבה בכיוונים שונים. תוך שימוש בתיאור גרפי אפשר לקבל תוצאות ונוסחות קירוב מענינות. במפתיע, התקבלה נוסחה פשוטה למדי הנותנת בקירוב טוב את התוצאות בתחום רחב של ערכים.

הערות: א) ניתן לבצע חלק מן הגרפים על נייר לוגריתמי ולהקל בכך על החישובים.

ב) לא מצאנו הסבר מתקבל על הדעת על-פי תורת ההסתברות לנוסחת הקשר הפשוט שבין x, n ו- Q . נשמח לקבל תגובות והצעות של קוראים.

ספרות:

1. Parzen E., Modern Probability Theory and its Applications,John Wiley, 1964

2. Courant R., Differential and Integral Calculus, Interscience, 1952

3. דבורצקי א', מבוא לתורת ההסתברות, הוצאת אקדמון, תשי"ח.