אליהו לוי

מאמרים במתמטיקה

לנוער חובב המתמטיקה

שפרסמתי בשנים 1992-1997 ב"אתגר-גליונות מתמטיקה"
שהוצא לאור ע"י הפקולטות והחוגים למתמטיקה במכון ויצמן למדע,
בטכניון - מכון טכנולוגי לישראל, ובאוניברסיטת חיפה

תוכן עניינים

המטריצות והמתמטיקה הסינית הקדומה

[1] בשם מטריצה קוראים פשוט למספרים רשומים בתבנית מלבנית. למשל: את כל האינפורמציה שבמערכת 2 משוואות ב 2 נעלמים, כגון:

$$\begin{array}{rcrcl} 2x & - & y & = & 3 \\ x & + & 3y & = & 19 \end{array} \qquad (*)$$

אפשר לכתוב בצורת מטריצה של המקדמים:

$$\left[\begin{array}{ccc} 2 & -1 & 3 \\ 1 & 3 & 19 \end{array}\right] \qquad (**)$$

ברור שכל מטריצה מתפרקת ל"שורות" אפקיות, או ל"עמודות" אנכיות, שכל אחת מהן היא סידרת מספרים. למשל במטריצה $(**)$ יש 2 שורות, בכל אחת 3 מספרים, ויש בה 3 עמודות, בכל אחת 2 מספרים. מטריצה כזו נקראת מטריצת 2×3.

נוכל לפתור את המשוואות $(*)$ ע"י "משחק" במטריצות, שבו מותר לנו לבצע פעולות ממעלה ראשונה בשורות. נסביר זאת: מותר לנו למשל להחליף במטריצה $(**)$ את השורה הראשונה בשורה שבה במקום הראשון (משמאל) יהיה $1 = 1 - 2$, במקום השני $-4 = -3 - 1$ ובמקום השלישי $-16 = 3 - 19$. כלומר: בדיוק מה שאנו עושים במקום הראשון (מחליפים את המספר שבמקום הראשון בשורה הראשונה באותו מספר ממנו מחסירים את המספר שבמקום הראשון בשורה השניה) עלינו לעשות גם במקום השני וכן הלאה. למה שעשינו קוראים בקיצור: החלפת השורה הראשונה בשורה הראשונה פחות השניה. בדומה לכך נוכל, למשל, להכפיל את השורה הראשונה ב 2, לחבר לשניה ואת התוצאה לשים במקום השניה. השורה השניה החדשה תהיה אז $5 \;\; 1 \;\; 25$, כי

$$2 \cdot 2 + 1 = 5 \qquad 2 \cdot (-1) + 3 = 1 \qquad 2 \cdot 3 + 19 = 25$$

פעולה מותרת אחרת: להכפיל (או לחלק) שורה מסויימת במספר קבוע שונה מאפס. למשל מותר לנו להחליף את השורה הראשונה בשורה $6 \;\; -2 \;\; 4$ (הכפלה ב 2). נדגיש שמותרות רק פעולות "ממעלה ראשונה": אסור, למשל, להעלות את כל אברי שורה בריבוע או להכפיל שתי שורות זו בזו. אחרי שביצענו פעולה אחת בשורות, אנו יכולים לבצע עוד פעולות כאלה, כאוות נפשנו. נקרא למשחק: משחק הפעולות בשורות.

למה זה טוב ? כי יש לנו שמורות שלא ישתנו ע"י משחק כזה. למשל אם x ו y פותרים את מערכת המשוואות של מטריצה, הם ימשיכו לפתור את מערכת המשוואות של המטריצות החדשות שהתקבלו ממשחק הפעולות בשורות (הוכח זאת לגבי הדוגמאות לפעולות בשורות שניתנו לעיל - כאן חשוב מאוד שלא מרשים פעולות שאינן ממעלה ראשונה). כדי לפתור את המשוואות עלינו רק לשחק במשחק, מתוך מטרה להגיע למטריצה שנותנת משוואות בעלות פתרון מיידי. נדגים זאת לגבי $(**)$ (רשמנו גם את הפעולות שהתבצעו: ש_1 מסמן את השורה הראשונה ו ש_2 את השניה):

$$\left[\begin{array}{ccc} 2 & -1 & 3 \\ 1 & 3 & 19 \end{array}\right] \xrightarrow[\text{במקום } \text{ש}_1]{\text{ש}_1 - 2 \cdot \text{ש}_2} \left[\begin{array}{ccc} 0 & -7 & -35 \\ 1 & 3 & 19 \end{array}\right] \xrightarrow[\text{במקום } \text{ש}_1]{(-\frac{1}{7}) \cdot \text{ש}_1} \left[\begin{array}{ccc} 0 & 1 & 5 \\ 1 & 3 & 19 \end{array}\right] \xrightarrow[\text{במקום } \text{ש}_2]{\text{ש}_2 - 3 \cdot \text{ש}_1} \left[\begin{array}{ccc} 0 & 1 & 5 \\ 1 & 0 & 4 \end{array}\right]$$

המשוואות של המטריצה האחרונה מימין הן $x = 4$, $y = 5$, והפתרון מיידי.

שאלה: האם שיטה זו שונה באמת משיטת החילוץ שנלמדת בבית הספר ?

באותה צורה אפשר לפתור מערכות של מספר כלשהו של משוואות ממעלה ראשונה במספר נעלמים.

וכאן נכנסת לתמונה המתמטיקה הסינית הקדומה. השימוש במטריצות היה נושא בו הצטיינו הסינים, אולי בהשפעת השימוש שלהם במקלות חישוב (מעין קדם-חשבוניות). הסינים לא נהגו לבצע חישובים ע"י כתיבה, אלא השתמשו במקלות על לוח. קבוצות של מקלות סימנו את סיפרת היחידות, העשרות, המאות וכו' והחישובים בוצעו ע"י העברת המקלות. כאשר קבוצות המקלות שסימנו מספרים הונחו בסדר מלבני - יש לנו מטריצה, והסינים פתרו בעזרת מטריצות כאלה מערכות משוואות ממעלה ראשונה, בדרך כמו זו שתארנו. דרך זו מתוארת כבר בספר הקלאסי של המתמטיקה הסינית: Jiu-zhang suan-shu (קרא: ג'יו-ג'אנג סואן-שו), שפירושו: "תשעה פרקים על אומנות המתמטיקה". ספר זה נערך בצורתו הסופית

[1] כל מה שייכתב בלשון זכר (כמו: פתור וכד') - הכוונה גם ללשון נקבה.

בתקופת שושלת חאן, לפני כ־2000 שנה, ומתמטיקאים סיניים מאוחרים יותר כתבו לו פירושים. מעניין שהמונח בו קוראים למשוואות בסין, fang cheng, אינו אלא השם המסורתי של שיטת הפתרון ע"י מטריצות, שנקראה: fang-cheng shui שפירושו בערך: "שיטה (אלגוריתם) של חישוב צעד־צעד בנתונים מסודרים על לוח המקלות בצורת ריבוע (או מלבן)".

באירופה הוכנסה המטריצה לראשונה כמושג מתמטי עצמאי באמצע המאה ה 19 ע"י המתמטיקאי הבריטי־יהודי סילבסטר (J. J. Sylvester), שהיה מחשובי מפתחי האלגברה "הגבוהה", אם כי שיטת החילוץ לפתרון מערכת משוואות, וכן מושג הדטרמיננט (ששייך למעשה לתורת המטריצות ושאולי חלק מהקוראים מכיר אותו), היו מוכרים לפני כן. עד מהרה פותחה תורה עשירה של המטריצות, כעצמים מתמטיים לכל דבר, וגם כיום עדיין חוקרים אותן וכן משתמשים בהן למטרות מגוונות שחלקן רחוקות מבעיית פתרון מערכת משוואות.

אבל נחזור לפתרון משוואות. באותו רעיון אפשר להשתמש גם עבור משוואות דיופנטיות, כלומר שבהן מחפשים פתרונות שלמים בלבד. כל ההבדל בשיטה הוא שבמשחק מותר רק לחבר או לחסר שורות מוכפלות במספרים שלמים או להכפיל שורה במספר שלם שאינו אפס (לפעולות כאלה נקרא: פעולות שלמות). גם בכך עסקו המתמטיקאים הסיניים, ובעיות מסוג זה ופתרונן בעזרת מטריצות מופיעים בספרו של המתמטיקאי Qin Jiu-shao (קרא: צ'ין ג'יו־שאו) משנת 1247 לספירת הנוצרים, ספר שגם הוא מחולק לתשעה פרקים (לפי צו המסורת ?)

צ'ין מטפל בבעיה שאפשר לנסחה כך (יש להדגיש שהסינים לא הכירו את הסימון האלגברי באותיות, שהוכנס לשימוש רק במאה ה 16 באירופה ע"י המתמטיקאי Vieta. בדומה למה שהיו עושים כל אלה שעסקו במשוואות בעולם עד Vieta, גם הסינים היו מנסחים בעיות מספריות ומראים כיצד לפתור אותן, והיה ברור ללומד ששיטת הפתרון אינה תלויה כלל בבחירת הנתונים):

נתונים מספרים טבעיים m ו a. מצא מספר שלם x כך שיתקיים $ax \equiv 1 \pmod m$ (זה סימון לקונגרואנציה מודולו m, שמשמעותה שההפרש בין שני האגפים מתחלק ב m, או, בלשון אחרת: שניהם נותנים אותה שארית בחלוקה ל m - הוכח שזה באמת אותו דבר!). כאן אם x פתרון יהיה כל מספר קונגרואנטי ל x מודולו m גם הוא פתרון (מדוע ?). משוואה זו היא דוגמא ל"משוואת קונגרואנציה".

שיטת הפתרון, שהיא בעיקרה השיטה המתוארת בספרו של צ'ין (אם כי אולי אינה מנומקת כמו שננמק אותה להלן) תהיה: לרשום את המטריצה $\begin{bmatrix} 1 & a \\ 0 & m \end{bmatrix}$ ולנסות לבצע פעולות שלמות בשורות עד שנקבל מטריצה שתהיה בה שורה מהצורה: $1\ b$. (הנחיה כיצד לעשות זאת: בכל שלב בחר פעולה שלמה בשורות כך שאחד המספרים שיהיו בעמודה הימנית יהיה קטן בערכו המוחלט משני המספרים שהיו בה בשלב הקודם - הבהר זאת). אז b יהיה פתרון. הוכחה: אם x פתרון, אזי $ax \equiv 1 \pmod m$ וכן $mx \equiv 0 \pmod m$ (מדוע ?) ותכונה זו - שהאיבר הראשון בכל שורה קונגרואנטי מודולו m לאיבר השני מוכפל ב x, לא תתקלקל ע"י המשחק שלנו. לכן במטריצה הסופית, בה יש שורה $1\ b$, יתקיים $b \equiv 1 \cdot x \pmod m$, וכיון ש x פתרון, גם b פתרון (וכן כל מספר קונגרואנטי ל b מודולו m).

נביא דוגמא פשוטה מספרו של צ'ין, שבלשוננו תיכתב כ $14 \cdot x \equiv 1 \pmod{19}$. דרך הפתרון תהיה (אצל צ'ין כתובה דרך הפתרון בצורה קצת שונה, בגלל שאין הוא משתמש בה במספרים שליליים - אילו פעולות (שלמות) בשורות בוצעו כאן ?

$$\begin{bmatrix} 1 & 14 \\ 0 & 19 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 14 \\ -1 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 4 \\ -1 & 5 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & 4 \\ -4 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 15 & 1 \\ -4 & 1 \end{bmatrix}$$

והפתרון הוא $x = 15$ (אין צורך בצעד האחרון: $x = -4$ הוא פתרון).

שאלות:

1. פתור בשיטה זו: א) $9x \equiv 1 \pmod{11}$, ב) $8x \equiv 1 \pmod{13}$, ג) $256x \equiv 1 \pmod{337}$. (כדאי לבדוק את הפתרונות)
2. נניח שמצאת פתרון x למשוואה $ax \equiv 1 \pmod m$. נסה לפתור בעזרתו משוואה כללית יותר: $ay \equiv b \pmod m$.
3. האם לכל משוואה $ax \equiv 1 \pmod m$ יש פתרון ? נסה למצא דוגמא בה אין פתרון. מה תוכל להגיד, אם מצאת פתרון x, על הצורה הכללית של כל הפתרונות ?
4. האם אתה רואה קשר בין השיטה שלעיל לבין האלגוריתם הידוע (של אוקלידס) למציאת המחלק המשותף הגדול ביותר של שני מספרים ? נסה להיעזר בכך כדי למצוא תנאי שצריכים לקיים m ו a שיהיה הכרחי ומספיק לכך שיש פתרון למשוואה $ax \equiv 1 \pmod m$.

5. צ'ין עוסק בבעיה אחרת, שבלשוננו תנוסח כך: יש למצוא y שלם שיקיים את מערכת משוואות הקונגרואנציה הבאות (כל המקדמים שלמים):

$$y \equiv b_1 \pmod{m_1}, \quad y \equiv b_2 \pmod{m_2}, \ldots, y \equiv b_n \pmod{m_n} \qquad (***)$$

אשר את השיטה הבאה למצוא פתרון (שהיא השיטה שבספרו של צ'ין): תהי m המכפלה $m_1 m_2 \cdots m_n$, ויהיו $a_1 = m/m_1, a_2 = m/m_2, \ldots a_n = m/m_n$ (ה a -ים שלמים!). נניח שמצאנו פתרונות $x_1, x_2, \ldots, x_n$ למשוואות:

$$a_1 x_1 \equiv 1 \pmod{m_1}, \quad a_2 x_2 \equiv 1 \pmod{m_2}, \ldots, a_n x_n \equiv 1 \pmod{m_n}$$

אז המספר y הבא ייתן פתרון למערכת $(***)$:

$$y = b_1 a_1 x_1 + b_2 a_2 x_2 + \ldots b_n a_n x_n$$

(שים לב שהצלחת השיטה תלויה באפשרות לפתור את המשוואות עבור ה x -ים !)

6. בהשפעת עיסוקם של הסינים בבעייה שבשאלה 5, נקרא כיום בספרות המתמטית בשם "משפט השאריות הסיני" המשפט הבא: אם $m_1, m_2, \ldots, m_n$ מספרים טבעיים זרים, כלומר לכל שניים מהם אין מחלק משותף גדול מ 1 , אזי עבור המשוואות $(***)$ יתקיים: א) יש תמיד פתרון, ו ב) פתרון זה יחיד עד כדי קונגרואנציה מודולו המכפלה $m = m_1 \cdot m_2 \cdot \cdots \cdot m_n$, כלומר: אם y פתרון כלשהו, אזי כל מספר קונגרואנטי ל y זה מודולו m הוא פתרון (זה קל) וגם ההפך: כל פתרון אחר הוא קונגרואנטי ל y זה מודולו m . האם תוכל להוכיח משפט זה ?

פרטים על הישגי הסינים הקדמונים במתמטיקה אפשר למצוא בספר:

Li Yan: *Chinese Mathematics*, Oxford University Press, 1987

על הזוכים בפרס וולף למתמטיקה 1992

כידוע, מדי שנה מחולק בישראל פרס קרן וולף למדענים ואמנים מצטיינים בעולם בשטחים שונים, גם במתימטיקה. השנה חולק הפרס למתימטיקה שווה בשווה בין פרופ' לנארט קארלסון (Lennart Carleson) מאוניברסיטת אופסלה, שוודיה ואוניברסיטת קליפורניה בלוס אנג'לס, ארה"ב ובין פרופ' ג'ון תומפסון (John G. Thompson) מאוניברסיטת קימבריג', בריטניה, שניהם מהשורה הראשונה של המתמטיקאים בעולם.

לנארט קארלסון נולד בשוודיה ב־1928. בגיל 22 כבר קיבל תואר דוקטור ובגיל 23 יצרה בשבילו הממשלה השוודית קתדרה מיוחדת של פרופסור מחקר באוניברסיטת אופסלה. בשנים 1982־1979 היה נשיא האיגוד העולמי למתמטיקה. שטח עבודתו של פרופ' קארלסון הוא האנליזה, שהיא הדיון בפונקציות, בקשר עם מושגים כמו גבולות, סכומים אינסופיים, נגזרות ואינטגרלים. הוא עבד בשטחים בהם יש קשר הדוק בין פונקציות ממשיות (כלומר מוגדרות על משתנים המקבלים ערכים במספרים ממשיים וגם ערכי הפונקציה ממשיים) לבין פונקציות קומפלכסיות (בהן באים המספרים הקומפלכסיים (המרוכבים) במקום הממשיים). קארלסון נודע כבעל יכולת להתמודד בשיטות מקוריות עם בעיות באנליזה שאחרים ניסו לפתרן ולא הצליחו.

אחת הבעיות המפורסמות שפתר קארלסון קשורה לטורי פוריה: המושג הפיסיקלי הבסיסי של פירוק קול לצלילים בגבהים שונים, של אור לצבעי הספקטרום ושל גלי רדיו לתדרים שונים נובע מעובדה מתמטית, שאומרת בערך שפונקציה המוגדרת על מספרים ממשיים (למשל פונקציה של הזמן t) מתפרקת באופן יחיד לצירוף של תנודות טהורות, כלומר סינוסואידות מהצורה $C\sin(\alpha+\omega t)$ כאשר ω נקראת התדירות. במקרה של פונקציה מחזורית $f(t)$, ונבחר את יחידת המידה של t כך שהמחזור יהיה 2π , כלומר יתקיים $f(t+2\pi)=f(t)$ לכל t , יופיעו רק הסינוסואידות שהן עצמן בעלות מחזור זה, כלומר בעלות ω שלם, (ונסמן n במקום ω). אז הפירוק שלנו צריך לייצג את f בעזרת ביטוי מהצורה

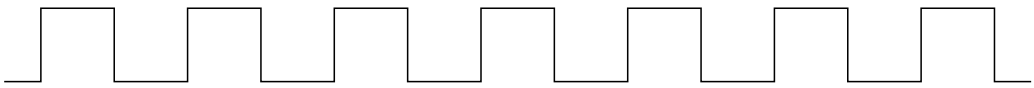
$$\sum_{n=0}^{\infty} C_n \sin(nt+\alpha_n)$$

או, מה שאקוויולנטי לכך,

$$\sum_{n=0}^{\infty} A_n \cos(nt) + B_n \sin(nt) \qquad (*)$$

ביטוי זה הוא טור אינסופי.

כבר במאה ה־18 הועלו מצד אחד נימוקים פיסיקליים לאפשרות של פירוק יחיד כזה, ומצד שני לא האמינו שאם f לא רציפה, למשל, אזי אפשר לפרק אותה לטור סינוסואידות. בתחילת המאה ה־19 הציע פוריה (Fourier) רעיון כיצד לבנות מכל פונקציה נתונה מערכת מקדמים A_n ו B_n שיתנו מועמד לפירוק יחיד כזה, ונתן דוגמאות מפתיעות שגם עבור פונקציות עם קפיצות, למשל פונקציית מדרגות כמו:



המועמד שלו "עובד". עבור פונקציה מחזורית, הטור $(*)$ בו A_n ו B_n נבחרו לפי הכלל שהציע פוריה, נקרא "טור פוריה" של הפונקציה.

אפשר להוכיח שבמובנים חשובים הוא נותן את ה"פירוק" המבוקש היחיד עבור פונקציות מחזוריות כלליות מאוד (אפילו לגמרי לא רציפות), אבל לגבי השאלה אם הפירוק לטור פוריה הוא גם במובן שסכום הטור שווה לפונקציה בכל t , המצב לא פשוט. (סכום טור אינסופי מוגדר כגבול, אם הוא קיים, של הסכומים החלקיים. יש לזכור שייתכן שלטור אינסופי לא יהיה סכום בכלל ושאם הטור תלוי ב t אזי קיום הסכום וערכו תלויים גם הם ב t).

כבר במאה ה־19 הוכיח דיריכלה (Dirichlet) שסכום טור פוריה של פונקציה מחזורית שווה אכן לפונקציה בכל t אם היא די "טובה" (למשל אם אפשר לחלק את המחזור למספר סופי של קטעים שבכל אחד מהם הפונקציה עולה או יורדת, ובנוסף לכך בכל נקודת אי־רציפות ערך הפונקציה הוא הממוצע בין הערכים להם היא שואפת כאשר שואפים לנקודה מימין ומשמאל - יש פונקציות, אפילו רציפות, שאינן כאלה !). בנוסף לכך מתקיים שאם t ערך בו יש לטור פוריה סכום ובנוסף לכך הפונקציה רציפה ב t אזי הסכום שווה ל $f(t)$. לכן עבור פונקציות רציפות הבעייה היא אם יש לטור בכלל סכום.

ניתנו דוגמאות לפונקציות רציפות שאין לטור פוריה שלהן סכום ב t ־ים מסויימים (אפילו באינסוף t ־ים עבור אותה פונקציה - כמובן שאלו פונקציות שלא מקיימות את תנאי דיריכלה), אבל באשר לשאלה עד כמה יכולה קבוצת t ־ים "סוררים" כאלה להיות גדולה כאשר f רציפה היתה הבעייה קשה לפיצוח זמן

רב. רק ב־1965 הוכיח קארלסון שאם f מחזורית ורציפה (ואפילו אם f מקיימת תנאים חלשים בהרבה מרציפות) יהיה סכום טור פוריה שלה שווה לערך הפונקציה עבור כל t פרט אולי ל t ־ים בקבוצת יוצאים מהכלל (קבוצה שתלויה בפונקציה f שלקחנו) וקבוצה זו בעלת אורך 0 (קבוצה בעלת אורך 0 לא יכולה, כמובן, להכיל אף קטע, אפילו קטן ביותר, אבל יכולה להיות בעלת אינסוף איברים).

למעשה היה ידוע (כבר קודם) שאילו משפט זה, שהוכיח לבסוף קארלסון, לא היה נכון, היה נובע מכך שאפשר למצוא דוגמא של פונקציה מחזורית רציפה "משוגעת" שבאף נקודה t אין לטור פוריה שלה סכום !

בשנים האחרונות עוסק קארלסון, עם תלמידו בנדיקס (M. Benedicks) בנושא המערכות הדינמיות, נושא שזכה בשנים האחרונות לפופולריות רבה ושכולל את מושג הכאוס (Chaos). מערכת דינמית היא קבוצה S עם פונקציה f המוגדרת בכל איבר x של S כך שלכל x כזה גם $f(x)$ שייך ל S. השם "מערכת דינמית" בא מכך שבפיסיקה מופיעות מערכות כאלה בהן S היא קבוצת המצבים האפשריים של מערכת פיסיקלית ו f מתארת את השתנות המערכת ביחידת זמן: אם כעת המערכת במצב x, אזי אחרי יחידת זמן היא תהיה ב $x_1 = f(x)$, אחרי עוד יחידת זמן ב $x_2 = f(f(x))$, כעבור עוד יחידת זמן ב $x_3 = f(f(f(x)))$, וכן הלאה. בהתאם לכך, בכל מערכת דינמית מוגדר **המסלול** של x כסדרה האינסופית x_n בה $x_0 = x$, $x_{n+1} = f(x_n)$. התנהגות מסלול זה יכולה להיות מסובכת ומפתיעה גם עבור פונקציה פשוטה כמו $f(x) = x^2+c$ (c קבוע). ברור שקל לתכנת מחשב כך שיחשב מסלול כזה, ומאז שבשנות ה־70 החלו בחישובים כאלה, והמחשבים הראו את התוצאות גם בצורה גרפית, נתגלו תופעות מופלאות והקוראים נתקלו בוודאי בתמונות בעלות היופי האסתטי המיוחד שיוצרים המחשבים, ושהודפסו גם בספרים ובפוסטרים. אבל תמונות אלו, שדיוקן הוא, כמובן, רק כמידת הדיוק של החישוב והגרפיקה של המחשב, אינן באות במקום הוכחה מדוייקת של משפט מתמטי על מערכות כאלה, ולכך תרמו ופרצו דרכים קארלסון ובנדיקס.

ג'ון תומפסון[2] הוא אחד החוקרים המבריקים והמקוריים ביותר בתורת החבורות. מחקריו של פרופ' תומפסון הצעידו את תורת החבורות קדימה ואיפשרו תוצאות שנחשבו לבלתי ניתנות להשגה עד לפני כ־30 שנה.

הוא נולד ב־1932 במדינת קנזס שבארצות הברית. בשנת 1951 החל בלימודי דת באוניברסיטת ייל, אבל לאחר זמן קצר החליף את נושא לימודיו למתמטיקה, ובשנת 1959 קיבל את תואר הדוקטור מאוניברסיטת שיקגו. מאז 1968 הוא נמצא באוניברסיטת קימבריג' באנגליה.

חבורה הינה קבוצה של איברים, G , עם פעולה בין אברי G הנותנת מכל שני איברים a ו b ב G איבר של G, והמקיימת את הדרישות: שיתקיים החוק האסוציאטיבי (חוק הקיבוץ); שיהיה ב G איבר ניטרלי (כלומר כזה שהפעלת הפעולה על איבר כלשהו a ועליו נותנת תמיד את a, למשל 0 לגבי חיבור או 1 לגבי כפל) ושעבור כל איבר a יהיה קיים ב G איבר a' כך שהפעלת הפעולה על a ו a' תיתן את האיבר הניטרלי (במקרה של חיבור $a' = -a$. במקרה של כפל $a' = 1/a$).

לסוג זה של מיבנה יש חשיבות עצומה במתמטיקה ובשימושיה בפיסיקה ובכימיה. כדוגמא לחבורה אפשר להביא את המספרים השלמים עם פעולת החיבור. זו חבורה אינסופית (כלומר קבוצת איבריה אינסופית). דוגמא אחרת מהווים המספרים הממשיים החיוביים עם פעולת הכפל. (אבל כל המספרים הממשיים לגבי הכפל, או כל המספרים הטבעיים לגבי הכפל או לגבי החיבור אינן חבורות ־ למה ?). אם ניקח את קבוצת השאריות (השלמות) מודולו n עם פעולת החיבור מודולו n נקבל חבורה סופית (כלומר כך שיש בקבוצה מספר סופי של איברים). בכל הדוגמאות האלה הפעולה של החבורה מקיימת את החוק הקומוטטיבי (חוק החילוף), כלומר תוצאת הפעולה על שני איברים אינה תלויה בסדר בו לוקחים אותם. חבורה כזו נקראת קומוטטיבית. אבל חבורה יכולה לא להיות כזו, וכדוגמא אפשר להביא חבורות של תנועות במישור או במרחב לגבי פעולת הרכבת תנועות (ביצוע תנועה אחת ואחר כך השניה), למשל חבורת 8 הסיבובים והשיקופים המעתיקים לעצמו ריבוע נתון במישור (אחד מאיברי החבורה הוא תנועת הזהות שאינה עושה כלום). הקורא מוזמן למצא מהם 8 סיבובים ושיקופים אלה ולבדוק שזו חבורה לא קומוטטיבית.

מחקריו של תומפסון התרכזו בעיקר בחבורות הסופיות, שחקירתן ומיונן מהוות נושא שמציג בעיות קשות. חבורות סופיות מסויימות נקראות **חבורות פשוטות** והן מהוות את "אבני הבניין" של כל החבורות הסופיות. יש להן תפקיד דומה במקצת לתפקיד המספרים הראשוניים באריתמטיקה. ואכן, החבורות הסופיות הפשוטות הקומוטטיביות אינן אלא חבורות השאריות מודולו p (עם פעולת החיבור) עבור p ראשוני. עבודותיו של פרופ' תומפסון איפשרו ופיתחו את הדרך למציאתן של כל החבורות הסופיות הפשוטות הלא־קומוטטיביות, השג מרשים שהושג בערך ב־1980.

כבר בעבודת הדוקטור שלו גילה תומפסון שהוא מתמטיקאי מבריק ומקורי, כאשר הוכיח השערה בתורת החבורות הסופיות שלא הצליחו לפתור במשך 50 שנה. יותר מאוחר הכניס פרופ' תומפסון שיטות מהפכניות שאיפשרו לו ולוולטר פייט (Walter Feit) להוכיח ב־1963 שכל חבורה פשוטה סופית לא קומוטטיבית היא בעלת מספר זוגי של אברים. הוכחת משפט זה משתרעת על 250 עמודים. כתוצאה ממשפט זה כל "אבני הבניין" הפשוטות של חבורה סופית כלשהי בעלת מספר איזוגי של איברים הן קומוטטיביות (ולכן הן חבורות שאריות מודולו ראשוניים). עבודות אלה שינו את פני תורת החבורות הסופיות מהקצה עד הקצה והפיחו בה רוח חדשה.

[2] אני מודה לפרופ' דוד צ'ילג, מהפקולטה למתמטיקה בטכניון, על עזרתו.

קו אופי מיוחד של מערכת החבורות הסופיות הפשוטות, שכאמור נמצאה במלואה בערך ב־1980, הוא שבנוסף לסדרות אינסופיות של חבורות הבנויות לפי כללים טבעיים (בדומה לסידרת החבורות הפשוטות הקומוטטיביות שהזכרנו) יש עוד 26 חבורות פשוטות "ספוראדיות". בגדולה שבהן יש

808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000

איברים ולא לחינם קוראים לה "המפלצת". עוד לפני שבנו חלק מחבורות ספוראדיות אלה, ביניהן המפלצת, בנו "קלסתרון מתמטי" שלהן, כלומר מערכת תכונות שחשדו שיש חבורה שמקיימת אותן, ורק אחר כך הוכיחו שקיימת חבורה יחידה המתאימה ל"קלסתרון". באשר למפלצת, עוד לפני שהוכח קיומה שמו לב במקרה לכך שהמספר $196883 = 47 \cdot 59 \cdot 71$ המופיע ב"קלסתרון" שלה, מופיע גם בשטח אחר לגמרי במתמטיקה, והדבר הביא את פרופ' תומפסון לנסח השערות על קשרים בין התחומים, שהוכחו אחר כך והביאו לפתיחת חלון לעובדות וקשרים מתמטיים חדשים, שקשורים גם לפיסיקה.

על השגיו זכה פרופ' תומפסון בפרס Fields, שהוא הפרס המכובד והחשוב ביותר במתמטיקה (כידוע אין פרס נובל למתמטיקה), בפרס Cole היוקרתי באלגברה ובשנה זו בפרס וולף.

טוב שם טוב משמן טוב או: משפטים בקומבינטוריקה

מוקדש לזכרו של פרופ' חיים חנני (1912-1991) שמהרצאותיו לקוח חלק מתוכן מאמר זה

לפעמים אפשר לראות מייד כיצד להוכיח משפט בעזרת משפט שנראה שונה לגמרי, אם רק יודעים לקרוא לדברים בשם הנכון, כלומר להשתמש בשפה הנכונה. במאמר זה נדגים זאת בעזרת שרשרת משפטים, חלקם מפתיעים, בקומבינטוריקה.

נקודת המוצא שלנו תהיה משפט דילוורת' (Dilworth) הוא דן בקבוצה סופית סדורה חלקית, מה שנקרא באנגלית פוסט Poset (Partially Ordered Set). (המונח פוסט בספרות המתמטית מתייחס לקבוצה סדורה חלקית כלשהי, לאו דווקא סופית, אבל במאמר זה נתכוון במונח פוסט תמיד לפוסט סופי.) זוהי קבוצה סופית P בה נתון יחס בין זוגות איברים ב P שיסומן ב $a > b$. למען הציוריות נניח ש P היא קבוצת אנשים ואת $a > b$ נבטא כ: "a שולט על b". מיחס זה דורשים את האכסיומות של סדר חלקי: טרנזיטיביות: $a > b \,\&\, b > c \Rightarrow a > c$ רפלכסיביות: לכל a $a > a$ (אם כך, כל אדם שולט על עצמו) ואנטי־סימטריות: $a > b \,\&\, b > a \Rightarrow a = b$. במקום לרשום $a > b$ אפשר לרשום $b < a$.

נציין במיוחד שני סוגים של תת־קבוצות A של P: A היא **שרשרת** אם כל שני איברים a, b בה נמצאים ביחס שליטה: או $a > b$ או $b > a$. כלומר זוהי תת־קבוצה בה כל אחד יודע את מקומו. הניגוד לכך היא תת־קבוצה A **בלתי תלויה** (ב"ת), שהיא תת־קבוצה בה בין אף $a \neq b$ אין יחס שליטה ־ יש דמוקרטיה מלאה. שרשרות וקבוצות בלתי תלויות אכן ייתכנו: בכל קבוצה סופית נתונה S נוכל להגדיר סדר חלקי (יחיד) כך ש S כולה תהיה בלתי תלויה וסדרים חלקיים (רבים) כך ש S כולה תהיה שרשרת (בדוק). דוגמא לפוסט P שאינו, בדרך כלל, לא כולו שרשרת ולא כולו קבוצה בלתי־תלויה: $P =$ אוסף כלשהו של תת־קבוצות של קבוצה סופית קבועה C ו $a > b$ פירושו: הקבוצה b חלקית לקבוצה a (או שווה לה).

הטענה הבאה מיידית: אם הצלחנו להציג פוסט P כאיחוד של k שרשרות (לאו דווקא זרות), אזי כל תת־קבוצה ב"ת I ב P מכילה לכל היותר k איברים (אחרת היו שני איברים ב I שייכים לאותה שרשרת, ואז אחד שולט על השני בניגוד לכך שהם איברים בקבוצה ב"ת). לשון אחרת: המספר הקטן ביותר של שרשרות שאיחודן $P \leq$ מספר האיברים בתת־קבוצה הב"ת של P בעלת מספר האיברים המכסימלי.

משפט דילוורת' טוען ששני מספרים אלה תמיד שווים. לשון אחרת: אם יש ב P תת־קבוצה ב"ת בת k אברים וזה המכסימום, אזי אפשר להציג את P כאיחוד של k שרשרות.

הוכחת משפט זה תובא בסוף המאמר.

בעיית השידוכים

משפט קומבינטורי זה, ומשפטים שנדון בהם בהמשך, ינוסחו באופן ציורי, אבל דבר זה לא צריך להטעות את הקוראים ־ אלה הם פשוט משפטים קומבינטוריים. את בעיית השידוכים מקובל לנסח בלשון האהבה: נתונות שתי קבוצות בנות n איברים: קבוצת בנים B וקבוצת בנות G. נתונים יחסי חיבה בין זוגות של בן ובת, כלומר נתון לנו אילו זוגות מחבבים זה את זו ואילו לא. למעשה פשוט נבחרה קבוצה כלשהי של זוגות {בן,בת} שהוגדרה כקבוצת הזוגות המחבבים (אנו מניחים שיחסי החיבה הם הדדיים). מטרתנו היא לחלק את $2n$ הצעירים ל n זוגות זרים שיבואו בברית הנישואין, כאשר התנאי הוא שנחתן רק זוג שמחבבים זה את זה. לחלוקה כזו נקרא שידוך. (כמובן, במקרים רבים אפשריים שידוכים שונים, למשל אם כל זוג {בן,בת} מחבבים זה את זו).

ברור שלא תמיד אפשרי שידוך, למשל אם אין כלל זוגות המחבבים זה את זו. אפשר לתת את התנאי ההכרחי הבא לקיום שידוך (לצערנו עלינו להשתמש בניסוחים שמפלים בין המינים): עבור כל תת־קבוצה B' של קבוצת הבנים נסמן ב $h(B')$ את קבוצת כל הבנות g כך שקיים בן כלשהו מ B' שמחבב את g. אם קיים שידוך, אזי $h(B')$ מכילה לפחות את כל הבנות ששודכו לאברי B' ולכן

$(*)$ עבור כל תת־קבוצה B' של B, מספר איברי $h(B') \geq$ מספר איברי B'

$(*)$ הוא תנאי שמערכת יחסי חיבה יכולה לקיים או לא, וכמו שראינו הוא תנאי הכרחי לאפשרות של שידוך. משפט השידוכים אומר שהתנאי ההכרחי $(*)$ הוא גם מספיק: אם $(*)$ מתקיים עבור כל תת־קבוצה B' של קבוצת הבנים, אז יש לפחות שידוך אחד.

להוכחת משפט השידוכים יש רק להשתמש בשפה הנכונה, שתהיה לצערי קצת מעליבה לאחד המינים: נגדיר יחס שליטה בקבוצת הבנים והבנות ע"י: כל בן או בת שולט(ת) על עצמו(ה), וחוץ מזה רק בת יכולה לשלוט על בן, וזה קורה אם ורק אם הם מחבבים זה את זו. נשאיר לקוראים לבדוק שזה אכן יחס סדר חלקי, כלומר טרנזיטיבי, רפלכסיבי ואנטיסימטרי. קבוצת $2n$ הצעירים היא כעת פוסט.

ננסה לראות מה משמעות המושגים שבמשפט דילוורת' במקרה זה: השרשרות הן כל הקבוצות בנות איבר אחד, וכן כל הזוגות {בן,בת} שמחבבים זה את זו. לכן אם הצלחנו להציג את קבוצת $2n$ הצעירים כאיחוד n שרשרות, אזי הן בהכרח n זוגות מחבבים שהם זרים זה לזה כקבוצות ויש לנו שידוך. לכן אם שידוך הוא בלתי אפשרי אזי הקבוצה אינה איחוד n שרשרות, לכן מספר השרשרות הדרוש כדי שאיחודן יהיה כל הקבוצה $< n$ ולכן לפי משפט דילוורת' יש קבוצה ב"ת I בת לפחות $n+1$ איברים. תהי $B' =$ קבוצת הבנים ב I, $G'' =$ קבוצת הבנות שאינן ב I . מכך שמספר איברי $I \leq n+1$ נובע שמספר איברי $G'' > B'$ מספר איברי B' וקל לראות שהעובדה ש I ב"ת גוררת ש $h(B')$ מוכלת ב G''. לכן $(*)$ לא מתקיים לגבי B' וגמרנו.

אהבה עם משקל

יש מקרה פרטי של משפט השידוכים, בו קל לוודא את התנאי $(*)$: נניח שעבור כל זוג של בן b ובת g יש לנו מספר ממשי $\ell(b,g) \geq 0$ שמודד את מידת החיבה ביניהם. אנו מניחים שהם אינם מחבבים אם ורק אם $\ell(b,g) = 0$ (כלומר את המבנה של יחסי החיבה העשרנו ע"י הכנסת מידה לחיבה). עבור כל בן b נקרא בשם כמות החיבה הכוללת שיש בליבו לסכום $\alpha(b)$ של $\ell(b,g)$ המסוכמים על כל הבנות g, ובצורה דומה תוגדר כמות החיבה הכוללת $\alpha(g)$ שיש בליבה של בת. כעת יש לנו המבחן הבא:

"משפט החיבה הכוללת הקבועה": אם קיים קבוע $\alpha > 0$ כך שכמות החיבה הכוללת של כל בן ושל כל בת שווה ל α אזי תמיד יש שידוך.

הוכחה: נוכיח שמתקיים התנאי $(*)$ של משפט השידוכים. נסמן ב $\#$ את מספר איברי קבוצה.

$$\alpha \cdot \#B' = \sum_{b\in B'} \alpha = \sum_{b\in B'} \sum_{g\in G} \ell(b,g) = \sum_{b\in B',\, g\in G} \ell(b,g) = \sum_{b\in B',\, b \text{ מחבב את } g} \ell(b,g) \leq$$
$$\leq \sum_{g\in h(B'),b\in B} \ell(b,g) = \sum_{g\in h(B')} \alpha = \alpha \cdot \#h(B')$$

ומכאן $\#B' \leq \#h(B')$ כלומר $(*)$ מתקיים.

עבור כל מערכת יחסי חיבה נתונה נוכל להגדיר משקלות באופן הבא: $\ell(b,g)$ יהיה 0 אם b ו g אינם מחבבים, ויהיה 1 אם הם מחבבים. אז מ"משפט החיבה הכוללת הקבועה" נובעת:

מסקנה: אם נתונה מערכת יחסי חיבה, בה לכל בן יש בדיוק k בנות שהוא מחבב, ולכל בת יש בדיוק k בנים שהיא מחבבת, $k > 0$ אזי תמיד יש שידוך.

בעיית הפרלמנט

נתונה מדינה בת $n \cdot k$ תושבים, שחלוקים בדעותיהם בשני נושאים: ענייני בטחון וענייני כלכלה. יש n דעות בענייני בטחון, בכל דעה מחזיקים k תושבים, ויש n דעות בענייני כלכלה, בכל דעה מחזיקים k תושבים (לכל תושב יש דעה אחת ויחידה בכל נושא). אין שום הנחות על קשר בין חלוקת הדעות בעיניני בטחון וכלכלה. אנו רוצים לבחור פרלמנט בעל n נציגים, ורוצים שבין n הנציגים ייוצגו הן כל n הדעות בענייני בטחון, והן כל n הדעות בענייני כלכלה (בהכרח ע"י נציג אחד לכל דעה).

העובדה המפתיעה היא שבלי כל הנחות נוספות תמיד קיים פרלמנט כזה.

הוכחה: כל מה שצריך הוא לדבר בשפה המתאימה: נתאר לנו מיתולוגיה בה יש n אלים המייצגים את n הדעות בענייני בטחון ו n אלות המייצגות את n הדעות בענייני כלכלה. יש ביניהם יחסי חיבה עם משקל באופן הבא: אל b ואלה g מחבבים זה את זו אם קיים תושב המחזיק גם בדעה שמייצג האל (בענייני בטחון) וגם בדעה שמייצגת האלה (בענייני כלכלה). מידת החיבה $\ell(b,g)$ תהיה שווה למספר התושבים המחזיקים בשתי הדעות. כעת גמרנו, כי ברור שמתקיימים תנאי "משפט החיבה הכוללת הקבועה" עם $\alpha = k$ ולכן יש שידוך. עבור כל אל ואלה ששודכו נבחר תושב בעל הדעות של הזוג וקבוצת n התושבים שנבחרו היא הפרלמנט שחפשנו.

הקוראים מוזמנים להוכיח חיזוק של מה שהוכחנו: שבהנחות שלנו אפשר לפרק את קבוצת $k \cdot n$ התושבים ל k פרלמנטים זרים.

ריבועי קסם

נקרא בשם "ריבוע קסם" למטריצה $n \times n$ של מספרים אישלילים, בה סכומי כל השורות וסכומי כל העמודות שווים לאותו מספר α.

מקרה פרטי של "ריבועי קסם", שנקרא להם: ריבועי קסם בסיסיים, הן מטריצות בהן קיים קבוע $\alpha > 0$ כך שלכל $i = 1, 2, \ldots, n$ השורה ה i כולה אפסים פרט למקום אחד $i, \sigma(i)$ שהוא α, ואותו דבר נכון לגבי העמודים. זה מחייב ש $i \neq j \Rightarrow \sigma(i) \neq \sigma(j)$ כלומר σ היא תמורה ב $\{1, 2, \ldots, n\}$.

נזכיר שסכום מטריצות $(a_{i,j})$ ו $(b_{i,j})$ מוגדר כמטריצה בה במקום i, j עומד $a_{i,j} + b_{i,j}$.

כתרגיל לקוראים נביא את הטענה הבאה, שאפשר להוכיח בעזרת "משפט החיבה הכוללת הקבועה":

משפט: כל ריבוע קסם ניתן להציג כסכום של ריבועי קסם בסיסיים (בדרך כלל יהיו הרבה הצגות שונות, אבל אחת לפחות קיימת).

הוכחת משפט דילוורת'

עלינו עוד לקיים את הבטחתנו ולהוכיח את משפט דילוורת' שעליו העמדנו את הכל.

ההוכחה תהיה באינדוקציה על מספר האיברים n בפוסט P: עבור $n = 0$ הטענה מיידית. יהיה נתון פוסט P בעל n איברים ותהי I תת־קבוצה ב"ת של P שמספר איבריה הוא המכסימלי האפשרי k. דבר זה אומר, בין השאר, שכל $a \in P$ חייב להיות ביחס שליטה עם איבר כלשהו של I (אם $a \in I$ אזי הוא ביחס שליטה עם עצמו ואם a לא ב I ואין לו יחס שליטה עם אף איבר ב I היינו יכולים לצרף אותו ל I). יתר על כן, אם a לא ב I חייבים כל יחסי השליטה שלו עם איברי I להיות באותו כיוון - לא ייתכן ש a שולט על $i \in I$ ונשלט ע"י $j \in I$, כי אז בגלל הטרנזיטיביות של היחס היה $j < i$ בעוד ש I ב"ת. לכן כל איברי P שאינם ב I נחלקים לקבוצה P^+ של אלה ששולטים על I וקבוצה P^- של אלה שנשלטים ע"י I.

נניח קודם שהקבוצות P^+ ו P^- אינן ריקות. אז המשלימות להן $P^- \cup I$ ו $P^+ \cup I$ הן בעלות פחות מ n איברים ונוכל לראות גם אותן כפוסטים. יתר על כן, אין בהן קבוצה ב"ת בת יותר מ k איברים כי ב P אין כזו. לפי הנחת האינדוקציה משפט דילוורת' נכון לגביהן ולכן גם את $P^- \cup I$ וגם את $P^+ \cup I$ אפשר להציג כאיחוד k שרשרות, וכל איבר ב I נמצא בדיוק באחת משרשרות אלה. אם עבור כל $i \in I$ נאחד את השרשרת מ $P^- \cup I$ ואת השרשרת מ $P^+ \cup I$ שמכילות את i נקבל k שרשרות ב P שאיחודן P (בדוק).

לכן: או P מקיימת את משפט דילוורת', אז תמיד P^+ או P^- ריקה. פירוש הדבר שכל קבוצה ב"ת I בעלת המספר המכסימלי k של איברים מורכבת או מאיברים a בעלי התכונה ש a אינו נשלט ע"י אף איבר שונה ממנו (נקרא לאיבר כזה "מלך". המינוח המתמטי המקובל הוא: "איבר מכסימלי". שימו לב שבפוסט יכולים להיות כמה איברים מכסימליים) או מאיברים a כך ש a אינו שולט על אף איבר שונה ממנו (נקרא להם "סמרטוטים". המינוח המקובל הוא: "איברים מינימליים").

יתר על כן, אם I מורכבת ממלכים, היא חייבת להיות קבוצת כל המלכים, כי כל מלך שאינו שייך לה נוכל לצרף לה ולקבל קבוצה ב"ת יותר גדולה. אותו דבר לגבי סמרטוטים. קבלנו איפוא שאו P מקיימת את משפט דילוורת' או יש בה לכל היותר שתי קבוצות ב"ת בעלות מכסימום איברים: קבוצת המלכים וקבוצת הסמרטוטים.

אני טוען שאם P לא ריקה יש זוג של מלך a וסמרטוט b כך ש $a > b$ (ייתכן גם $a = b$ - שימו לב שמלך יכול להיות סמרטוט!). כדי לקבל זוג כזה יש רק לצאת מאיבר כלשהו $c \in P$ ולהתחיל לעלות מ c במעלות השלטון עד שמגיעים למלך a וכן לרדת מ c במורד השליטה עד שמגיעים לסמרטוט b.

ניקח זוג של מלך a וסמרטוט b כך ש $a > b$. $\{a, b\}$ היא שרשרת. נוציא אותה מ P ונקבל פוסט יותר קטן P'. לפי הנחתנו הרסנו בכך את הקבוצות הב"ת היחידות ב P שהן בנות k איברים, ולכן ב P' כל קבוצה ב"ת היא בת פחות מ k איברים. לפי הנחת האינדוקציה P' מקיימת את משפט דילוורת'. לכן היא איחוד של לכל היותר $k - 1$ שרשרות. יחד עם השרשרת $\{a, b\}$ יש לנו את כל P, וגמרנו.

כיצד לנצח בשחמט

ברצוני להביא ולהוכיח משפט ידוע מתורת המשחקים בעל תוכן מפתיע - משחק השחמט (ומשחקים דומים לו) הוא בעיקרון "לא מעניין" במובן הבא: אחת משלוש האפשרויות הבאות מתקיימת: או יש ללבן איסטרטגיה (שאת ההוראות לה אפשר לתת אחת ולתמיד מראש), שאם ילך הלבן לפיה מובטח לו הניצחון, או יש לשחור איסטרטגיה שתבטיח את הניצחון לו, או יש לשניהם איסטרטגיות שיבטיחו לבעליהן ניצחון או תיקו, כך שאם ידבקו שניהם באיסטרטגיות אלה התוצאה תהיה בהכרח תיקו.

עם זאת, אל לחובבי השחמט להיבהל - למרות שאפשר להוכיח שקיימות איסטרטגיות כאלה, איש אינו יודע אותן, ואנו נסביר למה.

נתחיל מהגדרות:

קודם כל נניח שאם המשחק נמשך עד אינסוף, או - אם נקבע מספר מסעים מוגבל - אין נצחון לאף צד אחרי מספר מסעים זה, אזי התוצאה נחשבת לתיקו.

נניח שאנו נמצאים במצב כלשהו S של משחק שחמט. בשם **איסטרטגיה** של הלבן החל ממצב זה נקרא למערכת הוראות נתונה מראש הקובעת כיצד ינהג הלבן בכל מצב, בו תור הלבן לשחק, שייתכן בהמשך המשחק אחרי S , חוץ ממצב מט או מצב שאין ללבן שום מצב חוקי (שבשניהם, כידוע, המשחק מסתיים). נאמר שאיסטרטגיה כזו **מבטיחה ניצחון ללבן** אם כל עוד החל מ S ידבק הלבן באיסטרטגיה זו - מובטח לו הניצחון, לא חשוב מה יעשה השחור. בצורה דומה מוגדר מתי **איסטרטגיה כזו מבטיחה ללבן ניצחון או תיקו**. לגבי השחור הגדרות דומות.

כעת נוכיח את טענתנו: נתבונן במצב הפתיחה של המשחק. אם ללבן יש איסטרטגיה החל ממצב הפתיחה המבטיחה לו נצחון - גמרנו. אם לא, נעניק לשחור איסטרטגיה שתבטיח לו ניצחון או תיקו בצורה הבאה:

$(*)$ על השחור ללכת תמיד כך שהמצב שייוצר אחרי המסע של השחור לא יהיה מצב ממנו יש ללבן איסטרטגיה מנצחת (כלומר שמבטיחה ניצחון ללבן).

הקורא יכול לבדוק שאם אחרי מסע של השחור המשחק במצב (שנקרא לו S) בו אין ללבן איסטרטגיה מנצחת, אזי עבור כל מסע-נגד של הלבן יוכל השחור להגיב כך שעדיין לא תהיה ללבן איסטרטגיה מנצחת גם במצב שאחרי מסע השחור (כי אילו היה ללבן מסע שכל תגובה של השחור עליו תביא את הלבן לאיסטרטגיה מנצחת, היתה הבחירה במסע זה נותנת ללבן איסטרטגיה מנצחת גם במצב S). כיון שאנו מניחים שבמצב הפתיחה אין ללבן איסטרטגיה מנצחת, השחור יכול אכן לנהוג לפי $(*)$. אם כך ינהג השחור, אין שום אפשרות שהלבן ינצח, כי ברור שמצב בו הלבן ניצח הוא באופן טריביאלי מצב ממנו יש ללבן איסטרטגיה מנצחת, ואת זאת, הרי, השחור מונע.

הוכחנו, איפוא, שמתקיימת אחת משתי אפשרויות: או יש ללבן איסטרטגיה שתבטיח לו ניצחון, או יש לשחור איסטרטגיה שתבטיח לו ניצחון או תיקו. מסיבה דומה או יש לשחור איסטרטגיה שתבטיח לו ניצחון, או יש ללבן איסטרטגיה שתבטיח לו ניצחון או תיקו, והטענה הוכחה (שימו לב שלא ייתכן שגם ללבן וגם לשחור יש איסטרטגיה שמבטיחה ניצחון!).

כעת קל לראות מדוע, למרות שיש לנו הוכחה לקיום איסטרטגיות כאלה, אין איש יכול להצביע על איסטרטגיה כזו ולהשתמש בה באופן מעשי: האיסטרטגיה $(*)$ שהיצענו לשחור דורשת לדעת להכריע, עבור כל מצב S של המשחק, אם יש או אין ללבן איסטרטגיה מנצחת ממצב זה. בדיקה כזו דורשת את בדיקת כל האיסטרטגיות וההמשכים האפשריים של המשחק וכל עוד לא ימצא מישהו דרך גאונית לעשות זאת במעט חישובים, בדיקה כזו היא לגמרי לא מעשית.

לאמיתו של דבר, אנו הוכחנו שיש 3 אפשרויות (הלבן יכול לכפות ניצחון, השחור יכול לכפות ניצחון או שניהם יכולים לכפות תיקו), ועד כמה שידוע לי איש אינו יודע איזו אפשרות נכונה.

ולבסוף שתי הערות: קודם כל, ברור שלא היתה בהוכחתנו כל חשיבות לפרטי משחק השחמט ואותו דבר נכון לכל משחק בעל אותו אופי, למשל דמקה (הקורא מוזמן להגדיר באופן פורמלי את סוג המשחקים שלגביו ההוכחה תופסת). נוסף לכך, הראינו כאן דוגמא להוכחת קיום, המוכיחה שקיים לפחות עצם אחד בעל תכונה כלשהו אבל לא מראה כיצד לבנות עצם קונקרטי כזה. הוכחות חשובות ומפורסמות במתמטיקה הן הוכחות קיום כאלה, וכנגדן רבות גם הוכחות בניה, שכמצופה הן בדרך כלל יותר חשובות מבחינה מעשית.

גישה לא שגרתית ללוגריתמים הטבעיים ול e

במאמר זה נבנה את הלוגריתמים הטבעיים והמספר e בדרך שונה מזו המקובלת. בדרך נעשה משהו הפוך למה שעושים כאשר יוצאים ממשוואה רקורסיבית לסדרה ומחפשים ביטוי מפורש - אנו נצא מנוסחה מפורשת ונראה שכדאי דווקא לעבור לנוסחה רקורסיבית.

יהי x מספר ממשי חיובי. נתבונן בסדרה a_n הנתונה ע"י המשוואה הרקורסיבית:

$$a_0 = x \quad a_{n+1} = \sqrt{a_n}$$

לא קשה למצא ש $a_n = x^{1/2^n} = \sqrt[2^n]{x}$.

נתבונן כעת בסדרה b_n הנתונה ע"י

$$b_n = 2^n (a_n - 1) \tag{1}$$

ל b_n יש לנו נוסחה ישירה, אבל, כמו שנראה בהמשך, כדאי לנו דווקא למצא משוואה רקורסיבית שמקיימת b_n . לא קשה לעשות זאת. נבטא קודם את b_n לפי b_{n+1} (במקום להפך):

קודם כל $a_n = a_{n+1}^2$. לכן:

$$b_n = 2^n (a_n - 1) = 2^n \left(a_{n+1}^2 - 1\right) = 2^n (a_{n+1} - 1)(a_{n+1} + 1) = 2^n (a_{n+1} - 1)(a_{n+1} - 1 + 2) = \\ = 2^{-(n+2)} 2^{2(n+1)} (a_{n+1} - 1)^2 + 2^{n+1} (a_{n+1} - 1) = 2^{-(n+2)} b_{n+1}^2 + b_{n+1}$$

ומקבלים ביטוי ל b_n לפי b_{n+1}:

$$b_n = b_{n+1} + 2^{-(n+2)} b_{n+1}^2 \tag{2}$$

יש לנו כבר בונוס מכך שמצאנו את המשוואה (2) עבור b_n : מ (2) נובע ש $b_n \geq b_{n+1}$, כלומר הסדרה b_n יורדת (במובן החלש, כלומר לא־עולה. אם $x \neq 0$ היא עולה ממש). אם אנו רוצים לבטא את b_{n+1} לפי b_n , יש רק לפתור את המשוואה הריבועית (2) עבור b_{n+1} ומקבלים:

$$b_{n+1} = 2^{n+1} \left(-1 + \sqrt{1 + 2^{-n} b_n}\right) = \frac{2b_n}{1 + \sqrt{1 + 2^{-n} b_n}} \tag{3}$$

למשוואה הרקורסיבית (3) יש יתרון נוסף: הביטוי הישיר (1) הוא גרוע מאוד מבחינה חישובית: a_n שואף ל 1 (כי המעריך $1/2^n$ שואף לאפס) ולכן המספר $a_n - 1$ הוא קטן מאוד עבור n גדול, ואם יש לנו שגיאת העגלה ב a_n היא תהיה אחוז גדול מ $a_n - 1$ ולכן תשפיע הרבה על $b_n = 2^n (a_n - 1)$ עד כדי כך שמהר מאוד לא יהיה לנוסחה זו כל ערך לחישוב b_n . לביטוי האחרון ב (3) אין מגרעות כאלה. אמנם אם נשתמש בו הרבה פעמים עבור n־ים עוקבים תהיה הצטברות של שגיאת העגלה, אבל אין המצב קטסטרופלי כמו לגבי חישוב b_n בעזרת (1) .

הבטחנו קשר בין הסדרה b_n לביו לוגריתמים. קודם כל שימו לב שכל הסדרה b_n תלויה במספר החיובי x . לכן יותר נכון לכתוב $b_n(x)$. ראינו כבר ש $b_n(x)$ יורדת (במובן החלש). מכאן נובע קודם כל ש $b_n(x) \leq b_0(x)$ כלומר

$$b_n(x) \leq x - 1 \tag{4}$$

אנו רוצים למצא חסם תחתון ל $b_n(x)$. לשם כך נעזר בכך ש $a_n(x) = x^{1/2^n}$ ולכן

$$a_n(x) a_n(1/x) = 1$$

(אם רוצים אפשר להוכיח זאת באינדוקציה רק בעזרת המשוואה הרקורסיבית ל $a_n(x)$). מכאן:

$$b_n(x) = 2^n (a_n(x) - 1) = 2^n (a_n(x) - a_n(x) a_n(1/x)) = 2^n a_n(x) (1 - a_n(1/x))$$

כלומר

$$b_n(x) = -a_n(x) b_n(1/x) \tag{5}$$

אבל לפי (4) $b_n(1/x) \leq 1/x - 1$. לכן $-b_n(1/x) \geq 1 - 1/x$. מכאן ומ (5) מתקבל החסם התחתון ל $b_n(x)$ (שכן $a_n(x) > 0$):

$$b_n(x) \geq a_n(x)\,(1-1/x) \tag{6}$$

שימו לב שכיווני האישיוויונות ב (4) וב (6) הם אותם עבור $x>1$ ועבור $0<x<1$. מ (4) ו (6) אנו יכולים להסיק עכשיו שאם $x>1$ אזי לכל n $b_n(x)>0$ בעוד שאם $x<1$ אזי לכל n $b_n(x)<0$. אם $x=1$ ברור כבר מ (1) ש $b_n(x)=0$.

הקשר עם לוגריתמים ייוצר כאשר נשאיף $n\to\infty$. הסדרה $b_n(x)$ היא יורדת וחסומה (אם $x>1$ $0<b_n(x)\leq x-1$ ואם $x<1$ נובע מ (6) ומכך ש $0<a_n(x)<1$ ש $0>b_n(x)>1-1/x$). לכן יש לה גבול סופי כאשר $n\to\infty$. (כאן משתמשים בכך שאנו עובדים עם מספרים ממשיים, שאחת מתכונותיהם הבסיסיות היא המשפט האומר שלכל סדרה חסומה שהיא עולה או יורדת במובן החלש יש גבול) נסמן גבול זה ב $b(x)$. מה נוכל להגיד על $b(x)$?

קודם כל מ (4) ו (6) (ומכך ש $a_n(x)\to_{n\to\infty}1$ עבור כל x) נובעים אישיוויונות עבור $b(x)$:

$$1-x\geq b(x)\geq 1-1/x \tag{7}$$

מכאן

$$b(x)>0 \text{ עבור } x>1 \quad \text{ו} \quad b(x)<0 \text{ עבור } x<1 \tag{8}$$

ברור ש $b(1)=0$.

מהנוסחה $a_n(xy)=a_n(x)a_n(y)$ שנובעת מכך ש $a_n(x)=x^{1/2^n}$ (או, אם רוצים, אפשר להוכיח אותה באינדוקציה בעזרת המשוואה הרקורסיבית) אפשר לקבל נוסחה ל $b_n(xy)$:

$$\begin{aligned} &b_n(xy)=2^n\left(a_n(xy)-1\right)=2^n\left(a_n(x)a_n(y)-1\right)= \\ &=2^n\left(a_n(x)\left(a_n(y)-1\right)+a_n(x)-1\right)=a_n(x)b_n(y)+b_n(x) \end{aligned}$$

נשאיף $n\to\infty$. כיון ש $a_n(x)\to 1$ ו $b_n(x)\to b(x)$ מתקבל:

$$b(xy)=b(x)+b(y) \tag{9}$$

ל $b(x)$ יש, איפוא, תכונה של לוגריתם.

מ (9) נובע ש $b(x/y)=b(x)-b(y)$ ומכאן ומ (8) נובע שאם $x/y>1$ אזי $b(x)-b(y)>0$, כלומר $b(x)$ פונקציה עולה.

נבחר כעת מספר כלשהו $c>1$ ונתבונן ב $b(c^s)$. מ (9) נובע:

$$b(c^{s+t})=b(c^s c^t)=b(c^s)+b(c^t) \tag{10}$$

כלומר $f(s)=b(c^s)$ היא פונקציה עולה המקיימת את המשוואה הפונקציונאלית

$$f(s+t)=f(s)+f(t)$$

כידוע פונקציה כזו חייבת להיות מהצורה $f(s)=s\cdot f(1)$ (מדוע?) לכן

$$b(c^s)=s\cdot b(c) \tag{11}$$

מ (11) נובע ש $b(x)$ מקבלת כל ערך ממשי (כי אם נבחר $c>1$ כלשהו, נקבל ש b מקבלת כל ערך מהצורה $s\cdot b(c)$ כלומר כל ערך ממשי). $x\mapsto b(x)$ היא, איפוא, העתקה חד־חד־ערכית מהמספרים הממשיים החיוביים על כל המספרים הממשיים.

במיוחד קיים מספר יחיד, שנסמן אותו ב e, עבורו $b(e)=1$. אם נציב ב (11) e במקום c נקבל:

$$b(e^s)=s \tag{12}$$

כלומר:

$$b(x)=log_e(x) \tag{13}$$

$b(x)$ אינה אלא הלוגריתם לפי הבסיס e, ומסמנים אותה בדרך כלל ב $\ln x$ (לוגריתם טבעי).

נעיר שיש לנו דרך לחשב את $\ln x$ בעזרת הסדרה $b_n(x)$, כאשר מחשבים אותה ע"י הנוסחה הרקורסיבית (3) (הביטוי השני שם). למשל נחשב את $\ln 2$ ע"י חישוב $b_n(2)$:

מתחילים מ $b_0(2) = 2 - 1 = 1$ ומחשבים לפי הנוסחה

$$b_{n+1} = \frac{2 \cdot b_n}{1 + \sqrt{1 + 2^{-n} b_n}}$$

מקבלים את הסדרה:

1 .8284271 .7568285 .724062 .7083805 .7007088 .6969143 .6950275 .6940865
.6936166 .6933819 .6932646 .6932059 .6931766 .6931619 .6931545 .6931508
.693149 .6931481 .6931476 .6931475 .6931472 .6931471 .6931471 .6931471

ומכאן $\ln 2 = 0.693147$. היינו צריכים כ־20 צעדים כדי להגיע לדיוק של 6 ספרות אחרי הנקודה. יש דרכים לזרז את מהירות ההתכנסות ע"י מניפולציות בסדרה.

את אי השיוויונות (7) נוכל כעת לכתוב:

$$x - 1 \geq \ln x \geq 1 - 1/x \tag{14}$$

אם נציב $1 + y$ במקום x נקבל:

$$y \geq \ln(1 + y) \geq 1 - \frac{1}{1 + y} = \frac{y}{1 + y}$$

ומכאן:

אם $y > 0$ אזי $1 \geq \frac{\ln(1+y)}{y} \geq \frac{1}{1+y}$ ואם $-1 < y < 0$ אזי $1 \leq \frac{\ln(1+y)}{y} \leq \frac{1}{1+y}$

ואם נשאיף $y \to 0$ נקבל:

$$\frac{ln(1 + y)}{y} \to_{y \to 0} 1 \tag{15}$$

עובדה ממנה אפשר להסיק ש $\ln(x)$ גזירה ואת נוסחת הנגזרת.

אנו נשתמש ב (15) לקבל ביטוי ל e. נעלה e בחזקת שני אגפי (15) ונשתמש בכך שפונקציית החזקה רציפה (למעשה השתמשנו בכך כבר כאשר טענו ש $a_n(x) \to_{n \to \infty} 1$). נקבל

$$(1 + y)^{1/y} \to_{y \to 0} e$$

מה שנוכל לכתוב גם כ:

$$e = \lim_{r \to \pm\infty} \left(1 + \frac{1}{r}\right)^r$$

הערת סיכום אנו רצינו לבנות את הלוגריתם הטבעי. נשאיר לקוראים להסיק בשורה אחת ש $b_n(x) \to_{n \to \infty} \ln x$ כאשר מניחים שמושג הלוגריתם, כולל נוסחת הנגזרת של פונקציה לוגריתמית ומעריכית, ידועים.

ישרים על היפרבולואיד ומעגלים על טורוס

ניקח היפרבולה, בעלת המשוואה $\alpha y^2 - \beta x^2 = 1$, $\alpha, \beta > 0$ קבועים, ונסובב אותה סביב ציר x (שהוא הציר שלא חותך אותה - ההיפרבולה מורכבת משני ענפים משני צידי ציר זה). נקבל משטח שנקרא "היפרבולואיד חד-יריעתי" שהוא בעל המשוואה:

$$\alpha(y^2 + z^2) - \beta x^2 = 1 \tag{1}$$

(השם "חד יריעתי" נובע מכך שהיפרבולאיד זה מורכב מיריעה אחת, בניגוד להיפרבולואיד שנוצר מסיבוב ההיפרבולה סביב הציר השני שלה שחותך אותה, שמורכב משתי יריעות נפרדות).

ניקח מעגל במישור, ונסובב אותו סביב ישר באותו מישור שאינו חותך את המעגל. נקבל צורה דמויית כעך, שנקראת טורוס (Torus). ברור שעל הטורוס יש שתי משפחות של מעגלים: "קווי אורך" - המעגלים בהם יימצא המעגל המסתובב במהלך הסיבוב ו"קווי רוחב" - המסלולים המעגליים של כל אחת מנקודות המעגל.

מתברר ששני משטחים אלה צופנים הפתעות: על ההיפרבולואיד החד-יריעתי יש שתי משפחות של ישרים משופעים שכל אחת מהן עוטפת אותו מסביב (דרך כל נקודה בהיפרבולואיד עובר ישר אחד מכל משפחה). אם תבנו משטח גלילי ע"י שני מעגלי תיל-מתכת אופקיים מקבילים שמחוברים ביניהם בחוטים גמישים מתוחים מאונכים, ואם תחזיקו את המעגל התחתון קבוע ותסובבו את המעגל העליון סביב הציר המשותף לשני המעגלים - ייהפך המשטח הגלילי להיפרבולואיד חד-יריעתי שיהיה צר יותר ככל שתמשיכו לסובב את המעגל. החוטים יהוו אחת ממשפחות הישרים בהיפרבולאיד זה (המשפחה השניה תתקבל משיקוף החוטים בראי מאונך העובר דרך ציר המעגלים). מתקן כזה יש, למשל, במוזיאון המדע שבבנין הטכניון הישן בהדר בחיפה.

על הטורוס יש, מלבד שתי משפחות המעגלים ה"ברורות" שצוינו למעלה, עוד שתי משפחות מעגלים "אלכסוניים" (דרך כל נקודה בטורוס עובר מעגל אחד מכל משפחה), שכולם בעלי אותו רדיוס. מבחינה טופולוגית מעגלים אלה "חובקים" את הטורוס: אם היינו לוקחים טורוס מוצק מלא, והיינו שמים חוט לאורך מעגל כזה ומחברים את קצות החוט, היינו מקבלים לולאת חוט שהיא משולבת במרחב עם הטורוס - אם תופסים את הלולאה ומרימים אותה, תיווצר לולאה גדולה שממנה נתלה הטורוס המוצק, ואפשר להתהלך עם הטורוס כמו שמחזיקים תיק בידית.

את קיום המשפחות ה"מפתיעות" של ישרים ומעגלים אפשר להוכיח ע"י חשבונות בגאומטריה אנליטית במרחב. כאן אני רוצה להסביר את קיומם עם מינימום של חשבונות, וכן לקשר בין שני המשטחים: ההיפרבולואיד עם ישריו והטורוס עם מעגליו. אבל לשם כך עלי "לחיות" קצת במרחב 4-ממדי.

במרחב ה 4-ממדי (שהכל בו יהיה אנלוגי למרחב 2 או 3 -ממדי) נסמן את 4 הקואורדינטות הקרטזיות ב (x, y, z, u). נתבונן ב

$$\lambda(y^2 + z^2) = x^2 + u^2 \tag{2}$$

כאשר $\lambda > 0$ קבוע. (2) הוא "על-משטח" 3-ממדי בתוך המרחב ה 4-ממדי. נוכל להגיד עליו שני דברים: קודם כל, אם α ו β מספרים חיוביים כך ש $\lambda = \alpha/\beta$ ונחתוך את (2) ע"י ה"על-מישור ה 3-ממדי" $u = 1/\sqrt{\beta}$ נקבל בדיוק את המשטח (1). בנוסף לכך (2) הוא "חרוט" עם קודקוד בראשית הצירים, כי קל לראות שאם (x, y, z, u) נמצאת עליו, כלומר מקיימת את (2), אזי כל נקודה (tx, ty, tz, tu) עבור t ממשי כלשהו גם היא מקיימת את (2) כלומר נמצאת על המשטח, ופירוש הדבר שכל נקודה על הישר המחבר את (x, y, z, u) עם הראשית נמצאת על המשטח. (2) הוא, איפוא, ה"חרוט" במרחב ה 4-ממדי בעל "בסיס" שהוא ההיפרבולאיד (1) (למעשה יש ב (2) עוד ישרים יוצרים מקבילים ל"על-מישור ה 3-ממדי" שבו נמצא היפרבולואיד הבסיס, שכאילו מחברים את הקודקוד בראשית לנקודות באינסוף של ההיפרבולואיד).

נרשום את (2) בצורה:

$$\lambda y^2 - x^2 = u^2 - \lambda z^2$$

כלומר

$$(\sqrt{\lambda}y + x)(\sqrt{\lambda}y - x) = (u + \sqrt{\lambda}z)(u - \sqrt{\lambda}z) \tag{2'}$$

מ (2') רואים שה"על-חרוט" (2) מכיל, עבור כל t קבוע, את המישור ה 2-ממדי דרך הראשית שנתון ע"י:

$$\begin{cases} (\sqrt{\lambda}y + x) = t(u + \sqrt{\lambda}z) \\ (u - \sqrt{\lambda}z) = t(\sqrt{\lambda}y - x) \end{cases}$$

מדוע זה אכן מישור 2-ממדי?

כאשר t משתנה, מקבלים משפחה של מישורים 2־ממדיים דרך הראשית שמוכלים ב"על־חרוט". החיתוכים שלהם עם העל־מישור ה 3־ממדי של הבסיס (1) יתנו את אחת ממשפחות הישרים עליו. המשפחה השנייה מתקבלת ממשפחת המישורים ה 2־ממדיים על ה"על־חרוט":

$$\begin{cases} (\sqrt{\lambda}y + x) = t(u - \sqrt{\lambda}z) \\ (u + \sqrt{\lambda}z) = t(\sqrt{\lambda}y - x) \end{cases}$$

קבוצת הנקודות במרחק 1 מן הראשית במרחב ה 4־ממדי מהווה "על־כדור" שנתון ע"י המשוואה:

$$x^2 + y^2 + z^2 + u^2 = 1 \tag{3}$$

קל לראות שהחיתוך של ה"על־חרוט" (2) עם ה"על־כדור" (3) נתון ע"י:

$$\begin{cases} x^2 + u^2 = a^2 \\ y^2 + z^2 = b^2 \end{cases} \quad \text{כאשר} \quad a = \sqrt{\frac{\lambda}{1+\lambda}} \quad b = \sqrt{\frac{1}{1+\lambda}} \tag{4}$$

כלומר נקודה היא על החיתוך (4) אם קואורדינטות (x, u) שלה נמצאות על מעגל אחד וקואורדינטות (y, z) נמצאות על מעגל שני. ע"י קביעת (x, u) נקבל משפחה של מעגלים ברדיוס b שמכסה את (4) וע"י קביעת (y, z) נקבל משפחה אחרת של מעגלים ברדיוס a. אלה הם מעגלי "אורך ורוחב". בנוסף לכך שתי משפחות המישורים ה 2־ממדיים על (2) חותכות את (3), כלומר את (4), בשתי משפחות של מעגלים גדולים, כלומר בעלי רדיוס 1. סך הכל מצאנו על (4) ארבע משפחות של מעגלים.

הקוראים הרגישו בוודאי ש (4) הוא מעין טורוס. כדי לחזור לטורוס במרחב ה 3־ממדי בו דננו בהתחלה, נבצע הטלה סטיריאוגרפית במרחב ה 4־ממדי מהנקודה $N : x = y = z = 0, u = 1$ על העל־מישור ה 3־ממדי $u = 0$. (בהטלה סטיריאוגרפית רגילה, במרחב 3־ממדי, מטילים כדור מהקוטב הצפוני על מישור מקביל לקו המשווה. להעתקה כזו יש התכונה שהיא מעתיקה כל מעגל על הכדור למעגל במישור (ראו במאמרו של צחיק גלנדר: אינוורסיה, אתגר־גליונות מתמטיקה מס' 31, תמוז תשנ"ד - יוני 1994). הכל סימטרי לגבי סיבוב בקואורדינטות (y, z) ולכן נראה את כל התמונה אם ניקח קודם $y = 0$ ואחר כך נסובב. $y = 0$ נותן מ (3) כדור ברדיוס 1 סביב הראשית במרחב 3־ממדי עם קואורדינטות (x, z, u), מ (4) שני מעגלים ברדיוס a במישורים $z = \pm b$ אותם מטילים בהטלה סטיריאוגרפית (במרחב 3־ממדי) מ $N : x = z = 0, u = 1$ על המישור $u = 0$ והם יוטלו לשני מעגלים על מישור זה שמטעמי סימטריה הם בעלי אותו רדיוס. אחרי הסיבוב (שמחזיר אותנו ל 4 ממדים) נקבל מהם טורוס שהוא, איפוא, תמונת (4) ע"י ההטלה הסטיריאוגרפית במרחב ה 4־ממדי. ארבע משפחות המעגלים על (4) יתנו ארבע משפחות על הטורוס, ובמיוחד המעגלים הגדולים ברדיוס 1 יתנו את שתי המשפחות האלכסוניות "הלא־צפויות".

שתי תכונות של הכדור

במאמר זה נביא שתי תכונות מפתיעות של הכדור. דרך אגב, תכונות אלו מתקיימות רק במרחב תלת־ממדי. הקוראים מוזמנים לבדוק שלמעגל אין תכונות אנלוגיות (וגם לכדור במרחב n ממדי, $n > 3$ אין). האם תוכלו לתת "נימוק" לכך?

ב"כדור" אני מתכוון לפני הכדור ולא לכדור המלא.

ה**תכונה הראשונה** עליה מדובר די ידועה: נתון כדור, ונקרא לנקודה הגבוהה ביותר בכדור בשם הקוטב הצפוני ולנקודה התחתונה ביותר בשם הקוטב הדרומי. נתון גליל (כלומר משטח גלילי, לא גליל מלא) שגובהו כגובה הכדור, צירו הוא הקטע הישר המחבר את הקטבים והוא משיק לכדור בקו המשווה. אז הטענה היא:

אם מטילים את הכדור לגליל ע"י רדיוסים אופקיים היוצאים מציר הגליל, אזי בהטלה כזו נשמר השטח, כלומר כל יציר על הכדור מועתק ליציר על הגליל שהוא בעל אותו שטח (למרות שצורת היציר עלולה להתעוות לחלוטין).

בעובדה זו משתמשים בהכנת מפות של כדור הארץ בהן נשמר השטח, במחיר עיוות הצורה.

הוכחה: כאן מדובר בשטח של יציר על משטח עקום, ולכן כדי לדון במשפט כזה כמו שצריך באופן מדויק מבחינה מתמטית רצוי קודם להביא הגדרה של שטח יציר על משטח עקום (אפשר לתת הגדרות מיוחדות עבור כדור או גליל המסתמכות על הסימטריה שלהם, למשל עבור כדור יש יחס קבוע בין שטח יציר על פני הכדור ונפח פלח הכדור המלא שנוצר מהקטעים המחברים את נקודות היציר למרכז, אבל זה לא נראה לי כל־כך בריא...). ההגדרות הכלליות מסתמכות על החשבון הדיפרנציאלי והאינטגרלי. לא אביא אותן, ולכן לא אוכל לתת הוכחה לגמרי מדוייקת, אבל אעזר בעקרונות של החשבון הדיפרנציאלי והאינטגרלי לתת הצדקה לטענה שלעיל:

הרעיון הוא שיש לנו שני כאילו "פילוגי מסה" אנכיים: באחד מתאים לכל אינטרוול בין שני גבהים שטח הטבעת בכדור שבין שני גבהים אלה, ובשני מתאים לאינטרוול שטח הטבעת בגליל שביניהם. מטעמי סימטריה סיבובית, אם לא לוקחים טבעת שלמה אלא רק גזרה זוית שלה, יש להכפיל את השטח ביחס בין הזוית ו 2π - זה נכון הן בכדור והן בגליל. כעת מהשיקולים הרגילים של קירוב שטח ע"י "יציר מדרגות" משתכנעים שדי להוכיח שעבור כל אינטרוול גבהים שטחי הטבעות על הכדור ועל הגליל שווים, כלומר ששני "פילוגי מסה" אלה נותנים אותו ערך עבור כל אינטרוול גבהים. ולפי כלל בסיסי (שאינו למעשה אלא "המשפט היסודי של החשבון הדיפרנציאלי והאינטגרלי") מספיק לנו להשוות את הדיפרנציאלים של שני הפילוגים, כלומר להוכיח ששני הפילוגים שווים עבור אינטרוול גובה "קטן" dh בסביבת גובה כלשהו h כאשר מותר לנו לזרוק את כל האיברים פרט לקירוב הראשון ב dh.

במקום להשתמש בגובה h כפרמטר נשתמש בזוית ϕ הנוצרת בין הרדיוס אל המעגל על הכדור בגובה h ובין הרדיוס לקוטב הדרומי. אז (R הוא רדיוס הכדור והגליל):

$$h = -R\cos\phi$$

$$dh = R\sin\phi\, d\phi$$

ולכן שטח הטבעת הגלילית עבור $d\phi$ הוא $2\pi R \cdot R\sin\phi\, d\phi$. באשר לטבעת הכדורית, בקירוב ראשון ב $d\phi$ היא עיבוי ברוחב $R\, d\phi$ של מעגל על הכדור שרדיוסו $R\sin\phi$ ולכן הקפו $2\pi R\sin\phi$ ושטח הטבעת הוא $2\pi R\sin\phi \cdot R\, d\phi$. כלומר הדיפרנציאלים של שני "פילוגי המסה" שווים.

ה**תכונה השנייה** ידועה פחות: נתון אולם (למשל פלנטריום) גדול בצורת כדור. במרכזו יש ראי כדורי קטן, ומאירים אותו מלמעלה באלומת קרניים מקבילות, כך שהתפלגות עוצמת האור במישור אופקי היא אחידה, כלומר כמות אנרגית האור שעוברת דרך שטח חתך אופקי כלשהו פרופורציונית לשטח החתך. האור פוגע בראי הכדורי וחוזר לכל הכיוונים, ומאיר את הקיר הכדורי של האולם. הטענה היא:

האור המוחזר מהראי יאיר את הקיר הכדורי של האולם באופן אחיד (כאן מניחים שאפשר להזניח את גודל הראי לעומת גודל האולם).

שימו לב שרק החלק העליון של הראי יקלוט ויחזיר אור, אבל האור שפוגע כמעט בהשקה לראי יוחזר בזוית כמעט $180°$ ויאיר את הקיר כמעט מתחת לראי.

הוכחה: תסתמך על התכונה הראשונה. מטעמי סימטריה סיבובית די להוכיח את הטענה הבאה:

נתבונן בזוית ϕ, ונתבונן באור שפוגע בראי באזור של הראי שבתוך חרוט שקדקדו במרכז הראי, צירו פונה למעלה והוא בעל חצי זוית פתיחה ϕ. תהי E כמות האנרגיה שבאור זה ו S שטח הקיר אליו יגיע האור המוחזר. אז די להוכיח ש S פרופורציוני ל E בלי תלות ב ϕ כלומר ש S/E קבוע.

אבל E פרופורציוני לשטח החתך האופקי של האור, שהוא שטח העיגול המוגבל במעגל Γ בו חותך החרוט את הראי, שהוא (r = רדיוס הראי): $\pi(r\sin\phi)^2$ כלומר E פרופורציוני ל $\sin^2\phi$. כל קרן אור שפוגעת בראי תוחזר בכיוון שהוא השיקוף של כיוון הפגיעה לגבי הרדיוס. במיוחד האור שפוגע ב Γ יוחזר בזוית 2ϕ לגבי הכיוון למעלה, ויפגע בקיר במעגל $\tilde{\Gamma}$. לפי התכונה הראשונה S פרופורציוני להטלת החלק של הקיר שמעל ל $\tilde{\Gamma}$ על הציר האנכי, כלומר ל $1-\cos(2\phi)$. נותר רק להזכיר ש $1-\cos(2\phi)=2\sin^2\phi$.

שני פרקים על e ועל עושר ופרנסה לנצח

פרק ראשון: עושר ופרנסה לנצח

נתבונן בבעיה הבאה: לאדם יש חסכונות בבנק שערכם היום S. בכל שנה הוא מרויח (כלומר מכניס לבנק) סכום מסויים, ומוציא (מהבנק) סכום מסויים. נסמן את עודף ההוצאה על ההכנסה בשנה ב E (לגבי אדם עובד יש לקוות ש E שלילי, אבל אני בחרתי את נקודת ראותו של העצלן, או, להבדיל, של מי שיצא לפנסיה, עבורו E עלול להיות חיובי). עברו T שנים. אנו שואלים: כמה כסף יהיה לאדם אחרי T השנים, או, אם E חיובי והאדם יצא לפנסיה ורוצה לחיות מחסכונותיו ־ כמה שנים יוכל לחיות כך, כלומר מתי ייגמר כספו?

נניח שאנו מודדים הכל לפי הערך הריאלי של הכסף, כלומר מקזזים את האינפלציה (בישראל נקבל זאת, בערך, אם נחלק את הסכום בעלית המדד), ונניח שבמונחים ריאליים E הוא קבוע, ובנוסף לכך יש אחוז קבוע R של הריבית (הריאלית), כלומר אדם יכול להשקיע את כספו בלי סיכון ניכר כך שהכסף יעלה, ריאלית, R אחוז כל שנה. כדי לפשט את המתמטיקה נניח ש R אינו האחוז אלא האחוז מחולק ב 100. למשל 3% יתאים ל $R = 0.03$.

הגודל R מתייחס לריבית השנתית, כלומר לעליית ערך הכסף המושקע במשך שנה. אבל תהליך זה אינו קורה פעם בשנה דווקא, והתאור המתמטי הטוב ביותר של המציאות הוא תאור רציף: הכסף עולה באופן רציף כפונקציה אקספוננציאלית (מעריכית) של הזמן, והגודל הנוח כדי לבטא את עלית סכום הכסף (שנסמן גם אותו ב R) הוא

$$R = \frac{dS/dT}{S} \tag{1}$$

מה שאומר, כמו שידוע מתכונות פונקציות אקספוננציאליות:

$$S = C \cdot e^{RT} \tag{2}$$

כאשר C קבוע.

למעשה ה R האחרון אינו בדיוק ה R הקודם שלנו, כלומר הריבית השנתית. קל לראות שאם מתקיימת (2) אזי מידת העליה במשך שנה היא:

$$e^R - 1 \tag{3}$$

אבל אם גודל זה קטן לגבי 1, מה שקורה בדרך כלל (ערך טיפוסי, הוא, למשל, $R = 0.03$ שהוזכר למעלה) אפשר לזהות, בקירוב, את R עם (3).

לכאורה, הדרך לפנינו סלולה: יש לנו משוואה דיפרנציאלית:

$$\frac{dS}{dT} = RS - E$$

שלא קשה לפתור אותה ולענות על כל השאלות שהצגנו. אבל אני רוצה להציע דרך אחרת, שלא מסתמכת על תורת המשוואות הדיפרנציאליות ולא קשה לחשב בה אפילו בעל פה (ראו הפרק השני). למעשה זו אינה אלא דרך ציורית להציג שיטה לפתור משוואה דיפרנציאלית כזו.

אם לאדם יש סכום S בבנק, והריבית היא R, אזי בזמן dT נוסף לו סכום $RS\,dT$. אם גם עודף ההוצאות על ההכנסות בזמן זה הוא $RS\,dT$, כלומר אם בדיוק $E = RS$, אזי חסכונותיו של האדם **לא משתנים** ומובטחת לו פרנסה לנצח, ומצד שני חסכונותיו אינם גדלים. בדומה לכך, אם יש לאדם עודף הכנסות על הוצאות $-E$, הוא יכול להרשות לעצמו להיות בחוב $-S$ כאשר $E = RS$, והעודף יקזז בדיוק את הריבית של החוב כך שהחוב יישאר קבוע.

במקרה הכללי כאשר לא מתקיים, דוקא, $E = RS$, נציע לאדם לבצע את התחבולה המחשבתית הבאה: לדמות לעצמו בנק דמיוני שמלווה לו את הסכום $E/R - S$ (או אם $RS > E$, לווה ממנו את הסכום $S - E/R$). אנו מניחים שהכסף (או החוב) בבנק הפיקטיבי עולה בשיעור R שעליו דובר למעלה. בעולם החלומות נמצא כעת האדם בדיוק במצב $E = RS$ והוא יכול לחיות מחסכונותיו, שאינם משתנים (או לחיות עם חוב קבוע, שאינו משתנה). כלומר: הכסף האמיתי $+$ הפיקטיבי נשאר קבוע. מה שמשתנה הוא הכסף הפיקטיבי, הגדל כל הזמן: בפרק זמן T הוא גדל פי e^{RT}. אם כסף זה מושקע בבנק הפיקטיבי (כלומר אם היה $RS > E$) אשרי האדם וטוב לו: למעשה שום דבר לא מושקע בשום בנק, אלא החסכונות גדלים לנצח. אם, לעומת זאת היה $E < RS$ יוכל האדם לחלום שהוא מקבל הלוואה מבנק פיקטיבי, כל עוד סכום הלוואה זו קטן או שווה מהסכום הקבוע שיש כאילו לאדם, כלומר כל עוד סכום הכסף שיש לו באמת הוא אי־שלילי.

התמונה תובהר ע"י מספר דוגמאות. כדי לא להסתבך, החישובים שלנו יהיו מקורבים (ממילא כל ההנחות שלנו על ערכים קבועים של E ו R מתקיימים, אם בכלל, רק בקירוב):

דוגמא 1: אדם מרוויח כל חודש 1000 דולר יותר ממה שהוא מוציא. (נצא כאן מהנחה שהדולר שומר על ערכו, כלומר הוא מתאים כיחידה ריאלית). הריבית במשק היא $R = 0.03$ (כ־3% לשנה). אם יחסוך סכום זה, כמה יהיה לו אחרי 50 שנה?

פתרון: במשך 50 שנה גדל הכסף המושקע (סולידית, בלי סיכונים) בהתאם לריבית כלומר ב $e^{50\cdot 0.03} = e^{1.5}$. מספר זה הוא בקירוב 4.5 (ראו בפרק השני). מצד שני, עבור שנה $E = -12000$ ולכן יוכל גיבורנו להחזיק חוב דמיוני קבוע של $12000/0.03 = 400000$ דולר, ולכסות את הריבית מהכנסתו. בדמיון יש לו עכשיו 400000 דולר שבמשך 50 שנה ייהפכו ל $400000 \cdot 4.5 = 1800000$. אחרי ניכוי החוב הדמיוני יהיו לו אחרי 50 שנה 1400000 דולר.

דוגמא 2: אדם רוצה שיהיו לו אחרי 20 שנה 1000000 יחידות ריאליות, והריבית היא 4%. כמה עליו לחסוך כל חודש?

פתרון: במשך 20 שנה גדל הכסף פי $e^{20\cdot 0.04} = e^{0.8}$ שהוא בקירוב $2\frac{1}{4}$ (ראו פרק שני). הוא ייקח עכשיו הלוואה דמיונית, שכאשר תגדל פי $2\frac{1}{4}$ יתווסף לה המיליון שיהיה לו אחרי 20 שנה, מכאן שערכה $1000000 : (1\frac{1}{4}) \cdot 1 = 800000$ והריבית כל שנה תהיה $800000 \cdot 0.04 = 32000$ ולחודש כ 2650 יחידות ־ זהו הסכום שעליו לחסוך בחודש.

דוגמא 3: לאדם יש חסכון של 200000 יחידות ריאליות. הוצאותיו גדולות מהכנסתו ב 1000 יחידות לחודש. כמה זמן יוכל לחיות כך אם הריבית במשק תהיה א. 3% ב. 4%?

פתרון:
א. $R = 0.03$. עבור שנה $E = 12000$. הוא יוכל לחיות עם סכום קבוע $12000/0.03 = 400000$ לכן עליו לקחת הלוואה מבנק דמיוני 200000. הלוואה זו נושאת ריבית, והאדם יישאר עם כסף אמיתי עד שסכום ההלוואה יגיע ל 400000, כלומר יגדל פי 2. $\ln 2 \sim 0.7$ (ראה פרק שני) כלומר מספר השנים T מקיים $TR = 0.7$ ולכן $T = 0.7/R = 0.7/0.03 = 23$ ־ הוא יתרושש אחרי 23 שנים.

ב. $R = 0.04$ אז הסכום הקבוע צריך להיות $12000/0.04 = 300000$. עליו ללוות בדמיון 100000 וסכום ההלוואה יכול לגדול פי 3. $\ln 3 \sim 1.1$; $T = 1.1/0.04 = 27$. הוא יוכל לחיות כך 27 שנים.

דוגמה 4: בהנחה $R = 0.03$, איזה סכום של חסכון היה צריך להיות עכשיו לאדם מדוגמא 3 כדי שיוכל לחיות כך עוד 40 שנה? אם החסכון שלו עכשיו הוא, כמו בדוגמא 3, 200000 יחידות, והוא רוצה להתפרנס עוד 40 שנה, כמה מותר לו להוציא בחודש יותר ממה שהוא מרויח?

פתרון: במשך 40 שנה גדל הכסף פי $e^{40\cdot 0.03} = e^{1.2}$ שהוא בקירוב $3\frac{1}{3}$ (ראו פרק שני). הוא יכול ללוות עכשיו סכום פיקטיבי, כך שאם יגדל פי $3\frac{1}{3}$ הוא יהיה שווה לסכום האמיתי שלו עכשיו + הסכום הפיקטיבי = לסכום הקבוע. מכאן שהסכום הפיקטיבי, הסכום האמיתי שיש לו והסכום הקבוע (ששווה לסכומם) מתייחסים כמו $3 : 7 : 10 = 3\frac{1}{3} : 2\frac{1}{3} : 1$. במצב של $E = 12000$ צריך הסכום הקבוע להיות $= 12000/0.03$ 400000 ולכן ערך חסכונותיו עכשיו צריך להיות $400000 \cdot 7 : 10 = 280000$. אם החסכונות עכשיו הם רק 200000 צריך הסכום הקבוע להיות $285000 \sim 200000 \cdot 10 : 7$ ו $E = 285000 \cdot 0.03 = 8500$ (בשנה), כלומר הוא יכול להוציא כ 700 יחידות בחודש יותר מהכנסתו.

פרק שני: חישובים מקורבים של חזקות של e

כמו שראינו בפרק הראשון, כדי שהאדם שלנו שחולם על פרנסה ועושר לא ייאלץ לקחת נייר ועט או לגשת למחשב הוא זקוק לדרך לחשב בראשו בקירוב חזקות של e (ומכאן לוגריתמים טבעיים). נציע דרך לעשות זאת, בעקבות הפיסיקאי הדגול ריצ'רד פיינמן (Richard Feynman, ראו הפרק Lucky Numbers בסיפרו: Surely You're Joking, Mr. Feynman), אבל נסתפק רק בקירוב ראשון. עם זאת, צריך להיזהר לא לעשות חישובים עם יותר מדי שלבים, כי אז עלולה השגיאה להצטבר.

אנו מניחים שהעצלן שלנו זוכר שבקירוב $e = 2.7$, $\ln 10 = 2.3$, $e^3 = 20$. טענה אחרונה זו שקולה ל $\ln 2 = \ln 20 - \ln 10 = 0.7$.

כעת נוכל להמשיך: $\ln 4 = 1.4$ $\ln 8 = 2.1$ $\ln 5 = \ln 10 - \ln 2 = 1.6$.

$\ln 80 = \ln 10 + \ln 8 = 4.4$ **ואם נאמר ש** $3^4 = 81 \sim 80$ **אזי** $\ln 3 = 1.1$. **אז** $\ln 9 = 2.2$ $\ln 4.5 = \ln 9 - \ln 2 =$ $\ln 6 = \ln 2 + \ln 3 = 1.8$ 1.5

$7^2 = 49 \sim 50$ $\ln 50 = \ln 10 + \ln 5 = 3.9$ $\ln 7 = \frac{1}{2}\ln 50 = 1.95$.

$e^{0.2} = e^{2.3-2.1} = 10/8 = 1.25$ $e^{0.4} = e^{1.1-0.7} = 3/2 = 1.5$ $e^{0.6} = e^{2.8-2.2} = 4^2/9 = 1.8$
$e^{0.8} = e^{1.5-0.7} = 4.5/2 = 2\frac{1}{4}$ $e^{0.9} = e^{1.6-0.7} = 5/2 = 2.5$ $e^{1.2} = e^{2.3-1.1} = 10/3 = 3\frac{1}{3}$

וכו'

וכך אפשר להמשיך. אמנם עד כאן לא הגענו לסתירה, אבל כאמור השגיאה יכולה להצטבר. אם רוצים יותר ביטחון בחישוב, אפשר לעשות הערכת שגיאה: למעשה $e = 2.718\ldots$ כך שב $e = 2.7$ יש שגיאה של כ 0.7%; $\ln 10 = 2.30258\ldots$ כך שב $\ln 10 = 2.3$ יש שגיאה של כ $+0.002$ ולכן ב $e^{2.3} = 10$ השגיאה היא כ 0.2% (לפי הנוסחה $d(\ln x) = dx/x$); $\ln 2 = 0.693\ldots$ ולכן ב $\ln 2 = 0.7$ השגיאה היא כ -0.007 ולכן ב $e^{0.7} = 2$ השגיאה כ 0.7%.

כדי להעריך חסם לשגיאות בערכים שקבלנו אחרי חישובים, אפשר להעזר בנוסחת הדיפרנציאלים עבור מכפלה ועבור מנה (בדוק!):

$$\left|\frac{d(xy)}{xy}\right| \le \left|\frac{dx}{x}\right| + \left|\frac{dy}{y}\right|$$
$$\left|\frac{d(x/y)}{x/y}\right| \le \left|\frac{dx}{x}\right| + \left|\frac{dy}{y}\right|$$

מהי ההסתברות שקבוע פיסיקלי יתחיל בספרה 1?

לפני זמן רב נתקלתי במאמר, נדמה לי בעיתון מקצועי של מהנדסי חשמל, ששאל וענה על השאלה שבכותרת. לצערי איני זוכר היכן הופיע המאמר (ולכן איני יכול להביא דבר בשם אומרו) אבל נראה לי שאפשר לראות מה שיובא להלן כחלק מה"פולקלור המתמטי".

אפשר לבצע מחקר ניסויי: לקחת אוסף גדול של קבועים ממדעי הטבע ולראות איזה אחוז מהם מתחיל בכל אחת מ 9 הספרות. ניסוי כזה מראה שאין שוויון בין הספרות: הכי הרבה מתחילים ב 1, אחריו 2 וכן הלאה והכי פחות ב 9.

ממבט ראשון זו נראית שאלה שאין לנו מספיק נתונים לחקור אותה באופן תיאורטי, אבל מתברר שהנחה אחת, שמתקיימת במציאות בקירוב רב, מספיקה כדי לתת בדיוק את ההסתברות של כל סיפרה.

הנחה זו היא: אם נשנה את יחידת המידה (ס"מ, ק"ג וכו') בה צוינו הקבועים, לא תשתנה ההתפלגות של הספרות של הקבועים, כאשר מתעלמים מהנקודה העשרונית.

ההצדקה להנחה זו היא שיחידת המידה היא שרירותית לחלוטין, בעוד שהתפלגות ספרות הקבועים היא "חלק מהטבע", ולכן אחת לא צריכה להיות תלויה בשניה.

נתרגם את ההנחה לשפה יותר מתמטית: שינוי יחידת המידה פירושו כפל כל הקבועים במספר מסויים a, כלומר תוספת $\log_{10} a$ ללוגריתם שלהם. נזכור שאנו מתמקדים על הספרות ללא הנקודה, מה שמתאים למנטיסה של הלוגריתם, כלומר ללוגריתם מודולו 1. לכן מבחינתנו מה שחשוב הוא שאנו מוסיפים למנטיסה של הלוגריתמים של הקבועים את $\log_{10} a$ מודולו 1.

אבל אם כך ההסתברות שהמנטיסה היא בין 0.00 ו 0.01 צריכה להיות שווה להסתברות שהיא באיזשהו מאון אחר $[t, t+0.01)$ כי מאון זה מתקבל מהמאון $[0.00, 0.01)$ ע"י תוספת t מודולו 1. לכן ההסתברות שהמנטיסה היא במאון כלשהו חייבת להיות 0.01.

ומכאן לא ארוכה הדרך לקבל שההסתברות שהמנטיסה היא באינטרוול מסויים מודולו 1 היא בדיוק אורך האינטרוול.

ומכאן שההסתברות שהסיפרה הראשונה היא k $k = 1, 2, 3, 4, 5, 6, 7, 8, 9$ היא

$$\log_{10}(k+1) - \log_{10}(k)$$

והתשובה לשאלה שבכותרת: ההסתברות שהסיפרה הראשונה היא 1 שווה ל

$$\log_{10} 2 = 0.3010\ldots$$

כדאי לבדוק אם כל זה מתאים לניסוי.

ולסיום שאלה: לכאורה אותה אחידות בפילוג צריכה להתקיים גם אם כן נתחשב בנקודה העשרונית, האמנם?

מתמטיקה אצל בעלי התוספות

התלמוד נזקק לפעמים, לצרכי דיוניו ההלכתיים, לנושאים מתמטיים. "בעלי התוספות", מצרפת ואשכנז (גרמניה), שפירשו את התלמוד בימי הביניים, המשיכו את הדיון. נביא כאן את דיונם של בעלי התוספות בנושא גיאומטרי, כמו שהוא מופיע במסכת סוכה דף ח' עמ' א'. לשון התוספות מדברת בעד עצמה, רק סימני הפיסוק וההערות ב [ ] הם תוספת שלנו. הציורים הם בעקבות אלה שבהוצאות התלמוד הרגילות (דפוס וילנא).

בטקסט של התלמוד הבבלי, שדן בצורות האפשריות של סוכה, מופיעות שתי טענות גיאומטריות:

א. כמה מרובע יותר על העיגול רביע
(כלומר: עיגול החסום בריבוע - שיטחו $\frac{3}{4}$ משטח הריבוע, מה שמתאים ל $\pi = 3$)

ב. כל אמתא בריבועא אמתא ותרי חומשא באלכסונא
(כלומר: ריבוע שאורך צלעו 1 אמה - האלכסון שלו בעל אורך $1\frac{2}{5}$ אמות, כלומר $\sqrt{2} = 1\frac{2}{5}$).[3]

אנו יודעים ששתי טענות אלה אינן מדוייקות, למרות שהן נכונות בקירוב.

א. בעלי התוספות מקבלים את הטענה שהקף של מעגל גדול פי שלושה מקוטרו, טענה שאפשר להסיק גם מהפסוק בתנ"ך (מלכים א' ז' כ"ג וכמעט אותו פסוק בדברי־הימים ב' ד' ב'):

ויעש את הים מוצק עשר באמה משפתו עד שפתו עגל סביב וחמש באמה קומתו וקו שלשים באמה יסב אתו סביב

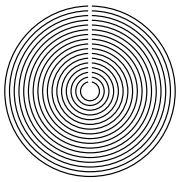
אבל התלמוד מתייחס לשטח ולא להיקף. התוספות אומרים על כך:

כמה מרובע יתר על העיגול רביע. אין להוכיח דבר זה מהא דטבלא מרובעת של שלש על שלש וחוט של שתים עשרה יסוב אותה, וטבלא עגולה של שלש חוט של תשע אמות יסוב אותה (דכל שיש בהקיפו שלשה טפחים יש בו רחב טפח כדאמר בשמעתין) דאין מביאין ראיה מחוט ההיקף הגדול רביע אצל רוחב המקום. דאטו טבלא עגולה של ד' על ד' אמות [ד' מסמן את המספר 4] - סלקא דעתך שאינה מחזקת אלא כטבלא של ג' על ג' מרובעים [ג'=3], לפי שהחוט המקיפו מדתו שוה? והלא כשתחלוק טבלא של ג' על ג' מרובע על ג' רצועות לאורך ושלש רצועות לרוחב - לא תמצא בה כי אם ט' [=9] של אמה על אמה; וטבלא עגולה של ד' על ד' על־כרחך יש בה י"ב [=12] רצועות של אמה על אמה, שהרי אם ריבוע של ד' על ד', כשנחלק ל ד' רצועות של רחב אמה לארכו וכן לרחבו תמצא בו י"ו [=16] רצועות של אמה על אמה, ומרובע אינו יתר על העיגול אלא רביע - נמצאת אתה אומר שהעגולה היא י"ב אמה על אמה [וקיבלנו סתירה מההנחה שיחסי ההיקף שווים בהכרח ליחסי השטח].

אלא ודאי אין ראיה מחוט של היקף כלל! ועוד תדע, דרצועה של ה' [=5] אמות אורך על רוחב אמה - חוט של י"ב [=12] אמה מקיפה; כשתצא לחלקה לרצועות של אמה על אמה - אין בה אלא חמש אמות. [והשווה עם ריבוע של 3 על 3 שיש לו אותו היקף]

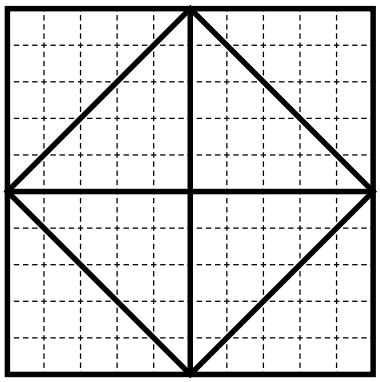
והיינו טעמא, לפי כשאתה מניח חוט בריבוע - הולך ומיצר לזויות; וכשאתה מניחו בעוגל - מרחיב והולך. ואם באנו לכוין החשבון דמרובע יתר על העגול, נוכל להוכיח בענין זה: שתעשה נקודה של משהו, ותקיפנה בחוטין הרבה סביב זה, סיבוב אחר סיבוב, עד שירחיבו ויגדל רוחב בעוגל טפח על טפח. ואחר כך תחתך החוטין מן הנקודה ולמטה, דהיינו מחצי רוחב עיגול ולמטה, ואחר שיחתכו יתפשטו כל החוטין מימין ומשמאל - ונמצא כל חוט הולך ומאריך מחבירו משהו מכאן ומשהו מכאן, עד שאתה מגיע לחוט העליון דארכו ג' [=3] טפחים שהוא חוט החיצון, שהוא מסבב טפח על טפח (דכל שיש ברוחבו טפח יש בהיקפו שלשה טפחים). נמצאו החוטין הללו סדורין לענין זה כמין רצועה רחבה באמצע חצי טפח, והיינו: כנגד הנקודה מכאן ומכאן כלה והולכת וצרה עד משהו. ואם באת לחזור ולחלק אותה באמצע, היינו: כנגד הנקודה, תמצא שתי רצועות שכל אחת ארכה טפח ומחצה, ומצד אחת רחבה חצי טפח, ומצד אחת כלה עד משהו. ועתה תצטרף אלו שתי הרצועות ושים הארוך כנגד הקצר - תמצא רצועה ארכה טפח ומחצה על רוחב חצי טפח. תחלוק אותה לשלש רצועות - תמצא בה שלש רצועות מחצי טפח על חצי טפח; ואילו רצועה מרובעת של טפח, כשתחלקנה שתי וערב, תמצא בה ארבע רצועות של חצי טפח על חצי טפח. הרי לך: מרובע יתר על העיגול רביע.





[3] "אמה" ו"טפח" הן יחידות אורך עתיקות.

ב. כל אמתא בריבוע אמתא ותרי חומשי באלכסונה. אין החשבון מכוון, ולא דק, דאיכא טפי פורתא! [= יש תוספת קטנה באלכסון]. שאם תעשה ריבוע של עשר על עשר ותחלק אותו שתי וערב - נמצא בתוכו ארבעה ריבועים של חמשה על חמשה. חזור וחלק אותם ריבועים לאלכסונים ההולך לצד אמצע של ריבוע גדול - תמצא בריבוע הפנימי חמשים אמה, שהרי הוא חציו של חיצון, שהרי חלקת הריבועים של ה' [=5] על ה', כל אחד לאלכסונו. ואם לא היה בו אלא לפי חשבון "אמתא ותרי חומשי", דהיינו ז' [=7] על ז', נמצא שאין בו חציו של חיצון, דריבוע של שבעה על שבעה אין בו אלא ארבעים ותשע רצועות של אמה על אמה, וראוי להיות חמשים, דהא הוא חציו של עשרה על עשרה, דעולה למאה רצועות של אמה על אמה.



האם יש דרך לבחור ראש־ממשלה בלי שיהיה כדאי "לרמות"?

[4] בשיטה שנהוגה בארצות רבות, וכעת גם בישראל,[5] לבחור ראש־ממשלה (או נשיא), בוחרים מי שקיבל הכי הרבה קולות (או, אם אף אחד לא קיבל יותר מאחוז מסויים (למשל 40%), עורכים סיבוב שני בין שני המועמדים המובילים). בשיטה כזו כדאי לבוחר לעתים קרובות לא להצביע בעד המועמד שהוא מעדיף, אלא, למשל, בעד מועמד אחר שלדעת הבוחר יש לו יותר סיכויים, כדי לחסום מועמד שלדעתו הוא עוד יותר גרוע.

בעיה כזו מתעוררת רק כאשר יש יותר משני מועמדים (או, מה ששקול מבחינה מתמטית, חבר בוחרים צריך לבחור בין יותר משתי אפשרויות). אם יש רק שתי אפשרויות, ומחליטים לפי הרוב, ברור שלא כדאי לאף אחד לא להצביע בעד האפשרות שהוא מעדיף. אותו דבר נכון כאשר מחליטים לפי רוב משוקלל (יש כאלה ששווים יותר) ־ בדוק!

במקרה שיש יותר משני מועמדים (או אפשרויות), יש שיטות המתחשבות בכל מערכת ההעדפות של הבוחרים (לא רק מי העדיף ביותר), למשל: כל בוחר נותן נקודות למועמדים השונים ובוחרים מי שקיבל מספר מקסימלי של נקודות.

בנושאים אלה החלו לטפל מבחינה מתמטית כבר במאה ה־ 18. ערב המהפכה הצרפתית הופיעו ספריהם של קונדורסה (Condorcet), **1785** [1] ובורדה (Borda), **1781** [2]. בורדה הוא שהציע את שיטת הנקודות. קונדורסה ניסה להציע הכרעה בין כל שני מועמדים ע"י הכרעת רוב לפי ההעדפה ביניהם, אבל עד מהרה גילה את "הפרדוקס של קונדורסה" או "הציקל של קונדורסה":
נניח שיש שלושה מועמדים a, b, c ושלושה בוחרים 1, 2, 3, שסדרי ההעדפות שלהם מתקבלים זה מזה ע"י תמורה ציקלית: 1: $a > b > c$, 2: $c > a > b$, 3: $b > c > a$. אז הרוב מעדיף a על b, הרוב מעדיף b על c והרוב מעדיף c על a ־ מקבלים סדר ציקלי ואין הכרעה לפי קונדורסה.

התברר שהשיטות שהוצעו סובלות מפרדוקסים, כגון (ראה [4]):

א. יש אפשרות שכדאי היה לבוחר לא להגיע לבחירות, אי־הצבעה היתה מביאה לתוצאה טובה יותר מבחינתו מאשר אם הוא מצביע לפי העדפתו.

ב. מועמד שמנצח לפי קונדורסה, כלומר שמנצח כל מועמד אחר לפי ההעדפה ביניהם, עלול להפסיד.

ג. אם מחלקים את הבוחרים למספר קבוצות (למשל אנשי הצפון ואנשי הדרום), אז יש אפשרות שמועמד מנצח לפי ההצבעה בכל קבוצה לחוד אבל מפסיד אצל כל הבוחרים.

ליתר דיוק, הקוראים מוזמנים לבדוק ש:
שיטת בורדה (שיטת הנקודות) לוקה ב־ ב׳ אך לא ב־ א׳ ולא ב־ ג׳
השיטה החדשה בישראל לוקה ב א׳, ב׳ ו־ג׳ (כל הכבוד לשיטה!)

אבל נחזור לבעיה שהוזכרה בהתחלה: בכל השיטות שהוצעו כדאי לפעמים לבוחר לא להצביע לפי העדפתו (נקרא לזה בלשון בוטה: "לרמות"). נביא שתי דוגמאות

דוגמא 1: ([3])
יש 6 מועמדים a, b, c, d, e, f. שני בוחרים (או, אם תרצו, שתי קבוצות בוחרים אחידות בעלות אותו גודל). העדפותיהם הן:
בוחר 1: $a > b > c > d > e > f$
בוחר 2: $c > e > b > f > d > a$
הבחירה נעשית לפי שיטת הנקודות (שיטת בורדה): המועדף ביותר מקבל 5 נקודות, זה שאחריו 4 נקודות וכן הלאה עד האחרון שמקבל 0.
אם הבוחרים יצביעו לפי דעותיהם האמיתיות, ייבחר c עם 8 נקודות (בדוק!). אבל 1 יכול "לרמות" ולדווח על העדפות:
$a > b > d > e > f > c$
ואז b ייבחר.
אם 2 ינחש ש 1 יעשה כך, הוא ידווח:
$e > c > f > d > b > a$
ויגרום לבחירתו של e. אם 1 ישתכנע שכך יעשה 2, הוא יבחר: $b > a > d > f > c > e$ ו b ייבחר.
וכן הלאה... עבור כל בחירה שיחליט אחד הבוחרים, אם הבוחר השני ינחש זאת, הוא יוכל לכפות את בחירת אחד משני העדיפים לו: a או b עבור 1, c או e עבור 2.

---

[4] אני מודה לפרופ' רון הולצמן, מהפקולטה למתמטיקה בטכניון, על הדרכתו ועזרתו.
[5] מאז הופיע המאמר לראשונה בוטלה בישראל השיטה ה"חדשה" בה נבחר ראש הממשלה ישירות ע"י הבוחרים.

מי שמכיר דוגמאות מתורת המשחקים, רואה שיש לנו כאן מה שנקרא משחק ללא נקודת שווי משקל. בכל אופן, הבוחר נאלץ כאן לנסות להיות שחקן ממולח.

דוגמא 2: [3],[5]
נניח שיש לנו ציקל קונדורסה שתואר למעלה: שלושה בוחרים 1, 2, 3 עם העדפות:
1: $a > b > c$
2: $c > a > b$
3: $b > c > a$.
כל בוחר מכניס לקלפי פתק עם שם אחד המועמדים. אם בשלושת הפתקים יש רק שני מועמדים, הבחירה לפי הרוב. אבל ל 1 יש זכות יתר: אם בשלושת הפתקים מופיעים כל 3 המועמדים, נבחר מי ש 1 בחר בו.

אנו מניחים שכולם יודעים מהן ההעדפות של כולם, וכולם יודעים שכולם יודעים זאת, וכולם יודעים שכולם יודעים שכולם יודעים וכן הלאה...

כל בוחר מנסה להיות רציונלי. עם זאת, מניחים שהבוחרים לא יוצרים קואליציות (קואליציה אפשרית היא, למשל, הסכם בין 2 ו 3 על הצבעה עבור c. מניחים שדבר כזה לא קורה).

נתחיל מ 1. ברור ש a היא הבחירה הטובה בשבילו: אם 2 ו 3 יסכימו, הם יכריעו, בלי תלות מה 1 בחר. אם 2 ו 3 לא מסכימים, ייבחר מי שבעדו הצביע 1. כיון ש 1 מעדיף את a, ברור שעליו לבחור a.

שיקול זה אינו תקף לגבי 2: ברור שלא כדאי לו לבחור b (שהוא הכי גרוע מבחינתו), אבל בגלל העדיפות שיש ל 1, לפעמים (למשל אם 1: b 3: c) כדאי לו להצביע c, בעוד שבמקרים אחרים (למשל 1: b 3: a) כדאי לו לבחור ב a. מצב דומה יש ל 3: לפעמים מוטב לו b ולפעמים c.

עם זאת, 3 יודע מה חושבים האחרים. הוא יודע שסביר ש 1 יצביע a ו 2 יבחר ב c או a. מכאן ש b לא ייבחר, ואין כל טעם לבחור b. "הבחירה האמתית" היא בין a ו c ול 3 כדאי לבחור c.

2 מודע לשיקוליהם של 1 ו 3, ולכן משער ש 1 יבחר a ו 3 c. לכן 2 מצביע c ו c נבחר - המועמד הגרוע ביותר מבחינתו של 1 !

יתרונו של 1 היה בעוכריו, כתוצאה מתכונות "המשחק" של ההצבעה.

אבל האם אפשר למנוע את הצורך ב"רמאות"?

מתברר שמתקיים המשפט המפתיע הבא:

משפט Gibbard-Satterthwaite (1973-1975)
נניח שצריך לבחור בין לפחות 3 אפשרויות (או מועמדים) - נסמן את קבוצת האפשרויות ב A. יש קבוצת בוחרים $\{1, 2, \ldots, N\}$ ולכל בוחר i יש "פונקצית תועלת" u_i (utility function) מ A למשיים. פונקציה זו אומרת כמה שווה כל מועמד עבור הבוחר i. נדרוש ש u_i מקבלת ערכים שונים עבור מועמדים שונים. יש לנו שיטת בחירה המתאימה לכל מערכת של פונקציות תועלת $u_1, \ldots, u_N$ איבר $S(u_1, \ldots, u_N) \in A$ - זהו המועמד שייבחר אם אלה יהיו פונקציות התועלת של הבוחרים.

אנו רוצים שייתקימו שתי הנחות:

לא כדאי "לרמות": $(*)$ אם הבוחר i ידווח על פונקציית תועלת v במקום u_i, ושאר הבוחרים לא ישנו את דיווחם, i לא ירוויח מכך. בנוסחה זה ייכתב כך:

$$u_i\big(S(u_1, \ldots, u_{i-1}, v, u_{i+1}, \ldots, u_N)\big) \le u_i\big(S(u_1, \ldots, u_{i-1}, u_i, u_{i+1}, \ldots, u_N)\big) \qquad (*)$$

לכל מועמד יש אפשרות תיאורטית להיבחר: במלים אחרות, הפונקציה S היא **על**.

הנחה שלישית מובלעת היא: **אוניברסליות:** הפונקציה S מוגדרת עבור **כל** מערכת פונקציות תועלת. אין מערכת שהיא "בלתי אפשרית".

משפט Gibbard-Satterthwaite טוען: הדרך היחידה לקיים תנאים אלה היא דרך הדיקטטורה: קיים בוחר i^* כך שעבור כל מערכת העדפות ייבחר מי ש i^* מעדיף ביותר:

$$S(u_1, \ldots, u_N) = a \text{ אם } u_{i^*} \text{ מקבלת ערך מקסימלי ב } a$$

כלומר: בכל שיטה אפשרית, שהיא אוניברסלית, אף מועמד לא נפסל מראש ושאינה דיקטטורה, יש מצבים בהם כדאי "לרמות".

יש משפטים דומים, בסיטואציות אחרות, שגם הם מראים שאם דורשים מספיק דרישות "טבעיות" מוכרחים לקבל דיקטטורה. החלוץ בכך היה K. Arrow בשנות ה 50-60, שהוא חתן פרס נובל לכלכלה.

הוכחת המשפט

אני מציע לקוראים לנסות, כתרגיל, להוכיח בעצמם את משפט G-S. למי שלא רוצה, אביא קוים להוכחה. ההוכחה תיעשה בשלבים:

נניח שנתונה שיטת בחירה S שמקיימת את התנאים שלנו. אנו רוצים להוכיח שקיים דיקטטור.

שלב א' נייחד את תשומת ליבנו לבוחר האחרון N. עבור כל מערכת קבועה של פונקציות תועלת של שאר הבוחרים $u_1,\ldots,u_{N-1}$ נסמן ב $T(u_1,\ldots,u_{N-1})$ את קבוצת כל המועמדים שיכולים להיבחר עבור כל האפשרויות של u_N. שימו לב שעבור $u_1,\ldots,u_{N-1}$ קבועות, אזי אם a נבחר, ו $x\in T(u_1,\ldots,u_{N-1})$ אזי אפשר לשנות את u_N לאיזשהו v כך ש x ייבחר, לכן $u_N(a)\geq u_N(x)$. לכן התנאי $(*)$ גורר:
עבור $u_1,\ldots,u_{N-1}$ קבועות, ועבור u_N כלשהי, המועמד שנבחר $S(u_1,\ldots,u_{N-1},u_N)$ הוא זה שבו u_N מקסימלית על הקבוצה $T(u_1,\ldots,u_{N-1})$.
לכן הפונקציה T קובעת את הפונקציה S.
שימו לב שערכה של T הוא תמיד קבוצה **לא ריקה**

שלב ב' ממסקנת שלב א' נובע, שהמועמד שנבחר תלוי רק בסדר ההעדפות של הבוחר N ולא בערכים הממשיים u_N עצמם. לשון אחרת: אם נשנה את u_N כך שסדר ההעדפות לא ישתנה (ישתנה רק בכמה N מעדיף מועמד אחד על אחר) לא ישתנה המועמד שנבחר. ברור שאותו דבר נכון עבור כל בוחר אחר i.
מסקנה: התנאי $(*)$ גורר, שתוצאת הבחירה לא תשתנה אם נשנה את הערכים הממשיים של "פונקציות התועלת" u_i, תוך שמירת סדרי ההעדפות.

שלב ג'
נתבונן בבוחר $i\leq N-1$, ונניח ש u_i מוחלף ב v שמתקבל בצורה הבאה: קובעים שני מועמדים $a,b\in A$ כך ש $u_i(a)>u_i(b)$ ואין אף מועמד אחר ביניהם, ו v מתקבל ע"י החלפת ערכי u_i ב a ו b: $v(a)=u_i(b)$, $v(b)=u_i(a)$, $v(x)=u_i(x)$ $(x\neq a,b)$.
נבדוק כיצד משתנה ערכה של T. לשם קיצור נסמן את ערכה של T עבור u_i ב T_0 ואת ערכה כאשר u_i מוחלף ב v ב T_1.
התנאי $(*)$ עבור הבוחר i גורר שעבור כל סדר העדפות של N, על המועמד העדיף על N ב T_0 (שהוא זה שייבחר, לפי שלב א') להיות לא גרוע (לגבי u_i) מאשר המועמד העדיף על N ב T_1. כיון שסדר ההעדפות של N הוא, מבחינתנו, שרירותי לחלוטין, הקורא יכול להשתכנע שאם $x\in T_0\cap T_1$, $y\in T_0\setminus T_1$, $z\in T_1\setminus T_0$ אזי חייב להיות $y>x>z$ לגבי u_i, אחרת אפשר לבנות סדר העדפות של N שיקלקל. אם נתבונן בבחירה כאשר פונקצית התועלת של i היא v, נקבל באופן דומה שעבור v $z>x>y$.
קבלנו, איפוא: במעבר מ u_i ל v הסדר של x,y,z מתהפך.
אבל שימו לב שהמועמדים היחידים שסדרם מתהפך הם a ו b.
ומכאן נקל לבדוק שיש רק ארבע אפשרויות:

$$T_1=T_0$$
$$T_0=\{a\},\ T_1=\{b\}$$
$$T_0=\{a,b\},\ T_1=\{b\}$$
$$T_0=\{a\},\ T_1=\{a,b\}$$

(שימו לב שערכה של T הוא תמיד קבוצה לא־ריקה, לכן T_0 ו T_1 אינן ריקות).

במיוחד קבלנו:

מסקנה: החלפת סדר העדיפויות של הבוחר i לגבי שני מועמדים סמוכים a ו b יכולה לשנות את ערכה של T רק אם ערכה $\{a,b\}$, $\{a\}$ או $\{b\}$. במיוחד, אי אפשר לשנות כך את T אם ערכה הוא קבוצה בת יותר משני איברים.

שלב ד'
כידוע, אפשר לעבור מכל סידור של K עצמים לכל סידור אחר ע"י ביצוע סדרת החלפות של זוגות איברים סמוכים. יתר על כן, אם הסדר היחסי של a ו b שווה בשני הסידורים, אפשר להניח שבאף אחת מההחלפות בסדרה לא הוחלפו a ו b. (הוכח טענות אלה!).

מטענות אלה ומהמסקנה שבסוף שלב ג', נובע:
נתבונן בבוחר i, $(i\leq N-1)$. נניח שנתונות פונקציות תועלת $u_1,\ldots,u_{N-1}$ עבורן לפונקציה T יש ערך T_0.

- אם T_0 בת יותר משני איברים, אזי כל שינוי ב u_i לא ישנה את ערכה של T. כיון שזה נכון לכל i, יוצא שאם T מקבלת עבור איזשהן פונקציות תועלת ערך שהוא קבוצה בת יותר משני אברים, אזי T פונקציה קבועה, וערכה הקבוע חייב להיות קבוצת כל המועמדים A ־ אחרת S לא תהיה על ־ ומכאן שהמועמד הנבחר הוא העדיף על N ־ N הוא דיקטטור.

- אם T מקבלת תמיד ערך שהוא קבוצה בת איבר אחד בלבד, אזי לדעתו של N אין כל חשיבות ־ ערכה של S הוא תמיד האיבר היחיד שבקבוצה שהיא ערכה של T. אפשר להשמיט את N בכלל. נוכל להניח באינדוקציה שטענת המשפט נכונה עבור $N-1$ בוחרים (מדוע הטענה נכונה עבור בוחר אחד?) ולכן יש דיקטטור בין הבוחרים $1,2,\ldots,N-1$.

- יש לטפל עוד במקרה ש T מקבלת ערך שהוא קבוצה בת שני איברים a ו b. אני רוצה להוכיח שאז הערך ש T מקבלת הוא תמיד תת־קבוצה של $\{a,b\}$. אז המועמד שייבחר יהיה תמיד a או b בניגוד לכך ש S היא על. כאן משתמשים בכך שיש לפחות שלושה מועמדים.

לפי המסקנה בסוף שלב ג' והאמור בתחילת שלב ד', אם T_0 בת שני איברים a ו b, אזי כל שינוי ב u_i שאינו משנה את הסדר היחסי ביניהם לא ישנה את T.

במיוחד אפשר לשנות את u_i, בלי לשנות את הסדר היחסי של a ו b, כך שהם יהיו שני המועמדים הגרועים עבור i. אותו דבר אפשר לעשות עבור כל הבוחרים $1,\ldots,N-1$, כך שעבור כולם a ו b יהיו הגרועים ביותר. ערכה של T יישאר $\{a,b\}$.

אבל, עבור בוחר i, a ו b הם הגרועים ביותר ו $T\subset\{a,b\}$, אזי כל שינוי של u_i לא יוכל להכניס ל T אף איבר $c\neq a,b$, אחרת ל N יוכל להיות סדר עדיפויות בו $c>a,b$ ואז יופר התנאי $(*)$ (ראה שלב א').

ע"י ביצוע שינויים כאלה עבור הבוחרים בזה אחר זה נוכל להגיע לכל מערכת פונקציות תועלת אפשרית, ויישמר המצב ש $T\subset\{a,b\}$, והוכחתי מה שרציתי.

מ.ש.ל.

מה לעשות?

תוצאת אי־אפשרות כמו משפט G-S נראית מייאשת, אבל אין היא סוף פסוק. אפשר לנסות לערער על סבירות הדרישות שלנו.
למשל:
האם באמת כל מערכת פונקציות תועלת היא אפשרית? בבחירות יש בדרך כלל (לפחות בקירוב) סדר של המועמדים לפי ימין־שמאל. אפשר להניח שנתון סידור של A (ימין־שמאל), ולהרשות רק פונקציות תועלת שלגבי סידור זה הן עולות, מגיעות למקסימום ויורדות (כלומר מניחים שכל בוחר מעדיף נקודה מסוימת בציר ימין־שמאל ומועמד הוא גרוע יותר לגביו ככל שהוא סוטה מנקודה זו). כלומר: מוותרים על האוניברסליות. אז אפשר להוכיח קיום מנצח במובן קונדורסה, במובן החלש (כלומר שמנצח ברוב מול כל מועמד אחר, עם אפשרות לתיקו), ואם מחליטים לבחור בו מקבלים שיטה בה לא כדאי "לרמות".

ביבליוגרפיה

[1] A. N. , Marquis de Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*, Paris.
[2] J. C. de Borda, 1781. *Memoirs sur les élections au scrutin*, Histoire de l'Académie Royale des Sciences, Paris.
[3] Hervé Moulin, *Fairness and strategy in voting*, in: Fair Allocation, Proceedings of Symposia in Applied Mathematics, Vol. **33**, American Mathematical Society, 1985.
[4] P. C. Fishburn, S. J. Brams, *Paradoxes of preferential voting*, Mathematics Magazine, Vol. **56**, No. **4**, September 1983.
[5] R. Farquharson, *Theory of Voting*, Yale University Press, New Haven, 1969.

משפט המספרים הראשוניים וחיקויים שלו

מי שמנסה לרשום את המספרים הראשוניים לפי הסדר

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, \ldots \tag{1}$$

(ובעזרת מחשבים עשו זאת עד מיליארד ויותר) רואה שתי תופעות:
א. סידרת הראשוניים נראית כמו משהו אקראי, ללא כל סדר ברור, בדומה לסדרה שבה המספר הבא נבחר כל פעם ע"י הגרלה.
ב. ככל שעוברים למספרים גדולים יותר, הסדרה נעשית בממוצע יותר דלילה: יש בה יותר רווחים גדולים (אם כי מממשיכים להופיע מדי פעם אפילו ראשוניים "תאומים" שהפרשם 2).

משפט המספרים הראשוניים

מתברר שאפשר למצוא חוקיות בהתנהגות זו: משפט, שזכה לכינוי הכבוד **משפט המספרים הראשוניים** אומר, שאם נסמן ב $\pi(N)$ את מספר המספרים הראשוניים הקטנים או שווים ל N, אזי, בהתאם ל־ ב„ היחס $\pi(N)/N$ שואף ל 0 כאשר $N \to \infty$, ובשאיפה זו יש חוקיות: מתקיים:

$$\frac{\pi(N)/N}{1/\ln N} \to 1 \tag{2}$$

($\ln$ מסמן לוגריתם טבעי, כלומר בבסיס e).
לשון אחרת: עבור N גדול, בערך $1/\ln N$ מבין המספרים מ 1 עד N הם ראשוניים, והדיוק ב"בערך" נעשה טוב יותר ויותר ככל ש N גדל.

רבים מהקוראים מכירים בוודאי את הנפה של ארטוסתנס, שהיא דרך לקבל את הסידרה (1): מתחילים מהמספר 2 ומוחקים את כל הכפולות שלו (פרט לעצמו). המספר הראשון שנשאר אחריו הוא 3, לכן מוחקים את כל הכפולות של 3 (פרט לעצמו). מאחר שהראשון שנשאר אחרי 3 הוא 5, מוחקים את כל הכפולות של 5 (פרט לעצמו) וכן הלאה. בדוק שאכן מתקבלת כך סידרת הראשוניים!

הוכחת משפט המספרים הראשוניים היא עמוקה ודורשת אמצעים מתקדמים. קבוצת חוקרים מהמחלקה המתמטית של מעבדות לוס אלאמוס בארה"ב, בראשות S. Ulam שרצו להשתעשע, ניסו לבדוק, מה יקרה אם נשנה את המרשם של נפת ארטוסתנס.

הם הגדירו את ה"מספרים בני המזל" (lucky numbers) בצורה הבאה: נתחיל מהאיזוגיים $1, 3, 5, \ldots$. המספר הראשון אחרי 1 הוא 3. לכן נשמיט את המספרים **במקומות** 9,6,3... בסדרה. 5, 11 וכו' יושמטו ויישארו $1, 3, 7, 9, 13, 15, 19, \ldots$. המספר הראשון אחרי 3 שנשאר הוא 7, ולכן נשמיט בסדרה החדשה את המספרים **במקומות** 21,14,7... כך יושמטו 19, 39, וכן הלאה. כעת יושמטו המספרים **שמקומותיהם** בסדרה החדשה מתחלקים ב 9, וכן הלאה. נקבל את סדרת המספרים בני המזל, שהיא שונה לגמרי מסדרת הראשוניים:

$$1, 3, 7, 9, 13, 15, 21, 25, 31, 33, 37, 43, 49, 51, 63, 67, 69, 73, 75, 79, 87, 93, 99, \ldots$$

אבל להפתעתם של אלה שהעלו את הרעיון, הם הצליחו להוכיח שגם לסדרה זו יש התכונה (2)!

במאמר זה ברצוננו לתאר (בלי הוכחה) שני סוגים של מרשמים, שכולם נותנים סדרות שמקיימות את (2). אפשר להתפתות ולזלזל בעומקו של משפט המספרים הראשוניים, כאילו הוא אינו מציין אלא תכונה של כל סדרה שנבנתה ע"י מרשם מסוג מסויים. אעיר רק, שההוכחה שידועה לי לכך שלמספרים הראשוניים יש התכונה (2) קשה ועמוקה בהרבה מההוכחות לגבי ה"וריאציות", למשל לגבי "המספרים בני המזל" שתוארו לעיל.

צפיפות של קבוצות של מספרים טבעיים

את התכונה (2) אפשר לנסח לגבי כל קבוצה אינסופית A של מספרים טבעיים (קבוצה כזו אפשר, כמובן, לסדר בסידרה עולה כמו את הראשוניים). לקבוצה כזו A או שיש או שאין תכונה זו. אפשר להוכיח שתכונה זו ניתן לתאר בצורה האקוויוולנטית הבאה:

נתאר לנו שחקן שמשחק נגד מכונת משחק. בכל פעם עליו לשלם סכום a ש"ח (a מספר ממשי חיובי) והמכונה מגרילה מספר טבעי. אם המספר שהוגרל אינו ב A השחקן לא מקבל כלום. אם המספר ב A הוא מקבל סכום השווה ל $\ln$ של המספר שהוגרל. השאלה היא: אם השחקן ישחק הרבה פעמים, האם הוא יפסיד או ירוויח? (כאשר מניחים שתוצאות משחקים שונים הן בלתי־תלויות.)

מתברר ש A תהיה בעלת התכונה (2) אם ורק אם מתקיים: השחקן ירוויח אם $a<1$ ויפסיד אם $a>1$, ולכן המשחק "הוגן" רק אם $a=1$.

אם רוצים לתאר את המשחק למי שלא שמע על לוגריתמים, אפשר לשנות ולאמר: אם המספר שהוגרל הוא ב A יקבל השחקן סכום בש"ח השווה למספר הספרות במספר. אז התכונה (2) שקולה לטענה שהמשחק "הוגן" רק אם $a=1/\ln 10$. למשל זהו המצב אם A היא קבוצת הראשוניים.

בוודאי רבים מהקוראים מזדעקים: מה פירוש הדבר "להגריל מספר טבעי"? אנו רוצים שכל מספר יהיה לו סיכוי שווה, ואז, מאחר שעבור כל מספר M יש אינסוף מספרים הגדולים מ M ורק מספר סופי של מספרים ה $\geq M$, יש סיכוי אפסי שייבחר מספר קטן מ M, וזאת עבור כל M!

ואכן, אין שום משמעות ל"הגרלת מספר טבעי" ולמרות זאת אפשר לתת משמעות לחישובים המתייחסים למשחק שלנו, ע"י הטריק המחשבתי הבא: נקח N גדול, ו"נזייף" את המשחק ע"י כך שלא ניתן למכונה לבחור מספר גדול מ N. אז הכל בסדר: כל מספר ב $\{1,\ldots,N\}$ ייבחר בהסתברות $1/N$, ואפשר לחשב את ה a ההוגן, שיהיה כעת תלוי ב N. נשאיר לקוראים כתרגיל להשתכנע בכך שבמקרה זה ה a ההוגן הוא:

$$a_N=\frac{\sum_{1\leq i\leq N, i\in A}\ln i}{N}$$

אם במקרה יש ל a_N גבול כאשר $N\to\infty$ נגדיר, בצדק רב, את ה a ההוגן עבור המשחק ה"אינסופי" בתור גבול זה: $a=\lim_{N\to\infty}a_N$.

אם A היא קבוצת הזוגיים, או קבוצת המספרים המתחלקים ב 1000000, ירוויח השחקן עבור כל a סופי: אז $\lim_{N\to\infty}a_N=\infty$. הסיבה היא של A יש "צפיפות חיובית" - האחוז של איברי A בין N המספרים הראשונים לא שואף ל 0 ואז יש "סיכוי חיובי" להגריל מספרים גדולים שיתנו לשחקן רווח עצום, כך שעבור כל מחיר שעליו לשלם הוא ירוויח אם ישחק הרבה פעמים.

דרך אגב, הצפיפות של קבוצה A של מספרים טבעיים מוגדרת כגבול, אם הוא קיים, של מספר איברי הקבוצה בין $1,2,\ldots,N$ מחולק ב N, כאשר משאיפים $N\to\infty$. לקבוצת המספרים הזוגיים יש צפיפות $\frac{1}{2}$ ואותה צפיפות יש לקבוצת המספרים הזוגיים הגדולים ממיליארד. לקבוצת המספרים המתחלקים ב 1000000 יש צפיפות $\frac{1}{1000000}$. אם עבור קבוצה A יש במשחק ערך "הוגן" a אזי צפיפותה 0 (אבל לא נכון שעבור כל קבוצה בעלת צפיפות 0 יש ערך הוגן).

סדרות בעלות התכונה שבמשפט המספרים הראשוניים

נביא (בלי הוכחה) שני מרשמים לסדרות כאלה, שמתקבלים ע"י שינוי של הנפה של ארטוסתנס ואפשר להוכיח שהן מקיימות את (2).

המרשם הראשון יהיה הסתברותי: מתחילים ב $\{2,3,4,\ldots\}$. בשלב הראשון של הנפה משמיטים חלק מהמספרים אחרי 2 לפי הגרלה: לכל מספר נותנים הסתברות $\frac{1}{2}$ להשמט (וההגרלות עבור מספרים שונים הן בלתי־תלויות). בשלב הבא מתבוננים מהו המספר הראשון שנשאר אחרי 2, נקרא לו x_2, ומשמיטים מספרים מהסדרה אחריו לפי הגרלה כמו קודם, אבל לכל מספר נותנים הסתברות $1/x_2$ להשמט, וכן הלאה (השוו עם נפת ארטוסתנס המקורית).

כאמור, הסדרה שמתקבלת תקיים את (2), אבל מאחר שהכל מקרי, יש להוסיף הסתייגות: הכוונה היא ש (2) מתקיים בהסתברות 1, לאו דווקא "תמיד".[6]

במרשם השני אין כל אקראיות. הוא הכללה של "המספרים בני המזל" שנזכרו בתחלת המאמר: יוצאים מקבוצה אינסופית כלשהי B של מספרים טבעיים, מסודרת בסידרה עולה. נניח שהמספר הראשון הוא x_1. אז בשלב הראשון מוחקים את כל המספרים **במקומות** שהם כפולות של x_1 כלומר את המספר שבמקום ה x_1, שבמקום ה $2x_1$ וכן הלאה. אם המספר הראשון שנשאר אחרי x_1 הוא x_2, מוחקים כעת מהסדרה החדשה את המספרים **במקומות** x_2, $2x_2$, $3x_2$ וכן הלאה. בשלב השלישי מסתכלים במספר x_3 שהוא הראשון שנשאר אחרי x_1 ו x_2 (הוכח שהם אכן לא יימחקו), וכן הלאה.

אפשר להוכיח שאם הקבוצה המקורית B היתה בעלת צפיפות חיובית כלשהי (ראו הגדרת צפיפות של קבוצה למעלה), אזי הקבוצה שתישאר בסוף תקיים את (2). שימו לב שהצפיפות של קבוצה יכולה להיות לא מוגדרת (נסו למצוא דוגמא) ולכן "בעלת צפיפות חיובית" אומר יותר מאשר "לא בעלת צפיפות אפס".

מה שמוזר הוא שההתנהגות הממוצעת של הסדרה הסופית לא תלויה כלל בצפיפות של הקבוצה המקורית B. דבר זה לא כל כך מפתיע, אם חושבים על כך שיש כאן שתי השפעות מנוגדות: אם בשלב ה k הקבוצה דלילה יותר, המספר x_k יהיה גדול יותר ולכן פחות מספרים יימחקו בשלב הבא והקבוצה תידלל פחות. מתברר שאם הקבוצה המקורית בעלת צפיפות חיובית יגיעו שתי השפעות אלה בסוף ל"איזון" שאינו תלוי בצפיפות של הקבוצה המקורית.

[6] ראו במאמר "האם הסתברות 1 פירושה וודאות" בחוברת זו

הוכחות (למתוחכמים, באנגלית)

Definition: We say that a set B of natural numbers has the **prime numbers property** if $\pi_B(N) := \#B \cap \{1, 2, \ldots, N\}$ satisfies:

$$\frac{\pi_B(N)/N}{1/\ln N} \to 1 \tag{1}$$

We wish to present proofs for the following theorems:

Theorem 1. Let $A = A_0$ be a subset of $\{2, 3, \ldots\}$ that has positive density in $\mathbf{N} = \{1, 2, \ldots\}$, i.e. the percentage of members of A in $\{1, 2, \ldots, N\}$ tends to some positive limit as $N \to \infty$. Perform the following process on A: Let a_1 be the smallest element in $A = A_0$. Remove the $m \cdot a_1$'th elements of A_0 (in ascending order) for $m = 1, 2, \ldots$ to obtain a set A_1. Clearly $a_1 \in A_1$. Let a_2 be the first element in A_2 after a_1. Remove the $m \cdot a_2$'th elements of A_1, $m = 1, 2, \ldots$ to obtain A_2. We have $a_1, a_2 \in A_2$. Let a_3 be the next element in A_2, Remove the $m \cdot a_3$'th elements of A_2 to get A_3, and so on. The intersection of all the A_j's is $B = \{a_1, a_2, a_3, \ldots\}$.
Then B has the prime numbers property.

Theorem 2. For A as in Theorem 1, perform a random process: instead of removing the $m \cdot a_j$'th elements at the j'th stage, remove from A_{j-1} each element after a_j independently with probability $1/a_j$ for removal. Then with probability 1 B has the prime numbers property.

We introduce some notations for the cases of both theorems:

$$P_n := \prod_{k=1}^{n} \left(1 + \frac{1}{a_k}\right) \tag{2}$$

$$p_n := \prod_{k=1}^{n} \left(1 - \frac{1}{a_k}\right) \tag{3}$$

$$\pi(n) := \max\{k : a_k \le n\} \tag{4}$$

$$r_n := \prod_{a_k \le n} \left(1 - \frac{1}{a_k}\right) = p_{\pi_n}. \tag{5}$$

Remark. (1) says that $\pi(n) \sim n/\ln n$, and one readily sees that this is equivalent to $a_n \sim n \ln n$, that is

$$\frac{a_n}{n \ln n} \to_{n\to\infty} 1 \tag{6}$$

Let $A_j = \{a_1^j, a_2^j, \ldots\}$ in ascending order. Define S_k^j by $a_k^j = a_{S_k^j}^0$ and S_k by $a_k = a_{S_k}^0$. In particular $S_k^0 = k$. We have

$$a_k^j = a_k, \quad S_k^j = S_k \qquad \text{if } k \le j+1, \tag{7}$$

In the case of Th. 1 we have

$$S_k^j = S^{j-1}_{k+\left\lfloor \frac{k}{a_j} \right\rfloor} \qquad \text{for all } k, j, \tag{8}$$

while for Th. 2

$$\Pr\left(a_{n+1} - a_n = m \,|\, a_1, \ldots, a_n \text{ fixed}\right) = p_n(1-p_n)^{m-1} \qquad m = 1, 2, \ldots \tag{9}$$

Pr denoting probability (in this case conditional probability).

Proof of theorem 1.

Claim. For $k' \le k$ one has

$$kP_{k'} \left(1 - \frac{1}{k}\right)^{k'} \le S_k \le kP_k. \tag{10}$$

Indeed, fixing k, define t_ℓ, $0 \le \ell \le k$ by

$$t_0 := k, \quad t_{\ell+1} := t_\ell + \left\lfloor \frac{t_\ell}{a_{k-\ell}} \right\rfloor, \quad 0 \le \ell \le k-1$$

From this and (8) one has by induction on ℓ:

$$S_k = S_k^k = S_{t_\ell}^{k-\ell}, \qquad 0 \le \ell \le k.$$

In particular $S_k = S_{t_k}^0 = t_k$.

Now one has

$$t_{\ell+1} \le t_\ell \left(1 + \frac{1}{a_{k-\ell}}\right) \tag{11}$$

$$t_{\ell+1} \ge t_\ell \left(1 + \frac{1}{a_{k-\ell}}\right) - 1 \ge t_\ell \left(1 + \frac{1}{a_{k-\ell}}\right)\left(1 - \frac{1}{n}\right) \tag{12}$$

since $t^\ell \ge n$.

Now (10) follows by (11) for $\ell = k-1, k-2, \ldots, 0$ for the upper bound and (12) for $\ell = k-1, k-2, \ldots, k-k'$ for the lower bound.

We shall use (10) in the sequel.

Lemma 1. Suppose $\Lambda > \lambda > 0$ are constants such that for k big enough

$$\lambda k P_k \le a_k \le \Lambda k P_k \tag{13}$$

Then

$$\Lambda^{-1} \le \liminf_{k\to\infty} \frac{P_k}{\ln k} \le \limsup_{k\to\infty} \frac{P_k}{\ln k} \le \lambda^{-1}, \tag{14}$$

which together with (13) implies

$$\frac{\lambda}{\Lambda} \le \liminf_{k\to\infty} \frac{a_k}{k \ln k} \le \limsup_{k\to\infty} \frac{a_k}{k \ln k} \le \frac{\Lambda}{\lambda}. \tag{15}$$

Proof of lemma 1. Let $u_k := P_{2^k}$. We have

$$\ln \frac{u_{k+1}}{u_k} = \sum_{n=2^k+1}^{n=2^{k+1}} \ln\left(1 + \frac{1}{a_n}\right).$$

By (13) and the fact that $a_n \ge n$ and $u_k = P_{2^k}$ is increasing one obtains (where $\ln 2$ comes from $\sum_{n=2^k}^{n=2^{k+1}} (1/n)$)

$$\frac{\ln 2}{\Lambda u_{k+1}} + O(2^{-k}) \le \ln \frac{u_{k+1}}{u_k} \le \frac{\ln 2}{\lambda u_k} + O(2^{-k}) \tag{16}$$

The sequence u_k is increasing, therefore either $u_k \to \infty$ or it has a finite limit. In the latter case $\ln(u_{k+1}/u_k)$ must tend to 0 while (16) would bound it away from 0, a contradiction. Hence $u_k \to \infty$. Then (16) gives $u_k/u_{k+1} \to 1$.

Let $w_k := \dfrac{k \ln 2}{u_k}$. (16) says that there is a $C > 0$ such that

$$-\lambda^{-1}\frac{w_k}{k} + \ln\frac{k+1}{k} - C \cdot 2^{-k} \le \ln \frac{w_{k+1}}{w_k} \le -\Lambda^{-1}\frac{w_{k+1}}{k+1} + \ln\frac{k+1}{k} + C \cdot 2^{-k} \tag{17}$$

Thus

$$-\Lambda^{-1}\frac{w_{k+1}}{k+1} + \ln\frac{k+1}{k} + C \cdot 2^{-k} < 0 \Rightarrow w_{k+1} < w_k$$

$$-\lambda^{-1}\frac{w_k}{k} + \ln\frac{k+1}{k} - C \cdot 2^{-k} > 0 \Rightarrow w_{k+1} > w_k,$$

or in other words:

$$w_{k+1} > \Lambda(k+1)\left(\ln\frac{k+1}{k} + C \cdot 2^{-k}\right) \Rightarrow w_{k+1} < w_k \tag{18}$$

$$w_k < \lambda k \left(\ln\frac{k+1}{k} - C \cdot 2^{-k}\right) \Rightarrow w_{k+1} > w_k, \tag{19}$$

The expressions in (18) and (19) tend to Λ and λ respectively. Hence for any $\varepsilon > 0$ there is a k_0 such that for $k \geq k_0$

$$w_{k+1} > \Lambda + \varepsilon \Rightarrow w_{k+1} < w_k \qquad w_{k+1} < \lambda - \varepsilon \Rightarrow w_{k+1} > w_k,$$

so for $k \geq k_0$

$$(\exists \ell > k) w_\ell > \Lambda + \varepsilon \Rightarrow w_k > w_{k+1} > \Lambda + \varepsilon$$
$$(\exists \ell > k) w_\ell < \lambda - \varepsilon \Rightarrow w_k < w_{k+1} < \lambda - \varepsilon.$$

Therefore either from some k onward $\lambda - \varepsilon \leq w_k \leq \Lambda + \varepsilon$, or from some k onward $w_k > \Lambda + \varepsilon$ and w_k decreases, or from some k onward $w_k < \lambda + \varepsilon$ and w_k increases. This being the case for any $\varepsilon > 0$ one deduces that either

$$\lambda \leq \liminf w_k \leq \limsup w_k \leq \Lambda \tag{20}$$

or w_k decreases from some place onward and tends to a limit $> \Lambda$ or increases from some place onward and tends to a positive limit $< \lambda$.

In any case, if a positive $L = \lim w_k$ exists (17) implies that for any $\varepsilon > 0$, from some k onward:

$$\left(-\frac{L}{\lambda} + 1 - \varepsilon\right)\frac{1}{k} \leq \ln \frac{w_{k+1}}{w_k} \leq \left(-\frac{L}{\Lambda} + 1 + \varepsilon\right)\frac{1}{k},$$

hence $L > \Lambda$ or $L < \lambda$ would make $\sum(\ln w_{k+1} - \ln w_k)$ divergent, contradicting the existence of a positive $\lim w_k$.

We conclude that (20) holds. Recalling that $w_k = \dfrac{k \ln 2}{u_k}$ and $u_k = P_{2^k}$, P_k increasing, one readily obtains (14). This concludes the proof of Lemma 1.

Now we are ready to prove Th. 1. Taking $k' = k$ in (10), noting that $\left(1 - \frac{1}{k}\right)^k \to 1/e$ and that since A has a density a_k/S_k is bounded on both sides by positive bounds, one obtains that (13) holds for some $\Lambda > \lambda > 0$, hence by the lemma one has (14) and (15). Thus $a_k = O(k \ln k)$ which implies that if one takes in (10) $k' = \lfloor \varepsilon k \rfloor$ for small fixed ε, still $P_{k'}/P_k \to 1$, yet $(1 - \frac{1}{k})^{k'}$ tends to $e^{-\varepsilon}$ which is close to 1. Since moreover a_k/S_k tends to the density of A, one finds that in (13) one may take Λ and λ as close as one wishes, which by (15) implies $a_k/(k \ln k) \to 1$.
QED

Proof of theorem 2. We need the[7]

Kolmogorov's inequality Let X_n, $n = 1, 2 \ldots$ be real stochastic variables with expectation 0 and finite variance $\sigma_n^2 = \sigma_{X_n}^2$ (= the expectation of X_n^2). Assume moreover that the sums:

$$S_n := \sum_{k=1}^{n} X_k \tag{21}$$

form a martingale, i.e. that the conditional expectation of X_n for knowledge of the values of $X_1, \ldots, X_{n-1}$ is also 0. Then

(i) For any $\varepsilon > 0$,

$$\Pr\left(\left(\max_{1 \leq k \leq n} |S_k|\right) \geq \varepsilon\right) \leq \frac{\sum_{k=1}^{n} \sigma_k^2}{\varepsilon^2}. \tag{22}$$

(ii) Consequently if the series $\sum \sigma_n^2$ converges, then $\sum X_n = \lim S_n$ converges a.s. (= almost surely, meaning: with probability 1).

Proof of (i). From the assumption that S_n form a martingale it follows that conditioned on any event concerning $X_1, \ldots, X_k$, the expectations $\mathbf{E}(X_j S_k) = 0$, $j > k$ (also $\mathbf{E}(X_j X_k) = 0$, i.e. these are orthogonal), implying for $n > k$ $\mathbf{E}(S_n^2) = \mathbf{E}(S_k^2) + \sum \mathbf{E}(X_j^2) \geq \mathbf{E}(S_k^2)$. Apply this to the events *the first time that* $|S_i| \geq \varepsilon$ *is for* $i = k$ (k fixed, $k = 1, 2, \ldots, n$). Conditioned on these events, $\mathbf{E}(S_k^2) \geq \varepsilon^2$ hence $\mathbf{E}(S_n^2) \geq \varepsilon^2$. Moreover, these events for $k = 1, \ldots, n$ partition the event of (22), so S_n^2 has expectation

[7] In these auxiliary facts I partly follow the classical book *Measure Theory* by P. R. Halmos.

$\geq \varepsilon^2$ relative to this event too hence its probability $\leq \varepsilon^{-2}$ times the integral of S_n^2 over the whole sample space, which is $\sum \sigma_j^2$ (since the X_j are orthogonal by the above).

Proof of (ii). Note that since the events in (22) decrease in n, one can take $n = \infty$ in (22). Now if $\sum X_n$ does not converge, the infimum (= limit) of the diameters of the "tails" $\{S_m, S_m + 1, \ldots\}$ is positive. If $\varepsilon > 0$ is less then it, then for every m

$$\sup_k |S_{m+k} - S_m| \geq \varepsilon$$

which by (22) has probability $\leq \varepsilon^{-2} \sum_k \sigma_{m+k}^2$, which thus bounds the probability that the limit of the diameters $> \varepsilon$. Taking $m \to \infty$ one finds that the last probability is 0 for any $\varepsilon > 0$, which proves (ii).

We need another
Fact. if x_n are real and $\sum (1/n) x_n$ converges then $y_n = (1/n) \sum_{k=1}^n x_k$ tends to 0 with n.

Indeed let $s_n = \sum_{k=1}^n (1/k) x_k$. We know $s_n \to L$ for some L. but then also the averages $v_N = (1/N) \sum_{n=1}^N s_n$ tend to L when $N \to \infty$ (for n bigger then some n_0, all s_n are "near L" and for N much bigger the "other" s_n's for $n \leq n_0$ have negligible contribution to the average). Now x_k appears with coefficient $1/k$ in all s_n's with $n \geq k$, thus appears in v_N with coefficient $(1/k)(1 - (k-1)/N)$. Hence $v_N = -y_N + (1 + 1/N) s_N$. Since $s_N \to L$, $v_N \to L$ we find $y_N \to 0$ as required.

So one has as corollary, that if X_n has sums which form a martingale and $\sum (\sigma^2/n^2) < 0$ then $(1/n) \sum X_n \to 0$ a.s.

Return now to the proof of Th. 2. We have (9), where, of course, now the a_k's, p_k are stochastic variables on an appropriate sample space. p_k is (always) decreasing, hence has a limit between 1 and 0.

Claim. p_k a.s. tends to 0.

Indeed, it suffices to prove that for any $\alpha > 0$, the probability for $(\forall k) p_k \geq \alpha$ is 0. For k fixed, the probability for $p_k \geq \alpha$ **and** $a_{k+1} - a_k > C \ln k$ ($C > 0$ fixed) is by (9) not greater than

$$(1 - \alpha)^{\lfloor C \ln k \rfloor} = e^{\lfloor C \ln k \rfloor \ln(1-\alpha)} \tag{23}$$

Choosing C such that $C(-\ln(1-\alpha)) > 1$, these for $k = 1, 2, \ldots$ form a convergent series, which implies that a.s. for only a finite set of k's both $p_k \geq \alpha$ **and** $a_{k+1} - a_k > C \ln k$. Thus assuming $(\forall k) p_k \geq \alpha$, we have a.s. $a_k = O(k \ln k)$, hence $\sum a_k^{-1}$ diverges and $\prod_{k=1}^\infty (1 - a_k^{-1}) = 0$, i.e. $p_k \to 0$. This proves the claim.

(9) gives the distribution of $a_{k+1} - a_k$, conditioned on knowing $a_1, \ldots, a_n$. We wish to "truncate" then, using a function $\eta_M : \mathbf{R} \to \mathbf{R}$ ($M > 0$ fixed) defined by:

$$\eta_M := \begin{cases} x & x \leq M \\ M & x \geq M \end{cases} . \tag{24}$$

Consider a stochastic variable ξ with distribution (as in (9))

$$\Pr(\xi = m) = p(1-p)^{m-1}, \qquad m = 1, 2, \ldots \tag{25}$$

with $0 < p < 1$ fixed. Such is the distribution of the first success in a sequence of independent trials, each with probability p for success (and that is why (9) has this form). It has expectation $1/(1-p)$, as one readily computes. The "truncated" $\eta_M(\xi)$, $M > 0$ an integer, has expectation $[1 - (1-p)^M]/p$ and takes, of course, values between 0 and M. Thus by (9), for any sequence $M_k > 0$ of integers, the stochastic variables

$$\eta_{M_k}(a_{k+1} - a_k) - \frac{1 - (1 - p_k)^{M_k}}{p_k} \tag{26}$$

will each has expectation 0 relative to knowledge of the predecessors. They are bounded by M_k. By the corollary above, if $\sum M_k^2/k^2 < \infty$ their averages tend a.s. to 0. For the "untruncated" variables one thus gets:

$$\liminf \left[\frac{a_k}{k} - \frac{1}{k} \sum_{\ell=1}^{k-1} \frac{1 - (1 - p_\ell)^{M_\ell}}{p_\ell} \right] \geq 0 \qquad \text{if } \sum \frac{M_k^2}{k^2} < 0 \tag{27}$$

and if $\sum \frac{M_k^2}{k^2} < 0$ and a.s. from some k onward $a_{k+1} - a_k \le M_k$ (hence $\eta_{M_k}(a_{k+1} - a_k) = a_{k+1} - a_k$) then

$$\frac{a_k}{k} - \frac{1}{k}\sum_{\ell=1}^{k-1} \frac{1-(1-p_\ell)^{M_\ell}}{p_\ell} \to 0 \tag{28}$$

In particular, take $M_\ell = M$ constant. Then $\sum M^2/k^2 < \infty$. We know that a.s. $p_\ell \to 0$. Then the summands in (27), hence their averages, tend to M, thus (27) says $\liminf(a_k/k) \ge M$. Since M is arbitrary, we get

$$\frac{a_k}{k} \to_k \infty \quad \text{a.s.} \tag{29}$$

This gives, for fixed M, for any sample point, $a_k \ge Mk$ from some k onward. Plugging that in (3) gives $p_k \ge Ck^{-1/M}$ for some C depending on the sample point, thus for any $\varepsilon > 0$ a.s. $p_k \ge k^{-\varepsilon}$ from some k onward.

Let $\beta > \alpha > 0$. For any $\varepsilon > 0$, For fixed k

$$\Pr\left(p_k \ge k^{-\alpha}\,,\, a_{k+1} - a_k > k^\beta\right) \le \left(1-k^{-\alpha}\right)^{\lfloor k^\beta \rfloor} = O\left(e^{-k^{\beta-\alpha-\varepsilon}}\right)$$

For small enough ε the last expressions form a convergent series in k. Hence a.s. from some k onward $p_k \ge k^{-\alpha} \Rightarrow a_{k+1} - a_k \le k^\beta$. We may conclude: for any $\varepsilon > 0$ a.s. $a_{k+1} - a_k = O(k^\varepsilon)$, hence a.s. $a_k = O(k^{1+\varepsilon})$. also we can write (see (4)): $a_{\pi(k)} \le k \le a_{\pi(k)+1} \le \pi(k) + \pi(k)^\varepsilon$ hence $\ln \pi(k) \sim \ln k$ a.s.

Recall the r_n (5), which are stochastic variables. Knowing the values of $r_2, \ldots, r_n$, i.e. the values of all a_k with $a_k \le n$ – knowing $\pi(n)$ and $a_1, \ldots, a_{\pi(n)}$, one has two possibilities:

1. $\pi(n+1) = \pi(n)$; $a_{\pi(n)+1} > n+1$; $r_{n+1} = r_n$ That has conditional probability $1 - r_n$.
2. $\pi(n+1) = \pi(n)+1$; $a_{\pi(n)+1} = n+1$; $r_{n+1} = r_n\left(1 - \frac{1}{n+1}\right)$ That has conditional probability r_n.

This makes the r_n a *Markov process*.

The r_n are always non-increasing, and we know that $r_n = p_{\pi(n)} \to 0$ a.s.

Consider r_{n+1}/r_n, which has value $1-1/(n+1)$ with (conditional) probability r_n and 1 with (conditional) probability $1 - r_n$. This means that the stochastic variables

$$\delta_{n+1} := 1 - \frac{r_{n+1}}{r_n} - \frac{1}{n+1} r_n \tag{30}$$

make martingale sums, and of course are bounded by $1/(n+1)$. By Kolmogorov's inequality, for integer $m > 0$ and $\frac{1}{2} > \varepsilon > 0$

$$\Pr\left((\exists m')\left|\sum_{n=2^m+1}^{m'} \delta_n\right| > 2^{(-\frac{1}{2}+\varepsilon)m}\right) = O\left(2^{-2\varepsilon m}\right)$$

The last expression forms a convergent series in m, hence for any $\frac{1}{2} > \varepsilon > 0$, a.s. for big enough m for any $m' > 2^m$

$$\left|\sum_{n=2^m+1}^{m'} \delta_n\right| \le 2^{(-\frac{1}{2}+\varepsilon)m}.$$

This means that for any $\varepsilon > 0$, a.s. $\sum \delta_n < \infty$ and

$$\sum_{n=k}^{\infty} \delta_{n+1} = O(k^{-\frac{1}{2}+\varepsilon})$$

(where the constant in O depends on the sample point).

Recalling (30), and the fact $x - 1 = \ln x + O(x^2)$ for $x \to 0$, one may say that a.s. for $k' > k$:

$$\ln \frac{r_{k'}}{r_k} = -\sum_{n=k}^{k'-1} \frac{1}{n+1} r_n + O\left(k^{-\frac{1}{2}+\varepsilon}\right) \tag{31}$$

Now we can imitate the proof of Lemma 1. If one writes

$$u_k = r_{2^k},$$

one has by (31)

$$\ln \frac{u_{k+1}}{u_k} = -\sum_{n=2^k}^{2^{k+1}-1} \frac{1}{n+1} r_n + O\left(2^{(-\frac{1}{2}+\varepsilon)k}\right) \tag{32}$$

and similarly to (16) one has

$$-(\ln 2)u_k + O\left(2^{(-\frac{1}{2}+\varepsilon)k}\right) \le \ln \frac{u_{k+1}}{u_k} \le -(\ln 2)u_{k+1} + O\left(2^{(-\frac{1}{2}+\varepsilon)k}\right), \tag{33}$$

and one proceeds in total analogy to the proof of Lemma 1., to find that a.s. $r_n \sim 1/\ln n$, thus $p_{\pi(n)} = r_n \sim 1/\ln n \sim 1/\ln \pi(n)$ hence finally $p_k \sim 1/\ln k$.

Still we have to deduce the asymptotics of a_k. To this end, take in (27),(28) M_k satisfying (for some fixed $\eta < 1$):

$$\frac{M_k}{\ln k} \to_k \infty, \quad \sum_k \frac{M_k^2}{k^2} < \infty, \quad \sum_k \left(1 - \frac{\eta}{\ln k}\right)^{M_k} < \infty$$

For example, take $M_k = (\ln k)^3$ for $k \ge 3$.

Then by (28) a.s. if from some ℓ onward $a_{\ell+1} - a_\ell \le M_\ell$ then

$$\frac{a_k}{k} - \frac{1}{k}\sum_{\ell=1}^{k-1} \frac{1-(1-p_\ell)^{M_\ell}}{p_\ell} \to 0 \tag{34}$$

But a.s. $p_\ell \sim 1/\ln \ell$ thus $M_\ell/\ln \ell \to \infty$ implies $(1-p_\ell)^{M_\ell} \to 0$ making the summands in (34) $\sim \ln \ell$, so do their averages, hence $a_k \sim k \ln k$, all that if $a_{\ell+1} - a_\ell \le M_\ell$ from some ℓ onward, which we have to prove that it holds a.s. But that follows from the fact that for fixed ℓ the probability of that not holding and moreover $p_\ell \ge \eta/\ln \ell$ (which we know holds a.s. from some ℓ onward – note that $\eta < 1$) does not exceed

$$\left(1 - \frac{\eta}{\ln \ell}\right)^{M_\ell}$$

which we assumed form a convergent series.
QED

מתמטיקאי בלאס־ווגאס

הקדמה

מתמטיקאי נקלע לעיר ההימורים לאס־ווגאס. הוא נכנס לקזינו, ומתיישב ליד מכונת הימורים. לשם פשטות נניח שבכל משחק המכונה מגרילה שני ערכים 0 או 1, בהסתברות שווה $\frac{1}{2}$. כל משתתף מפקיד בכל משחק סכום כסף ומהמר על אחת משתי התוצאות. מי שהימר נכון, מקבל בחזרה מה שהפקיד וזוכה בסכום כסף נוסף השווה לזה שהפקיד. מי שטועה, מפסיד את הסכום שהפקיד.

הקוראים יסכימו בוודאי שמשחק כזה נראה הוגן - לשחקן יש אותו סיכוי להפסיד מה שהפקיד או להרוויח סכום דומה.[8]

אבל, כאמור, המשתתף שלנו הוא מתמטיקאי ומצופה ממנו שיכיר, או ימציא, שיטה להיטיב את סיכוייו, וכך אחרי הרבה משחקים לצאת בעושר.

אם המשתתף יכול לקבוע מהו סכום הכסף שהוא מפקיד הוא יכול להציע את השיטה הבאה: [9] יש להגדיל את הסכום המופקד באופן פרוע, למשל בכל משחק להפקיד פי עשרה מסכום כל ההפקדות שלו במשחקים הקודמים. אם לעולם הוא לא ינחש נכון, הוא יפסיד. אבל ההסתברות שכך יקרה היא 0 (ראו [2] (לקריאה נוספת)). אם נתעלם מאפשרות כזו, הוא פעם ינחש נכון, ואז יקבל בחזרה הרבה יותר ממה שהפסיד! כעת יוכל להפסיק, או להמשיך באותה התנופה.

כמובן ששיטה זו אינה מעשית. היא מניחה שלמשתתף יש כסף (וזמן) בלתי מוגבל. מי שכספו מוגבל ויקשיב לעצה כזו - יש לו הסתברות קרובה ל 1 להפסיד את כל כספו (מי שרוצה, שינסה לחשב הסתברות זו).

מעתה נניח שסכום הכסף המופקד קבוע, אבל בכל משחק אפשר להשתתף או לעמוד מהצד. למשל אפשר לחכות עד שתהיה סדרה ארוכה של תוצאות 0 ואז להמר על 1. (זה דומה קצת למצבו של משקיע קטן בבורסה).

באופן יותר פורמלי, נקרא בשם **איסטרטגית הימורים** לכלל הקובע, בהתאם לתוצאות שהתקבלו עד עכשיו, אם להמר או לא, ועל מה להמר.

האם המתמטיקאי יכול להציע איסטרטגית הימורים שתשפר את סיכוייו?

אכן, המתמטיקאי יכול לאמר משהו, וזה המשפט הבא:

משפט: שום איסטרטגית הימורים לא תשנה כהוא זה את הסתברויות הזכיה של השחקן.

כאן חשובה מאוד העובדה שהנחנו שההחלטה להמר או לא תלויה **רק בעבר**. ברור שמי שניחן בראיית עתידות יכול להחליט להמר על מה שיוגרל ולזכות תמיד.

בהמשך המאמר ברצוני להבהיר ולהוכיח משפט זה.

מושגי יסוד בהסתברות

בדיון הסתברותי דנים במערכת, שקורים בה דברים שונים ולדברים שיכולים לקרות יש הסתברויות. משהו שיכול לקרות נקרא **מאורע** ואפשר לבטא אותו כטענה, למשל: "במשחק הראשון יצא 0" או: "באף משחק לא יצא 0", אבל נוח לראות את המאורעות כתת־קבוצות של קבוצה הנקראת "מרחב המדגם" (sample space). איבריו של מרחב המדגם הן כל האפשרויות לתאור מלא של כל מה שמעניין אותנו על המערכת. למשל כאשר דנים בזריקה אחת של קוביה, מרחב המדגם הוא קבוצת 6 התוצאות האפשריות. עבור ניסוי של 3 זריקות מטבע, מרחב המדגם הוא קבוצת $8 = 2^3$ שלישיות התוצאות האפשריות.

עבור משחק ההימורים שלנו מרחב המדגם הוא אינסופי; איבריו הן כל הסדרות $x_1, x_2, x_3, \ldots$ המתארות את תוצאות ההגרלות (x_i יכול להיות 0 או 1). למשל הסדרה $1, 1, \ldots$ מתארת את האפשרות בה בכל ההגרלות יצא 1. תאור כזה: "בכל ההגרלות יצא 1" הוא תאור מלא של כל מה שמעניין אותנו, ולכן זה איבר במרחב המדגם. דוגמה לאיבר אחר היא הסדרה $1, 0, 1, 0, \ldots$ שמתאימה לתאור: "בהגרלות האיזוגיות יצא 1 ובהגרלות הזוגיות יצא 0". מאורע כמו "במשחק הראשון יצא 0" יפורש כקבוצת כל איברי מרחב המדגם, בהם אכן במשחק הראשון יצא 0, כלומר מאורע זה הוא קבוצת כל הסדרות בהן $x_1 = 0$.

[8] מבחינה זו הסיפור שלנו אינו מתאים למציאות - במתקני הימורים, כמו אלה שבלאס ווגאס, יש תמיד לבית העסק עדיפות בתוחלת, כך שבטווח זמן ארוך הוא ייצא ברווח.

[9] ראו הדיון ב"פרדוקס סנט־פטרסבורג" במאמרו של רון אהרוני שהופיע בכותרת "פרדוקסים הסתברותיים" בגליון 39-40 של "אתגר־גליונות מתמטיקה" נובמבר 1996 (שבו התפרסם לראשונה גם המאמר הנוכחי), ובכותרת "לרמות את אלת המזל", בגליון 15 של "גליליאו" מרץ/אפריל 1996.

בין המאורעות ישנו המאורע הוודאי, שהוא מרחב המדגם כולו, והמאורע הבלתי־אפשרי שהוא הקבוצה הריקה $\emptyset$. במאורעות, שהן תת־קבוצות של מרחב המדגם, מוגדרות הפעולות הבוליאניות: איחוד, חיתוך, משלים, וגם פעולות בוליאניות אינסופיות - איחוד או חיתוך של אינסוף מאורעות. כמו כן מוגדר ביניהן יחס ההכלה (להגיד שמאורע A מכיל מאורע B זה להגיד ש B גורר את A, מדוע?) את המשלים של מאורע A = המאורע האומר "A לא קרה" נסמן ב A'.

למשל יהי A המאורע: בשלש ההגרלות הראשונות יצא 1, שהוא קבוצת כל הסדרות בהן $x_i = 1, i = 1, 2, 3$. המשלים שלו A' הוא המאורע: לא בכל שלש ההגרלות הראשונות יצא 1. A הוא חיתוך שלושת המאורעות: $A_1 = \{x_1 = 1\}$, $A_2 = \{x_2 = 1\}$ ו $A_3 = \{x_3 = 1\}$ כאשר {טענה} מסמן את קבוצת כל הסדרות שמקיימות את הטענה. A' הוא **איחוד** שלושת המאורעות $\{x_1 = 0\}$, $\{x_2 = 0\}$ ו $\{x_3 = 0\}$.

שני מאורעות שהחיתוך שלהם הוא $\emptyset$, כלומר שבלתי אפשרי שיתקיימו ביחד, נקראים מאורעות זרים. למשל: "בארבע ההגרלות הראשונות יצא 0" ו "מספר ה 1־ים שיצאו בארבע ההגרלות הראשונות הוא איזוגי".

מניחים שלמאורעות יש הסתברויות. אם אנו רוצים להיות במסגרת של תורת ההסתברות אין עלינו כל הגבלה בבחירת ההסתברויות, פרט לאקסיומות הבאות:

1. ההסתברות של כל מאורע A היא מספר ממשי $p = \Pr(A)$ המקיים $0 \le p \le 1$.
2. למאורע הוודאי (= כל מרחב המדגם) יש הסתברות 1 ולמאורע הבלתי־אפשרי - הסתברות 0.
3. אם נתונה סדרה סופית או אינסופית של מאורעות זרים $A_1, A_2, \ldots$ ו A הוא האיחוד שלהם, אזי $\Pr(A) = \sum_i \Pr(A_i)$.

מאקסיומות אלה נובע ש $\Pr(A) + \Pr(A') = 1$. מדוע?

כאמור, כל עוד מתקיימות אקסיומות אלו אנו במסגרת תורת ההסתברות, ויש אפשרויות שונות לתת הסתברויות למאורעות כך שהאקסיומות תתקיימנה. למשל: מכונה המגרילה באופן הוגן, מכונה המגרילה באופן לא הוגן (עדיפות ל 1, למשל), מכונה ללא זכרון (התוצאות בלתי־תלויות) או מכונה עם זכרון (למשל מכונה החוזרת תמיד פעמיים על אותה תוצאה) ואפשרויות עוד יותר מלאכותיות.

אבל אנו נניח שמדובר במכונה עם הסתברות $\frac{1}{2}$ לכל תוצאה, ושהתוצאות בהגרלות שונות בלתי תלויות. פרוש הדבר ש

$$\Pr\{x_1 = 0\} = \tfrac{1}{2},\ \Pr\{x_1 = 1\} = \tfrac{1}{2},$$
$$\Pr\{x_1 = 0, x_2 = 0\} = \left(\tfrac{1}{2}\right)^2,\ \Pr\{x_1 = 0, x_2 = 1\} = \left(\tfrac{1}{2}\right)^2,$$
$$\Pr\{x_1 = 1, x_2 = 0\} = \left(\tfrac{1}{2}\right)^2,\ \Pr\{x_1 = 1, x_2 = 1\} = \left(\tfrac{1}{2}\right)^2$$

וכן הלאה: אם $k = 1, 2, \ldots$ ו $a_1, a_2, \ldots, a_k$ מספרים נתונים השווים כל אחד ל 0 או 1 אזי

$$\Pr\{x_1 = a_1, x_2 = a_2, \ldots, x_k = a_k\} = \left(\frac{1}{2}\right)^k \tag{1}$$

הסתברות מותנית ואי־תלות

נניח שאנו עובדים עם מערכת הסתברותית, כלומר עם מרחב מדגם עם הסתברויות למאורעות בו, ונודע לנו שמאורע A קרה. למשל, במקרה של ההגרלות, נודע לנו שבשתי ההגרלות הראשונות יצא 0 (זה יהיה המצב שלנו בעתיד, אחרי שבוצעו שתי ההגרלות הראשונות, אם באמת יצא בשתיהן 0). ל A היתה בתחילה הסתברות $p = \Pr(A)$ (בדוגמא שלנו $p = \frac{1}{4}$), אבל אחרי שאנו יודעים ש A קרה עלינו לשנות את מערכת המושגים שלנו, ולתת ל A הסתברות 1!. מערכת ההסתברויות החדשה שלנו אחרי שאנו יודעים ש A קרה תיקרא "הסתברויות מותנות ע"י A" או "הסתברויות יחסיות ל A" והן תהיינה פונקציה חדשה $\Pr_A$ על המאורעות. $\Pr_A$ תוגדר כך: אם B מאורע אחר כלשהו, יהיה כעת B "אקוויולנטי" ל $A \cap B$ (כי אנו הרי יודעים עכשיו ש A נכון!) ומגדירים:

$$\Pr_A(B) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

(כדי שהגדרה זו תהיה בעלת טעם, צריך ש A לא יהיה בעל הסתברות 0).

הסימון המקובל בתורת ההסתברות לערכה של ההסתברות המותנית $Pr_A(B)$ הוא $Pr(B \mid A)$ קרי: ההסתברות של B כאשר נתון ש A קרה.

למשל: ההסתברות המותנית של כל מאורע שכולל את A (כלומר התרחשותו גוררת את התרחשות A) תהיה 1. ההסתברות המותנית של מאורע שסותר את A תהיה 0.

ההסתברות המותנית ביחס למאורע הוודאי היא ההסתברות הרגילה. (בדוק!)

שני מאורעות A ו B נקראים **בלתי תלויים** אם ידיעת אחד מהם אינה משנה את הסתברותו של השני. ליתר דיוק: אם מתקיים התנאי הבא:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

השקול לכל אחד משני התנאים הבאים (בהנחה ש A ו B אינם בעלי הסתברות 0):

$$\Pr(B \mid A) \ \big(= \Pr_A(B)\big) = \Pr(B)$$
$$\Pr(A \mid B) \ \big(= \Pr_B(A)\big) = \Pr(A)$$

בדומה לכך, n מאורעות $A_1, A_2, \ldots, A_n$ נקראים בלתי תלויים אם עבור כל בחירה $A_1 =: \widetilde{A_1}$ או $(A_1)'$, $A_2 =: \widetilde{A_2}$ או $(A_2)'$, ... , $A_n =: \widetilde{A_n}$ או $(A_n)'$, יתקיים:

$$\Pr\left(\widetilde{A_1} \cap \widetilde{A_2} \cap \cdots \cap \widetilde{A_n}\right) = \Pr(\widetilde{A_1}) \cdot \Pr(\widetilde{A_2}) \cdot \cdots \cdot \Pr(\widetilde{A_n})$$

(שימו לב: זה לא אותו דבר כמו להגיד שכל שניים מהם בלתי תלויים.)

למשל, הקוראים יכולים לבדוק שנוסחת ההסתברות (1) שקולה לתנאים: כל תוצאה בכל הגרלה היא בעלת הסתברות $\frac{1}{2}$ ומאורעות האומרים מהן התוצאות בהגרלות שונות הם בלתי תלויים.

את הוכחת משפט העזר הבא נשאיר כתרגיל לקוראים:

משפט עזר: אם $A_1, A_2, \ldots, A_k$ הם מאורעות זרים, ו B מאורע שההסתברות המותנית שלו לגבי כל אחד מהמאורעות A_k היא אותו ערך p, אזי גם ההסתברות המותנית שלו לגבי איחוד המאורעות $A_i,\ i = 1 \ldots k$ היא p (איחוד זה הוא, כמובן, המאורע: "אחד מ $A_1 \ldots, A_k$ קרה"). אותו דבר נכון לגבי סדרה אינסופית של מאורעות זרים $A_i,\ i = 1, 2, \ldots$.

הוכחת משפט העזר היא שימוש ישיר בהגדרת הסתברות מותנית ובאקסיומה 3, ועבור סדרה אינסופית - גם בתכונות פשוטות של טורים מתכנסים של מספרים חיוביים.

הוכחת המשפט שלנו: כל איסטרטגיית הימורים לא תשנה את סיכויי המהמר.

איסטרטגית הימורים של השחקן שלנו תבוטא ע"י סידרה $\tau(1), a_1, \tau(2), a_2, \ldots$ בה השחקן מהמר פעם ראשונה במשחק ה $\tau(1)$ על התוצאה a_1, פעם שניה במשחק ה $\tau(2)$ על התוצאה a_2 וכו'. ה τ-ים וה a-ים הן פונקציות של תוצאות ההגרלות **בעבר**. למשל אם האיסטרטגיה היא: המר על 1 אחרי שיצאו 3 פעמים 0, יהיה, עבור כל איבר $x = (x_1, x_2, \ldots)$ של מרחב המדגם, $\tau(1) =$ ה i הראשון כך ש $x_{i-1} = x_{i-2} = x_{i-3} = 0$ ו $a(1) = 1$. $\tau(2)$ יהיה (עבור כל x) ה i השני בעל תכונה זו, וגם $a(2) = 1$ וכן הלאה.

נסמן $y_i = x_{\tau(i)}$ זוהי תוצאת ההגרלה בפעם ה i בה השחקן שלנו שיחק. y_i מקבלת ערך 0 או 1 בהתאם לנקודה במרחב המדגם, והשחקן שלנו היה רוצה למצוא איסטרטגיה עבורה ל y_i תהיה, למשל, הסתברות גדולה להיות 1 ואז יוכל להחליט להמר על 1. אילו היה גיבורנו ניחן בידיעת העתיד, היה קל מאוד למצוא $\tau(i)$ כך ש $x_{\tau(i)}$ יהיה אפילו בוודאות 1, למשל $\tau(i) =$ הפעם ה i בה ייצא 1!

הקוראים יכולים להשתכנע שאת המשפט שלנו נוכל לנסח כעת כך:

משפט (ניסוח שני): עבור כל איסטרטגיה, האומרת שיש להשתתף במשחקים ה $\tau(1), \tau(2), \ldots$, כאשר כל $\tau(i)$ הוא פונקציה של תוצאות ההגרלות **בעבר**, יהיו ההסתברויות של ערכי ה $y_i = x_{\tau(i)}$ נתונות ע"י נוסחה זהה ל (1):
אם $k = 1, 2, \ldots$ ו $a_1, a_2, \ldots, a_k$ מספרים נתונים השווים כל אחד ל 0 או 1 אזי

$$\Pr\{y_1 = a_1, y_2 = a_2, \ldots, y_k = a_k\} = \left(\frac{1}{2}\right)^k \qquad (1')$$

ניסוח שקול ל $(1')$ הוא

$(1'')$ למאורע $\{y_k = 1\}$ יש הסתברות מותנית $\frac{1}{2}$ לגבי כל מאורע מהצורה:

$$A = \{y_1 = a_1, y_2 = a_2, \ldots, y_{k-1} = a_{k-1}\} \qquad (*)$$

(מדוע זה שקול?)

נוכיח כעת את $(1'')$. נעזר במשפט העזר שלנו. הוא אומר שנוכל לפרק את המאורע A מ $(*)$ לסדרה סופית או אינסופית של מאורעות זרים, ודי יהיה להוכיח של $\{y_k = 1\}$ יש הסתברות מותנית $\frac{1}{2}$ לגבי כל אחד מהם.

נתחיל מפירוק A למאורעות: A_j = "A קרה ו $\tau(k) = j$", לשון אחרת: "A קרה והפעם ה k שמהמר החליט לשחק היא המשחק ה j", נקבל פירוק לאינסוף מאורעות $A_1, A_2, \ldots$ שחלקם יכול להיות ריק (נתעלם מאלה שריקים).
הייתרון של פירוק זה הוא שכאשר אנו דנים בהסתברות מותנית לגבי A_j, כלומר אנו בהנחה ש $\tau(k) = j$, המאורע $\{y_k = 1\}$, כלומר $\{x_{\tau(k)} = 1\}$, ניתן להחלפה במאורע $\{x_j = 1\}$.
לכן עלינו לבדוק שההסתברות המותנית של $\{x_j = 1\}$ לגבי A_j היא $\frac{1}{2}$.
אבל אני טוען ש A_j מתייחס רק לתוצאות ההגרלות $1, 2, \ldots, (j-1)$. כי כדי לקבוע שקרה A_j יש לקבוע קודם כל ש $\tau(k) = j$ וזה תלוי בתוצאות ההגרלות **לפני** ההגרלה ה j, וכעת כשאנו בהנחה $\tau(k) = j$ יש עוד לקבוע שקרה A, שתלוי בתוצאות הגרלות ב $\tau(1), \tau(2), \ldots, \tau(k-1) < \tau(k) = j$ כלומר בתוצאות הגרלות שלפני ה־j.[10]
מצאנו, איפוא, ש A_j תלוי רק בתוצאות ההגרלות $1, 2, \ldots, (j-1)$. אבל זה אומר ש A_j הוא איחוד זר של מאורעות מהצורה:

$$\{x_1 = b_1, x_2 = b_2, \ldots, x_{j-1} = b_{j-1}\}$$

כאשר ה b_i־ים הם 0 או 1.
והעובדה שלגבי מאורעות כאלה יש ל $\{x_j = 1\}$ הסתברות מותנית $\frac{1}{2}$ נובעת מ (1).

ולבסוף, שאלה לקוראים: אני דנתי בדוגמא מאוד מיוחדת של מכונת הימורים, שבה איסטרטגיית הימורים לא תועיל. נסו להכליל זאת למצבים יותר כלליים (ויותר שכיחים בחיים).

לקריאה נוספת

[1] William Feller: An Introduction to Probability Theory and Its Applications, Vol. I

[2] המאמר: האם הסתברות 0 פירושה ודאות, חוברת זו.

[10] הדיון כאן אינו מדוייק לגמרי. כדי לדייק לגמרי יש להגדיר באופן פורמלי מה פירוש הדרישה "$\tau(k)$ תלויה רק בהגרלות שהיו בעבר", דבר שנמנעתי ממנו כאן.

הגיאומטריה של חלוקת המושבים בכנסת בין הסיעות

[11] [12] לא, אין זה מאמר על פוליטיקה, וגם לא על הארכיטקטורה של אולם הכנסת, אלא על הבעיה הבאה:

בשיטת הבחירות היחסית הנהוגה בישראל, צריכה מפלגה שקבלה חלק x מן הקולות $(0 \le x \le 1)$ לקבל $120x$ מושבים בכנסת. אלא דא עקא - יש אילוץ אחר: מספר המושבים של המפלגה צריך להיות שלם. כלומר: למפלגה בודאי מגיעים $\lfloor 120x \rfloor$ מושבים (ב $\lfloor y \rfloor$ מסמנים את המספר השלם הגדול ביותר שאינו גדול מ y), אבל אחרי שלכל מפלגה i, שקבלה חלק x_i מהקולות, הוענקו $\lfloor 120x_i \rfloor$ המושבים שמגיעים לה, נותרים עוד מספר מושבים שיש לחלק בין המפלגות, שכן נשארו **עודפים** $120x_i - \lfloor 120x_i \rfloor$. כיצד לעשות זאת תוך שמירה על כללי הצדק וההגינות?

לשאלה זו אין תשובה חד־משמעית. דרך מתבקשת היא לתת את המושב הראשון הפנוי למפלגה בעלת העודף הגדול ביותר (שהוא כמובן מספר בין 0 ו 1), את המושב הבא למפלגה בעלת העודף השני בגודלו וכן הלאה עד שמחלקים את כל המקומות הפנויים. בתקופת גיבושה של החוקה האמריקאית הוצעה שיטה זו ע"י המילטון (Hamilton). נקרא לשיטה זו: **שיטת המילטון**[13]

שיטה זו היתה נהוגה בישראל בעבר, אבל היא שונתה בשנת 1970 בערך (חוק באדר־עופר) לטובת שיטה אחרת, מסובכת יותר, המעדיפה יותר מפלגות גדולות על קטנות. שיטה זו, הנהוגה בישראל עד היום, הוצעה ע"י ג'פרסון (Jefferson), גם הוא מאבות החוקה האמריקאית ואח"כ נשיא, שהיה בר־פלוגתא של המילטון בנושאים רבים. בשיטה זו, אם המפלגה ה i קיבלה a_i קולות, יש למצוא z ממשי כך ש

$$\sum_i \left\lfloor \frac{a_i}{z} \right\rfloor = 120$$

והמפלגה ה i תקבל $\left\lfloor \frac{a_i}{z} \right\rfloor$ מושבים[14]. נשאיר לקוראים לבדוק שלמרות ש z אינו נקבע באופן יחיד, הערכים $\left\lfloor \frac{a_i}{z} \right\rfloor$ נקבעים באופן יחיד, וששיטה זו אכן מעדיפה מפלגות גדולות על קטנות לעומת שיטת המילטון.

נעיר, שלשם פשטות התעלמנו כאן מקיום אחוז חסימה והסכמי עודפים.

נציג את הבעיה שלנו בצורה הבאה:
נניח שיש N מפלגות המשתתפות בבחירות. יש לנו קבוצה E של כל הוקטורים האפשריים של תוצאות הבחירות, מותאמות ל 120 חברי הכנסת:

$$E = \left\{ \bar{v} = (v_1, v_2, \ldots, v_N) \mid v_i \ge 0,\ i = 1, \ldots, N,\ \sum_{i=1}^{N} v_i = 120 \right\}$$

אבל חלוקות מושבים בכנסת יכולות להעשות רק ע"י אברי התת־קבוצה של E, שאבריה הם הוקטורים $\bar{v}$ שכל רכיביהם **שלמים**. נסמן תת־קבוצה זו ב E_0.

הבעיה שלנו היא: נתון איבר של E, מצא איבר של E_0 שיכול לייצג אותו באופן ההוגן ביותר.

בעיות כאלה נקראות בעיות **אפרוקסימציה (קירוב)**.

דוגמא אחרת לבעיית אפרוקסימציה: אסטרונום גילה כוכב־שביט והוא ערך עליו N מדידות, שמהוות וקטור v. לא כל וקטור של N תוצאות אפשרי לפי התיאוריה: השביט חייב לנוע לפי חוקי הכבידה של ניוטון; אם מתחשבים רק בכבידת השמש - לפי חוקי קפלר. מצד שני ברור שבתוצאותיו של האסטרונום נכנסו שגיאות מדידה. אם נסמן ב E את קבוצת כל וקטורי התוצאות וב E_0 את קבוצת הוקטורים שאפשריים לפי חוקי ניוטון, אזי אפשר לראות כסביר לאמר שהתנועה האמיתית של השביט מתוארת ע"י וקטור $v' \in E_0$ שהוא הקירוב הטוב ביותר ל v מבין אברי E_0 - זוהי בעיית אפרוקסימציה, והאנלוגיה לבעיית המושבים בכנסת ברורה.

אלא שכמו במקרה של הכנסת, המושג "קירוב טוב ביותר" אינו חד־משמעי. באסטרונומיה ובמדעים אחרים נוקטים לעתים קרובות, בעקבות המתמטיקאי הידוע גאוס (1855-1777), ב**שיטת הריבועים המינימליים:** מחפשים וקטור $v' \in E_0$ עבורו יהיה מינימום ל:

$$(v'_1 - v_1)^2 + (v'_2 - v_2)^2 + \cdots + (v'_N - v_N)^2 \qquad (1)$$

[11] הרעיונות במאמר זה באים בחלקם מהרצאה שנתן איתן לפידות, אז דוקטורנט בטכניון, בשנת 1970 בערך.

[12] אני מודה לפרופ' רון הולצמן, מהפקולטה למתמטיקה בטכניון, על הערותיו ועזרתו.

[13] בארצות הברית יש לשאלות מתמטיות כאלה מעמד משפטי והן מגיעות עד בתי המשפט, שכן יש חובה לפעול לפי החוקה, ואם החוקה אומרת, למשל ש"המדינות צריכות להיות מיוצגות לפי מספר תושביהן" יכול בית המשפט להתבקש להכריע אם שיטה מסויימת מקיימת תנאי זה.

[14] תאורטית ייתכן ש z כזה אינו קיים בגלל בעיות תיקו (למשל: 7 מפלגות, $a_1 = a_2 = \ldots = a_7$) אבל במציאות הסיכוי שזה יקרה זניח.

(כאן ה v_i ־ים נתונים וה v'_i־ים משתנים על E_0). הסיבה ללקיחת חזקה 2 ב (1) היא רק הנוחיות המתמטית. חזקה 2 מופיעה בנוסחת המרחק בגיאומטריה האנליטית הרגילה, ולכן הגדרת מרחק כזו נותנת מבנה עשיר כמו בגיאומטריה האוקלידית הרגילה שיותר קל לעבוד בו.
למשל נניח $N = 3$ ונתאר את וקטור תוצאות הבחירות (v_1, v_2, v_3) כנקודה במרחב, שחייבת להימצא במשולש

$$v_1 + v_2 + v_3 = 120,\ v_1 \geq 0,\ v_2 \geq 0,\ v_3 \geq 0$$

E_0 ־ החלוקות האפשריות של מושבים בכנסת תתואר כסריג נקודות בתוך משולש זה. (1) תתאים לנוסחת ריבוע המרחק האוקלידי הרגיל והבעיה היא למצוא נקודה בסריג שהיא קרובה ביותר לנקודה נתונה במשולש.

אבל באותה מידה יכולנו לחפש מינימום ל:

$$|v'_1 - v_1| + |v'_2 - v_2| + \cdots + |v'_N - v_N| \tag{2}$$

מדוע צריך כאן ערך מוחלט? (העובדה שב (1) אפשר לוותר על הערך המוחלט היא אחד היתרונות שבחזקה 2).

המרחק (2) הוא "ישיר" יותר (אין חזקה שניה) אבל קשה יותר לעבוד איתו. הוא מתאר את ה"גיאומטריה של נהגי טכסי" (בעיר כמו ניו־יורק), שיכולים לנוע רק לאורך קוים מקבילים לצירי הקואורדינטות, ולגביהם אכן זהו מרחק הנסיעה הקצר ביותר בין שתי נקודות (בדוק!).

לכל נוסחת מרחק יש צורה אחרת ל"כדור". ב"כדור" הכוונה, כמובן, לקבוצת הנקודות שמרחקן ממרכז מסויים קטן מ r. עבור "הגיאומטריה של נהגי הטכסי" (2) הכדור יהיה: במישור ($N = 2$) ־ ריבוע שאלכסוניו מקבילים לצירים, ובמרחב ($N = 3$) ־ תמניון (אוקטאדר).

מרחק "טבעי" אחר הוא

$$\max_{1 \leq i \leq N} |v'_i - v_i| \tag{3}$$

עבור מרחק זה, הכדור הוא קוביה שצלעותיה מקבילות לצירים.

נוסחות "מרחק" שונות יתנו, בדרך כלל, תשובות שונות לגבי הקירוב הטוב ביותר.[15]

תרגיל: נתבונן במקרה הפשוט ש E_0 היא קבוצת הסדרות הקבועות $v_1 = v_2 = \cdots = v_N$. הוכח שהקירוב הטוב ביותר לוקטור כלשהו $\bar{v}$ יהיה:
בשיטת הריבועים המינימליים (1) ־ הסדרה הקבועה השווה לממוצע של אברי $\bar{v}$.
ב"מרחק" (2) ־ הסדרה הקבועה השווה לחציון של אברי $\bar{v}$. (החציון מוגדר כאותו מספר μ עבורו מספר אברי הסדרה הגדולים מ μ שווה למספר אברי הסדרה הקטנים מ μ. ייתכן שהחציון לא נקבע באופן חד־ערכי, ואז כל הסדרות הקבועות השוות לכל אחד מהחציונים יתנו פתרון לבעיה ־ כולן יהיו באותו מרחק מ $\bar{v}$ ומרחק זה יהיה מינימלי).
ב"מרחק" (3) ־ הסדרה הקבועה השווה לממוצע בין האיבר המכסימלי והמינימלי ב $\bar{v}$.
(אם קשה לך להוכיח עבור N כללי, נסה להוכיח עבור $N = 3$. אולי שאלה זו תהיה קלה יותר אחרי קריאת מאמר זה בשלמותו).
בנה דוגמא בה שלושת הקירובים הטובים ביותר האלה שונים זה מזה.

נחזור לבעיית המושבים בכנסת. מהו הקירוב הכי טוב בשיטת הריבועים המינימליים? במרחקים (2) ו (3)?

כאמור, הבעיה שלנו היא הבאה: יש לנו קבוצה (כל תוצאות הבחירות האפשריות, מותאמות ל 120 חברי כנסת):

$$E = \left\{ \bar{v} = (v_1, v_2, \ldots, v_N) \mid v_i \in \mathbf{R},\ v_i \geq 0,\ i = 1, \ldots, N,\ \sum_{i=1}^{N} v_i = 120 \right\}$$

ותת־קבוצה שלה (כל החלוקות האפשריות של מושבים בכנסת):

$$E_0 = \left\{ \bar{n} = (n_1, n_2, \ldots, n_N) \in E \mid \text{ שלמים } n_1, \ldots, n_N \right\}$$

נתון לנו $\bar{v} \in E$ ואנו רוצים למצוא $\bar{n} \in E_0$ שנותן מינימום למרחק בין $\bar{v}$ ו $\bar{n}$, כאשר לוקחים אחת השיטות שהזכרנו למדידת מרחק. אנו נדון במרחק האוקלידי (1) וב"מרחק נהגי הטכסי" (2). את הדיון במרחק המכסימום (3) נשאיר לקוראים.

[15] בכל סוגי המרחק הנדונים במאמר זה, הגאומטריה היא "הומוגנית", במובן הבא: הזזה של המרחב לא משנה את המרחק בין נקודות, במלים אחרות: המרחק בין נקודות תלוי רק בוקטור המחבר אותן. אבל אפשריים בהחלט גם מרחקים ("גאומטריות") שאינם כאלה ־ חישבו, למשל על מרחק שמוגדר כזמן ההליכה הקצר ביותר בין שתי נקודות על שטח אדמה שחלקו סלול, חלקו אדמת חול, חלקו מיוער וחלקו בצה.

שימו לב שבניגוד לבעיות מינימום שפותרים אותן בעזרת נגזרת, כאן הקבוצה E_0 עליה מחפשים מינימום היא בדידה (דיסקרטית), אפילו סופית. עם זאת, נמצא את המינימום בעזרת אנאלוג בדיד לשיטת הנגזרת: נבצע שינוי במשתנה הבדיד $\bar{n}$ ונאמר שאם זה מינימום למרחק אזי השינוי במרחק המתאים לשינוי במשתנה חייב להיות אי־שלילי.

קודם כל כיון שהקבוצה E_0 סופית, בוודאי יש $\bar{n} \in E_0$ שנותן מינימום (מה שלא היה מובן מאליו עבור קבוצה אינסופית: התבוננו למשל בבעיית המינימום של הפונקציה $f(x) = 1/x$ כאשר x עובר על כל המספרים החיוביים!). לכן אם נמצא תנאים שחייב לקיים $\bar{n}$ שנותן מינימום, כך שקיים $\bar{n}$ יחיד המקיים תנאים אלה, או קיימים כמה $\bar{n}$ ־ים שמקיימים את התנאים ושנותנים כולם אותו ערך למרחק מ $\bar{v}$, אזי בוודאי $\bar{n}$ או $\bar{n}$ ־ים אלה נותנים את המינימום המבוקש.

נחפש תנאים כאלה, שכאמור מזכירים את שיטת התאפסות הנגזרת: נניח ש $\bar{n}$ נותן מינימום למרחק. בלי הגבלת הכלליות נניח שסידרנו את האינדקסים כך ש

$$v_1 - n_1 \geq v_2 - n_2 \geq \cdots \geq v_N - n_N \tag{4}$$

כיוון שסכום ה n_i ־ים שווה לסכום ה v_i ־ים, חייבים להיות בין ההפרשים ב (4) גם חיוביים וגם שליליים (אלא אם כן $\bar{v} = \bar{n}$, כלומר $\bar{v} \in E_0$ ואין מה להוכיח) לכן $v_1 - n_1 > 0$, $v_N - n_N < 0$.
נשתמש בכך, שמאחר ש $\bar{n}$ נותן מינימום למרחק, המרחק עבור $n_1 + 1, n_2, \ldots, n_N - 1$ יהיה יותר גדול (או שווה) מהמרחק עבור $n_1, n_2, \ldots, n_N$ (שימו לב ש $n_N \geq 1$ כי $v_N - n_N < 0$).

כעת נבדיל בין נוסחאות המרחק השונות:

- המרחק האוקלידי (1) ־ שיטת הריבועים המינימליים.

השינוי בריבוע המרחק האוקלידי (1) כאשר עוברים מ $n_1, n_2, \ldots, n_N$ ל $n_1 + 1, n_2, \ldots, n_N - 1$ הוא:

$$\begin{aligned}&\big((v_1 - n_1 - 1)^2 + (v_2 - n_2)^2 + \cdots + (v_N - n_N + 1)^2\big) - \big((v_1 - n_1)^2 + (v_2 - n_2)^2 + \cdots + (v_N - n_N)^2\big) = \\ &= 2\left[(v_N - n_N + 1) - (v_1 - n_1)\right]\end{aligned}$$

וזה חייב להיות אי־שלילי, כלומר

$$v_N - n_N + 1 \geq v_1 - n_1 \tag{5}$$

מכאן נוכל להסיק כמה מסקנות:

א. זיכרו ש $v_1 - n_1 > 0 > v_N - n_N$. לכן (5) גורר ש $1 > v_1 - n_1$ ו $-1 < v_N - n_N$. כל ההפרשים ב (4) חייבים להיות בין 1 ו -1 (אם $\bar{n}$ נותן מינימום).

ב. לכן אם $v_i - n_i \geq 0$ (כלומר $v_i \geq n_i$) אזי $n_i = \lfloor v_i \rfloor$ ־ המפלגה i לא נהנתה מהעודף שלה; ואם $v_i - n_i < 0$ (כלומר $v_i < n_i$) אזי $n_i - 1 = \lfloor v_i \rfloor$ ־ המפלגה i קבלה מושב נוסף הודות לעודף שלה.
אם

$$v_1 - n_1 \geq \cdots \geq v_k - n_k \geq 0 > v_{k+1} - n_{k+1} \geq \cdots \geq v_N - n_N$$

אזי רק הסיעות $k+1, \ldots, N$ קבלו מושב נוסף הודות לעודף שלהן.

ג. (5) אומר, כעת, שאם $1 \leq i \leq k < j \leq N$ אזי $v_j - n_j + 1 \geq v_i - n_i$, כלומר $v_j - \lfloor v_j \rfloor \geq v_i - \lfloor v_i \rfloor$. פירוש הדבר הוא שכל המפלגות שזכו במקום נוסף הן בעלות עודף גדול או שווה מהסיעות שלא זכו במקום נוסף.

ד. מסקנה: $\bar{n}$ נבנה בשיטת המילטון שתוארה בתחילת המאמר. אם יש תיקו, כלומר לכמה מפלגות יש אותו עודף ויש להכריע ביניהן, אזי יש כמה אפשרויות ל $\bar{n}$ שנותנות אותו ערך למרחק המינימלי.

- "מרחק נהגי־הטכסי" (2).

השינוי במרחק (2) כאשר עוברים מ $n_1, n_2, \ldots, n_N$ ל $n_1 + 1, n_2, \ldots, n_N - 1$ הוא:

$$\begin{aligned}&(|v_1 - n_1 - 1| + |v_2 - n_2| + \cdots + |v_N - n_N + 1|) - (|v_1 - n_1| + |v_2 - n_2| + \cdots + |v_N - n_N|) = \\ &= (|v_1 - n_1 - 1| - |v_1 - n_1|) + (|v_N - n_N + 1| - |v_N - n_N|)\end{aligned}$$

וכמו קודם, צריך להתקיים, אם ב $\bar{n}$ מתקבל מינימום:

$$(|v_1 - n_1 - 1| - |v_1 - n_1|) + (|v_N - n_N + 1| - |v_N - n_N|) \geq 0 \tag{6}$$

נזכיר ש $v_1 - n_1 \geq 0$, $v_N - n_N \leq 0$ לכן (6) שקול ל:

$$(|v_1 - n_1 - 1| - (v_1 - n_1)) + (|v_N - n_N + 1| + (v_N - n_N)) \geq 0$$

כלומר

$$(|v_1 - n_1 - 1| - (v_1 - n_1 - 1)) + (|v_N - n_N + 1| + (v_N - n_N + 1)) - 2 \geq 0$$

מתקיים $|x| - x = 2\max(-x, 0)$, $|x| + x = 2\max(x, 0)$ ולכן מתקבל:

$$\max\big(1 - (v_1 - n_1), 0\big) + \max\big(1 + (v_N - n_N), 0\big) \geq 1 \tag{7}$$

וכיון ש $1 - (v_1 - n_1) < 1$, $1 + (v_N - n_N) < 1$, (7) יכול להתקיים רק אם

$$\big(1 - (v_1 - n_1)\big) + \big(1 + (v_N - n_N)\big) \geq 1$$

שאינו אלא (5)

אחרי שקבלנו גם כאן את (5), מתקבלות אותן מסקנות א'-ד' ויש לנו אותו פתרון - שיטת המילטון.

לסיכום, עבור בעיית המושבים בכנסת, נותנים המרחק האוקלידי ו"מרחק נהגי הטכסי" אותו מינימום. זו תופעה לא רגילה, כי בדרך כלל בבעיה של נקודה קרובה ביותר הפתרון יהיה שונה עבור כל הגדרת מרחק (ראו למשל בתרגיל לעיל).

לקריאה נוספת

על בעיית חלוקת מושבים - שיטות נוספות שהוצעו, גישות נוספות להשוואה בין השיטות ע"י שימוש בכלים מתמטיים, ואנקדוטות היסטוריות הקשורות לנושא, אפשר למצוא בספר:

M. L. Balinski and H. P. Young, *Fair Representation: Meeting the Ideal of One Man, One Vote*, Yale University Press, New Haven, 1982.

נוסחת נפח אוניברסלית

נתבונן בטענה הבאה:

נפח גוף, המוגבל בין שני מישורים מקבילים שווה ל:
ששית שטח הבסיס העליון של הגוף (במישור המקביל העליון שיסומן ב α)
+ ששית שטח הבסיס התחתון של הגוף (במישור המקביל התחתון שיסומן ב β)
+ ארבע ששיות שטח החתך האמצעי של הגוף
כל זה כפול המרחק בין המישורים α ו β.

נוסחה זו נכונה עבור: (בדוק!)

- מנסרה או גליל (α ו β - מישורי הבסיסים)
- פירמידה או חרוט (β - מישור הבסיס, α -מישור מקביל לו דרך הקודקוד)
- פירמידה קטומה או חרוט קטום (α ו β - מישורי הבסיסים)
- כדור (α ו β מישורים משיקים מקבילים)
- החלק של כדור שבין שני מישורים מקבילים
- החלק של אליפסואיד, פרבולואיד או היפרבולואיד שבין שני מישורים מקבילים

מדוע אותה נוסחה טובה לכולם? נסו למצוא הוכחה שעובדת עבור כל המקרים האלה (רמז: הנפח הוא אינטגרל של שטח החתך, ושטח החתך בכל המקרים האלה היא פונקציה של הגובה z שהיא פולינום ממעלה 2 לכל היותר. מה יקרה עבור גוף ששטח החתך שלו הוא פולינום ממעלה 3? ממעלה 4?)

מיצאו דוגמה לגוף עבורו נוסחה זו אינה נכונה.

המשוואה $x^y = y^x, x \neq y$

נתבונן במשוואה $x^y = y^x, x \neq y$ כאשר מניחים $x, y > 1$.
פתרון מפורסם הוא, למשל, $x = 2\,,\ y = 4$.

גישה ראשונה
נכתוב את המשוואה בצורה $x^{1/x} = y^{1/y}$. עלינו רק לחקור את הפונקציה $x^{1/x}$ באינטרוול $1 \leq x < \infty$.
דבר זה עושים בקלות בעזרת חשבון דיפרנציאלי. (את הנגזרת של $f(x) = x^{1/x}$ מוצאים ע"י כתיבתה כפונקציה מורכבת $g(h(x))$ עם $g(x) = e^x$, $h(x) = (lnx)/x$).
מקבלים שפונקציה זו, שערכה ב $x = 1$ הוא 1, עולה ב $1 \leq x \leq e$, יש לה מקסימום ב e והיא יורדת ב $e \leq x \leq \infty$ ושואפת ל 1 כאשר $x \to \infty$.
לכן: לכל $x \neq e$ יש פתרון יחיד $y \neq x$. אין פתרון אם $x = e$.

גישה שנייה
נכתוב $t = y/x$. המשוואה מקבלת את הצורה:

$$(tx)^x = x^{tx} \Leftrightarrow t^x x^x = x^{tx} \Leftrightarrow t^x = x^{(t-1)x} \Leftrightarrow t = x^{t-1} \Leftrightarrow$$
$$(*) \qquad x = t^{\frac{1}{t-1}}, \quad y = t^{\frac{t}{t-1}}$$

שימו לב שאצלנו $t \neq 1$. אם $t \to 1$ הביטויים $(*)$ שואפים ל e (בידקו!).

בדרך זו אפשר גם לחקור פתרונות x, y רציונליים. אז $t = p/q$ רציונלי ולפי $(*)$:

$$x = \left(\frac{p}{q}\right)^{\frac{1}{\frac{p}{q}-1}} = \left(\frac{p}{q}\right)^{\frac{q}{p-q}}$$
$$y = \left(\frac{p}{q}\right)^{\frac{p}{p-q}}$$

למשל מתקבלים פתרונות רציונליים עבור $p = q + 1$:

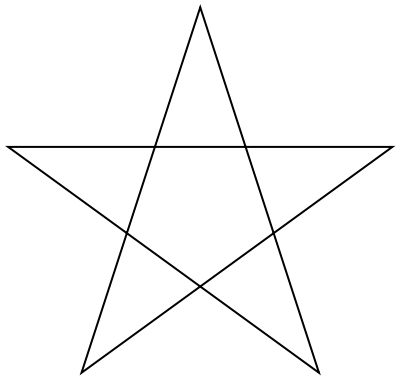
$$x = \left(\frac{q+1}{q}\right)^q, \quad y = \left(\frac{q+1}{q}\right)^{q+1} \quad q \text{ טבעי}$$

האם אלה הם הפתרונות הרציונליים היחידים?

הפנטאגרמה וגילוי האירציונליות

כידוע, היתה לתגלית שקיימים מספרים אירציונליים השפעה מכריעה על המחשבה המתמטית היוונית. תגלית זו, שנעשתה כנראה במאה החמישית לפני הספירה, מיוחסת לפיתגוראיים, שהיו כת חצי מתמטית וחצי מיסטית, שאת ייסודה מייחסים לפיתגורס, שלא ידוע אם הוא אינו בעצמו דמות אגדית.

ממבט ראשון סביר להניח שהמספר האירציונלי הראשון שנתגלה הוא $\sqrt{2}$, אבל היסטוריונים אחדים סבורים שאולי היה זה $\sqrt{5}$. הסבה לסברה זו היא החשיבות הרבה שנודעה במיסטיקה של הפיתגוראים לכוכב המחומש המשוכלל, הפנטאגרמה:



ומהתבוננות בכוכב זה אפשר להגיע למסקנה ש $\sqrt{5}$ הוא אירציונלי.

כיצד? (רמז: נסו להוכיח בדרך השלילה, בעזרת הפנטאגרמה, שלא ייתכן שצלעות משולש בעל זויות $72°, 72°, 36°$ כולן כפולות שלמות של אותה יחידת אורך).

נשמח לקבל את תגובותיכם.

דע מה שתשיב - $\pi > 3$

אם ישאלך השואל: מנין לך ש π אינו 3 אלא הוא גדול מ 3?

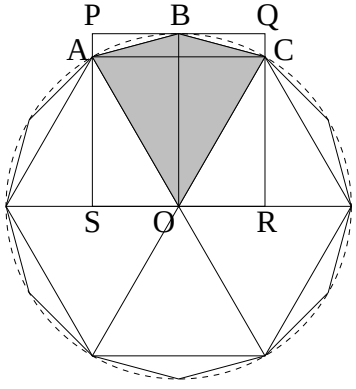
תשובה אפשרית לכך היא:

נחלק מעגל בעל רדיוס 1 (ולכן העיגול בעל שטח π) ל 12 חלקים שווים וע"י כך נחסום בו מצולע משוכלל בעל 12 צלעות.

שטח המצולע קטן משטח העיגול.

אני טוען ששטח המצולע הוא בדיוק 3.

זה נובע מכך ששטח כל אחד מ 6 המרובעים מטיפוס $ABCO$ הוא חצי שטח הריבוע $PQRS$ ששיטחו 1 ($AC = AO = OC = OB = 1$) כי AOC משולש שווה־שוקיים בעל זוית ראש $60°$ ולכן שווה־צלעות).



$\pi > 3$ - הערה לחוברת הקודמת

בחוברת הקודמת הצעתי לשכנע מי ששואל מדוע $\pi > 3$, בעזרת הוכחה שהשוותה שטח עיגול עם שטח מצולע משוכלל בעל 12 צלעות החסום במעגל (ראו בציור). בתגובות שקבלתי הציעו לעשות זאת יותר פשוט: משושה חסום במעגל בעל רדיוס 1 הוא בעל צלע השווה לרדיוס, ולכן בעל הקף 6, וזה קטן מהקף המעגל שהוא 2π.

כמובן, הוכחה אחת מתייחסת ל π של השטח והשנייה ל π של ההקף. העובדה שזה אותו π בשתי הנוסחאות מצריכה הוכחה נפרדת, ובוודאי רבים מהקוראים מכירים הוכחות או כמעט־הוכחות לכך.

משהו שאפשר לאמר בזכות ההוכחה שהבאתי בחוברת הקודמת (המתייחסת לשטח) לעומת ההוכחה הפשוטה המתייחסת להקף הוא: השומע החריף עלול לשאול: מניין לכם שאכן הקף המעגל גדול מהקף המשושה? זה נובע מהטענה שקו ישר הוא הדרך הקצרה ביותר בין שתי נקודות, טענה שכוללת בתוכה את המשפט שבמשולש סכום שתי צלעות גדול מהצלע השלישית (מה שנקרא: אי שוויון המשולש). בהוכחה שהובאה בחוברת הקודמת (ראו בציור) העובדה ששטח העיגול (בעל רדיוס 1) גדול משטח המצולע בעל 12 הצלעות (ששיטחו 3) נובעת מכך ששטח העיגול הוא שטח המצולע ועוד שטחי 12 הכיפות שנשארות אם נוציא את המצולע מהעיגול. אי אפשר לאמר דבר דומה כדי להוכיח שקשת במעגל ארוכה מהמיתר המחבר את קצותיה.

אינדוקציה, אבל לאו דווקא במספרים טבעיים

אקסיומת האינדוקציה ושיטת ההוכחה באינדוקציה ידועות כמשהו ששייך באופן מובהק למספרים הטבעיים (ראו מאמרו של שי גירון ב"אתגר־גליונות מתמטיקה" חוברת 36, ספטמבר 1995, בה התפרסם לראשונה גם מאמר זה). אני רוצה להראות במאמר זה שאפשר למצוא "אינדוקציה" גם במקומות אחרים.

אקסיומות פאנו

אם רוצים לבסס את המספרים הטבעיים בצורה אקסיומטית (בדומה לביסוס אקסיומטי של הגיאומטריה) מקובל ללכת בדרכו של פאנו (Peano). פאנו שם לב שאפשר לבסס על מושג אחד את כל המושגים המתייחסים למספרים טבעיים (למשל החיבור, הכפל, היחס: "גדול מ־", ואחר כך כמובן מושגים כמו התחלקות, מספרים ראשוניים וכו'). מושג זה הוא פעולת החשבון: "הוסף 1 למספר a" או, בלשון אחרת: "קח את המספר העוקב ל a". בניגוד לארבע פעולות החשבון הידועות, שהן בינריות (**משני** מספרים מקבלים תוצאה, למשל סכומם, או הפרשם או מכפלתם), פעולתו של פאנו, שנקרא לה "העוקב" היא אונרית (unary) (ממספר **אחד** a מקבלים תוצאה ־ העוקב שלו $a+1$). את האקסיומות של פאנו, שמהן אפשר לקבל את כל תורת המספרים הטבעיים, אפשר לנסח כך:

יש לנו קבוצה N של עצמים שייקראו: "מספרים טבעיים". יש ב N פעולה אונרית, שנקרא לה "עוקב" ונסמן את העוקב של מספר טבעי a ב a' (בכך שאמרנו שהפעולה היא ב N, נאמר שכמובן $a' \in \mathrm{N}$ אם $a \in \mathrm{N}$). N ופעולת העוקב $a \mapsto a'$ צריכות לקיים את האקסיומות:

Pea1 הפונקציה $a \mapsto a'$ היא חד־חד־ערכית, לשון אחרת: לכל $a, b \in \mathrm{N}$

$$a \neq b \Rightarrow a' \neq b'$$

Pea2 יש לפחות מספר אחד שאינו עוקב של אף מספר אחר (כלומר פונקציה זו **אינה** על).

Pea3 (אקסיומת האינדוקציה): קיים מספר טבעי (כלומר איבר ב N), שנסמן אותו ב 1, שלגביו מתקיימת התכונה הבאה:

אם M תת־קבוצה של N, שהיא בעלת שתי התכונות הבאות:

א. $1 \in M$

ב. M מקיימת: אם מספר טבעי כלשהו a שייך ל M, אזי גם העוקב שלו a' שייך ל M.

אז M חייבת להיות כל N.

כמו שטענו למעלה, די בשלש אקסיומות אלה כדי להגדיר ולהוכיח את כל המושגים והתכונות של המספרים הטבעיים. התחלת הדרך היא קצת קשה ומופשטת: יש להגדיר ולהוכיח (באופן נכון מבחינה לוגית, כלומר בלי הנחת המבוקש, מוקש שקל מאוד ליפול בו!) מושגים כמו: יחס הסדר "$a > b$", פעולת החיבור, פעולת הכפל ואת תכונותיהם הבסיסיות.

שימו לב: כמו שהאקסיומות נוסחו כאן, את המושג של המספר הראשון 1 לא לוקחים כמושג בסיסי אלא מגדירים אותו (הכל ברוח הכלל שיוצאים ממושגים בסיסיים ואקסיומות שיש בהן פחות ככל האפשר): כתרגיל נסו להוכיח שמהאקסיומות נובע שקיים מספר טבעי יחיד 1 בעל התכונה שבאקסיומת האינדוקציה, ושהוא המספר הטבעי היחיד שאינו עוקב של מספר כלשהו. כתרגיל שני נסו להוכיח שלכל מספר a מתקיים $a \neq a'$.

פיתוח של תורת המספרים הטבעיים בעזרת אקסיומות פאנו (בניסוח קצת שונה משלנו) אפשר למצוא בספר:

E. Landau: Foundations of Analysis

זה לא כל כך פשוט...

מי שנחשף קצת ללוגיקה המתמטית יאמר כעת, בצדק, שהרושם שיצרתי זה עתה מוטעה: שלש האקסיומות מספיקות, אם יש ברשותנו בולדוזר כבד ומאיים: מושג הקבוצה והרשות לעשות פעולות של תורת הקבוצות באופן חופשי. למשל את אקסיומת האינדוקציה אי אפשר לנסח בלי שיהיה ברור לנו מה זה "תת־קבוצה כלשהי של N". מי ששמע על הסתירות שנתגלו בראשית המאה כאשר דנים בקבוצות באופן חפשי, ועל הבעיות והפרדוקסים שמופיעים (ראו מאמרו של רון אהרוני: הפרדוקסים, אתגר־גליונות מתמטיקה 35־34, יוני 1995) יאמר אולי ששימוש בכלי כבד כזה מבטל את כל הרושם של פשטות שיצרתי זה עתה. ואכן, הלוגיקאים המתמטיים רואים את N כמערכת פשוטה בהרבה מתורת הקבוצות, ומעדיפים לבסס אותה

רק על לוגיקה בסיסית ללא פעולות חפשיות בקבוצות. נושא זה הוא מאבני הפנה של הלוגיקה, (ומתקשר באופן מפתיע לתורת האלגוריתמים ולמדעי המחשב). חשוב לציין, שאז **אי אפשר לנסח את אקסיומת האינדוקציה במלואה**. אפשר רק לנסח משהו כמו (שהוא בפרוש טענה חלשה יותר):

Ind': יהי $P(\cdot)$ ביטוי שמבטא תכונה שיכולה להיות למספרים טבעיים, מבוטאת ע"י המושגים הבסיסיים (מושגי הלוגיקה, מושג העוקב, $<$, החיבור, הכפל - ראו להלן). (למשל: $P(x)$ הוא $x \cdot x \geq x$) אז אם

$$P(1)$$

ו

$$\forall a\,(P(a) \Rightarrow P(a+1))$$

אזי

$$\forall a\,P(a)$$

(כמובן שאת ה $\forall$ יש לקרוא: לכל מספר **טבעי**)

אקסיומה זו (שהיא למעשה מערכת של אקסיומות, עבור כל ביטוי P) מספיקה למטרות הרגילות, אבל כאמור היא אינה נותנת את אקסיומת האינדוקציה Pea3. (למי שיודע: מספר הביטויים P הוא רק בן־מניה, בעוד שמספר התת־קבוצות של N הוא עוצמת הרצף, שהיא הרבה יותר גדולה. לכן כמעט כל תת־קבוצה M אינה מוגדרת ע"י ביטוי) למשל, אם מסתפקים בה, אי אפשר להגדיר את החיבור והכפל רק בעזרת העוקב, ולכן דרושים מושגים בסיסיים נוספים על העוקב, בעוד שבעזרת Pea1,2,3 (ושימוש חפשי בקבוצות) אפשר להסתפק בפעולת העוקב (מה בדבר היחס $a > b$? כיצד אפשר להגדיר אותו אם החיבור ישנו?)

וכדי שלא תהיה טעות: הגישה של הלוגיקאים (או לפחות כמעט כולם) אינה שאסור לדון בקבוצות כמו שמקובל במתמטיקה, אלא שיש לראות את תורת הקבוצות כעמוקה ומסובכת יותר מתורת המספרים הטבעיים, ולכן אין הרבה טעם לבסס את המספרים הטבעיים על תורת הקבוצות.

וכמו בהרבה מקרים במתמטיקה, הופכים אי־נוחות ליתרון. הלוגיקאי אברהם רובינזון (ששהה תקופה מסויימת בארץ) ניצל את העובדה שאפשר להשתמש באקסיומה Ind' במקום אקסיומת האינדוקציה Pea3 לכל הוכחה "רגילה", אבל עדיין אין הן אקוויוולנטיות (ולכן אפשרית מערכת שמקיימת את Ind' אבל שונה באופן מהותי מ N), כדי לבנות את ה"אנליזה הלא־סטנדרטית", בה המערכת ה"אחרת" (בלשון הלוגיקאים: מודל אחר לאקסיומות של תורת המספרים הפורמלית) היא כזו בה מוסיפים לאברי N ה"רגילים" (ה"סטנדרטיים") מספרים אינסופיים, וכתוצאה מזאת מוסיפים ל $\mathbf{R}$ (הממשיים) מספרים אינסופיים ומספרים קטנים־לאינסוף (אינפיניטסימליים), ועדיין נשאר נכון כל מה שמוכח בעזרת האקסיומות בגרסתם של הלוגיקאים המתמטיים (שלא עושים בה פעולות "חפשיות" בקבוצות), כולל Ind'. בתורה זו אפשר להפוך טענות גבול (כמו בהגדרת הנגזרת בחשבון הדיפרנציאלי) לטענות על ה"אינפינטסימלים" שהן מאוד אינטואיטיביות. אנו מקווים להביא באחת החוברות הבאות מאמר המתאר תורה מעניינת זו ביתר פירוט.

"אינדוקציה" במבנים אלגבריים

אחרי ששלמנו "מס שפתיים" ללוגיקה, נמשיך להשתמש בקבוצות באופן המקובל במתמטיקה. (אם רוצים, זה ניתן לתיאור לוגי פורמלי אם מתארים את הלוגיקה עצמה - דברים כמו: "או", "ו", "אם", "לכל" "קיים" וכו' - באופן פורמלי ומבססים את הפעולות בקבוצות על אקסיומות). נשים לב שהצורה בה נוסחו אקסיומות פאנו Pea1,2,3 היא מאוד "אלגברית", משהו דומה למושג החבורה: יש קבוצה עם פעולה אונרית (כמו שבחבורה יש פעולה בינרית) ודורשים שיתקיימו אקסיומות. האם יש אנלוגיה לאקסיומת האינדוקציה במערכות אלגבריות אחרות?

נתבונן, למשל במספרים הרציונליים עם ארבע פעולות החשבון (החילוק, כמובן, לא מוגדר אם המחלק הוא 0). הטענה הבאה נכונה (מדוע?)

טענה 1: נקבע מספר רציונלי r_0 השונה מ 0. תהי M קבוצה של מספרים רציונליים המקיימת:

א. $r_0 \in M$

ב. עבור כל שני מספרים רציונליים x ו y ב M, מתקיים: $x+y$, $x-y$, xy שייכים ל M, ואם, בנוסף לכך, $y \neq 0$ אזי גם x/y שייך ל M

אז M היא קבוצת כל המספרים הרציונליים.

האנלוגיה לאקסיומת האינדוקציה ברורה: במקום פעולה אונרית אחת יש לנו ארבע פעולות בינריות, ובמקום 1 בא r_0.

דוגמא אחרת:

טענה 2: תהי M קבוצה של מספרים שלמים המקיימת:

א. $1 \in M$

ב. עבור כל שני מספרים שלמים x ו y ב M, גם $x+y$ ו $x-y$ שייכים ל M

אזי M היא קבוצת כל המספרים השלמים.

(האם תוכלו לקבל טענה דומה עם אותו 1 אבל עם פעולות אונריות במקום $+$ ו $-$?)
כל טענה כזו נותנת דרך הוכחה ב"אינדוקציה". למשל: אנו רוצים להוכיח את הטענה הבאה:

טענה 3: אם $f : \mathbf{R} \to \mathbf{R}$ פונקציה מונוטונית (עולה או יורדת) המקיימת:

$$f(x+y) = f(x) + f(y) \qquad \text{לכל } y, x \text{ ממשיים} \tag{1}$$

אזי

$$f(x) = x \cdot f(1) \qquad \text{לכל } x \text{ ממשי} \tag{2}$$

הוכחה:
אנו רוצים להשתמש ב"אינדוקציה" על המספרים הרציונליים, לפי טענה 1, עם $r_0 = 1$. קודם כל נחזק את הטענה שלנו (כמו באינדוקציה רגילה, לפעמים כדי להוכיח טענה באינדוקציה צריך לחזק אותה - ראו במאמרו של שי גירון שנזכר לעיל).
נתבונן בטענה

$$f(xy) = x \cdot f(y) \tag{3}$$

נוכיח ש (3) נכונה עבור x, y **רציונליים**. נעשה זאת ע"י הוכחת הטענה: "(3) נכונה עבור כל y רציונלי" באינדוקציה על x רציונלי, כאשר האינדוקציה היא לפי טענה 1 (עם $r_0 = 1$).

עבור $x = 1$ הטענה (3) מיידית.
נניח ש (3) נכונה עבור x ו x' רציונליים מסויימים, ועבור כל y רציונלי. תפקידנו הוא להוכיח אותה עבור $x + x'$, $x - x'$, xx', x/x' (אם $x' \neq 0$) ועבור כל y רציונלי. אבל זה נובע מ:

$$f\big((x+x')y\big) = f(xy + x'y) \overset{\text{לפי (1)}}{=} f(xy) + f(x'y) \overset{\text{לפי הנחת האינדוקציה}}{=} xf(y) + x'f(y) = (x+x')f(y)$$

$$f(xy) = f\big((x-x')y + x'y\big) = f((x-x')y) + f(x'y)$$
$$f((x-x')y) = f(xy) - f(x'y) = xf(y) - x'f(y) = (x-x')f(y) \quad \text{ולכן}$$

$$f(xx'y) \overset{\text{לפי הנחת האינדוקציה}}{=} xf(x'y) \overset{\text{לפי הנחת האינדוקציה}}{=} xx'f(y)$$

$$x'f\big((x/x')y\big) = f\big(x'(x/x')y\big) = f(xy) = xf(y)$$
$$f\big((x/x')y\big) = (x/x')f(y) \quad \text{ולכן}$$

כעת אנו יודעים ש (3) נכונה אם x, y רציונליים. נציב כעת $y = 1$ ונקבל את (2) עבור x רציונלי. כדי להוכיח את (2) עבור x ממשי כאשר היא ידועה עבור x רציונלי משתמשים במונוטוניות של f ובכך שבין כל שני מספרים ממשיים יש מספר רציונלי:
ניפטר קודם כל מהמקרה $f(1) = 0$. אז לפי מה שהוכחנו $f(x) = 0$ לכל x רציונלי, והודות למונוטוניות של f - גם לכל x ממשי. לכן נניח $f(1) \neq 0$.
נניח שהיה קיים x ממשי עבורו $f(x) \neq x \cdot f(1)$. בלי הגבלת הכלליות נניח $f(x) > x \cdot f(1)$. אז קיים מספר רציונלי r כך ש

$$x \cdot f(1) < r \cdot f(1) = f(r) < f(x) \tag{4}$$

נעיר שאם נציב ב (1) $x = y = 0$ נקבל ש $f(0) = 0$. לכן:

$$\text{אם } f \text{ עולה אזי } f(1) > 0 \text{ ואם } f \text{ יורדת אזי } f(1) < 0 \tag{5}$$

נשאיר לקוראים לבדוק ש (5) סותרת את (4).

מ.ש.ל.

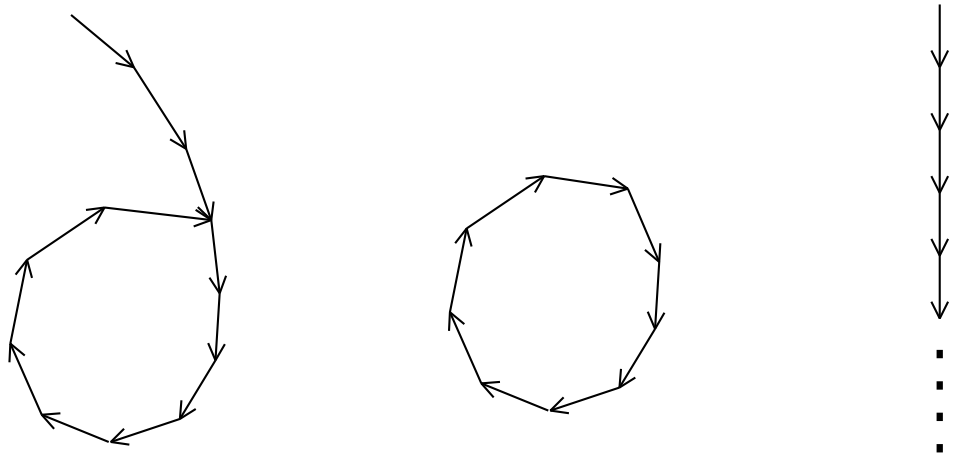
אפשר לנסח את התופעה שלנו כך: יש לנו מבנה אלגברי, כלומר קבוצה S עם פעולות, ויש לנו איבר (המספר 1 באקסיומת האינדוקציה) או כמה איברים, כך שכל תת־קבוצה M של S שמכילה איברים אלה ושהיא סגורה לגבי הפעולות היא בהכרח הקבוצה S כולה.
למצב כזה יש מונח מקובל במתמטיקה: אומרים שאיברים אלה **יוצרים** את S במובן של המבנה האלגברי הנדון. אקסיומת האינדוקציה הרגילה ניתנת לניסוח כך:

המספר 1 יוצר את קבוצת המספרים הטבעיים N במובן של המבנה האלגברי הנתון ע"י פעולת העוקב.

לעומת זאת 1 **לא** יוצר את קבוצת המספרים השלמים $\mathbf{Z}$ במובן של המבנה האלגברי הנתון ע"י פעולת העוקב $a \mapsto a+1$ (מדוע?). עבור אילו מבנים אלגבריים ב $\mathbf{Z}$ 1 כן יוצר את $\mathbf{Z}$? מה בדבר 0 כיוצר? עבור אילו מבנים אלגבריים ב N, פרט למבנה העוקב, 1 הוא יוצר? עבור כל דוגמא של יוצר, נסחו את "אקסיומת האינדוקציה" המתאימה ונסו להשתמש בה להוכחת טענות.

מערכות עוקב עם אינדוקציה

נשאלת כעת שאלה: מהם כל המבנים, הדומים ל N בכך שהם קבוצה עם פעולה אונרית "עוקב", והמקיימים את אקסיומת האינדוקציה, כלומר יש בהם איבר יחיד 1 שיוצר אותם. נשאיר את פתרון שאלה זו לקוראים. רמז לפתרון ניתן בציור הבא:



האם תוכלו להוכיח טענות מעניינות בעזרת אינדוקציה במבנה כזה?

האינדוקציה והסדר

נחזור למספרים הטבעיים ולאקסיומות $\mathrm{Pea1,2,3}$. אחרי שיש לנו המושג: $x \geq y$[16] אפשר להוכיח בעזרת אקסיומת האינדוקציה את הטענה:

קיום איבר קטן ביותר: בכל קבוצה לא ריקה M של מספרים טבעיים יש איבר (בהכרח יחיד) שהוא קטן ביותר.

טענה זו אפשר להסיק באופן "נקי" (אבל עם הרבה "פעולות בקבוצות") מאקסיומת האינדוקציה באופן הבא:

הוכחה: נניח ש M היתה קבוצה לא ריקה של מספרים טבעיים שאין בה איבר קטן ביותר. נתבונן בקבוצה K שמורכבת מכל המספרים הטבעיים הקטנים מכל איברי M.
נוכיח ש K מקיימת את דרישות אקסיומת האינדוקציה:
קודם כל יש להוכיח ש $1 \in K$. לשם כך נשים לב ש $1 \notin M$, כי אחרת היה 1 בוודאי האיבר הקטן ביותר ב M. כיון ש 1 אינו ב M, הוא בודאי קטן מכל איברי M, ולכן שייך ל K.
נניח כעת ש $a \in K$. אנו רוצים להוכיח ש $a' \in K$. אבל נזכור ש a קטן מכל איברי M. לכן כולם גדולים או שווים מ a'. לכן אילו היה $a' \in M$ הוא היה האיבר הקטן ביותר בה, מה שלא נכון. מכאן $a' \notin M$ ולכן a' קטן מכל איברי M, כלומר $a' \in K$.
לפי אקסיומת האינדוקציה נובע כעת ש $K = \mathrm{N}$ אבל ברור ש K זרה ל M ולכן מקבלים ש $M = \emptyset$, בניגוד להנחה.

מ.ש.ל.

אבל האם גם ההפך נכון? האם מקיום איבר קטן ביותר נובעת אקסיומת האינדוקציה?
התשובה היא **שלילית** ונוכיח זאת כמו שמוכיחים שאקסיומת המקבילים בגאומטריה האוקלידית לא נובעת משאר האקסיומות: בונים מערכת "לא אינדוקטיבית" בה בכל קבוצה לא־ריקה יש איבר קטן ביותר אבל לא מתקיימת אקסיומת האינדוקציה. מערכת כזו תהיה, למשל:

$$\{1, 2, 3, \ldots, \infty, \infty+1, \infty+2, \ldots\}$$

עם הסדר (ופעולת העוקב) הברורים. בדוק!

[16] למי שסקרן כיצד אפשר להגדיר מושג כזה בעזרת הכלים של פאנו נציע את ההגדרה הבאה: $x \geq y$ אם קיימת פונקציה $f : \mathrm{N} \to \mathrm{N}$ המקיימת: $f(1) = y$; $\forall a\,(f(a') = f(a)')$ וקיים $z \in \mathrm{N}$ כך ש $f(z) = x$. מי שרוצה, שינסה להסיק את תכונות הסדר מהגדרה כזו ומהאקסיומות.

השם "לא אינדוקטיבית" למערכות כאלה הוא בהחלט לא מתאים: עבור קבוצה סדורה S (כלומר עם יחס סדר טרנזיטיבי, רפלקסיבי ואנטי־סימטרי - לאלה שמכירים מונחים אלה) התכונה שבכל תת־קבוצה לא ריקה יש איבר קטן ביותר אקוויוולנטית ל"תכונת האינדוקציה" הבאה:

תהי M תת־קבוצה של S בעלת התכונה הבאה:
עבור כל $a \in S$, אם כל איברי S הקטנים מ a שייכים ל M אזי גם $a \in M$
אזי M היא כל S.

את הוכחת האקוויוולנטיות (שמשתמשת ברעיונות דומים להוכחה שבתחילת סעיף זה) נשאיר לקוראים. נציין רק שאין לנו צורך בדרישה $1 \in M$ (1 מסמן, כמובן, את האיבר הקטן ביותר ב S כולה!), כי כיון שקבוצת איברי S הקטנים מ 1 היא $\emptyset$, וכל דבר מתקיים עבור כל איברי הקבוצה הריקה, לכן תמיד כל איברי S הקטנים מ 1 שייכים ל M ולפי הנחתנו נובע מכך ש $1 \in M$.

לקבוצות סדורות בהן בכל תת־קבוצה לא ריקה יש איבר קטן ביותר קוראים, בעקבות גאורג קנטור (Cantor), מייסד תורת הקבוצות, בשם **קבוצות סדורות היטב**, וקנטור הראה שיש להן תפקיד מרכזי בתורת הקבוצות כולה, ואחת הסיבות העיקריות לכך היא האפשרות לעשות אינדוקציה בקבוצה כזו.

ומה בדבר המספרים הממשיים?

אם מחפשים דוגמא למערכת סדורה בה **לא** בכל קבוצה לא ריקה (אפילו אם היא חסומה מלמטה) יש איבר קטן ביותר, מופנית תשומת הלב מייד למספרים הממשיים: קבוצה כמו הקטע $(0,1]$ (אבל נסו להסביר זאת למי שלא למד מתמטיקה!), או כמו $\{1, 1/10, 1/100, 1/1000, \ldots\}$ (זיכרו את פרדוקס "אכילס והצב" של זנון).

אבל המספרים הממשיים מספקים תחליף חלקי (שלא קיים במערכת המספרים הרציונליים!). מי שלמד כיצד בונים את תורת המספרים הממשיים באופן מתמטי מדוייק, שמע על מושגי הסופרמום והאינפימום. המצב הוא שבקבוצה M לא־ריקה של מספרים ממשיים לא חייב להיות איבר קטן ביותר, אבל אם היא חסומה מלמטה, כלומר יש לה חסם תחתון, (חסם תחתון ל M מוגדר כמספר ממשי שהוא קטן או שווה מכל איברי M) אזי **בקבוצת החסמים התחתונים** חייב להיות איבר **גדול** ביותר. החסם התחתון הגדול ביותר נקרא האינפימום (infimum) של M (ומסומן $\inf M$) וזה מה שיש כתחליף לאיבר קטן ביותר ב M. (אם במקרה יש בקבוצה איבר קטן ביותר, אזי איבר זה יהיה האינפימום שלה, מדוע?) בדומה לכך אם M חסומה מלמעלה יש לה חסם עליון קטן ביותר שנקרא הסופרמום (supremum) שלה ומסומן $\sup M$. (דרך אגב, העובדה שקיים תמיד סופרמום אקוויוולנטית לעובדה שקיים תמיד אינפימום, מדוע?)

כתרגיל נציע לקוראים למצוא מהם האינפימום והסופרמום של הקבוצות ללא איבר קטן ביותר שהוזכרו בתחילת סעיף זה. כתרגיל נוסף, נסו למצא קבוצה של מספרים **רציונליים** ש**בקבוצת המספרים הרציונליים** יש לה חסם תחתון אבל אין לה אינפימום.

אבל האם, ברוח מה שנאמר בפרק הקודם, נוכל להפוך את קיום האינפימום לטענת "אינדוקציה"?

ננסה לעשות זאת: תהי M קבוצה שאנו רוצים למצוא תנאים שיבטיחו ש $M = \mathbf{R}$. נתבונן במשלימה $M^c = \mathbf{R} \setminus M$ כלומר בקבוצת כל המספרים הממשיים שאינם ב M. אנו רוצים לכפות עליה להיות ריקה. אנו יודעים שאם M^c אינה ריקה, אזי אם יש לה חסם תחתון אז יש לה אינפימום. בשפה של M: אם יש איבר $a \in \mathbf{R}$ שכל איבר קטן ממנו שייך ל M, אזי קיים איבר $a = \inf M^c$ כך שכל איבר קטן מ a שייך ל M, אבל אין אף $b > a$ כך שכל איבר קטן מ b שייך ל M. לכן התנאים הבאים יבטיחו ש $M^c = \emptyset$ כלומר ש M היא כל $\mathbf{R}$:

טענה 4: תהי M קבוצה של מספרים ממשיים. נניח ש M מקיימת:

א. קיים a ממשי כך שכל מספר ממשי קטן מ a שייך ל M.

ב. אם a הוא כזה שכל מספר ממשי קטן ממנו שייך ל M, אזי קיים $b > a$ בעל אותה תכונה.

אזי M היא קבוצת כל המספרים הממשיים.

כמובן, טענת אינדוקציה זו אינה אלא ניסוח "הפוך" של קיום אינפימום, ולכן מה שאפשר להוכיח בעזרתה אפשר להוכיח גם בעזרת קיום אינפימום, אבל מעניין לפעמים לנסח הוכחות בממשיים בשפת "אינדוקציה" כזו ולהשוות עם הוכחות אנלוגיות בטבעיים.

דוגמא

אם רוצים להוכיח את הטענה הבאה באופן מדוייק, עושים זאת ע"י אינדוקציה:

טענה 5: תהי נתונה סדרת מספרים ממשיים (כלומר נתונה $f : \{0, 1, 2, \ldots\} \to \mathbf{R}$) כך שלכל $k = 1, 2, \ldots$ מתקיים $f(k) \geq f(k-1)$.
אזי לכל $k \in \mathbf{N}$ מתקיים $f(k) \geq f(0)$.

הוכחה: באינדוקציה על k.

נוכיח כעת ב"אינדוקציה ממשית" את הטענה הבאה:

טענה 6: תהי נתונה $f:[0,\infty)\to\mathbf{R}$ רציפה, וכן גזירה מימין בכל $[0,\infty)$ כך שהנגזרת הימנית מקיימת:

$$f'(x)>0 \qquad x\in[0,\infty)$$

אז לכל $x>0$ מתקיים $f(x)\geq f(0)$.

הוכחה:
תהי M קבוצת כל המספרים הממשיים x כך שאו $x\leq 0$ או $f(x)\geq f(0)$. עלינו להוכיח ש $M=\mathbf{R}$ ואת זאת נעשה ב"אינדוקציה": נוכיח ש M מקיימת את א' ו ב' מטענה 4. א' ברור (קחו $a=0$), וב' נובע מהגדרת הנגזרת מימין ומתנאי המשפט: אם $a\geq 0$ ו $x<a\Rightarrow x\in M$, כלומר מתקיים $f(x)\geq f(0)$ עבור $0\leq x<a$, אזי מהרציפות נובע $f(a)\geq f(0)$ ואם בנוסף לכך הגבול

$$\lim_{y\to a^+}\frac{f(y)-f(a)}{y-a}$$

קיים וחיובי, אזי עבור $y>a$ די קרוב ל a יהיה

$$\frac{f(y)-f(a)}{y-a}>0$$

ומכאן $f(y)>f(a)\geq f(0)$. לכן קיים $b>a$ כך ש $y<b\Rightarrow y\in M$, כלומר מתקיים ב'.
מ.ש.ל.

כתרגיל נציע לקוראים להסיק מטענה 6 שאם פונקציה רציפה וגזירה מימין, והנגזרת הימנית אי־שלילית אזי הפונקציה עולה (במובן החלש כלומר לא־יורדת). שימו לב שבהוכחה זו לא משתמשים במשפט לגרנז' על ערך הביניים ובמשפט רול.

כתרגיל נוסף אנו שואלים: האם טענה 4 תישאר נכונה אם נחליף בה "קטן" ב"קטן או שווה"?

האם הסתברות 1 פירושה וודאות ?

הפילוסוף פרנץ ברנטאנו (F. Brentano) ניסה להוכיח במאה ה־19 שתהליך אקראי הוא בלתי אפשרי.[17] את טיעונו ניתן לנסח בצורה הבאה:

נתבונן בתופעה, נניח התפרקות רדיואקטיבית בגוש חומר נתון, שעלולה לקרות בכל רגע t בזמן (ייתכן גם שייקרו מספר התפרקויות ברגעים שונים במשך הזמן). נניח שההסתברות שהתפרקות תקרה בדיוק בזמן t מסויים היא p, ולמען הפשטות נניח שעבור כל t יש אותו p . נוסף לכך אנו מניחים שמה שקורה בזמנים שונים בלתי תלוי במובן של תורת ההסתברות. זה אומר שאם t_1 ו t_2 שני רגעי זמן, אזי יש הסתברות p^2 שבשניהם תהיה התפרקות, הסתברות $(1-p)^2$ שבשניהם לא תהיה התפרקות, הסתברות $p(1-p)$ להתפרקות ברגע הראשון ולא בשני ואותה הסתברות $p(1-p)$ להתפרקות ברגע השני ולא בראשון. בדומה לכך מה שקורה בשלושה רגעים שונים הוא בלתי תלוי, לכן יש הסתברות $(1-p)^3$ שלא תהיה התפרקות באף אחד מ 3 הרגעים. ואותו דבר לגבי n רגעים.

כעת אנו מגיעים למסקנה הבאה: נבדיל בין שתי אפשרויות:

אפשרות א: $p>0$. אז $0<1-p<1$. אם נבחר איזשהם n רגעים, ההסתברות שלא תהיה התפרקות באף אחד מהם היא $(1-p)^n$. גם אם p קטן ולכן $1-p$ קרוב ל 1 , יהיה, אם נקח n גדול מאוד, $(1-p)^n$ קטן כרצוננו. אם E קבוצת רגעי זמן אינסופית, למשל אינטרוול זמן קטן כרצוננו, אזי לכל n טבעי יש קבוצה E_n בת n רגעי זמן המוכלת ב E, ואם לא קרתה התפרקות ב E אזי לא קרתה התפרקות ב E_n וההסתברות לכך קטנה מ $(1-p)^n$. קבלנו שההסתברות שלא היתה התפרקות ב E קטנה מכל אחד מהמספרים $(1-p)^n$, ששואפים לאפס כאשר $n\to\infty$. הסתברות זו היא איפוא 0. לכן בכל אינטרוול קטן כרצוננו יש הסתברות 1, כלומר וודאות, שתהיה התפרקות ־ כלומר יש התפרקות כל הזמן.

אפשרות ב: $p=0$. אז עבור כל רגע זמן t יש הסתברות 0 שיש התפרקות, כלומר התפרקות היא בלתי אפשרית באף רגע זמן ־ לעולם לא תהיה התפרקות.

תהליך אקראי, כמו התפרקות רדיואקטיבית שרואים בטבע, שמתקיימת בקירוב את הנחות האי־תלות שהנחנו, הוא, איפוא, בלתי אפשרי.

עם זאת, יש מודל מתמטי ידוע מאוד שיש לו כל התכונות של אי־תלות שהוזכרו, ושיש בו אקראיות מלאה, ושמשמש, בין השאר, לתאור של התפרקות רדיואקטיבית, שנקרא תהליך פואסון (Poisson). יש מודלים אחרים, בעלי תכונות אחרות, אבל בכולם התהליך אקראי מאוד.

מה קורה, איפוא?

תהליך פואסון מאמץ את הפתרון הבא לבעייתו של ברנטנו: עבור כל t מסויים יש הסתברות 0 שההתפרקות תקרה בדיוק באותו t, **למרות שזה אפשרי**. עבור אינטרוול זמן E יש הסתברות חיובית, תלויה באורך האינטרוול, למאורע שתהיה התפרקות באותו אינטרוול, והיא נתונה ע"י נוסחתו של פואסון (שנותנת גם את ההסתברות שתהיינה k התפרקויות באינטרוול זה, עבור $k=0,1,2,\ldots$). עבור אינטרוולים זרים בעלי אותו אורך יש לנו אי־תלות כמו קודם: אם יש הסתברות p שתהיה התפרקות באחד מהם, אזי יש הסתברות p^2 שתהיינה התפרקויות בשניהם, הסתברות $(1-p)^2$ שלא תהיה התפרקות באף אחד מהם, הסתברות $p(1-p)$ שתהיה התפרקות בראשון ולא בשני ואותה הסתברות שתהיה התפרקות בשני ולא בראשון.

כלומר: כדי שיתגברו על בעיות כמו זו של ברנטנו, היה על המתמטיקאים להרשות תופעה מוזרה: מאורעות בעלי הסתברות 0 שהם אפשריים, מאורעות בעלי הסתברות 1 שאינם הכרחיים (למשל המאורע שלא תהיה התפרקות ברגע זמן נתון t). כמו כן היה עליהם להרשות משפחה של מאורעות, שכל אחד מהם בעל הסתברות 0 אבל ההסתברות שאחד מהם לפחות יקרה היא חיובית, ואפילו 1 (המאורעות שתהיה התפרקות ברגע מסויים t).

אפשר לתת לכך דוגמא הרבה יותר פשוטה: נתון ריבוע בעל צלע 1 ובוחרים באופן מקרי נקודה בריבוע (למשל ע"י כך שנותנים לחלקיק אבק ליפול על הריבוע באופן מקרי). אנו מניחים אחידות: שההסתברות שהנקודה שנבחרה שייכת לחלק E של הריבוע היא שיטחו של E. למשל יש הסתברות $\frac{1}{3}$ שתיבחר נקודה במלבן שהוא השליש העליון של הריבוע; אם נחלק את הריבוע ל 4 ריבועים חופפים ע"י חציתו לאורך ולרוחב, יש הסתברות $\frac{1}{4}$ שתיבחר נקודה בריבוע השמאלי העליון, וכו'. אז עבור כל נקודה P נתונה ההסתברות שהיא תיבחר היא אפס , למרות שאיזושהי נקודה חייבת להיבחר! יתר על כן: ההסתברות שתיבחר נקודה באלכסון הריבוע היא אפס! (כי שיטחו של האלכסון הוא 0).

למעשה המצב של מאורע בעל הסתברות 1, אבל שיכול לא לקרות, הוא לחם חוקו של חוקר ההסתברות. למשל: משפט מרכזי שנקרא "החוק החזק של המספרים הגדולים" אומר בין השאר כך: נניח שאנו מבצעים

[17] ראו: שמואל הוגו ברגמן, מבוא לתורת ההיגיון, מוסד ביאליק 1964, עמ' 367.

ניסוי, ויש הסתברות p שיצליח והסתברות $1-p$ שייכשל, $0<p<1$ (למשל זורקים מטבע לא הוגנת, כאשר צד אחד נחשב להצלחה והשני לכשלון. המטבע תהיה הוגנת אם $p=\frac{1}{2}$). המשפט אומר שמתקיימת טענה שכולנו מצפים לה: שאם נבצע ניסוי כזה באופן בלתי תלוי סדרה אינסופית של פעמים אזי יש הסתברות 1 שהיחס בין מספר ההצלחות ב־N הנסיונות הראשונים ובין מספר הנסיונות N ישאף ל p.

עם זאת, אפשרי גם, למשל, שלא יהיו הצלחות כלל בכל הסדרה האינסופית, כלומר שאיפת היחס ל p אינה ממש הכרחית. קיים גם המצב הבא: כאמור, היחס שואף בהסתברות 1 ל p, אבל זה יכול לקרות ע"י אינסוף אפשרויות שונות של סדרת תוצאות של הנסיונות. לכל סדרה קונקרטית של תוצאות יש הסתברות 0 (את זה אפשר להוכיח בדיוק בעזרת השיקול של ברנטנו ־ הקוראים מוזמנים לנסות). אם כך גם לכך שתתקבל סדרה מסויימת עם יחס ששואף ל p יש הסתברות 0, בדיוק כמו לכך שתתקבל סדרת התוצאות בה אין הצלחות כלל ־ למרות זאת יש הסתברות 1 שתתקבל איזושהי מהסדרות עם יחס שואף ל p.

שתי הדוגמאות האחרונות, בחירת נקודה וביצוע ניסוי אינסוף פעמים, אינן למעשה כה שונות. בחירת נקודה בריבוע היא בחירת שני מספרים x,y בין 0 ו 1 ־ הקואורדינטות של הנקודה. כל מספר כזה אפשר לכתוב כשבר עשרוני אינסופי, כך שלמעשה כל בחירת מספר בין 0 ו 1 היא בחירת סדרה אינסופית של ספרות $0,\ldots,9$, מה שנוכל לראות כחזרה אינסוף פעמים על ניסוי שנותן כתוצאה אחת מ־10 ספרות אלה. הקוראים יכולים לבדוק שנקבל אותן הסתברויות אם נבחר מספר בהסתברות אחידה בקטע $[0,1]$ או אם נבחר את 10 הספרות כך ש: א) לכל ספרה תהיה אותה הסתברות $1/10$; ב) בחירת הספרות השונות תהיה בלתי תלויה.[18] מהחוק החזק של המספרים הגדולים נובע כעת:

משפט (של בורל Borel): אם בוחרים מספר בקטע $[0,1]$ בהסתברות אחידה (כלומר ההסתברות שהמספר שנבחר יהיה בקטע מסוים שוה לאורך הקטע), אזי יש הסתברות 1 שהמספר שנבחר הוא כזה שלכל אחת מ 10 הספרות $k=0,\ldots,9$ האחוז של הסיפרה k בין N הספרות הראשונות של המספר שואף ל 10% כאשר $N\to\infty$.

אין צריך לומר שכמעט כל מספר שנתקלים בו אינו בעל התכונה שהאחוז של כל ספרה שואף ל 10% (למשל כל שבר עשרוני סופי אינו כזה). עד כמה שידוע לי השאלה אם π, למשל, הוא כזה, היא פתוחה ואיש אינו יודע איך לגשת אליה (אפשר לבדוק בין עשרות מיליארדי הספרות הראשונות של π שחושבו, אבל כמובן זה אינו אומר דבר על מה שקורה כאשר $N\to\infty$).

כפי שציננו, המתמטיקאים נאלצו להרשות משפחה אינסופית של מאורעות, שכל אחד בעל הסתברות 0 אבל אחד מהם חייב לקרות. עבור מספר סופי של מאורעות אחת האכסיומות הבסיסיות של תורת ההסתברות גוררת שדבר כזה לא קורה: אם $E_1,E_2,\ldots,E_n$ מאורעות שכל אחד בעל הסתברות 0 אזי גם המאורע "אחד מ $E_1,E_2,\ldots,E_n$ קרה" גם הוא בעל הסתברות 0. זה נכון אפילו לגבי סדרה אינסופית של מאורעות, אבל לא לגבי משפחה של מאורעות E_t שתלויה ב t ממשי, או "רציף" כפי שראינו (למשל $E_t=$ "ההתפרקות קרתה בזמן t" או $E_t=$ "בחרנו את הנקודה t").

לבסוף נאמר, שהמצב שתואר כאן, שנוגד את האינטואיציה, קיים במרחב מדגם שהוא "אינסופי מאוד" או "רציף". (מרחב המדגם הוא קבוצת כל התוצאות האפשריות. למשל בזריקה אחת של קוביה מרחב המדגם הוא קבוצת שש התוצאות האפשריות. במקרה של ביצוע ניסוי אינסוף פעמים מרחב המדגם הוא קבוצת כל הסדרות האפשריות של תוצאות. במקרה של ההתפרקות הרדיואקטיבית מרחב המדגם הוא אוסף כל הקבוצות האפשריות של רגעי הזמן בהם היתה התפרקות. מהו מרחב המדגם במקרה של בחירת נקודה בריבוע?) מרחבים "אינסופיים מאוד" כאלה הם תמיד אידיאליזציה של ה"טבע הנסיוני" (מי יכול לצייר נקודה חסרת גודל? מי יכול לבצע ניסוי אינסוף פעמים? בטבע התפרקות רדיואקטיבית לוקחת זמן ואינה קורית ב"נקודת זמן מדוייקת"). מתברר שכאשר דנים במרחב מדגם סופי (למשל בשאלות המפורסמות על זריקות קוביה, הוצאת כדורים משק וכד'. זה נכון גם אם מרחב המדגם הוא בן־מניה) אנו פטורים מצרה זו בכלל, ונוכל להניח בשקט שרק המאורע הריק, כלומר שהוא בלתי אפשרי, הוא בעל הסתברות 0 ולכן הסתברות 1 פירושה ודאות ממש.

[18] יש אמנם הבדל קטן: שבר עשרוני סופי כמו 0.368 מתקבל גם מבחירת הספרות $0.3680000\ldots$ וגם מבחירת ספרות אחרת $0.3679999\ldots$. אבל ההסתברות שייבחר מספר שהוא שבר עשרוני סופי היא 0 ולכן אפשר להתעלם מתופעה זו.

עולם השניים

ננסה להסתפק רק בשני מספרים: 0 ו 1. שאר המספרים לא יהיו קיימים. לעולם המורכב משני מספרים אלה נקרא: "עולם השניים" או בקיצור: "עולם ה 2".

פעולת הכפל אינה בעיה: היא אינה מוציאה אותנו מהקבוצה $\{0,1\}$. אבל בקבוצה זו יש לה פירוש מיוחד: אם מסדרים את 0 ו 1 בסדר הרגיל: $0 < 1$, אזי הכפל הוא אותו דבר כמו פעולת המינימום (בידקו!):

$$xy = \min(x,y) \quad x,y = 0,1$$

באשר לחיבור, הבעיה היחידה היא ש $1+1$ אינו מוגדר (2 לא קיים!). יש שתי אפשרויות: להגדיר אותו כ 0 או כ 1.

במקום לבחור בין שתי אפשרויות אלה, נאמץ את שתיהן: הפעולה שנותנת 1 אינה אלא פעולת המכסימום $\max(x,y)$ ואת הפעולה שנותנת 0 נסמן ב $+$. זהו החיבור מודולו 2, כלומר הוא נותן, אם לא את התוצאה הרגילה, לפחות את השארית בחלוקת תוצאה זו ב 2.

יש לנו כעת שלש פעולות. לשם נוחיות נשתמש בסימון (המקובל לפעמים במתמטיקה):

$$x \wedge y = \min(x,y) = xy \quad x \vee y = \max(x,y)$$

ולא לחינם דומים סימנים אלה לסימני החיתוך והאיחוד של קבוצות: נקח קבוצה לא ריקה כלשהי S (אפשר להסתפק בקבוצה בת איבר אחד) ונזהה את 0 עם הקבוצה הריקה $\emptyset$ ואת 1 עם ה"שלם" S. בידקו שהפעולות $\wedge$ ב $\vee$ מתאימות בדיוק לפעולות החיתוך והאיחוד.

שני ה"חיבורים" שלנו $+$ ו $\vee$ מקיימים את חוקי הפעולות: החוק הקומוטטיבי (חוק החילוף), החוק האסוציאטיבי (חוק הקיבוץ) ויחד עם הכפל את החוק הדיסטריבוטיבי (חוק הפילוג). אפשר להוכיח זאת ע"י בדיקת כל האפשרויות, או להעזר ב"פירושים" שנתננו לשני החיבורים שלנו: החיבור $+$ (וכן הכפל) נותנים תמיד את התוצאה הרגילה מודולו 2, ולחיבור $\vee$ עם הכפל ישנו הפירוש של הקבוצות S ו $\emptyset$.

החיסור וחוקי השדה

נתמקד כעת בפעולה שנותנת $1+1=0$.

סיבה טובה לסמן דווקא את הפעולה שנותנת $1+1=0$ ב $+$ היא שיש לנו פעולה הפוכה - חיסור, כלומר עבור כל a ו b בעולם ה 2 יש למשוואה:

$$x + a = b$$

פתרון יחיד x, שראוי לסמנו ב $x = b - a$. אין זה כך לגבי המשוואה $x \vee a = b$ (מיצאו דוגמה!). אבל החיסור קצת מאכזב - הוא אותו דבר כמו החיבור: בידקו שעבור כל a ו b:

$$a - b = a + b$$

למעשה היינו צריכים לצפות לכך, כי אם ישנם לנו חוקי הפעולות, אזי מכך ש $1+1=0$ נובע (כיצד?) $-1 = 1$ ולכן $a - b = a + (-1)b = a + 1 \cdot b = a + b$.

אין גם בעיה עם החילוק - חילוק ב 0 אסור וחילוק ב 1 בודאי לא מעורר בעיות.

עולם ה 2 הוא איפוא **שדה**, כלומר יש בו ארבע פעולות חשבון (בחילוק מניחים שהמחלק שונה מ 0) שמקיימות את החוקים הפורמליים הבאים, כמו במספרים הרציונליים או הממשיים: החוקים הקומוטטיביים והאסוציאטיביים (חוקי החילוף והקיבוץ) של החיבור ושל הכפל, החוק הדיסטריבוטיבי (חוק הפילוג) של הכפל והחיבור, העובדה שחיסור הופך חיבור וחילוק הופך כפל, והתכונות של איברים מיוחדים הנקראים 0 ו 1: ש $x \cdot 1 = x$ עבור כל x בקבוצה, ומזה מוכיחים ש $x + 0 = x$ ו $0 \cdot x = 0$. (כל מערכת בה מוגדרות ארבע פעולות חשבון בדרך כלשהי, אבל כך שמתקיימים חוקים

פורמליים אלה, נקראת **שדה**, ולחוקים פורמליים אלה קוראים: **חוקי השדה**. כך המספרים הרציונליים הם שדה, המספרים הממשיים הם שדה והמספרים הקומפלקסיים הם שדה, הכל עם הפעולות הרגילות, אבל המספרים הטבעיים או השלמים אינם שדה.) בניגוד למספרים הטבעיים, לא היינו צריכים לשם כך להוסיף שום דבר - לא "שליליים" ולא "שברים".

כאמור, שדה זה הוא שדה השאריות מודולו 2. אפשר להוכיח, בעזרת משפטים מתורת המספרים, שעבור כל p ראשוני השאריות מודולו p עם החיבור והכפל מודולו p מהווים שדה (אבל לא עבור p פריק). דרך אגב, עבור $p = 2$, ולא עבור $p > 2$, הכפל מודולו 2 הוא הכפל הרגיל בלי שינוי.

חזקות ו"קומפלקסיים"

ומה בדבר אלגברה ממעלה שניה, שלישית וכו'?

נרשום x^2, x^3 וכו' כקיצורים ל $x \cdot x$, $x \cdot x \cdot x$ וכו'. (נדגיש שזה רק סימון ואין אנו מכניסים, חס וחלילה, את ה"מספרים" הלא קיימים $.2, 3, \ldots$) חזקות אלה מרשימות באמת, שכן מתקיים:

$$x^n = x$$

ולכן לא קשה להוציא שרשים.

בכל זאת יש משוואות ריבועיות שאין להן פתרון. למעשה יש משוואה אחת כזו:

$$x^2 + x + 1 = 0$$

(בידקו! מדוע אי אפשר לפתור משוואה זו בשיטה הרגילה של השלמה לריבוע?)

וזה מאפשר לנו למצוא, סוף סוף, סיבה להרחיב את עולם ה 2. נזכיר שהעובדה שבין המספרים הממשיים אין פתרון למשוואה $x^2 + 1 = 0$, כלומר $x^2 = -1$ דחפה אותנו לבנות את המספרים הקומפלקסיים (מרוכבים), ע"י "הוספת" מספר חדש i שהוא דווקא פתרון למשוואה $x^2 + 1 = 0$; התברר לנו שאפשר להוסיף i כזה, ואת כל ה"מספרים החדשים" שמתבקשים ממנו, בלי סתירה, ולקבל הרחבה של הממשיים שקראנו לאיבריה: המספרים הקומפלקסיים. ננסה, בדומה לכך, להוסיף לעולם ה 2 מספר j שיקיים

$$j^2 + j + 1 = 0$$

עם דרישה שהחוקים הפורמליים (חוקי השדה) של הפעולות יישמרו.

תופעה ראשונה שתתגלה לנו היא שחזקות כבר אינן טריביאליות, בפרט $j^2 \neq j$. זה מוכרח להיות, כי:

$$j^2 = j^2 + 0 = j^2 + (j^2 + j + 1) = (1 + 1)j^2 + j + 1 = 0 \cdot j^2 + j + 1 = j + 1$$

ו $j + 1 \neq j$ כי אחרת היה $1 = 0$.

ומקבלים קבוצה של ארבעה איברים שונים $\{0, 1, j, j + 1\}$ (מדוע כולם חייבים להיות שונים?) ואם אנו רוצים שיישמרו חוקי הפעולות, חייבים לוחות החיבור והכפל שלהם להיות:

$+$	0	1	j	$j+1$
0	0	1	j	$j+1$
1	1	0	$j+1$	j
j	j	$j+1$	0	1
$j+1$	$j+1$	j	1	0

$\times$	0	1	j	$j+1$
0	0	0	0	0
1	0	1	j	$j+1$
j	0	j	$j+1$	1
$j+1$	0	$j+1$	1	j

כתרגיל בנו את לוח החילוק (מה בדבר החיסור?).

רואים שחוקי השדה אינם מצריכים הוספת איברים נוספים על ארבעה אלה. עם זאת לא הוכחנו שבהמשך החישוב לא נקבל סתירה מהפעלת חוקי הפעולות ב"עולם הארבעה" זה. אפשר להוכיח זאת, וכן להוכיח

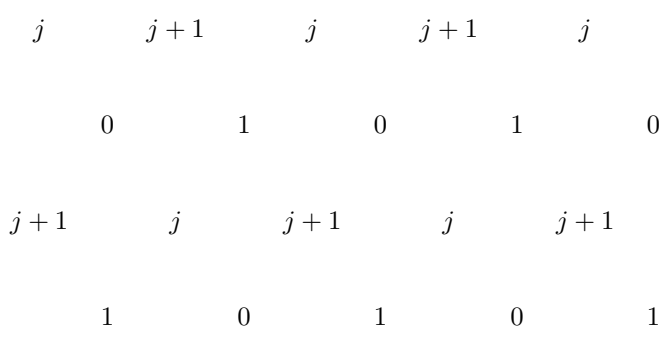
שקבוצת ארבעה אברים אלה היא שדה (כלומר שחוקי השדה מתקיימים תמיד). קבוצה זו מתייחסת לעולם ה 2 בערך כמו שהמספרים הקומפלקסיים (מרוכבים) מתייחסים למספרים הממשיים.

דרך אגב, מהו הפתרון השני של המשוואה הריבועית $x^2+x+1=0$? (מדוע בכלל יש לצפות שיהיו לה שני פתרונות?)

נוכל להשתכנע יותר בחוסר הסתירה שב"עולם הארבעה" $\{0,1,j,j+1\}$ אם נתאר אותו באופן גיאומטרי, כמו שעושים לגבי המספרים הקומפלקסיים.

נתחיל מ"ציר המספרים של עולם ה 2". זהו ציר הנקודות השלמות, בו במקום לסמן את הנקודות ב $\ldots,-2,-1,0,1,2,\ldots$ כמקובל, נסמן אותן לסירוגין ב $\ldots,1,0,1,0,1,0,\ldots$. ציר זה יתייחס לפעולות בעולם ה 2 כמו שציר המספרים הרגיל מתייחס לפעולות בממשיים: למשל כדי לחבר $1+1$ בודקים איזו הזזה מזיזה את הנקודה 0 לנקודה 1 (לא חשוב איזו נקודה 0 ואיזו נקודה 1) ורואים לאן היא מזיזה את הנקודה 1. כדי להכפיל $1\cdot 1$ מנפחים את הציר, תוך השארת הנקודה 0 במקום והעברת הנקודה 1 לנקודה 1 ורואים לאן עברה הנקודה 1 (כמו שבממשיים הרגילים כדי להכפיל 2.5×3.2 מנפחים את הציר, תוך השארת 0 במקום, כך ש 1 תעבור ל 2.5 ורואים לאן עוברת 3.2 וזה $2.5\times 3.2=8$).

כדי לקבל את "מישור המספרים של עולם הארבעה" עוברים למישור ובמקום ציר השלמים בישר (שאפשר לקבלו מהקדקדים של ריצוף הישר בקטעים שווים), מרצפים את המישור במשולשים שווי צלעות, לוקחים את קבוצת הקדקדים, ורושמים בהם את ארבעת אברי הקבוצה $\{0,1,j,j+1\}$ לסרוגין:



הפעולות מתוארות באופן גיאומטרי כמו במספרים הקומפלקסיים: בשביל לחבר $a+b$ לוקחים את ההזזה המקבילה של המישור שמעתיקה $0\mapsto a$ ורואים לאן היא מעתיקה את b, וזה $a+b$ (לא חשוב איזו נקודה a, איזו נקודה b $\ldots$). בשביל להכפיל $a\cdot b$, בודקים איזה סיבוב סביב 0 עם ניפוח המישור מהמרכז 0 מעתיק $1\mapsto a$ ורואים לאן הוא מעתיק את b, וזה $a\cdot b$.

אפשר להשתמש במודל גיאומטרי זה, או אם תרצו, בבניית "עולם הארבעה" כאשר יוצאים מהקומפלקסים (הבהירו!) כדי להוכיח את חוקי השדה ב"עולם הארבעה" $\{0,1,j,j+1\}$.

"עולם הארבעה" הוא דוגמא לשדה המכיל בתוכו את עולם ה 2. כמו שראינו, בשדה כזה החזקות כבר לא טריביאליות: $j^2=j+1\neq j$ לכן מתחיל להיות מענין לחקור את האלגברה של החזקות. ואכן, בכל שדה שמכיל את עולם ה 2 מתקיים:

$$-x=0\cdot x-x=(1+1)x-x=x$$
$$x-y=x+(-y)=x+y$$
$$(x+y)^2=x^2+xy+yx+y^2=x^2+(1+1)xy+y^2=x^2+0\cdot xy+y^2=x^2+y^2$$

וכיוון שגם $(xy)^2=x^2y^2$, אפשר להוכיח מכאן שעבור כל פולינום $P(x,y,z,\ldots)$ עם מקדמים 0 או 1 מתקיים:

$$(P(x,y,z,\ldots))^2=P(x^2,y^2,z^2,\ldots)$$

(למי שמכיר מונח זה: ההעתקה $f(x)=x^2$ היא הומומורפיזם מהשדה לעצמו.)

נעיר, שיש הבדל חשוב בין "עולם הארבעה" לבין הקומפלקסיים הרגילים - הארבעה $\{0,1,j,j+1\}$ לא מקיימים את ה"משפט היסודי של האלגברה", האומר שבין הקומפלקסים יש פתרון לכל משוואה אלגברית. למשל למשוואה

$$x^3+x+1=0$$

אין פתרון ב"עולם הארבעה" (בידקו!).

לכן אפשר "להוסיף" "מספר" u כך ש

$$u^3 + u + 1 = 0$$

ולקבל שדות אחרים המכילים את "עולם ה 2".

נושא זה נחקר בענף של האלגברה שנקרא תורת השדות הסופיים, בהם מופיעים בדרך כלל מספרים ראשוניים p כלשהם במקום 2. שדות אלה נקראים גם "שדות גלואה", כי כבר המתמטיקאי הידוע גלואה (E. Galois, 1811-1832), שנהרג בדו־קרב בגיל 21, חקר אותם בעודו נער.

(אפשר להוכיח שאם n טבעי, אזי: אם $n > 1$ הוא חזקת ראשוני, כמו $2 = 2^1$, $4 = 2^2$, $8 = 2^3$, $3 = 3^1$, $25 = 5^2$, אזי קיים שדה בעל n איברים, וכל שני שדות כאלה הם איזומורפיים, כלומר נבדלים רק בשמות שנותנים לאיברים, כך שלמעשה יש רק שדה אחד בעל n איברים. אם n אינו חזקת ראשוני (למשל אם $n = 6$ או $n = 10$) אזי לא ייתכן שדה בעל n איברים.)

תרגיל: האם אפשר להרחיב, בצורה טובה, את יחס הסדר $0 < 1$ ואת הפעולות $\wedge$ ו $\vee$ מעולם ה 2 $\{0,1\}$ ל"עולם הארבעה" $\{0,1,j,j+1\}$?

עוד על גיאומטריה

אפשר להסתכל על הקשר בין עולם ה 2 לבין גיאומטריה גם אחרת: המישור או המרחב מתוארים בעזרת גיאומטריה אנליטית תוך שימוש במספרים ממשיים. אם נחקה את הגיאומטריה האנליטית, אבל במקום מספרים ממשיים נקח את איבריו של שדה אחר, למשל שדה הקומפלקסים, או שדה סופי, במיוחד עולם ה 2, נקבל מערכת נקודות, ישרים וכו' שיהיה לה דמיון מה למישור או המרחב "הרגילים" אבל גם שוני. במקרה של שדה סופי יהיה ב"מישור" או "מרחב" כזה מספר **סופי** של נקודות.

כדוגמה נקח את המרחב (התלת ממדי) מעל עולם ה 2. יש בו 8 נקודות, המתוארות ע"י הקואורדינטות שלהן (לשם קיצור נכתוב, למשל, 011 במקום $(0,1,1)$):

$$000\ 001\ 010\ 011\ 100\ 101\ 110\ 111$$

נתמקד ב**מישורים דרך 000**. הם נתונים ע"י משוואות ממעלה ראשונה ב x, y, z ללא איבר חפשי ויש 7 מישורים כאלה. כולם מכילים את 000 ובכל מישור כזה יש 4 נקודות. הנקודות שאינן 000 בשבעת המישורים הן:

$x = 0$	001	010	011
$y = 0$	001	100	101
$z = 0$	010	100	110
$x + y = 0$	110	001	111
$x + z = 0$	101	010	111
$y + z = 0$	011	100	111
$x + y + z = 0$	011	101	110

הישרים דרך 000 מכילים כל אחד 2 נקודות, שאחת מהן 000. לכן הם נמצאים בהתאמה חד־חד ערכית עם 7 הנקודות שאינן 000.

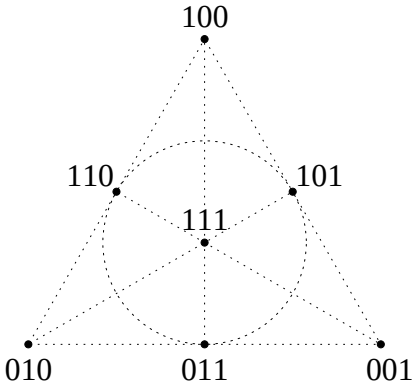
אם נחליט לקרוא רק ל 7 הנקודות שאינן 000 בשם **נקודות**, ולשלישיות הנקודות שאינן 000 על 7 המישורים בשם "**ישרים**" דווקא, אזי מהעובדה שהם התקבלו ממישורים וישרים בגאומטריה האנליטית המרחבית שלנו נובע שדרך כל שתים מהנקודות עובר ישר יחיד וכל שני ישרים נחתכים בנקודה (כי דרך כל שני ישרים שונים העוברים דרך הראשית עובר מישור יחיד, וכל שני מישורים שונים המכילים את הראשית נחתכים בישר).

קבלנו "מישור פרוייקטיבי סופי": בקומבינטריקה קוראים בשם "מישור פרוייקטיבי" לקונפיגורציה כלשהי הבנויה מקבוצה כלשהי שאיבריה נקראים "נקודות" עם משפחת תת־קבוצות שנחשבות ל"ישרים" כך שדרך כל שתי נקודות עובר ישר יחיד וכל שני ישרים נחתכים בנקודה - אין "ישרים מקבילים". כאשר מדובר

במישור פרוייקטיבי סופי, דורשים גם שמספר הנקודות על כל ישר יהיה קבוע, ושווה גם למספר הישרים העוברים דרך כל נקודה.

למי שמכיר: דוגמא למישור פרוייקטיבי אינסופי נותנת הגאומטריה הפרוייקטיבית המישורית ה"רגילה". אם תרצו, בגאומטריה זו הנקודות הם הישרים דרך הראשית במרחב תלת ממדי רגיל, והישרים הם המישורים דרך הראשית. (באופן דומה אפשר לבנות "מישור פרוייקטיבי" בעזרת כל שדה שהוא).

עולם ה 2 נתן לנו, איפוא, מישור פרוייקטיבי בעל 7 נקודות עם 7 ישרים בעלי 3 נקודות כל אחד. נתאר אותו בציור (אנו מוכרחים לצייר אחד ה"ישרים" כמעגל):



אפשר להוכיח שאי אפשר לצייר מישור פרוייקטיבי זה (או איזשהו מישור פרויקטיבי סופי אחר עם "ישר" המכיל יותר משתי נקודות) במישור "הרגיל" בלי לעקם חלק מה"ישרים". זה נובע מהמשפט האומר, שבהינתן n נקודות במישור, לא כולן על ישר אחד, קיים תמיד ישר העובר דרך שתיים מהן בדיוק - ראו: פ. ארדש: על בעיות אחדות בגיאומטריה אלמנטרית, אתגר-גליונות מתמטיקה 40-39, נובמבר 1996.

הפעולות $\wedge$ ו $\vee$

נחזור לעולם ה 2 עצמו ונעבור לפעולת "החיבור" השנייה $\vee$, כלומר המכסימום, ונסמן עכשיו את הכפל ב $\wedge$, כלומר המינימום. כאמור, לפעולת "חיבור" זו אין חיסור, אבל יש לה תכונה אחרת:

יש שתי תמורות בקבוצה $\{0,1\}$: תמורת הזהות $(1 \mapsto 1, 0 \mapsto 0)$ והתמורה $(1 \mapsto 0, 0 \mapsto 1)$. נסמן תמורה אחרונה זו ב $x \mapsto x'$. כך:

$$0' = 1, \quad 1' = 0$$

אם רוצים "נוסחה", יש לנו:

$$x' = x + 1$$

(מדוע?)

בעוד שפעולות חשבון רגילות הן **בינריות** - נותנות תוצאה משני מספרים או עצמים אחרים, $'$ היא פעולה **אונרית** - נותנת תוצאה מ"מספר" אחד.

במודל של $1 = S$ קבוצה ו $0 =$ הקבוצה הריקה $\emptyset$, יש ל $'$ משמעות של פעולת לקיחת **המשלים** של קבוצה.

וקל לראות שהתמורה הלא-טריביאלית היחידה הזו הופכת $\vee \leftrightarrow \wedge$, $\leq \leftrightarrow \geq$ (וכמובן $0 \leftrightarrow 1$), כלומר מתקיימים חוקי דה-מורגן:

$$(x \wedge y)' = x' \vee y'$$
$$(x \vee y)' = x' \wedge y'$$

וכן

$$x \leq y \Leftrightarrow x' \geq y'$$

(וכמובן מתקיים החוק $(x')' = x$.)

ומכך נובעת הטענה המטא־מתמטית הבאה:

עקרון דואליות: אם טענה כלשהי נכונה בעולם השניים, אזי הטענה המתקבלת ממנה ע"י החלפת:

$$\wedge \leftrightarrow \vee \quad 0 \leftrightarrow 1 \quad <\leftrightarrow>$$

נכונה גם היא.

הוכחה בערך: הטענה המתקבלת ע"י ההחלפה היא הטענה המקורית, רק שבכל מקום מציבים x' במקום x. (הבהירו).

למשל, הזכרנו קודם של"חיבור" $\vee$ יש חוק דיסטריבוטיבי (חוק פילוג) עם הכפל $(c\ b\ a)$ בעולם ה 2):

$$a(b \vee c) = (ab) \vee (ac)$$

או, בסימון הכפל ע"י $\wedge$:

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$$

בהתאם לעקרון הדואליות אנו מקבלים חוק דיסטריבוטיבי דואלי

$$a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$

כדי להוכיח חוק זה מהחוק הקודם באופן ישיר, משתמשים בהצבות $'$ ובחוקי דה־מורגן שהובאו לעיל:

$$a \vee (b \wedge c) = a \vee (b' \vee c')' = (a' \wedge (b' \vee c'))' =$$
$$= ((a' \wedge b') \vee (a' \wedge c'))' = (a' \wedge b')' \wedge (a' \wedge c')' = (a \vee b) \wedge (a \vee c)$$

אלגבראות בוליאניות

בעולם ה 2 יש לנו מבנה של הפעולות $\wedge$ (הכפל), $\vee$, הפעולה האונרית $'$ ויש גם נוסחאות מיוחדות שמקיימים 0 ו 1:

$$0 \wedge x = 0 \quad 0 \vee x = x$$
$$1 \vee x = 1 \quad 1 \wedge x = x$$
$$x \wedge x' = 0 \quad x \vee x' = 1$$

לקוראים רבים ידוע בוודאי שלפעולות אלו בעולם ה 2 $\{0,1\}$ (שנקראות: פעולות **בוליאניות** על שם G. Boole שפיתח את ה"אלגברה של הלוגיקה" במאה ה־19) אפשר לתת פירוש בלוגיקה: אברי עולם ה 2 הן טענות, כמו "הירח נראה עגול". כל הטענות האמיתיות הן בעלות ערך האמת "אמיתי" המזוהה עם 1 והטענות השקריות הן בעלות ערך האמת "שקרי", המזוהה עם 0.

הפעולה האונרית $'$ תהיה אז פעולת השלילה: (הירח נראה עגול)$'$ היא הטענה: הירח לא נראה עגול, שהיא, כמובן, אמיתית אם הטענה: "הירח נראה עגול" שיקרית ושיקרית אם הטענה: "הירח נראה עגול" אמיתית.

הפעולה $\vee$ מחברת טענות ע"י "ו": (הירח נראה עגול)$\wedge$(השעה היא חצות) היא הטענה: "הירח נראה עגול והשעה היא חצות". בידקו שערך האמת שלה אכן נתון ע"י $\wedge$ של ערכי האמת של שתי הטענות.

הפעולה $\wedge$ מחברת טענות ע"י "או" (במובן של או ... או ... או שניהם): "הירח נראה עגול או השעה היא חצות (או שניהם)". בידקו את ערכי האמת.

בשפות תיכנות, כמו פסקל, יש משתנים בוליאניים, שיכולים לקבל את שני הערכים "אמיתי" ו"שקרי" והשפה מכילה את הפעולות הבוליאניות במשתנים כאלה. הם משמשים לתיאור תנאים לוגיים.

בתורת המיתוג של מעגלים חשמליים משתמשים בשערים לוגיים, שמתאימים לפעולות הבוליאניות, וגורמים לכך שהטענה "יש סיגנל ב z" תהיה פעולה בוליאנית נתונה בטענות: "יש סיגנל ב x" ו"יש סיגנל ב y".

אבל בכל הדיון הזה בפעולות הלוגיות, למרות שטענו שאנו מדברים על עולם ה 2, יש לנו למעשה קבוצה יותר גדולה בה מוגדרות הפעולות הבוליאניות והתקיימו החוקים שלהן (קומוטטיביות, אסוציאטיביות, דיסטריבוטיביות, חוקי דה־מורגן וכו'). כי טענה כמו "הירח נראה עגול" אינה שקולה לאחד משני ערכי האמת "אמתי" או "שקרי" כלומר לאחד משני האיברים 0 ו 1 בעולם ה 2. טענה כזו לפעמים נכונה (בלילה, באמצע חודש הירח, כלומר למשל החודש העברי או המוסלמי), ולפעמים לא. עם זאת עדיין נוכל להסכים שהטענה: "הירח נראה עגול" שקולה ל: "לא נכון שהירח לא נראה עגול" בהתאם לחוק הבוליאני $(x')' = x$. ויהיו טענות שיהיו אכן שקולות ל"אמיתי", כלומר אמיתיות באופן לוגי, כמו: "הירח נראה עגול

או הירח לא נראה עגול", בהתאם לחוק הבוליאני $x \vee x' = 1$. בדומה לכך החוק $x \wedge x' = 0$ יגרור שהטענה "הירח נראה עגול והירח לא נראה עגול" שקרית באופן לוגי, כלומר שקולה ל"שקרי".

מערכת הטענות היא, איפוא, דוגמא להרחבה של עולם ה 2 שנקראת **אלגברה בוליאנית** (שימו לב לשימוש, הנראה מוזר אך מקובל במתמטיקה, במלה "אלגברה" לציון קבוצה עם פעולות):

הגדרה: **אלגברה בוליאנית** היא קבוצה B בה מוגדרות, באופן כלשהו, פעולות בינריות $\wedge$ ו $\vee$, פעולה אונרית $'$ ויש בה שני אברים מיוחדים הנקראים 0 ו 1 (לכן היא מכילה את עולם ה 2) וכך שכל משוואה עם אגפים הבנויים מהסימנים $\wedge$, $\vee$, $'$, 0, 1 ומסוגריים, ושמתקיימת זהותית בעולם ה 2, מתקיימת זהותית גם ב B.

מתברר שאפשר להוכיח (באופן לא פשוט) שכדי שתנאי זה יתקיים די לדרוש שיתקיימו ב B החוקים הבוליאניים שהוזכרו במאמר זה.

בנוסף לאלגברה הבוליאנית של הטענות, דוגמא מוכרת לקוראים רבים היא האלגברה הבוליאנית של התת־קבוצות של קבוצה S, עם $\vee =$ איחוד, $\wedge =$ חיתוך, $' =$ משלים, $1 =$ הקבוצה "השלמה" S ו $0 =$ הקבוצה הריקה.

אפשר לקשר בין שתי דוגמאות אלו: מערכת טענות לוגיות כמו "הירח נראה עגול" שהן לעתים נכונות ולעתים לא, נוח לתאר ע"י קבוצה S של "עולמות אפשריים", כאשר בעולם אפשרי נקבע עבור כל טענה כזו אם היא נכונה או לא. למשל עבור טענות כמו "הירח נראה עגול" ו"עכשיו שעת חצות" אפשר לתאר את מערכת העולמות האפשריים ע"י קואורדינטות שמציינות את: היום בחודש הירח, השעה, הקביעה אם השמש כבר שקעה. כעת נזהה כל טענה עם קבוצת העולמות האפשריים בהם היא נכונה (ושתי טענות ייחשבו לשקולות אם קבוצות אלו שוות לגביהן). ע"י כך יתאימו הפעולות הבוליאניות בטענות לפעולות הבוליאניות בקבוצות (בידקו!).

ואם כבר התחלנו לקשר אלגבראות בוליאניות שונות זו לזו, נתבונן בפונקציות המוגדרות על קבוצה כלשהי S, אבל במקום ערך מספרי הן מקבלות ערך בעולם ה 2, או, מה שהוא אותו דבר - בעולם ערכי האמת "אמיתי" ו"שקרי". בפונקציות אלו נעשה את הפעולות הבוליאניות כמו שעושים פעולות אלגבריות בפונקציות מספריות - פועלים על ערכי הפונקציה בכל מקום. נקבל אלגברה בוליאנית של פונקציות. האם קיבלנו משהו חדש? (היעזרו בכך כדי להוכיח שהתת־קבוצות של קבוצה אכן מקיימות את הדרישות של ההגדרה של אלגברה בוליאנית...)

ולסיום: מאחר שבאלגברה בוליאנית מתקיימות כל הזהויות שמתקיימות בעולם ה 2, אפשר להגדיר בכל אלגברה כזו, למשל את היחס $\leq$ (שבעולם ה 2 היה יחס הסדר הרגיל בין 0 ו 1). אכן, בכל אלגברה בוליאנית מתקיים (כי בעולם ה 2 זה מתקיים):

$$(x \vee y) \wedge x = x \qquad (x \wedge y) \vee x = x$$

ומכאן נובע ש $x \vee y = y \Leftrightarrow x \wedge y = x$ ואם זה מתקיים, נגדיר ש $x \leq y$ (בידקו עם 0 ו 1). באלגברה הבוליאנית של הקבוצות זהו היחס: "הקבוצה x מוכלת (או שוה) לקבוצה y". מהי משמעות יחס זה באלגברה הבוליאנית של הטענות?

ואפילו את החיבור $+$ נוכל להגדיר באלגברה בוליאנית כלשהי, כי בעולם ה 2 הוא נתון ע"י (בידקו):

$$x + y = (x \wedge y') \vee (x' \wedge y)$$

ובכל אלגברה בוליאנית יתקיים החוק הדיסטריבוטיבי של החיבור $+$ והכפל $\wedge$ (כי זה מתקיים בעולם ה 2):

$$x \wedge (y + z) = x \wedge y + x \wedge z$$

מהי משמעות פעולת $+$ באלגברה הבוליאנית של הקבוצות? האם אלגברה בוליאנית עם פעולות החיבור והכפל האלה היא שדה? (בידקו לגבי האלגברה הבוליאנית של התת־קבוצות של קבוצה סופית).

בעיית פרמה נפתרה סוף סוף ע"י אנדרו ויילס

[19] ב 23.6.93 קיבלו המתמטיקאים ברחבי העולם הודעות נרגשות בדואר אלקטרוני: בעיית פרמה נפתרה! Andrew J. Wiles , מתמטיקאי אנגלי בן 40 הנמצא כיום באוניברסיטת פרינסטון בארצות הברית, נתן באותו שבוע סידרת הרצאות באוניברסיטת קימבריג', אנגליה, על עקומים אליפטיים, תבניות מודולריות והצגות גלואה, בלי להודיע מראש מה הוא טומן באמתחתו, ובסוף ציין שמסקנה צדדית ממשפט שהוכיח היא המשפט האחרון של פרמה. רק קומץ משומעי ההרצאות חשדו שבכך מדובר. בחודשים שעברו מאז שככה קצת ההתלהבות. ויילס כתב את הוכחתו ב־200 עמודים, ובדיקות ע"י מומחים כבר גילו שיש בה פערים, כלומר שיקולים שצריך להשלים, ורק אם הם יושלמו לשביעות רצון המבינים בדבר תוכל קהיליית המתמטיקאים לאמר שמשפט פרמה הוכח[20] (המתמטיקאי היפני Yoichi Miyaoka הציג ב־1988 "הוכחה" למשפט, ובדיקה העלתה שהוכחתו פגומה. כל זה מעורר הרהורים פילוסופיים על הקשר בין הוכחות מתמטיות וידיעה בטוחה). עם זאת, אין ספק שבעבודתו של ויילס יש תוצאות חשובות ולא פסה התקווה שהפערים יושלמו ותהיה בידנו הוכחה למשפט פרמה. בכל מקרה, כדאי להציץ בנושאים המתמטיים שסביב הסיפור.

רבים מהקוראים יודעים בוודאי מה טוען משפט (לשעבר השערת) פרמה: שאם n טבעי גדול מ 2 לא קיימים מספרים שלמים חיוביים X, Y, Z כך ש $X^n + Y^n = Z^n$. טענה זו כה פשוטה שכל תלמיד בית־ספר יכול להבין מה היא אומרת, ולמרות זאת במשך 350 שנה ניסו המתמטיקאים ללא הצלחה להוכיח או להפריך אותה. אבל אפילו בהשוואה לבעיות אחרות בתורת המספרים שקל לנסחן והן פתוחות עד היום, אין כמו בעיית פרמה מבחינת ההיסטוריה המיוחדת שלה.

היסטוריה זו מלאה עובדות פיקנטיות: פייר דה פרמה (Pierre de Fermat) $(1601 - 1665)$, שלמרות שהיה מהמתמטיקאים החשובים ביותר במאה ה־17 היה במקצועו שופט, התעניין מאוד בתורת המספרים. בשם זה קוראים לענף המתמטיקה העוסק בשאלות בהם חשוב לנו אם מספר הוא שלם, רציונלי וכד'. פרמה מצא תוצאות בסיסיות בשטח זה, אבל אם הוא ה"אבא" של תורת המספרים, אזי ה"סבא" היה דיופנטוס (Diophantos), מתמטיקאי יווני מאלכסנדריה במאה השלישית. הוא היה מיוחד בין היוונים בכך שעסק בעיקר באלגברה ולא בגיאומטריה. בסיפרו Arithmetica דן בבעיית מציאת פתרונות שלמים (או רציונליים) למשוואות (כיום נקראות משוואות בהן מחפשים פתרונות שלמים בשם: משוואות דיופנטיות). בעייה זו שונה לגמרי באופייה ממציאת פתרונות ממשיים, ולדוגמה ניקח בעייה שבה מטפל דיופנטוס - שלשות פיתגוראיות, כלומר שלשות מספרים X, Y, Z שהן צלעות של משולש ישר־זוית, מה שאומר, לפי משפט פיתגורס, ש $X^2 + Y^2 = Z^2$. אם מחפשים פתרונות ממשיים הבעייה טריוויאלית, אבל דיופנטוס חיפש X, Y, Z שלמים. אז הבעייה יותר קשה, אבל הוא מביא את הפתרון המלא, שבוודאי ידוע לרבים מהקוראים: X, Y, Z הן כל השלשות מהצורה:

$$X = k(a^2 - b^2), \quad Y = k(2ab), \quad Z = k(a^2 + b^2) \qquad \text{שלמים } k, a, b$$

פתרון זה הביא מייד איש תורת המספרים כמו פרמה לשאול מה קורה עם מעריך גדול מ־2. הוא ניסה כנראה למצוא פתרונות, וכמובן לא הצליח. על מחשבותיו על כך ידוע לנו מהערה שנמצאה אחרי מותו בשוליים של עותק הספר הנזכר של דיופנטוס שהיה ברשות פרמה. הוא כתב שם בלטינית: "אי אפשר לכתוב חזקה שלישית כסכום שתי חזקות שלישיות, חזקה רביעית כסכום שתי חזקות רביעיות או באופן כללי חזקה גדולה משתיים כסכום חזקות מאותו סוג. מצאתי לעובדה זו הוכחה מיוחדת במינה. השוליים קטנים מדי מלהכיל אותה".

הערת שוליים זו נכתבה כנראה ב 1637, ומאז לא הפסיקה הבעייה להציק למתמטיקאים: האם טענתו של פרמה נכונה או לאו? האם אפשר למצוא הוכחה "מיוחדת במינה" כמו שטען פרמה? איש לא הצליח לענות על כך. הבעייה זכתה לפרסום גם בין לא־מתמטיקאים. חובבים רבים הקדישו לה שנים בתקווה למצוא הוכחה "פשוטה ואלמנטרית". היא מוזכרת אפילו בספר ההרפתקאות לנוער "השד מהשביעית" של קורנל מקושינסקי. ב־1907 הציע המיליונר הגרמני־יהודי Wolfskehl פרס של 100,000 מרק למי שיכריע את הבעיה, דבר שגרר המוני "הוכחות" ממתמטיקאים והדיוטות כאחת ואפילו איומים בתביעות משפטיות. פרשה זו נסתיימה כאשר האינפלציה בגרמניה אחרי מלחמת העולם הראשונה איפסה את ערך הפרס.

אבל גם במסגרת ההיסטוריה של המתמטיקה עצמה שמור לבעיה זו מקום מיוחד. לכשלעצמה אין לבעיה זו שום חשיבות מיוחדת, ויש הרבה בעיות דומות שקשה להכריע, במיוחד בתורת המספרים. מתמטיקאים דגולים כמו גאוס והילברט טענו שאין טעם להקדיש לה זמן (כנראה אחרי שהם עצמם ניסו ולא הצליחו). הדברים אותם רואים רוב המתמטיקאים כמתמטיקה (טהורה) בעלת חשיבות הן תורות כלליות, במיוחד תורות המגלות נקודות ראות חדשות במבנה מתמטי קיים, או המקשרות, לעתים באופן מפתיע, בין מבנים שנראים שונים, וכך פותחות דרכים מסוג חדש לפתור בעיות.

---

[19] אני מודה לפרופ' שי הרן, מהפקולטה למתמטיקה בטכניון, על הערותיו.

[20] באוקטובר 1994 הראה ויילס כיצד להימנע מהפער שנתגלה, תוך פישוט הטיעון (לשלב אחד תרם גם ריצ'רד טיילור) וכיום ברור שמשפט פרמה־ויילס הוכח.

הנסיון לפתור את בעיית פרמה הביא ליצירת תורה כזו במאה ה־19: תורת המספרים האלגבריים (שננסה להסביר בהמשך), והפתרון של Wiles בא אחרי ש Kenneth A. Ribet , בעקבות רעיונות של Frey , Serre ואחרים, הראה ב־1987 שמשפט פרמה נובע מהשערה, הנקראת השערת Taniyama-Shimura, שאם היא נכונה היא יוצרת קשר חשוב בין שתי תורות: תורת העקומים האליפטיים ותורת התבניות המודולריות (בהמשך ננסה להסביר זאת). השערה זו היא כן משהו שנחשב כמתמטיקה חשובה, ויתר על כן ־ המתמטיקאים רצו מאוד שהיא תהיה נכונה. Wiles , שעוד בנערותו ניסה "לגלות" את הוכחתו האבודה של פרמה, החליט להקדיש את זמנו להוכחת משפט פרמה בעזרת משפטו של Ribet. אחרי 6 שנים הצליח להוכיח את השערת Taniyama-Shimura, אמנם באופן חלקי, אבל מספיק בשביל משפט פרמה. הוא נעזר בתורה של Barry Mazur שמקשרת בין שתי התורות שנזכרו לעיל: העקומים האליפטיים והתבניות המודולריות לבין תורת הצגות גלואה. מיותר לציין שכל התורות האלו לא היו קיימות בזמנו של פרמה, ולמעשה פותחו רק בעשרות השנים האחרונות. את ההוכחה של השערת Taniyama-Shimura יש לראות כחשובה הרבה יותר ממשפט פרמה, ובניגוד לו, היא לא סיום תקופה אלא התחלת תקופה ־ הכנסת סדר והבנה יותר טובה של התורות שנזכרו, ותקווה להגדיל את ההבנה ע"י הוכחת עוד השערות "בוערות".

ננסה כעת להסביר חלק מהרעיונות והתורות המתמטיים שהיוו את 350 שנות ההיסטוריה של משפט פרמה:

אנו רוצים להוכיח שעבור $n > 2$ טבעי אין פתרון בשלמים שונים מ 0 למשוואה:

$$X^n + Y^n = Z^n$$

נעיר קודם שאם X, Y, Z פתרון עבור $n = km$ אזי X^k, Y^k, Z^k יהיו פתרון עבור m. לכן די להוכיח את משפט פרמה עבור n שהוא ראשוני איזוגי או 4 (כי כל מספר גדול מ 2 מתחלק בראשוני איזוגי או ב 4). נוסף לכך כל פתרון אפשר לחלק במחלק המשותף המכסימלי ולכן די להוכיח שאין פתרון בו אין גורם משותף ל X, Y, Z.

שיטת הירידה האינסופית של פרמה

אחד מרעיונותיו של פרמה בתורת המספרים נקרא "שיטת הירידה האינסופית" היא אומרת שאם אנו רוצים להוכיח שלא קיים מספר טבעי n בעל תכונה מסויימת, די להוכיח שאם קיים n טבעי כזה אזי קיים n טבעי יותר קטן (מדוע שיטה זו תקפה?). בעזרת שיטה זו הוכיח פרמה את משפטו עבור $n = 4$: הוא הוכיח אפילו שלא קיימים X, Y, Z שלמים כך ש $X^4 + Y^4 = Z^2$.

ואכן, אחרת היו X^2, Y^2, Z מהווים שלשה פיתגוראית. יתר על כן, נוכל להניח שאין להם גורם משותף. אז לפי הפתרון עבור שלשות פיתגוראית שהובא למעלה:

$$X^2 = a^2 - b^2, \quad Y^2 = 2ab, \quad Z = a^2 + b^2$$

כאשר a ו b זרים, ובעלי זוגיות שונה (אחרת X, Y, Z היו זוגיים). אילו a היה זוגי, היו b ו X איזוגיים ולכן $b^2, X^2 \equiv 1 \pmod 4$ ומכאן $a^2 = X^2 + b^2 \equiv 2 \pmod 4$ וזה לא ייתכן.

לכן b זוגי ו a איזוגי. השלשה X, b, a פיתגוראית ללא גורם משותף, בה b זוגי ו a, X איזוגיים. לכן

$$X = r^2 - s^2, \quad b = 2rs, \quad a = r^2 + s^2$$

אז

$$Y^2 = 2ab = 4rs(r^2 + s^2)$$

וכל שניים מ $r, s, r^2 + s^2$ זרים. לכן כל אחד מהם הוא ריבוע:

$$r = u^2, \quad s = v^2, \quad r^2 + s^2 = t^2$$

ומכאן מקבלים פתרון חדש למשוואה המקורית:

$$u^4 + v^4 = t^2$$

לפי שיטת הירידה האינסופית, ההוכחה תושלם אם נוכיח ש $t < Z$. אבל

$$Z = a^2 + b^2 > a^2 = (r^2 + s^2)^2 = t^4$$

והטענה הוכחה.[21]

[21] ראו הוכחה אחרת במאמרו של ו. מנביץ ב"אתגר־גליונות מתמטיקה" חוברת 29־30, מרס 1994, בה התפרסם לראשונה גם מאמר זה.

המשך המאמר עלול להיות קצת קשה להבנה, ומניח במקומות אחדים שהקורא/ת שמע/ה על מושגים כמו: טופולוגיה, חבורה ועוד.

שלמים אלגבריים

נותר להוכיח שאין פתרון ל $(*)$ עבור $n=p$ ראשוני איזוגי. עבור $n=3$ הטענה הוכחה במאה ה־18 ע"י אוילר (Euler), אבל לשם כך הוא נאלץ לצאת מקבוצת המספרים השלמים ולדון במספרים מהצורה $a+b\omega$ כאשר a ו b שלמים ו ω הוא המספר הקומפלכסי (מרוכב) $\omega=\frac{1}{2}(-1+\sqrt{3}i)$ שהוא אחד השורשים השלישיים הקומפלכסיים של 1. בקבוצת מספרים אלה אפשר לעשות פעולות חשבון בדומה למספרים השלמים. זו מעין הרחבה של השלמים ל"שלמים כלליים יותר". המספר $z=a+b\omega$ מקיים משוואה אלגברית בה משווים לאפס פולינום עם מקדמים שלמים ומקדם ראשון 1:

$$z^2-(2a-b)z+(a^2-ab+b^2)=0$$

ומספרים כאלה נקראים **שלמים אלגבריים**. מערכת יותר פשוטה הם **השלמים של גאוס** $a+bi$, a,b שלמים (איזו משוואה אלגברית עם מקדמים שלמים מקיים $a+bi$?). בדומה לכך מספר המקיים משוואה אלגברית בה משווים לאפס פולינום עם מקדמים רציונליים נקרא **מספר אלגברי**, למשל המספרים הנתונים ע"י הביטויים לעיל אם נניח a,b רציונליים במקום שלמים (מקדמים רציונליים אפשר להכפיל במכנה משותף ולקבל מקדמים שלמים, אבל המקדם הראשון לא יהיה 1). קבוצה של מספרים אלגבריים שהם צירופי מספר סופי מהם עם מקדמים רציונליים ושביצוע כפל שני איברים בה משאיר אותנו בקבוצה (למשל המערכות שציננו זה עתה עם a,b רציונליים) נקראים **שדות מספרים אלגבריים**. מתברר שאז גם חילוק שני איברים שונים מאפס בקבוצה ייתן איבר בקבוצה (בדוק לגבי $a+b\omega$ ולגבי $a+bi$).

בדרך דומה טיפלו במחצית הראשונה של המאה ה־19 ב $p=5,7$, בעזרת שדות אחרים של מספרים אלגבריים (Dirichlet, Legendre, Lamé), אבל לא הצליחו להתקדם יותר, ובמיוחד - לא היתה שיטה כללית לטפל ב p כלשהו.

עם זאת היה אמצעי אחד בו השתמשו בכל ההוכחות שנעזרו בשלמים אלגבריים בהצלחה: בשדות המספרים האלגבריים שהשתמשו בהם, היתה לקבוצת השלמים האלגבריים (כלומר אלה שהם פתרון משוואה עם מקדמים שלמים והמקדם הראשון 1) תכונת הפריקות החד־ערכית לגורמים ראשוניים כמו בשלמים הרגילים, ובתכונה זו השתמשו באופן חזק. (הראשוניים עליהם מדובר כאן הם השלמים האלגבריים בשדה שאין להם פירוק לא טריביאלי לשלמים אלגבריים באותו שדה. פירוק נחשב לטריביאלי אם אחד הגורמים הוא מספר כמו ±1, $\pm i$ שגם ההפכי הכפלי שלו שלם אלגברי. למשל בין השלמים של גאוס $2+i$, $1+i$, 3 ו 7 הם ראשוניים בעוד ש 5 פריק, כי $5=(2+i)(2-i)$).

בשנת 1847 סבר קומר (Kummer) שיש לו הוכחה עבור כל p . הרעיון שלו היה להיעזר במספר הקומפלכסי $\zeta_p=\cos\frac{2\pi}{p}+i\sin\frac{2\pi}{p}$. ל ζ_p יש התכונה ש $\zeta_p{}^p=1$ ושהשורשים הקומפלכסיים ה p -ים של 1 הם בדיוק החזקות של ζ_p , שמהווים קבוצה של בדיוק p איברים שונים: $1,\zeta_p{}^1,\zeta_p{}^2,\ldots,\zeta_p{}^{p-1}$ (בדוק!). קבוצת הצירופים של אלה עם מקדמים רציונליים היא שדה מספרים אלגבריים שנקרא השדה הציקלוטומי המתאים ל p (ציקלוטומי ביוונית פרושו: "מחלק מעגל". מקור שם זה הוא בכך ש ζ_p והחזקות שלו מתוארות במישור המספרים הקומפלכסיים כנקודות המחלקות את מעגל היחידה ל p קשתות שוות. במיוחד בעיית חלוקת מעגל ל p חלקים שווים ע"י סרגל ומחוגה שקולה לבעייה לבטא את ζ_p ע"י שורשים ריבועיים).

העזרה שהשדה הציקלוטומי נותן לבעיית פרמה היא שכיוון ששורשי המשוואה $x^p-1=0$ הם $1,\zeta_p,\zeta_p{}^2,\ldots,\zeta_p{}^{p-1}$, הפולינום x^p-1 הוא מכפלת הגורמים ממעלה ראשונה $x-\zeta_p{}^k$ עבור $k=0,1,\ldots,p-1$ ומכאן שאם $X^p+Y^p=Z^p$ אזי $X^p=Z^p-Y^p$ הוא מכפלת הגורמים $Z-\zeta_p{}^kY$. קומר ובני דורו היו בטוחים שלגבי כל שדה מספרים אלגבריים נכון משפט הפריקות החד ערכית לגורמים ראשוניים של השלמים האלגבריים בו, ואז אם X,Y,Z שלמים נוכל מהעובדה שמכפלת הגורמים היא חזקה p־ית ומהעובדה שהם זרים להסיק שכל אחד הוא חזקה p־ית, ואז יכול היה קומר להוכיח את משפט פרמה.

קומר הגיש הוכחה זו למורו דיריכלה (Dirichlet) , וזה העיר שאין זה מובן מאליו שבכל שדה מספרים אלגבריים יש פריקות חד ערכית. כשבדק זאת קומר מצא להפתעתו שבדרך כלל אין פריקות חד ערכית. דוגמא הוא שדה המספרים מהצורה $a+b\sqrt{5}i$, a,b רציונליים, שהשלמים האלגבריים בו יתקבלו אם a,b שלמים. כאן $21=3\cdot7=(1+2\sqrt{5}i)\cdot(1-2\sqrt{5}i)$ ולהפתעת הכל כל ארבעת הגורמים הם ראשוניים בשדה זה.

תגלית זו, שלכאורה היתה מפח נפש, היתה התחלת התורה המודרנית של מספרים אלגבריים. לקומר היה רעיון "מטורף", ליצור בכח פירוק חד־ערכי ע"י כך שלא לוקחים את השלמים האלגבריים הראשוניים עצמם, אלא "מספרים אידיאליים" שמוגדרים בצורה מתאימה כך ששלם אלגברי ראשוני יכול להיות פריק ל"אידיאלים" ואז אפשר להשיג פריקות חד ערכית. תורה זו קיבלה אחר כך צורה נוחה וטבעית ע"י דדקינד (Dedekind) והפכה את מושג ה"אידיאל" לאבן פינה באלגברה המופשטת כולה. הושגה הבנה

טובה של חוקי המבנה של שדות מספרים אלגבריים. במיוחד עבור כל שדה כזה יש מספר טבעי h , הנקרא Class Number , שמתאר באופן כמותי בכמה האידיאלים הם תוספת על המספרים ה"אמיתיים". תהיה פריקות חד־ערכית רק אם $h = 1$, מה שהוא נדיר.

"הוכחתו" של קומר למשפט פרמה שנזכרה לעיל היתה תקפה אילו לכל השדות הציקלוטומיים היה $h = 1$. קומר גילה, לצערו, שאין זה כך. למרות זאת הוא הצליח להוכיח את משפט פרמה עבור **כל p שאינו מחלק את h של השדה הציקלוטומי שלו**. p־ים כאלה נקראים **רגולריים**. ידוע שיש אינסוף ראשוניים שאינם רגולריים. הראשוני הראשון שאינו רגולרי הוא 37, ועד 100 יש רק עוד שניים כאלה: 59 ו 67. למרות זאת הצליח קומר להוכיח את משפט פרמה גם עבור הראשוניים הלא־רגולריים הקטנים מ 164.

בכך היתה בודאי התקדמות גדולה לעומת המצב הקודם. במשך מאה השנים שאחרי קומר הצליחו בשיטות כאלה להוכיח את משפט פרמה עד מעריך $150,000$ (לבסוף בעזרת מחשבים), אבל לא היתה שיטה כללית, והבעיה לא היתה קשורה לשום הבנה כללית, ולכן היא נדחפה לחצר האחורית של המתמטיקה, למרות הפרסום שזכתה לו בקהל הלא־מתמטי, שעליו דובר לעיל. מתמטיקאים רבים איבדו את התקווה שהבעיה תיפתר בעתיד הנראה לעין.

לעומת זאת כאמור תורת המספרים האלגבריים הפכה לתורה עם הבנה ברורה של המבנה הבסיסי, והיא השתלבה בנושאים אחרים בהם האלגברה היא השחקן העיקרי. הנושא סביבו בנויה הוכחתו של Wiles הוא נושא כזה:

הגאומטריה האלגברית

הקוראים בוודאי יודעים שבגאומטריה האנליטית (אותה, אגב, ייסדו דקרט (Descartes) ואותו פרמה במאה ה־17) מתוארים עקומים ע"י משוואות. הגאומטריה האלגברית עוסקת בעקומים (או משטחים וכד') המתוארים ע"י משוואות פולינומיאליות במספר משתנים. עקומים כאלה נקראים **עקומים אלגבריים**. נדון במקרה של שני משתנים, כלומר במישור.

נניח שנתונה משוואה פולינומיאלית כזו $P(x, y) = 0$. בגאומטריה האנליטית הרגילה אנו מחפשים נקודות ממשיות, כלומר x, y ממשיים, אבל איש אינו מונע אותנו מלחפש פתרונות x, y רציונליים, או שלמים, או קומפלכסיים, או מספרים בשדה מספרים אלגבריים, למשל. אומרים שפתרון (x, y) כזה מהווה "נקודה רציונלית" או "שלמה" או "קומפלכסית" וכו'. לדוגמא למשוואה $x^2 + y^2 + 1 = 0$ אין פתרונות ממשיים אבל יש פתרונות קומפלכסיים. לכן משוואה זו מתארת "עקום ממשי ללא נקודות" אבל "עקום קומפלכסי עם נקודות". את משפט פרמה נוכל לנסח כטענה האומרת שאין נקודות רציונליות על העקום $x^n + y^n = 1 \quad n > 2$ פרט לנקודות הטריביאליות על הצירים. בדרך כלל יש טעם להתעניין בנקודות רציונליות רק אם מקדמי P רציונליים. אז נאמר שהעקום **אריתמטי** וע"י הכפלה במכנה משותף נוכל להניח שכל המקדמים שלמים. (עם זאת, יש טעם רב להתעניין בנקודות ממשיות או קומפלכסיות בעקום אריתמטי!)

כמו בהרבה מצבים באלגברה, המצב הוא פשוט יותר אם נחפש x, y קומפלכסיים. אז נוצר קשר מפתיע עם הענף במתמטיקה שהוא לכאורה ההפך מאלגברה ־ הטופולוגיה . משתנה אחד קומפלכסי מתאר מישור. מתברר שיהיה זה טבעי להוסיף לו נקודה ב ∞ , ו"לסגור" אותו כך שייהפך למשהו שהוא מבחינה טופולוגית כדור (אפשר לתאר סגירה זו ע"י הטלת סטיריאוגרפית ־ הטלת המישור על כדור המשיק למישור בקוטב הדרומי כשמרכז ההטלה הוא הקוטב הצפוני. נקודות המישור יתאימו לכל נקודות הכדור פרט לקוטב הצפוני ו ∞ תתאים לקוטב הצפוני).

בדומה לכך נתבונן בעקום המתואר ע"י משוואה פולינומיאלית, ונניח שאין בעקום נקודות סינגולריות , כלומר נקודות בהן העקום חותך את עצמו או יש חוד (כמו למשל ל $x^3 = y^2$ ב 0). נוסיף נקודות מתאימות ב ∞ (למשל לפרבולה יש נקודה אחת באינסוף, להיפרבולה שתיים. גם לאליפסה יש שתי נקודות באינסוף, אבל הן לא־ממשיות). אז תהווה קבוצת הנקודות הקומפלכסיות משטח. מבחינה טופולוגית יכול משטח זה להיות או כדור, או המשטח החיצוני של כעך, שאותו אפשר לתאר ככדור שהוסיפו לו ידית (משטח כזה נקרא טורוס Torus), או המשטח החיצוני של כעך עם שני חורים ־ כדור עם שתי ידיות, וכן הלאה. מבחינה טופולוגית נקבע המשטח במלואו ע"י מספר הידיות g , שנקרא הגנוס genus של המשטח ושל העקום האלגברי. עבור משטח שהוא מבחינה טופולוגית כדור $g = 0$, עבור טורוס $g = 1$ וכו'. עבור העקום הקשור לבעיית פרמה $x^n + y^n = 1$ מתברר ש $g = \frac{(n-1)(n-2)}{2}$. עבור כל עקום ממעלה 1 או 2 (למשל כל חתכי החרוט) הגנוס הוא 0.

אחת התגליות החשובות ביותר במאה ה־19, שנעשתה בעיקר ע"י רימן (Riemann) , היתה החשיבות העצומה של הגנוס. אפילו בעיית מציאת האינטגרל של הפונקציה האלגברית שהגרף שלה הוא העקום תלויה בגנוס. המצב הוא קל ביותר עבור גנוס 0 , והסיבה שכל כך קל לפתור את משוואת פרמה עבור $n = 2$ היא שאז הגנוס הוא 0 . בדומה לכך, לפונקציות אלגבריות שהגרף שלהן הוא עקום מגנוס 0 (למשל אלה שמכילות שורש של פולינום ממעלה 2) יש אינטגרל שהוא פונקציה אלמנטרית.

אבל כשרוצים לחשב את האורך של אליפסה מופיע אינטגרל של פונקציה שהגרף שלה הוא עקום אלגברי ממעלה 3 (או ממעלה 4 שאפשר להביא למעלה 3 ע"י פעולות באינטגרל). אם בעקום ממעלה 3 אין סינגולריות, הגנוס שלו הוא 1 , וכתוצאה מכך קוראים לעקומים מגנוס 1 **עקומים אליפטיים**.

הגנוס נראה במבט ראשון כמשהו טופולוגי לעילא, שהוא רלבנטי רק לגבי קבוצת הנקודות הקומפלכסיות בעקום, אבל המחקרים המעמיקים מאז רימן (ולפניו Abel ו Jacobi) קישרו בין הגנוס ומספר כה רב של תכונות אלגבריות שאין זה מפתיע שאפשר להגדיר את הגנוס באופן אלגברי טהור, ואז אין זה מפתיע שהגנוס יהיה רלבנטי לשאלות הדנות בנקודות **הרציונליות** של עקום אריתמטי.

דוגמא לרלבנטיות כזו היא השערת Mordell , שנוסחה על ידו ב 1922 והוכחה ב 1983 ע"י המתמטיקאי הגרמני הצעיר Faltings. היא טוענת שאם עקום אלגברי אריתמטי הוא מגנוס גדול מ 1 אזי יש בו רק מספר סופי של נקודות רציונליות. מכאן ינבע שלמשוואת פרמה עבור $n > 3$ יש רק מספר סופי של פתרונות.

כדאי להעיר ש Faltings נעזר בכך שלפניו הראו שהשערת Mordell נובעת מהשערות אחרות ו Faltings הוכיח את ההשערות האחרות. אותו דבר עשה Wiles לגבי משפט פרמה. פתרון הבעיות נעשה ע"י הוספת חוליות בשרשרת המורכבת ממשפטים מהסוג "מהשערה א' נובעת השערה ב'".

השערת Mordell מתייחסת לגנוס $1 <$. גנוס 0 הוא, כאמור, קל. הוכחתו של Wiles תלויה בתורת העקומים מגנוס 1 , כלומר העקומים האליפטיים. לפני שנדון בהם נזכיר מושג בסיסי כאשר עובדים עם פונקציות קומפלכסיות:

פונקציות אנליטיות

הקוראים יודעים אולי כיצד המספרים הקומפלכסיים מפשטים בהרבה את האלגברה בכך שהם מבטיחים פתרון לכל משוואה אלגברית (המשפט היסודי של האלגברה). אבל במאה ה־19 התברר שהם צופנים הפתעות לא פחות גדולות גם באנליזה. רק טבעי הוא לנסות להכליל את מושג הפונקציה לפונקציה שמוגדרת בקבוצת מספרים קומפלקסיים ומקבלת ערכים קומפלקסיים, למשל פונקציות הנתונות ע"י פולינומים או מנות של פולינומים. אפשר לנסות להרחיב גם פונקציות מעריכיות או טריגונומטריות לתחום הקומפלקסי וחלק מהקוראים יודעים בודאי איך עושים זאת. ניתן להרחיב גם את מושג הנגזרת, למשל.

ואז מתברר שבניגוד גמור למצב בפונקציות ממשיות, יש סוג של פונקציות קומפלקסיות, שנקראות **פונקציות אנליטיות**, שמתבלט באופן חד: מצד אחד כל פולינום הוא פונקציה אנליטית וכל פונקציה שמתקבלת באופן "הגון" ע"י ארבע פעולות החשבון, פונקיות סתומות, גבולות, טורים, נגזרות, אינטגרלים ועוד מפונקציות אנליטיות היא אנליטית. יתר על כן, כל פונקציה גזירה (פעם אחת) במובן הקומפלכסי היא אנליטית. למשל הפונקציות הטריגונומטריות, המעריכיות וכו' שמרחיבים לתחום הקומפלכסי הן "כמובן" אנליטיות. מבחינה זו מושג זה הוא רחב מאוד. מצד שני מקיימות הפונקציות האנליטיות דרישות מחמירות ביותר, למשל לכולן יש באופן אוטומטי נגזרות מכל סדר במובן הקומפלקסי, אפשר לפתח אותן לטור חזקות, יש נוסחאות שכולן חייבות לקיים, יש להעתקה שהן מגדירות תכונות גאומטריות מיוחדות, והתכונה שהיא אולי הראויה ביותר לציון ־ פונקציה אנליטית המוגדרת בתחום מסויים אפשר להרחיב לתחום גדול יותר רק בצורה אחת, אם בכלל, כך שגם הפונקציה המורחבת תישאר אנליטית. לכן אזור קטן בתחום הגדרת הפונקציה קובע לגמרי את התנהגות הפונקציה בכל מקום שאפשר להמשיך אותה אליו.

עקומים אליפטיים

נתון עקום אליפטי, למשל עקום ממעלה 3 ללא סינגולריות. קבוצת הנקודות הקומפלקסיות שלו היא, מבחינה טופולוגית, טורוס. נזכיר שזהו משטח הפנים של כעך עם חור אחד (וכן של ספל עם ידית אחת).

יש דרך אחרת, הנראית שונה לגמרי, לקבל טורוס: נקח מישור, למשל המישור הקומפלכסי, ונבחר בו שני וקטורים u ו v שונים מ 0 שאינם באותו ישר (הוקטורים לאו דוקא מאונכים ולאו דוקא בעלי אותו אורך). נתבונן בסריג Λ , שהוא קבוצת הוקטורים מהצורה $mu+nv$, m,n שלמים. נוכל לבנות את המקבילית שקדקד אחד שלה בראשית ושתי צלעותיה לאורך u ו v (נקרא לה: המקבילית הבסיסית) ולרצף את המישור כולו ע"י הזזות של מקבילית זו באברי Λ , כך שכל נקודה במישור תהיה באחת המקבילות וקבוצות נקודות הפנים של המקבילות יהיו זרות. אנו ניקח את המישור מודולו Λ , כלומר נזהה שתי נקודות במישור שהוקטור המחבר אותן שייך ל Λ (בדומה לחשבון מודולו m , בו מזהים שני מספרים שהפרשם הוא כפולה של m). עבור כל נקודה במישור קיימת נקודה z במקבילית הבסיסית שמזוהה איתה (בדומה למצב בחשבון מודולו m בו אפשר לקחת נציג בין 0 ו m). אם z פנימית למקבילית היא לא מזוהה עם אף נקודה אחרת במקבילית. בשפת המקבילית יהיה זיהוי נקודות שהפרשן u או v . כמובן שכל ארבעת קדקדי המקבילית מזוהים.

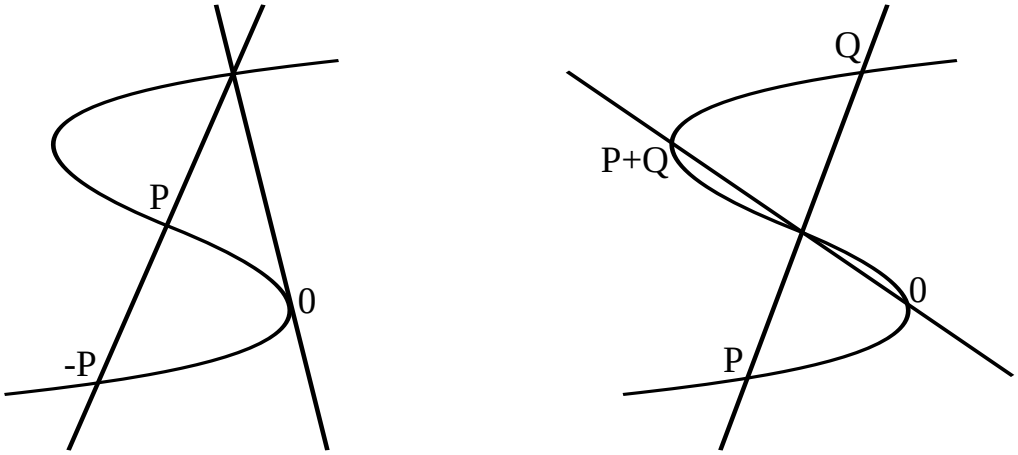
אם נדביק נקודות מזוהות, נקבל משטח שהוא קבוצת נקודות המישור עם הזיהוי (שהיא אותו דבר כמו קבוצת נקודות המקבילית הבסיסית עם הזיהוי) ואם הקורא יבדוק כיצד מזוהות נקודות השפה של

המקבילית הבסיסית הוא ייווכח שמשטח זה יהיה טורוס, אותו עוטף המישור אינסוף $\times$ אינסוף פעמים. מבחינה טופולוגית אין שום חשיבות לבחירת הוקטורים u ו v - תמיד יתקבל טורוס, אבל מבחינות אחרות עליהן נרמוז להלן יש הבדל בין זוגות וקטורים שונים.

אלא שהטורוסים שקבלנו מהמישור הקומפלכסי מודולו סריג הם באופן טבעי חבורה קומוטטיבית, כלומר יש בהם פעולת חיבור - החיבור של המספרים הקומפלכסיים מודולו Λ , שמקיימת את אכסיומות החבורה הקומוטטיבית.

אם נתון עקום אליפטי, אפשר לעטוף את משטח הנקודות הקומפלכסיות בו אינסוף $\times$ אינסוף פעמים ע"י מישור וכך לזהות אותו עם מישור מודולו סריג. ברור שאין קושי לעשות זאת מבחינה טופולוגית ואז כל זוג וקטורים u ו v יהיה טוב, שכן מבחינה טופולוגית אין כל הבדל בין בחירות שונות של u ו v . אבל אפשר להוכיח שאפשר לבצע זאת כך שתתקיים דרישה נוספת חזקה מאוד: שהצורה בה המישור עוטף את העקום תהיה כך, שפולינומים בקואורדינטות x, y על העקום ייהפכו, אחרי ההעברה למישור, לפונקציות **אנליטיות**. זאת אפשר לעשות רק עבור וקטורים u ו v בעלי זוית ויחס אורכים המותאמים לעקום האליפטי (ע"י נוסחאות לא פשוטות). ע"י עיטוף כזה תעבור פעולת החיבור מהמישור מודולו הסריג לעקום, ותהפוך את (קבוצת הנקודות הקומפלקסיות) בעקום ל**חבורה קומוטטיבית**.

המעניין הוא, שגם את פעולת החיבור שקבלנו כעת בין (הנקודות הקומפלכסיות) בעקום האליפטי אפשר להגדיר באופן אלגברי טהור (יש רק לבחור באופן שרירותי מה תהיה נקודת האפס). הדבר נובע ממשפט האומר שאם E עקום אליפטי במישור ו C עקום אלגברי אחר כלשהו ממעלה n באותו מישור אזי הסכום, לגבי פעולת החבורה ב E , של הנקודות בהן C חותך את E (כולל נקודות באינסוף) תלוי רק במעלה n והוא שווה ל $n \cdot \alpha$ כאשר α איבר קבוע בחבורה התלוי רק בבחירת נקודת האפס. במיוחד סכום נקודות החיתוך של E עם ישר אינו תלוי בישר ושווה ל α . (אם בנקודת חיתוך העקומים E ו C משיקים או יש ל C, למשל, חוד יש לקחת את הנקודה מספר פעמים בהתאם לסדר ההשקה - עבור השקה רגילה, לא בנקודות פיתול, יש לקחת את הנקודה 2 פעמים). אם העקום האליפטי ממעלה 3, אזי ישר החותך אותו בשתי נקודות P_1 ו P_2 (או משיק לו בנקודה שאינה נקודת פיתול) חייב לחתוך אותו בנקודה שלישית P_3, ואת קואורדינטות נקודה זו אפשר למצוא ע"י ארבע פעולות החשבון ממקדמי העקום וקואורדינטות P_1 ו P_2 (איך?). בעזרת בניה כזו אפשר לבנות, בשימוש במשפט שצוטט לעיל, את הסכום של שתי נקודות ואת מינוס נקודה (כאשר נתונה נקודת ה 0). בניה כזו מתוארת בציורים הבאים. הסבר.



מהעובדה שאפשר לבטא את פעולות החבורה בעקום אליפטי ממעלה 3 ע"י ארבע פעולות החשבון נובע שקבוצת הנקודות הרציונליות בעקום אליפטי אריתמטי סגורה לגבי החיבור והמינוס ומהווה גם היא חבורה קומוטטיבית. בניגוד למקרה של גנוס $1 <$, בו לפי השערת Mordell שהוכיח Faltings יש מספר סופי של נקודות רציונליות, במקרה של עקום אליפטי כלומר גנוס 1, חבורת הנקודות הרציונליות יכולה להיות סופית או לא. "גודלה" של חבורה זו היא בעייה קשה.

דרך אגב, כאשר מתארים עקומים אליפטיים אריתמטיים ע"י משוואה עם מקדמים שלמים אפשר לקחת הכל מודולו מספר ראשוני p ולקבל עקום אליפטי מודולו p שיהיה חבורה קומוטטיבית סופית. בשנים האחרונות השתמש Lenstra בעקומים־חבורות כאלה כדי לפתח אלגוריתמים טובים לבדיקה אם מספר גדול נתון הוא ראשוני ולפירוק מספרים גדולים לגורמים ראשוניים. הייתרון של עקומים אליפטיים מודולו p למטרה זו הוא שהם נותנים, עבור אותו p , חבורות שונות שיש בהן מספר שונה של איברים בעוד ששיטות קודמות היו מוגבלות ע"י מבחר מצומצם של חבורות עבור כל p , למשל החבורה הכפלית של השאריות מודולו p שנותנת רק חבורה אחת בת $p-1$ איברים. לנושא זה יש חשיבות מעשית, שכן מספרים ראשוניים גדולים משמשים בשיטות הצפנה, שכל ערכן תלוי בקיום או אי־קיום אלגוריתמים טובים לבדיקת ראשוניות ולפירוק לגורמים ראשוניים של מספרים גדולים.

פונקציות־L

לאוילר, שאותו הזכרנו כבר, היה הרעיון לחקור את המספרים הראשוניים ע"י פונקציה. אם $s>1$ הטור $\sum_{n=1}^{\infty}\frac{1}{n^s}$ מתכנס, ולכן מגדיר פונקציה של s שמסמנים אותה ב $\zeta(s)$, והיא נקראת פונקציית ζ (זיתא). הרעיון של אוילר היה שמשפט הפירוק החד־ערכי של המספרים הטבעיים לגורמים ראשוניים ניתן לקידוד בעזרת פונקציה זו כך:

$$\sum_{n=1}^{\infty}\frac{1}{n^s}=\sum_{k_1,k_2,\ldots}\frac{1}{\left(2^{k_1}\cdot 3^{k_2}\cdot 5^{k_3}\cdots\right)^s}=\left(\sum_{k=0}^{\infty}\frac{1}{2^{ks}}\right)\cdot\left(\sum_{k=0}^{\infty}\frac{1}{3^{ks}}\right)\cdot\left(\sum_{k=0}^{\infty}\frac{1}{5^{ks}}\right)\cdots=$$
$$=\frac{1}{1-\frac{1}{2^s}}\cdot\frac{1}{1-\frac{1}{3^s}}\cdot\frac{1}{1-\frac{1}{5^s}}\cdot\frac{1}{1-\frac{1}{7^s}}\cdots$$

בעזרת זה הוכיח אוילר, למשל, שהטור $\sum_{p}\frac{1}{p}$ מתבדר. זה חיזוק של משפט אוקלידס שיש אינסוף מספרים ראשוניים.

רימן, שגם אותו הזכרנו, הרחיב את הגדרת הפונקציה בכך שלקח s קומפלקסי (כאמור לעיל יש הגדרה ל n^s כאשר s קומפלכסי). כיוון שהפונקציה מוגדרת ע"י טור "הגון", זו תהיה אז פונקציה אנליטית של s , מה שמאפשר להשתמש באוסף המרשים של תכונות של פונקציות אנליטיות ולהגיד דברים רבים על המספרים הראשוניים. אבל מה שמעניין הוא שבעוד שהטור מתכנס רק בתחום מסויים של s־ים, אנו מוכיחים קודם כל שפונקצית ζ ניתנת להמשכה (שהיא כאמור יחידה) לכל המישור הקומפלקסי כמעט, והתכונות החשובות של הראשוניים תלויות במה שקורה באזורי ההמשכה בהם ההגדרה המקורית ע"י הטור (שבה הופיעו הראשוניים) חסרת משמעות.

מתברר ששיטה זו ־ קידוד תכונות אריתמטיות ע"י פונקציות אנליטיות ושימוש בתכונות פונקציות אנליטיות אלה לקבלת תוצאות אריתמטיות ־ ניתנת להכללה למקרים אחרים, ובמיוחד לעקומים אלגבריים אריתמטיים. לעקום כזה יש משוואה עם מקדמים שלמים ואת האינפורמציה של מה שקורה מודולו p עבור כל הראשוניים p אפשר לאסוף לפונקציה אנליטית, שנקראת פונקצית ζ או, בצורה אחרת, פונקצית־L של העקום.

פונקצית־L של עקום אליפטי עשירה באינפורמציה על העקום, לפעמים אפילו יותר ממה שהיה אפשר לצפות. הבעיה היא שקשה לחקור אותה, במיוחד קשה לקבוע האם תמיד יש לה, כמו לפונקצית ζ הרגילה, המשכה לכל המישור הקומפלקסי. בהנחה שיש המשכה, השערה של Birch ו Swinnerton-Dyer (שהיא עדיין מאוד פתוחה) אומרת שהתנהגות פונקצית־L של העקום בנקודה $s=1$ (שאינה בתחום ההגדרה המקורי) קובעת מה גודל חבורת הנקודות הרציונליות של העקום.

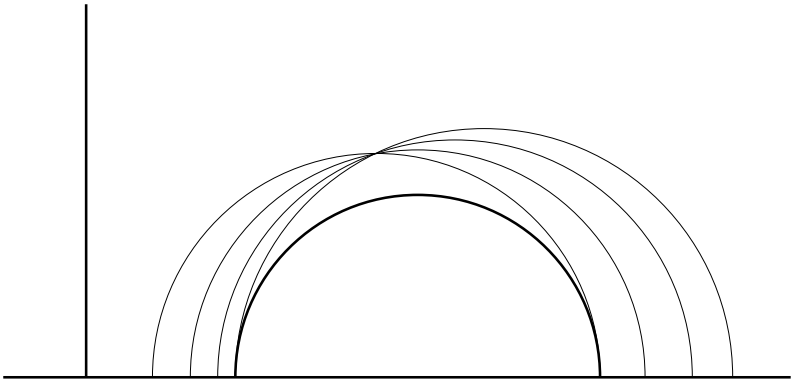
תבניות מודולריות

ניזכר במישור עם הסריג Λ. מצב כזה מביא מייד למושג של פונקציות מחזוריות לגבי Λ, כלומר פונקציות שיש להן אותו ערך בנקודות שהוקטור המחבר אותן שייך ל Λ. פונקציות כאלה נוכל לראות כפונקציות על הטורוס שהוא המישור מודולו Λ. למשל אם עוטפים את (הנקודות הקומפלכסיות) של עקום אליפטי ע"י מישור קומפלכסי מודולו סריג Λ יהפוך כל פולינום בקואורדינטות x,y בעקום לפונקציה אנליטית שהיא מחזורית לגבי Λ. לפעמים יש צורך לדון בפונקציות שאין להן מחזוריות מלאה, ומסתפקים בכך שהזזה באיבר של Λ משנה את הפונקציה באופן פשוט (למשל מכפילה אותה בפונקציה פשוטה).

המישור שלנו הוא, כמובן, המישור האוקלידי, והוא עני בסוגים של מחזוריות: המחזוריות היחידה המעניינת בו היא או מחזוריות בכפולות שלמות של וקטור מסויים (כמו המחזוריות של גל מישורי) או מחזוריות ע"י סריג הנקבע ע"י שני וקטורים u ו v . במיוחד אין במישור כזה ישועה לגבי עקומים מגנוס גדול מ 1 . עושר גדול הרבה יותר יש במישור של הגאומטריה הלא־אוקלידית ההיפרבולית, בה לא מתקיימת אכסיומת המקבילים של אוקלידס ודרך כל נקודה עוברים אינסוף ישרים שאינם חותכים ישר נתון. מישור כזה, שנקרא המישור ההיפרבולי, אפשר לראות כמבנה מתמטי מופשט, שאפשר לתת לו מודל (בו הישרים "יתעוותו") בצורה הבאה:

נקודות המישור ההיפרבולי יהיו נקודות חצי המישור העליון (שמעל לציר x) ללא ציר x עצמו. הישרים ההיפרבוליים יהיו הן חצאי־הישרים המאונכים לציר x והן כל חצאי־המעגלים שמרכזם בציר x . המודל ישמור על הזויות (למשל חצי־ישר ℓ מאונך לציר x ייצור זוית ישרה עם חצי־מעגל שמרכזו M בציר x אם ורק אם הישר של ℓ עובר דרך M). האורך ההיפרבולי של קו $x(t),y(t)$ יהיה $\int\frac{ds}{y(t)}$ כאשר ds הוא אלמנט האורך הרגיל (האוקלידי) בחצי המישור. כתוצאה מכך יהיה ציר x במרחק היפרבולי אינסופי

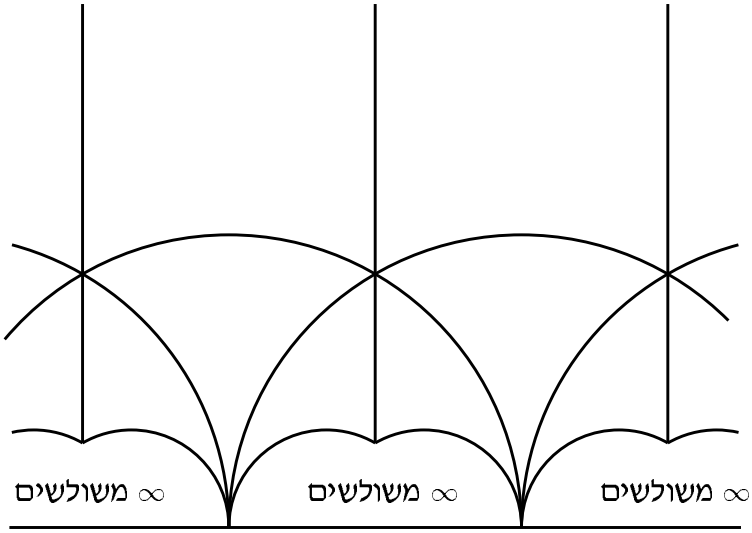
מכל נקודה היפרבולית. בציור רואים כיצד דרך נקודה נתונה עוברים אינסוף ישרים היפרבוליים שאינם חותכים ישר נתון.



נוכל לדבר על פונקציות אנליטיות על המישור ההיפרבולי - למטרה זו נראה אותו פשוט כחצי המישור הקומפלקסי.

כאמור, במישור ההיפרבולי יש עושר של סוגי מחזוריות, או, בלשון אחרת, של חבורות בדידות (דיסקרטיות) של תנועות (בדומה לחבורת ההזזות בסריג Λ במקרה האוקלידי לעיל). את המישור האוקלידי ריצפנו ע"י מקביליות וקיבלנו מישור העוטף טורוס, את המישור ההיפרבולי אפשר לרצף ע"י מצולעים בעלי מספר גדול כרצוננו של צלעות ולעטוף משטחים בעלי גנוס גדול מ 1 .

אחת החבורות הדיסקרטיות של תנועות במישור ההיפרבולי נקראת החבורה המודולרית. היא מתאימה לריצוף המישור ההיפרבולי, כמו שמתואר בציור, ע"י משולשים בעלי קדקד באינסוף



(של המישור ההיפרבולי, כלומר על ציר ה x במודל או בנקודת האינסוף של חצי המישור, למשל למשולשים בציור שמוגבלים ע"י קשת ושני ישרים מקבילים יש קדקד ב ∞). ברור שמשולשים כאלה הם יצור שאינו קיים במישור אוקלידי. פונקציות אנליטיות שהן מחזוריות לגבי החבורה המודולרית, או משתנות באופן מסויים פשוט כאשר מבצעים תנועה בחבורה (ושמקיימות עוד תנאי התנהגות טובה כאשר שואפים לנקודה באינסוף של המשולש) נקראות **תבניות מודולריות**. כתוצאה מצירוף האנליטיות והמחזוריות של תבניות מודולריות יש להן תכונות חזקות מאוד. מצד שני פונקציות ממקורות שונים ומשונים מתגלות כתבניות מודולריות. (לדוגמא, הטור $\sum_n p(n)z^n$ בו $p(n)$ הוא מספר האופנים להציג את n כסכום לא מסודר של מספרים טבעיים חיוביים, למשל $7 = 3 + 2 + 2$, מביא לתבנית מודולרית. בעזרת זה אפשר לקבל אינפורמציה על $p(n)$).

Eichler ו Shimura הראו דרך להציג עקומים אליפטיים מסויימים ע"י תבניות מודולריות. הצגה כזו ניתנת למעשה ע"י מעין הטלה מחזורית (המחזוריות היא לגבי תת־חבורה של החבורה המודולרית) מהמישור ההיפרבולי אל (קבוצת הנקודות הקומפלקסיות של) העקום האליפטי. את ההטלה הזו מאפיינת תבנית מודולרית, ואם תבנית זו היא במובן מסויים רציונלית, אז העקום הוא אריתמטי ואת הכל אפשר לעשות באופן אריתמטי, כלומר בעזרת אלגברה עם המספרים הרציונליים. יתר על כן, אז פונקצית־L של העקום ניתנת באופן פשוט ע"י התבנית המודולרית. במקרה כזה מבטיחות תכונות המחזוריות של התבנית המודולרית שלפונקצית־L זו יש המשכה לכל המישור הקומפלקסי, והתכונות החזקות של תבניות מודולריות נותנות מכשיר רב־כח לחקירת פונקצית־L זו והעקום האליפטי.

הבעיה היא רק: האם כל עקום אליפטי אריתמטי מתקבל בצורה כזו? עקום המתקבל כך נקרא עקום מודולרי, והשערת Taniyama-Shimura, שהוזכרה בתחילת המאמר, אומרת שכל עקום אליפטי אריתמטי הוא מודולרי. ברור שהמתמטיקאים העוסקים בנושא קיוו שהשערה זו נכונה, וכמובן לא היו מנסחים אותה אילולא הראה הניסיון שאיש לא הצליח לבנות עקום אליפטי לא מודולרי. בנוסף לכך, השערה זו אפשר לראות כמקרה פרטי במערכת השערות גדולה הרבה יותר של Langlands, ולכן עוד יותר חבל היה אילו היא היתה נופלת.

והנה כמו שציינו בתחילת המאמר Frey ב־1986 (ולפניו Hellegouarch עוד ב־1971) בנו מכל פתרון שלם a, b, c של משוואת פרמה $a^p + b^p = c^p$ עקום אליפטי אריתמטי

$$y^2 = x(x - a^p)(x - c^p)$$

שמתורת העקומים האליפטיים נראה שהוא לא יכול להיות מודולרי, מה שהוכיח K. Ribet ב־1987. אם כך, אי נכונות השערת פרמה גוררת קיום דוגמא לעקום אליפטי לא מודולרי. עד אז לא היתה שום סיבה להאמין בנכונות טענתו של פרמה, פרט להערת השוליים של פרמה ולכך שהיא הוכחה למעריכים כה גדולים שכל דוגמה נגדית צריכה להיות ענקית. עכשיו היה לפחות רצוי מאוד שהיא תהיה נכונה. כעת Wiles הוכיח את השערת Taniyama-Shimura, אמנם בהגבלה מסויימת על העקום האליפטי אבל העקום של Frey מקיים הגבלה זו, ובכך הוכח משפט פרמה.

לא דנתי בנושא אחר ששיחק תפקיד חשוב ־ הצגות גלואה. ברור שמנקודת ראותו של המתמטיקאי הוכחת השערת Taniyama-Shimura והרעיונות הקשורים לכך חשובים בהרבה ממשפט פרמה. כאמור, מה שהוכח הוא מקרה פרטי של השערות מרחיקות לכת שמקווים להוכיח אי־פעם. עבודתו של Wiles אין פירושה דווקא סיום תקופה, אלא גם התחלה.