אוניברסיטת תל-אביב
הפקולטה למדעים מדויקים ע"ש ריימונד ובברלי סאקלר
בית-הספר למדעי המתמטיקה
בית-הספר למדעי המחשב

מבוא למתמטיקה בדידה

פרופ' א. אברון

אוניברסיטת תל-אביב
הפקולטה למדעים מדויקים ע"ש ריימונד ובברלי סאקלר
בית-הספר למדעי המתמטיקה
בית-הספר למדעי המחשב

מבוא למתמטיקה בדידה

פרופ' א. אברון

נדפס בישראל תשס"א
Printed in Israel 2001

הקדמה

הספר שלפניכם מבוסס על קורס המועבר זה מספר שנים בבתי הספר למדעי המחשב ולמתמטיקה של אוניברסיטת תל-אביב. הספר מתחלק לשני חלקים. נושאי החלק הראשון מסווגים בדרך-כלל כ"לוגיקה ותורת הקבוצות". עם זאת, מטרתו העיקרית היא להוות מבוא ללימודי המתמטיקה בכלל. יש בדרך-כלל פער ניכר בין מה שמוקנה לתלמיד בבית-הספר העל-יסודי ובין מה שמצופה ממנו לדעת ולהבין, עת הוא מתחיל את לימודיו באוניברסיטה. החלק הראשון מהווה אפוא ניסיון לגשר על פער זה על-ידי לימוד מעמיק ויסודי של הלשון המתמטית, כמו גם מושגיה היסודיים של המתמטיקה המודרנית ושיטות ההוכחה שלה. תקוותי הגדולה היא, שהסטודנט ייעזר בחומר הנלמד כאן לצורך הבנת כל קורס מתמטי אחר, שהוא לומד או ילמד בעתיד.
חלקו השני של הספר מוקדש לשני נושאים קלאסיים של המתמטיקה הבדידה: קומבינטוריקה סופית ותורת הגרפים. חלק זה מוצג כהמשך ישיר של הראשון, ומיושמת בו המסגרת שנלמדה בחלק הראשון. עם זאת, מודגשים בו גם הקשרים עם ענפים אחרים של המתמטיקה (חשבון אינפיניטסימלי ואלגברה לינארית). בהתאם למטרה הפדגוגית של הספר, גם בחלק זה נשמרת רמה גבוהה של דיוק מתמטי, וניתנות הוכחות מלאות וריגורוזיות.

שני דברים ייחודיים לספר זה, שכדאי לציינם, הם:

(1) בספר מושם דגש רב על הצד הלשוני. כך מוקדשים בו פרקים מיוחדים לצורות הסימון השונות של קבוצות ופונקציות, ולשימוש הנכון בהן. פרק חשוב במיוחד מוקדש לשימוש הנכון במשתנים במתמטיקה. למיטב ידיעתי, אין פרק כזה בשום ספר לימוד אחר, למרות שמרבית השגיאות הלוגיות הנעשות על-ידי סטודנטים (ולא רק סטודנטים) נובעות מאי-הבנת נושא זה דווקא. סיבה נוספת לחשיבות הרבה המוענקת לצד הלשוני היא, שמרבית הסטודנטים בקורס הם ממדעי המחשב, ואלה עתידים לעבוד הרבה עם שפות פורמליות (כולל, כמובן, שפות תכנות).

(2) בספר נעשה שימוש מסיבי בסימון-למדא עבור פונקציות. הדבר מאפשר, למשל, הבחנה בין הביטוי "x^2" המסמן מספר, ובין הביטוי "$\lambda x.\, x^2$" המסמל את הפונקציה המתאימה לכל x את ריבועו. כמו כן מאפשר סימון-למדא תיאור קל של פונקציות, שהתחום (או הטווח) שלהן הוא קבוצה של פונקציות, כמו גם זיהוי קל של "משתנים", לעומת "פרמטרים", בתוך ביטויים מורכבים. כדאי

לציין, שצורות הסימון הנהוגות בטקסטים מתמטיים הן בחלקן מיושנות מאוד, ומובילות כמעט בהכרח לשגיאות ולחוסר הבנה. התפתחותם של הלוגיקה ושל מדעי המחשב הולידה צורות סימון מודרניות טובות ומדויקות הרבה יותר. קרוב לוודאי, שצורות אלו ייתפשו בעתיד, לאט ובהדרגה, את מקומן של הישנות (עם זה, בספר נלמדות גם צורות הסימון הקלאסיות, כדי שהתלמיד יוכל לקרוא ללא בעיות גם טקסטים אחרים). אני מאמין, שעצם הידיעה איך צריך לכתוב מסייעת מאוד גם כאשר אין כותבים כך בפועל. בנוסף, סימון-למדא מהווה את הבסיס של שפות תכנות פונקציונליות דוגמת SCHEME (השפה בה משתמשים כיום בקורס המבוא הראשון למדעי המחשב בביה"ס למדעי המחשב של אוניברסיטת תל-אביב, כמו גם באוניברסיטאות רבות אחרות ברחבי העולם).

לנוחות קוראי הספר, סומנו פרקים או סעיפים קשים יותר על-ידי כוכבית. על סעיפים אלו ניתן לדלג. בנוסף הוכנסו קטעים מסוימים בספר בין סימני אחוז כפולים: %% .. %%. קטעים אלו נועדו לקוראים מתקדמים בלבד. חלקים גדולים מהחומר סוכמו בטבלאות. רשימה מלאה של הטבלאות ניתן למצוא מיד לאחר תוכן העניינים.

הערה נוספת: הפרקים הראשונים, המוקדשים ללוגיקה, אינם קורס בלוגיקה מתמטית ואינם מהווים תחליף לקורס כזה. מטרתם רק לצייד את הסטודנט בכלים לוגיים פורמליים, שיסייעו לו בהמשך.

לסיום, ברצוני להודות לבתי הספר למתמטיקה ולמדעי המחשב על העזרה והתמיכה שהגישו לי בכתיבת ספר זה, בהדפסתו ובהוצאתו לאור.

ארנון אברון
בית הספר למדעי המחשב
אוניברסיטת תל-אביב

תוכן העניינים

רשימת הטבלאות

א. פרקי יסוד בלוגיקה

א.1. הצרנה

במתמטיקה עוסקים בהוכחות, והוכחות הינן סדרות של היסקים. מתוך הנחות מסוימות אנו מסיקים מסקנות. ממסקנות אלו אנו מסיקים עוד מסקנות, וכך אנו ממשיכים, עד שעולה בידינו להסיק את מה שרצינו. מה שמצדיק את ההיסקים בכל שלב ושלב הינם חוקי הלוגיקה – תורת ההגיון. כך, אם הראינו על מספר x מסוים, שהוא או גדול מ-1 או קטן מ-1– (למשל: הראינו ש- $x^2 > 1$), ואחר-כך אנו מראים שאינו יכול להיות קטן מ-1– (יען כי, לדוגמה, קיים y כך ש- $x = y^2$), או אז ביכולתנו להסיק, בעזרת ***כלל לוגי מתאים*** (עם שם נאה: "סילוגיזם דיסיונקטיבי"), ש- x גדול מ-1. לימוד כללי ההיסק הבסיסיים של הלוגיקה הינו אחת המטרות העיקריות של חלק זה. ברם, ממש כפי שהנוסחה לפתרון משוואה ריבועית הינה חסרת ערך לפתרון בעיה, כל עוד לא תורגמו נתוניה ללשון המשוואות, החושפת את הקשרים האריתמטיים ביניהם – ממש כך ידיעת כללים לוגיים עלולה שלא להועיל, בטרם תורגמו טענות הכתובות בשפות כמו עברית או אנגלית ללשון, בה נחשפים היחסים הלוגיים בין מרכיבי הטענה.

המשפט האחרון עלול להיראות סתום במקצת. ננסה להבהירו באמצעות דוגמה:

(1) **"למספר שלילי אין שורש"**

זוהי דוגמה לטענה מתמטית, המנוסחת באמצעות משפט בשפה העברית. בניגוד למקרים רבים אחרים, הנוסח הוא קצר וממצה. עם זאת, למי שמושג השורש (או המושג של מספר שלילי) אינו ברור (למשל: מחשב מתחיל), משפט זה הוא חסר מובן. חשוב עוד יותר הוא, שגם למי שמושגים אלו כן נהירים, עלולות להיות בעיות חמורות, עת יצטרך לטפל בטענה באופן אפקטיבי. לדוגמה, אם יתבקשו לנסח משפט, המביע את שלילתה של הטענה הנ"ל (כלומר משפט המבהיר, מה פירוש הדבר ש- (1) אינו נכון), עלולים רבים מהקוראים לסבור בטעות, שהתשובה הנכונה היא: "למספר שלילי יש שורש". תשובה כזו נובעת באופן טבעי מהמבנה התחבירי של השפה העברית, ומכך שהיפוכו של "אין" הוא "יש". כדי להיווכח, שתשובה זו אכן מוטעית, די אם נשנה את (1) ל: "למספר לא חיובי אין שורש". משפט זה הוא שקרי, כי 0 הינו מספר לא חיובי, שיש לו שורש (ולכן הוא מהווה דוגמה נגדית). עם זאת, גם המשפט "למספר לא חיובי יש שורש" הוא כמובן שקרי, ומכאן שאינו יכול להיות שלילת קודמו. שלילת טענה מסוימת הינה נכונה בדיוק כאשר אותה טענה אינה נכונה.

למעשה, שלילת (1) הינה: "קיים מספר שלילי בעל שורש". במקרה זה די אולי בהגיון אינטואיטיבי כדי להגיע לכך. עבור טענות מורכבות יותר (וכאלו יש למכביר במתמטיקה) תהליך ניסוח השלילה בצורה מועילה הוא קשה הרבה יותר-אם הוא נעשה באופן אינטואיטיבי. הוא נעשה קל ומכני, אם מבצעים תחילה תרגום לשפת הלוגיקה, ומפעילים אחר-כך את הכללים הלוגיים הרלוונטיים. הדבר דומה, שוב, למה שקורה בבעיות אלגבריות. בעיות פשוטות אפשר לפתור ישירות, ללא משוואות, בעזרת הגיון בריא. הדבר נעשה בלתי אפשרי, כשעוסקים בבעיות מורכבות יותר.

נפנה עתה לניתוח בשלבים של תוכן המשפט בדוגמה (1) ולתרגומו לשפה נוחה לטיפול, המשקפת תוכן זה.

ראשית, "שורש של מספר" הוא מספר, שאם נעלה אותו בריבוע נקבל את המספר הראשון. מספר שלילי הוא מספר קטן מאפס. אם ננסח את המשפט (1) במונחים יסודיים של כפל, שוויון, אי-שוויון וכו' (זה אינו תמיד נחוץ, אך מועיל כאן) נקבל:

(2) **"למספר הקטן מאפס אין מספר, שאם נעלה אותו בריבוע נקבל את המספר הראשון"**

זה בקושי מצלצל כעברית. בשלב זה יעדיף לכן המתרגם להחליף את המלה "אין" בביטוי כמו "אי אפשר למצוא". זה קריא יותר, ואפילו מקובל למדי בטקסטים מתמטיים, הנועדים לקוראים בעלי בגרות מתמטית[1]. עם זאת, אין ניסוח כזה מדויק, כי הנקודה פה אינה יכולתנו או אי יכולתנו למצוא שורש. ייתכן שחלק מהקוראים אינו מסוגל למצוא בכוחות עצמו (ללא עזרת מחשבון) את השורש של 12.71. מסקנת אותם קוראים מעובדה זו לא תהיה, ששורש כזה אינו קיים. הנקודה האמיתית כאן היא זו של *קיום* או אי-קיום השורש. מכאן שניסוח נכון יותר וטוב יותר ל- (2) הוא:

(3) **"למספר הקטן מאפס לא קיים מספר, שאם נעלה אותו בריבוע נקבל את המספר הראשון".**

זה עדיין מסורבל, ובהחלט לא קל להבנה. מי פה המספר הראשון ומי השני? טוב, כאן אפשר להסתדר עם זה (עם קצת רצון טוב). אך מה עם משפטים מסובכים יותר, בהם מדובר, נאמר, על ארבעה מספרים? ניסוחים מהסוג של (3) ייהפכו עבורם חיש מהר לחסרי תקווה! הפתרון לבעיה זו בכל טקסט מתמטי ראוי לשמו (וגם בשפות תכנות)

---

[1] לא לבלבל עם "בגרות במתמטיקה"!

הוא הכנסת *מזהים* (identifiers) עבור המספרים השונים בהם מדובר. למזהים אלו קוראים במתמטיקה בדרך-כלל *משתנים* (variables). על-ידי שימוש בהם נוכל לנסח את (3) כך:

(4) **"אם a מספר הקטן מאפס, אז לא קיים מספר b, כך שאם נעלה את b בריבוע, נקבל את a".**

זה כבר טוב לאין-ערוך, ובהחלט ניתן להבנה. עם זאת, כדאי לנצל את ההצלחה ולהשתמש במשתנים שהכנסנו לצורך ניסוח קצר וברור יותר של חלקי המשפט השונים:

(5) **"אם $a < 0$, אז לא קיים b, כך ש- $a = b^2$".**

בניסוח כזה, קרוב לוודאי, נמצא משפטים כאלה בטקסטים מתמטיים. השימוש בשפה העברית צומצם כאן למינימום של "אם", "אז", "לא", "קיים", "כך ש-". אלו כולם מושגים *לוגיים* המופיעים בכל ענפי המתמטיקה (וגם מחוץ לה, כמובן). כדאי מאוד לעשות כאן צעד נוסף ולהכניס סימונים קצרים מיוחדים גם עבור מושגים לוגיים בסיסיים אלו. בקורס זה נשתמש בסימנים הבאים:

(i) את הצירוף "אם – אז..." נחליף בסימן $\Rightarrow$. במלים אחרות: במקום "אם A אז B" נכתוב $B \Leftarrow A$ או $A \Rightarrow B$. $\Rightarrow$ נקרא קֶשֶר *הגרירה* (אימפליקציה בלע"ז), ומשפט מהצורה $A \Rightarrow B$ נקרא משפט גרירה. $A \Rightarrow B$ הוא גם תרגום עבור צירופים כמו "A גורר את B", "B נובע מ- A" ועוד.

(ii) במקום המלה "לא" נשתמש לעתים קרובות בסימן $\neg$. $\neg$ נקרא קֶשֶר *השלילה* (negation), ו- $\neg A$ יחליף גם צירופים כמו: "אין זה נכון ש- A" ואחרים.

(iii) במקום המלים "גם", "וגם" או סתם "ו-", נשתמש בסימן: $\wedge$. $\wedge$ נקרא בשם קֶשֶר *הקוניוקציה* (conjunction), ומשפט מהצורה $A \wedge B$ נקרא משפט קוניוקציה.

(iv) עבור המלה "או" נכניס את הסימן $\vee$. $\vee$ נקרא קֶשֶר *הדיסיונקציה* (disjunction), ומשפט מהצורה $A \vee B$ נקרא משפט דיסיונקציה.

(v) במקום הצירוף "אם ורק אם" (אם"ם, בקיצור) נשתמש בסימן $\Leftrightarrow$. $\Leftrightarrow$ נקרא קֶשֶר *השקילות* (אקוויוולנציה בלע"ז), ומשפט מהצורה $A \Leftrightarrow B$ נקרא משפט שקילות. $A \Leftrightarrow B$ הוא תרגום גם עבור צירופים כמו "A שקול ל- B", "A ו- B הם שקולים", "A הוא תנאי הכרחי ומספיק עבור B" ועוד.

(vi) במקום המלה "לכל" נשתמש בסימן $\forall$. ליתר דיוק: במקום, למשל, "לכל x מתקיים ש- ...", נכתוב $\forall x\,(\ldots)$. $\forall$ נקרא הכמת ה***אוניברסלי*** (כולל), ומשפט מהצורה $\forall x\,(\ldots)$ נקרא משפט אוניברסלי.

(vii) במקום המלים "יש", "קיים" וכו' נשתמש בסימן $\exists$. ליתר דיוק: במקום, למשל, "קיים x כך ש-...", נכתוב $\exists x\,(\ldots)$. $\exists$ נקרא הכמת ה***ישי*** (אקזיסטנציאלי), ומשפט מהצורה $\exists x\,(\ldots)$ נקרא משפט ישי.

טבלה א.1 מסכמת את מה שצריך לדעת על הקשרים והכמתים (כולל סימונים אלטרנטיביים מקובלים) והאופן בו מכניסים אותם למשפטים – מה שנקרא בטבלה "שימוש תחבירי".

טבלה א.1: הקשרים והכמתים העיקריים

שם	*מלה*	*סימן*	*סימונים אלטרנטיביים*	*שימוש תחבירי*	*ניסוחים בעברית*
גרירה (אימפליקציה)	אם ... אז	$\Rightarrow$	$\supset$, $\rightarrow$	$(A)\Rightarrow(B)$	אם A אז B, A גורר את B, B נובע מ- A
שלילה	לא	$\neg$	$\sim$	$\neg(A)$	אין זה נכון ש- A, ... לא ___
קוניוקציה	וגם	$\wedge$	&	$(A)\wedge(B)$	A וגם B, גם A וגם B, A ו- B
דיסיונקציה	או	$\vee$		$(A)\vee(B)$	A או B, A ו/או B, או ש- A או ש- B
שקילות (אקויולנציה)	אם ורק אם (אם"ם)	$\Leftrightarrow$	$\equiv$, $\leftrightarrow$	$(A)\Leftrightarrow(B)$	A אם ורק אם B, A שקול ל- B, A הוא תנאי הכרחי ומספיק ל- B
כמת אוניברסלי	כל, לכל	$\forall$	$(\ )$	$(x)(A)$, $\forall x(A)$	לכל x מתקיים ש- A
כמת אקזיסטנציאלי	קיים, יש	$\exists$		$\exists x(A)$	קיים x כך ש- A, יש x כך ש- A

הערות

א. לא ניכנס בשלב זה למהות המושגים של "קשר" ו"כמת", שהוזכרו למעלה. דבר זה נעשה בקורס בלוגיקה.

ב. בהגדרות האחרונות השתמשנו באותיות A ו- B כמשתנים עבור משפטים או נוסחאות.

ג. בהמשך הקורס נשתמש בניסוח המשפטים שלנו (בדרך-כלל) בצירופים בשפה העברית, לצורך הנוחות הפסיכולוגית של הקריאה. בסימונים המקוצרים נשתמש בכל עת, שהדבר ייראה מועיל.

השלב של הכנסת סימונים מקוצרים כרוך, למעשה, במשהו עמוק הרבה יותר מאשר עצם הקיצור. נשים לב, למשל, שהתאמנו סימון יחיד ($\Rightarrow$) עבור הצירוף "אם ___ אז ...", הניתן בעברית על-ידי שתי מלים, *שאינן* נכתבות זו אחר זו! ברור לכן, ש*התחביר* של טענות, הנכתבות בעזרת סימנים אלו, אינו זהה לזה של הטענות המקבילות בעברית. הכנסת הסימנים מלווה אם כן בניסוח כללי תחביר מדויקים, המורים איך יש להשתמש בהם. זה מוביל באופן טבעי ליצירת שפה *פורמלית* מדויקת, דומה מבחינות רבות לשפות תכנות, אליה מתרגמים טענות מתמטיות, הכתובות בשפה יותר מילולית. לתהליך תרגום זה קוראים בשם *הצרנה* (פורמליזציה). הצורה המוצרנת של טענות מתמטיות היא בדרך-כלל קצרה ותמציתית הרבה יותר מזו הלא-מוצרנת. מה שחשוב יותר הוא, שהצורה המוצרנת חושפת את המבנה ה*לוגי* האמיתי של טענות. יתר על כן, השימוש בצורות מוצרנות מאפשר ניסוח מדויק של חוקים לוגיים ושימוש בהם (בדומה לניסוח חוקים אלגבריים ושימוש בהם לצורך טיפול במשוואות ובזהויות). צורות מוצרנות הן גם הכרחיות לצורך טיפול ממוחשב בידע.

תיאור מלא ומדויק של שפות לוגיות פורמליות ניתן בקורס בלוגיקה מתמטית. בקורס המבוא הנוכחי נסתפק בדוגמאות ובהבנה אינטואיטיבית של התחביר. כדוגמה ראשונה נצרין באופן מלא את (5) למעלה:

$$\forall a((a < 0) \Rightarrow (\neg\exists b(b^2 = a))) \qquad (6)$$

אין ספק, שריבוי הסוגריים כאן מפריע מאוד לעיניים (ולהבנה) של בני-אנוש (למחשב, אגב, אין שום בעיה עם זה). לכן, ממש כמו באלגברה, שם אנו רושמים $a \cdot b + a \cdot c$ (או אפילו $ab + ac$) במקום $(a \cdot b) + (a \cdot c)$, גם כאן יש הסכמים שונים (אותם לא

נפרט במדויק) המאפשרים השמטת חלק ניכר מהסוגריים. משפטים כמו (6) נכתוב, למשל, בעתיד בצורה הבאה:

$$\forall a(a < 0 \Rightarrow \neg\exists b.\ b^2 = a) \quad (7)$$

הערה:
הנקודה אחרי b כמוה כסוגר שמאלי. הסוגר הימני המתאים מושמט, ומקומו נקבע על-פי ההגיון. (ליתר דיוק: הוא או בסוף המשפט או לפני הסוגר הימני הראשון, שהסוגר השמאלי המתאים לו נמצא לפני נקודה זו.)

את (7) יש לקרוא בעברית כך: "לכל a, אם a קטן מאפס, אז לא קיים b, כך ש- b^2 שווה ל- a".

לעתים קרובות מקצרים אפילו את (7) וכותבים רק:

$$\forall a < 0\ \neg\exists b.\ b^2 = a \quad (8)$$

את (8) יש לקרוא בעברית כך: "לכל a קטן מאפס לא קיים b, כך ש- b^2 שווה ל- a" או, בעברית טובה יותר: "לשום a הקטן מאפס לא קיים b, כך ש- b^2 שווה ל- a". צורה (8) הינה לכן לא רק קצרה יותר מ- (7), אלא אף קרובה יותר לעברית. בעברית אכן מקובל להגיד, בדרך-כלל, "לכל a קטן מאפס..." ולא "לכל a, אם a קטן מאפס אז...". אולם יש לזכור תמיד, ש- (7) הוא המשקף נכונה את המבנה הלוגי של המשפט, ו- (8) אינו אלא קיצור של (7). כללית, משפט מהצורה "$\forall a < b\ldots$" הינו תמיד קיצור של $\forall a(a < b \Rightarrow \ldots)$, ו- "$\exists a < b\ \ldots$" הינו תמיד קיצור של $\exists a(a < b \wedge \ldots)$. כדאי לשנן עובדות אלה היטב. (שוב, אנלוגיה מתחום האלגברה עשויה לעזור: בנוסחאות הקשורות במשוואות ריבועיות אנו כותבים לעתים קרובות "Δ" במקום $b^2 - 4ac$. אנו זוכרים אבל תמיד קיצור של מה הוא Δ, ושלצורך שימוש בנוסחאות יש בדרך-כלל להציב את $b^2 - 4ac$ במקום Δ.)

טבלה א.2 מרכזת את שמונת הניסוחים שמצאנו למשפט לעיל. מופיע שם ניסוח נוסף, תשיעי, אליו נגיע בהמשך (פרק א.3).

טבלה א.2: תשעה ניסוחים לטענה אחת

(1)	למספר שלילי אין שורש
(2)	למספר הקטן מאפס אין מספר, שאם נעלה אותו בריבוע נקבל את המספר הראשון
(3)	למספר הקטן מאפס לא קיים מספר, שאם נעלה אותו בריבוע נקבל את המספר הראשון
(4)	אם a מספר הקטן מאפס, אז לא קיים מספר b, כך שאם נעלה את b בריבוע נקבל את a
(5)	אם $a < 0$ אז לא קיים b כך ש- $b^2 = a$
(6)	$\forall a((a < 0) \Rightarrow (\neg\exists b(b^2 = a)))$
(7)	$\forall a(a < 0 \Rightarrow \neg\exists b.\ b^2 = a)$
(8)	$\forall a < 0\ \neg\exists b.\ b^2 = a$
(9)	$\forall a < 0\ \forall b.\ b^2 \neq a$

הערות

א. אם נשווה את (7) למשפט המקורי (1), ניווכח, שהניסוח המקורי בעברית לא כלל שום גרירה (ושום אם ____ אז ...). גם המלה "לכל" לא הופיעה שם כלל. הכנסת ה"אם ____ אז ..." נעשתה באופן טבעי למדי במעבר מ- (3) ל- (4), בעוד "לכל" הוכנס באופן מפורש בהצרנה (בצורה של "$\forall$"), כשעברנו מ- (5) ל- (6). הרהור קצר בעניין יבהיר, שמשפט (5) מתייחס אכן לכל a. כשאנו אומרים "אם a קטן מאפס אז ל- a יש תכונה P", כוונתנו היא, *שכל* מספר a שננסה לבדוק, אם יתברר שהוא קטן מאפס, אזי ניווכח שיש לו גם התכונה P. השימוש ב"אם" בשפה העברית יש לו לעתים קרובות ביותר משמעות אוניברסלית, דהיינו איזה "כל" שהוא מסתתר מאחוריו. תהליך ההצרנה חושף "כל" זה. במקרה הנוכחי הוא חשף גם, שמדובר בכלל בגרירה. בכל זה אין מקריות. כבר הדגשנו, שתהליך ההצרנה חושף את המבנה הלוגי הפנימי ("מבנה התשתית", בלשונו של הבלשן הנודע נועם חומסקי) המסתתר מאחורי המבנה החיצוני ("מבנה השטח") של

טענות בעברית (או כל שפה טבעית אחרת). מבנה לוגי פנימי זה אינו דומה בהכרח למבנה הטענה בעברית!

ב. דוגמה נוספת להבדל בין התחביר המקורי של משפט לבין תחביר גרסתו המוצרנת נותן קשר השלילה ($\neg$), המקביל למלה "לא". כשאנו אומרים בעברית ש- "x הוא לא גדול מ- y" (או ש- "x אינו גדול מ- y", שזו עברית יפה יותר), ה"לא" מופיע באמצע המשפט, אחרי ה- "x". בשפה המוצרנת כותבים פשוט $\neg(x > y)$ או רק $\neg x > y$. סימן השלילה מופיע לפני כל מה שהוא בא לשלול. דבר זה משקף נכונה את המבנה הלוגי ומונע טעויות.

ג. התחביר של (6) (שהוא (7) כתוב בצורה מלאה) אינו היחיד הבא בחשבון. בשפת התכנות LISP וקרובות משפחתה (כמו Scheme) כותבים, למשל, $(+, a,b)$ במקום $a + b$ ו- $(\Rightarrow,A,B)$ במקום $(A) \Rightarrow (B)$. (6) או (7) ייכתב שם בדרך-כלל כך:

$$(\forall, a, (\Rightarrow, (<, a, 0)\ , (\neg, (\exists, b, (=, (*\,*, b, 2), a))))$$

לאמתו של דבר, צורה זו אכן משקפת טוב עוד יותר את המבנה התחבירי האבסטרקטי של המשפט ומקלה על המכניזציה של עבודת הניתוח התחבירי (parsing). עם זאת, למי ש- LISP אינה שפת אמו, צורה זו קשה יותר להבנה.

ד. ב- (6) (או (7)) מופיעות הנוסחאות $a < 0$ ו- $b^2 = a$. המשפט כולו נבנה מנוסחאות אלו בעזרת הסימונים עבור המושגים הלוגיים וכללי התחביר שלהם. אנו אומרים לכן, ש- $b^2 = a$ ו- $a < 0$ הן הנוסחאות ***האטומיות*** בפסוק הנ"ל. כשאנו עושים הצרנה, עלינו להחליט גם, מה נרצה לראות בגדר נוסחה אטומית ומה כנוסחאות הדורשות ניתוח נוסף. לא תמיד עומד לרשותנו סימון סטנדרטי ומקובל, כמו במקרה של $a^2 = b$ ו- $a < 0$. כאשר זה המצב עלינו להכניס סימונים משלנו.

כל זה נשמע, קרוב לוודאי, די מעורפל. כיוון שאין זה קורס בלוגיקה, ומטרתנו כאן היא שימוש בהצרנות ככלי-עזר, לא נלמד את הנושא באופן שיטתי, אלא נסתפק בדוגמאות, שיבהירו (כך אנו מקווים) את העניין וידגימו, איך מתבצעות הצרנות בפועל. הדוגמאות מורכבות ממשפט בעברית המלווה בהצרנה מתאימה.

♦ **דוגמה 1:**

כל בני האדם הם בני תמותה.

$$\forall x(\text{human}(x) \Rightarrow \text{mortal}(x))$$

♦ **דוגמה 2:**

יהודי הוא מי שאמו יהודייה, או עבר גיור.

$$\forall x(\text{Jew}(x) \Leftrightarrow (\text{Jew}(\text{mother\_of}(x)) \vee \text{Guyar}(x)))$$

♦ **דוגמה 3:**

לכל משוואה ריבועית, שבה הדיסקרימיננטה אינה שלילית, יש פתרון.

$$\forall a \forall b \forall c((a \neq 0 \wedge \neg (b^2 - 4a \cdot c < 0)) \Rightarrow \exists x . a \cdot x^2 + b \cdot x + c = 0)$$

♦ **דוגמה 4:**

דרך שתי נקודות עובר קו ישר.

$$\forall A \forall B(\text{point}(A) \wedge \text{point}(B) \wedge A \neq B \Rightarrow \exists \ell\, (\text{line}(\ell) \wedge \text{on}(A, \ell) \wedge \text{on}(B, \ell)))$$

♦ **דוגמה 5:**

לשני ישרים יש לכל היותר נקודה אחת משותפת.

$$\forall \ell \forall m(\text{line}(\ell) \wedge \text{line}(m) \wedge \ell \neq m \Rightarrow \neg \exists A \exists B(A \neq B \wedge \text{point}(A) \wedge \text{point}(B) \wedge$$
$$\wedge \text{on}(A, \ell) \wedge \text{on}(B, \ell) \wedge \text{on}(A,m) \wedge \text{on}(B,m)))$$

כדאי לשים לב, איך בכל ההצרנות כמעט מלווה המלה "כל" ($\forall$) בגרירה ($\Rightarrow$), בעוד מושג הקיום ($\exists$) מלווה בקוניוקציה ($\wedge$).

הערה:

אמצעי נפוץ כדי למנוע התארכות טענות (כמו בדוגמה 5) הוא שימוש במשתנים מסוג מיוחד עבור עצמים מסוג מיוחד. כך, אם נסכים שאותיות לטיניות קטנות תשמשנה כמשתנים עבור ישרים, בעוד אותיות לטיניות גדולות תשמשנה כמשתנים עבור נקודות, נוכל לקצר את ההצרנה בדוגמה 5 ל-

$$\forall \ell \forall m(\ell \neq m \Rightarrow \neg \exists A \exists B(A \neq B \wedge \text{on}(A, \ell) \wedge \text{on}(B, \ell) \wedge \text{on}(A,m) \wedge \text{on}(B,m)))$$

עקרון הקיצור כאן הוא ש- $\forall \ell \ldots$ הוא קיצור של $\forall \ell (\text{line}(\ell) \Rightarrow \ldots$, $\exists \ell \ldots$ הוא קיצור של $\exists \ell (\text{line}(\ell) \wedge \ldots)$, ו- $\forall A \ldots$ הוא קיצור של $\forall A(\text{point}(A) \Rightarrow \ldots)$. אמצעי זה אינו זר לקוראים, הרגילים מן הסתם לשימוש באותיות i, j, k, ℓ, m, n כמשתנים עבור מספרים טבעיים, α, β, γ כמשתנים עבור זוויות וכו'. כדאי לשים לב, שאם נרחיב את ההצרנה המקוצרת של דוגמה 5 לפי הכללים לצורה מלאה, לא נקבל בדיוק את ההצרנה המקורית, אלא משהו השקול לה לוגית (על סמך כללים שנלמד בהמשך).

♦ **דוגמה 6**

לכל מספר ראשוני ניתן למצוא מספר ראשוני גדול יותר.

בדוגמה זו ניתנות 3 הצרנות, מותאמות לדרגת הפירוט אליה נרצה להגיע. תרגום מיידי ייראה כך:

$$\forall n(\text{prime}(n) \Rightarrow \exists k(\text{prime}(k) \wedge n < k)) \qquad \text{(I)}$$

בהצרנה זו מבוטאת העובדה ש- n הוא מספר ראשוני על-ידי נוסחה אטומית: $\text{prime}(n)$. את "ניתן למצוא" תרגמנו ל- "$\exists$" (כלומר "קיים"); הקיצורים שלנו אינם משקפים את ההבדל הדק הקיים כאן, ומרבית הטקסטים המתמטיים בלאו הכי מתעלמים ממנו. כדאי גם לשים לב לכך, שבהצרנה (I) שוב מתורגמת המלה "לכל" לשילוב של $\forall$ ו- $\Rightarrow$, בעוד "קיים" – לשילוב של $\exists$ ו- $\wedge$.

הבעיה בהצרנה זו הינה, שמושג "המספר הראשוני" אינו נחשב בדרך-כלל למושג בסיסי, אלא למושג, *שמגדירים* אותו בעזרת מושגים בסיסיים יותר. מספר ראשוני מוגדר כמספר גדול מאחד, שמתחלק רק בעצמו ובאחד. הצרנת הגדרה זו היא:

$$\text{Prime}(n) =_{Df} \forall i(i \setminus n \Rightarrow i = 1 \vee i = n) \wedge n > 1$$

בשורה האחרונה, "\" הוא סימן מקובל ל"מחלק את" (כלומר $i \setminus n$ פירושו "i מחלק את n" או "n מתחלק ב- i"). הסימן $=_{Df}$ פירושו, שאגף ימין מהווה הגדרה של מה שכתוב בצד שמאל. אם נציב הגדרה זו ב- (I) למעלה, נקבל את ההצרנה היותר מפורטת הבאה:

$$\text{(II)} \quad \forall n((\forall i(i \setminus n \Rightarrow i = 1 \vee i = n) \wedge n > 1) \Rightarrow \exists k(\forall i(i \setminus k \Rightarrow$$
$$\Rightarrow i = 1 \vee i = k) \wedge k > 1 \wedge n < k))$$

בהצבה זו החלפנו, כמובן, לא רק את $\text{prime}(n)$ בהגדרתו, אלא גם את $\text{prime}(k)$.

גם מושג ההתחלקות אינו נחשב, בדרך-כלל, מושג בסיסי. אנו אומרים שמספר n מתחלק במספר k, אם n שווה למכפלה של k באיזשהו מספר (שלם). נצרין זאת:

$$k \setminus n =_{Df} \exists i(n = k \cdot i)$$

אם נציב הגדרה זו ב- (II), נקבל את ההצרנה המפורטת והבסיסית מאוד הבאה:

$$\text{(III)}\qquad \forall n((\forall i(\exists j(n = i{\cdot}j) \Rightarrow i = 1 \vee i = n) \wedge n > 1) \Rightarrow$$
$$\Rightarrow \exists k(\forall i(\exists j(k = i{\cdot}j) \Rightarrow i = 1 \vee i = k) \wedge k > 1 \wedge n < k))$$

אם נכניס למקומם את כל הסוגריים שהשמטנו (דבר שעלול להיות חיוני פה כדי להבטיח קריאה נכונה), נקבל:

$$\forall n(((\forall i((\exists j(n = i{\cdot}j)) \Rightarrow ((i = 1) \vee (i = n)))) \wedge (n > 1)) \Rightarrow$$
$$\Rightarrow (\exists k(((\forall i((\exists j(k = i{\cdot}j)) \Rightarrow ((i = 1) \vee (i = k)))) \wedge (k > 1)) \wedge (n < k))))$$

יש שתי הערות, שחשוב לציין ביחס להצרנה האחרונה. ראשית, כשהצבנו את ההגדרה $i\backslash n$ ב- (II) שינינו את המשתנה i שהופיע בהגדרה למשתנה אחר: j. הדבר נעשה כיוון שהמשתנה i היה כבר "תפוס" בתוך (II) (ובתוך $i\backslash n$). העיקרון כאן הוא, שאין הבדל בין

$$k\backslash n =_{Df} \exists i(n = k \cdot i)$$

לבין

$$i\backslash n =_{Df} \exists j(n = i \cdot j)$$

ממש כשם שאין הבדל בין הזהות האלגברית $a^2 - b^2 = (a - b)(a + b)$ ובין הזהות $x^2 - y^2 = (x - y)(x + y)$ שינוי כזה של שמות משתנים, כדי למנוע שימוש באותו משתנה במובנים שונים (במסגרת אותו קונטקסט), נעשה תכופות במתמטיקה. כך כבר בבית-הספר התיכון נהוג לייעץ לתלמידים מתחילים, שכאשר מתבקשים הם לפרק את $4a^2 - 9b^2$ לפי הנוסחה $a^2 - b^2 = (a - b)(a + b)$, כדאי שישנו תחילה את הנוסחה ל- $x^2 - y^2 = (x - y)(x + y)$ ואז יציבו $x = 2a$, $y = 3b$. זאת, כיוון שבהצבה ה"ישירה" $a = 2a$, $b = 3b$ יש למשתנה a (למשל) מובנים שונים בשני האגפים של $a = 2a$. (לשינוי המשתנים כאן קוראים כלל α, ועוד נדון בו בהרחבה בהמשך.)

הערה שנייה, מובנת הרבה יותר, היא, ש- (III) הוא בלתי ניתן לקריאה ולהבנה על-ידי בני-תמותה רגילים. זוהי בדיוק הסיבה מדוע אנו משתמשים ב*הגדרות* בתחומי המדע השונים. עקרונית, ניתן לבטא כל משפט בעזרת מושגי היסוד שלנו, ואין שום צורך בהגדרות. מעשית, נקבל כך משפטים מדרגת סיבוכיות כזו, שלא נוכל בשום פנים להבינם, או אפילו לנסחם ללא שגיאות. בדוגמה הנוכחית, למשל, נתחיל תמיד בהצרנה (I), ורק אם יתעורר הצורך נציב את ההגדרה של $\mathrm{prime}(n)$ או של $i\backslash k$.

א.2. המשמעות של הקשרים והכמתים

הסמנטיקה של הקשרים והכמתים (דהיינו: מובנם) מוסברת בדרך-כלל על-ידי תיאור התנאים לכך, שטענה הנבנית בעזרתם תהיה נכונה (לתנאים אלו קוראים תנאי האמת של הטענה (truth conditions)). תיאור תנאים אלו פשוט מאוד ואינטואיטיבי במקרים של $\wedge$, $\vee$ ו- $\neg$:

1. $A \wedge B$ נכונה כאשר (ורק כאשר) שני מרכיבי הקוניוקציה, A ו- B הינם נכונים.
2. $A \vee B$ נכונה כאשר (ורק כאשר) לפחות אחד משני המרכיבים של הדיסיונקציה (דהיינו A ו- B) הינו נכון. (משמעות "או" בטקסטים מתמטיים היא לכן מה שמצוין לפעמים ב"ו/או" בחוזים משפטיים).
3. $\neg A$ נכון כאשר A אינו נכון (ורק אז).

אפשר לסכם את היחס בין נכונות/אי נכונות $A \wedge B$, $A \vee B$ ו- $\neg A$ לבין נכונות מרכיביהם בעזרת הטבלאות הבאות, המכונות "טבלאות אמת". בטבלאות אלו "t" הוא קיצור של "true" (אמיתי) ו- "f" הוא קיצור של "false" (שקרי).

A	$\neg A$
t	f
f	t

A	B	$A \vee B$
t	t	t
t	f	t
f	t	t
f	f	f

A	B	$A \wedge B$
t	t	t
t	f	f
f	t	f
f	f	f

כדי להבין את הטבלה, ניקח לדוגמה את השורה השלישית בטבלה של $\vee$. משמעותה היא, שאם A טענה שקרית ו- B טענה אמיתית, אז $A \vee B$ הינה טענה אמיתית.

הרעיון שעומד מאחורי טבלאות האמת הוא, שערך האמת של משפט מורכב תלוי אך ורק בערכי האמת של מרכיביו ולא בשום דבר אחר. עקרון זה נראה ברור למדי עבור הקשרים $\wedge$, $\vee$ ו- $\neg$. אך מה בדבר קשר הגרירה? האם גם כאן תלויה נכונות או אי-נכונות $A \Rightarrow B$ אך ורק בערכי האמת של A ו- B? במלים אחרות: האם גם את המובן של $\Rightarrow$ ניתן לתאר בעזרת טבלת אמת? אינטואיטיבית, התשובה עלולה במבט

ראשון להיות שלילית. כדי להצדיק את טבלת האמת, שבכל זאת נביא בהמשך, יועיל אם נברר תחילה את המשמעות האינטואיטיבית של הכמתים.

4. $\forall yP$ נכון בתחום מסוים, אם לכל אלמנט באותו תחום יש התכונה המתוארת על-ידי P. לדוגמה $\forall z(z^2 \neq -1)$ נכון בתחום של המספרים הממשיים, אך אינו נכון בתחום המספרים המרוכבים (כי אין זה נכון, למשל, ש- $i^2 \neq -1$). אינטואיטיבית מייצג $\forall$ מעין קוניוקציה אינסופית. כך המובן של $\forall x(x + 1 > x)$ בתחום המספרים הטבעיים זהה ל-

$$(0 + 1 > 0) \wedge (1 + 1 > 1) \wedge (2 + 1 > 2) \wedge (3 + 1 > 3) \wedge \ldots$$

5. $\exists yP$ נכון בתחום מסוים, אם קיים אלמנט באותו תחום עם התכונה המתוארת על-ידי P. כך לדוגמה $\exists x(x \cdot 2 = 3)$ נכון בתחום של המספרים הממשיים (ל- $x = 3/2$ התכונה המבוקשת), אך לא בתחום המספרים הטבעיים. אינטואיטיבית מייצג $\exists$ מעין דיסיונקציה אינסופית. כך, לדוגמה, נכונות $\exists z(z^2 - 7z + 12 = 0)$ בתחום המספרים הטבעיים זהה ל-

$$(0^2 - 7 \cdot 0 + 12 = 0) \vee (1^2 - 7 \cdot 1 + 12 = 0) \vee (2^2 - 7 \cdot 2 + 12 = 0) \vee \ldots$$

הפסוק לכן נכון בתחום זה, כי המרכיבים הרביעי והחמישי של "דיסיונקציה אינסופית" זו הינם נכונים (אחד מהם היה מספיק, כמובן).

6. נפנה עתה לקשר הגרירה. נתחיל במספר דוגמאות.

א. "אם ריבוע של מספר אחד גדול או שווה מריבוע של מספר אחר, אז או ששניהם שליליים, והראשון גדול או שווה מהשני, או ששניהם חיוביים, והשני גדול או שווה מהראשון" (*שקרי* במספרים הממשיים!).
הצרנה: $\forall x \forall y(x^2 \geq y^2 \Rightarrow ((x \geq 0 \wedge y \geq 0 \wedge x \geq y) \vee (x \leq 0 \wedge y \leq 0 \wedge x \leq y)))$

ב. "אם n מספר זוגי, אז $2^n - 1$ מתחלק ב-3" (אמיתי בטבעיים).
הצרנה: $\forall n((\exists j.n = 2j) \Rightarrow (\exists k.2^n - 1 = 3k))$
או: $\forall n(2 \backslash n \Rightarrow 3 \backslash 2^n - 1)$

ג. "אם $n > 1$, אז $n^2 > 0$" (אמיתי בטבעיים).
הצרנה: $\forall n(n > 1 \Rightarrow n^2 > 0)$

ד. "אם $n > 0$, אז $n^2 > 1$" (שקרי בטבעיים).
הצרנה: $\forall n(n > 0 \Rightarrow n^2 > 1)$

בכל הדוגמאות הללו פורק בתהליך ההצרנה הצירוף "אם ___ אז ... *לשילוב* של הקשר $\Rightarrow$ והכמת $\forall$. כבר בפרק הקודם הדגשנו, שזוהי תופעה כללית. פרט לדוגמאות בספרי לימוד של לוגיקה, לעולם לא נמצא בטקסטים מתמטיים נורמליים טענות מוזרות כמו: "אם $3 \neq 2$ אז $5 = 5$". למשפטי אם-אז יש תמיד איזשהו תוקף אוניברסלי, ולהצרנתם יש בדרך-כלל הצורה:

$$\forall x_1 \forall x_2 \ldots \forall x_n (P \Rightarrow Q).$$

עתה, לפי המשמעות של $\forall$, טענה מהסוג $\forall x(P(x) \Rightarrow Q(x))$ (למשל) נכונה בתחום מסוים, אם לכל איבר a בתחום, הטענה $P(a) \Rightarrow Q(a)$ היא נכונה. כך, למשל, הטענה מדוגמה (ג) נכונה במספרים הטבעיים, אם ורק אם כל אחד מאינסוף הפסוקים הבאים הינו אמיתי:

$$\begin{array}{ll} \text{(i)} & 0 > 1 \Rightarrow 0^2 > 0 \\ \text{(ii)} & 1 > 1 \Rightarrow 1^2 > 0 \\ \text{(iii)} & 2 > 1 \Rightarrow 2^2 > 0 \\ & \vdots \end{array}$$

עתה הטענה המובעת בדוגמה (ג) היא ללא ספק נכונה אינטואיטיבית.[2] כיוון שכך, ברור, שעלינו לקבל ש- $0 > 1 \Rightarrow 0^2 > 0$ נכון גם כן. ברם, הסיבה היחידה האפשרית לנכונות פסוק זה היא העובדה, שהרישא שלו, $0 > 1$, הינה שקרית. בדיקת המקרים האחרים (ודוגמאות אחרות) מראה, שתמיד טענה מהצורה $\forall x(P(x) \Rightarrow Q(x))$ מתקבלת כנכונה בתחום מסוים, אם לכל a בתחום, או ש- $P(a)$ שקרי, או ש- $Q(a)$ אמיתי. במלים אחרות, $\forall x(P(x) \Rightarrow Q(x))$ שקול ל- $\forall x(\neg P(x) \vee Q(x))$. זה מוביל באופן בלתי נמנע לזיהוי $A \Rightarrow B$ עם $\neg A \vee B$ ולטבלת האמת הבאה עבור $\Rightarrow$:

A	B	$A \Rightarrow B$
t	t	t
f	t	t
t	f	f
f	f	t

[2] היא לא *האינפורמטיבית* ביותר האפשרית, כי $\forall n(n > 0 \Rightarrow n^2 > 0)$ ללא ספק אינפורמטיבית יותר ועדיין נכונה. עובדה זו אינה גורעת אבל במאומה מאמיתות הטענה בדוגמה זו.

פסוק (i) למעלה מהווה דוגמה לשורה האחרונה בטבלה זו, פסוק (ii) – לשורה השנייה, ופסוק (iii) – לשורה הראשונה. אשר לשורה השלישית, היא היחידה בטבלה, שהינה מובנת מאליה. כדוגמה ספציפית יכול לשמש הפסוק בדוגמה (ד) למעלה ($(\forall n(n > 0 \Rightarrow n^2 > 1)$). ברור שפסוק זה שקרי, כי 1 מהווה דוגמה נגדית: $1 > 0 \Rightarrow 1^2 > 1$ אינו נכון לפי השורה השלישית בטבלת האמת של $\Rightarrow$, וגם לא לפי כל אינטואיציה שהיא בדבר גרירה.

הפירוש של $\Rightarrow$ לפי טבלת האמת האחרונה נתקל בדרך-כלל בתמיהה ובחוסר הסכמה על-ידי מי שפוגש בו לראשונה. מקור העניין הוא באי-הבנה. למעשה, הקשר $\Rightarrow$, בו אנו משתמשים בלוגיקה המתמטית (והידוע בשם: "גרירה מטריאלית"), ***אינו*** תרגום מלא ומדויק של "אם... אז" בעברית, אלא לכל היותר אפרוקסימציה ראשונה שלו. תרגום מדויק ניתן, כמעט תמיד, על-ידי שילוב של $\forall$ ו- $\Rightarrow$. גם כאן תהליך ההצרנה חושף מבנה פנימי עמוק יותר, המוסתר עת משתמשים בשפה טבעית. במקרה זה מדובר בפירוק מושג הגרירה למושגים בסיסיים (ופשוטים) יותר.

7. לבסוף, משמעות $\Leftrightarrow$ היא פשוט גרירה בשני הכיוונים: $A \Leftrightarrow B$ נכון אם (ורק אם) $A \Rightarrow B$ ו- $B \Rightarrow A$ נכונים שניהם. במלים אחרות, $A \Leftrightarrow B$ הוא קיצור של $(A \Rightarrow B) \wedge (B \Rightarrow A)$. קל למצוא מזה את טבלת האמת המתאימה ל- $\Leftrightarrow$, ואנו משאירים זאת לקורא.

טבלאות אמת אינן רק כלי להסבר המובן של קשרים, אלא גם כלי שניתן להשתמש בו לצורך היסקים ולצורך בדיקת טיעונים. הדבר אפשרי, כאשר ההיסקים והטיעונים בהם מדובר מתבססים בעיקר על הקשרים, בעוד לכמתים תפקיד משני (אם בכלל). העיקרון פשוט. באופן כללי, אנו אומרים, שטענה B נובעת לוגית מטענות $A_1, \ldots, A_n$ (סימון: $A_1, \ldots A_n \vdash B$) אם אחרי שהצרנו את כולן בשפה פורמלית מתאימה, מקבלים ש- B חייב להיות נכון בכל פירוש, לפיו $A_1, \ldots, A_n$ נכונים כולם. כאשר רק הקשרים מעורבים, "פירוש" אינו יותר מאשר מתן ערכי אמת לנוסחאות האטומיות השונות. טבלאות האמת מאפשרות אז לנו לבדוק באילו פירושים (אם בכלל) $A_1, \ldots, A_n$ יוצאים נכונים כולם, ואם B אכן נכון גם הוא בכל הפירושים הללו.

נביא מספר דוגמאות ספציפיות, שמהן יהיה ברור, כיצד הדבר נעשה באופן שיטתי.

♦ **דוגמה 1:**

ראובן ושמעון דנים בסוגייה החשובה: מי זכה באליפות ישראל בכדורגל בשנת 1979?
ראובן: אני לא זוכר מי זו היתה בדיוק, אבל אני יודע בוודאות, שזו היתה אחת הירושלמיות, הפועל או בית"ר. את הפקק, שנתקעתי בו באותו ערב בדרך לירושלים, לא אשכח לעולם!
שמעון: לא ייתכן, שזו היתה הפועל ירושלים. במחצית השנייה של שנות השבעים היא שיחקה בכלל בליגה הארצית, לא הלאומית!
ראובן: אם כך, היתה זו בית"ר ירושלים!

בעזרת ניתוח אנליטי הסיק ראובן מסקנה לוגית מהנתונים, עליהם היתה הסכמה. נוכל לבדוק את תקפות מסקנתו באופן הבא. נסמן:

P – "בית"ר ירושלים זכתה באליפות ב-1979"

Q – "הפועל ירושלים זכתה באליפות ב-1979"

בסימונים אלו, האינפורמציה שסיפק ראובן היא: $P \vee Q$
האינפורמציה שסיפק שמעון היא: $\neg Q$
המסקנה של ראובן היא: P

לדעתו של ראובן, לפי הסיפור, $P \vee Q\ ,\ \neg Q \vdash P$ (כלומר: P נובע לוגית מ- $P \vee Q$ ומ- $\neg Q$). נוכל לאמת את צדקתו על-ידי טבלאות אמת, בהן נבדוק את כל האפשרויות למתן ערכי אמת ל- P ול- Q, וחישוב ערכי האמת של $\neg Q$, $P \vee Q$ ו- P בכל אחת מהן:

P	$\neg Q$	$P \vee Q$	Q	P	
t	f	t	t	t	
t	t	t	f	t	*
f	f	t	t	f	
f	t	f	f	f	

אנו רואים, שהשורה היחידה, שבה הנתונים $P \vee Q$ ו- $\neg Q$ נכונים שניהם, היא השורה השנייה, ובאמת גם P נכונה בשורה זו. מכאן, שראובן צדק.

הערות

א. באומרנו, שהטבלה מראה שראובן צדק, כוונתנו רק להיסק הלוגי שלו. הטבלה אינה מוכיחה, שאכן בית"ר ירושלים זכתה באליפות ב-1979. היא מוכיחה רק, שאם הנתונים נכונים, אז זה המצב. גם אם אחד מהשניים טעה באינפורמציה

שסיפק, גם אז היתה עדיין המסקנה נובעת לוגית מהנתונים (אך לא היתה אז נכונה בהכרח). רק כאשר אנו יודעים בוודאות, שהנתונים נכונים, משמשת הטבלה ערובה לכך, שגם המסקנה נכונה.

ב. העמודה האחרונה בטבלה למעלה היא חזרה על העמודה הראשונה, ובמקרים מעשיים היינו מוותרים עליה ומתייחסים לעמודה הראשונה גם כעמודת המסקנה.

♦ דוגמה 2:

אוהד של אברטון מסביר למה אינו יכול לסבול את אוהדי ליברפול: "אם מישהו הוא סוציאליסט, אז הוא אוהב את הצבע האדום. אוהדי ליברפול אוהבים את הצבע האדום. מכאן שהם סוציאליסטים!"

אם נסמן כאן ב- A את הטענה, שאוהד ליברפול אוהב את הצבע האדום, וב- B את הטענה שאוהד ליברפול הוא סוציאליסט, הרי ברור, שבאופן בסיסי כוונת אותו אוהד חכם היא, שכיוון ש- $B \Rightarrow A$ ו- A נכונים, אז גם B נכון. במלים אחרות, לדעתו $B \Rightarrow A, A \vdash B$.

כדי להיווכח, שלא זה המצב, די אם ניקח את הפירוש, שבו A מקבל את הערך t, ו- B את הערך f. לפי טבלת האמת של $\Rightarrow$, $B \Rightarrow A$ מקבל את הערך t בפירוש זה, ולכן זהו פירוש, שבו שתי ההנחות נכונות, והמסקנה לא. נשים לב, שלא היינו חייבים כאן לבדוק את כל השורות של טבלת האמת. כדי *להפריך* טיעון, די שנספק שורה אחת המפריכה אותו. רק כדי *לאמת* אותו יש לבדוק את כולן.

האוהד שלנו יוכל לטעון אבל, כמו כל פוליטיקאי מצוי, ששמנו בפיו דברים, אותם הוא לא אמר. הוא מעולם לא אמר ש- $B \Rightarrow A$. זוהי רק פרשנות מסולפת של דבריו. פורמלית – הצדק אתו. נעשה לכן הצרנה *מדויקת* של דבריו. לצורך זאת נשתמש בקיצורים הבאים:

$\mathrm{S}(x)$ פירושו ש- x הוא סוציאליסט

$\mathrm{R}(x)$ פירושו ש- x אוהב את הצבע האדום

$\mathrm{L}(x)$ פירושו ש- x הוא אוהד של ליברפול

הטיעון של אוהד אברטון הוא:

$$\forall x(\mathrm{S}(x) \Rightarrow \mathrm{R}(x)) \quad , \quad \forall x(\mathrm{L}(x) \Rightarrow \mathrm{R}(x)) \ \vdash \ \forall x(\mathrm{L}(x) \Rightarrow \mathrm{S}(x))$$

בדיקת טיעון זה, כמו שהוא, בעזרת טבלאות אמת בלבד, אינה אפשרית, כיוון שהכמת האוניברסלי מעורב פה בצורה חזקה. אולם ניתן להראות (וזה ברור למדי באופן אינטואיטיבי), שאפשר כאן להתרכז בבן-אדם אחד (נקרא לו משה), לנסח את הנתונים והמסקנה ביחס אליו בלבד, והטיעון המקורי יהיה תקף אם ורק אם הטיעון ביחס למשה תקף. לאור זאת עלינו לבדוק אם:

$$(*) \quad S(Moshe) \Rightarrow R(Moshe) \ , \ L(Moshe) \Rightarrow R(Moshe) \vdash L(Moshe) \Rightarrow S(Moshe)$$

לגבי טיעון זה ניתן להפעיל את שיטת טבלאות האמת, כיוון שהכמתים אינם מעורבים בו. טבלת אמת מלאה תכיל פה 8 שורות (ראה דוגמה 3), אולם די שנציין, כי אם ניתן ל- R(Moshe) את הערך t, ל- L(Moshe) את הערך t ול- S(Moshe) את הערך f, נקבל פירוש, שבו ההנחות ב- (*) נכונות אך המסקנה שקרית.

♦ דוגמה 3:

הראה ש- $\neg C, \neg A \Rightarrow B \, , A \Rightarrow B \ \vdash B \vee C$

נבדוק בעזרת טבלת אמת, המכסה את כל האפשרויות:

	A	B	C	$\neg C$	$\neg A$	$A \Rightarrow B$	$\neg A \Rightarrow B$	$B \vee C$
	t	t	t	f	f	t	t	t
*	t	t	f	t	f	t	t	t
	t	f	t	f	f	f	t	t
	t	f	f	t	f	f	t	f
	f	t	t	f	t	t	t	t
*	f	t	f	t	t	t	t	t
	f	f	t	f	t	t	f	t
	f	f	f	t	t	t	f	f

בשתיים בדיוק משמונה השורות נכונות כל ההנחות (השנייה והשישית). בשתיהן נכונה גם המסקנה, ולכן היא אכן נובעת מההנחות.

הערה:

דוגמה זו מרמזת על חולשתה של השיטה גם באותם מקרים, בהם ניתן להפעיל אותה. כל תוספת של פסוק אטומי (כמו A, B ו- C בדוגמה זו) מכפילה גם את מספר השורות ולכן גם את העבודה הנדרשת. באופן מעשי, השיטה ישימה רק כשמספר קטן של נוסחאות אטומיות מעורב בטיעון.

א.3. שקילויות לוגיות

שתי נוסחאות נקראות ***שקולות לוגית***, אם בכל פירוש בו נכונה האחת, נכונה גם השנייה. שימוש בשקילויות לוגיות (דהיינו: בעובדה, שנוסחה מסוימת שקולה לוגית לנוסחה אחרת) היא אחת הדרכים היעילות ביותר לטיפול בטענות לצורך טיעונים (ולצרכים אחרים). צורת השימוש בהן דומה לצורת השימוש בזהויות באלגברה. באלגברה, אם יש זהות בין שני ביטויים (כמו $(a+b)^2$ ו- $a^2+2ab+b^2$), אזי ניתן להחליף כל אחד מהם בשני בכל קונטקסט אלגברי. עיקרון דומה, הנקרא "עיקרון החלפת אקוויוולנטים", קיים בלוגיקה:

אם שתי נוסחאות הינן שקולות לוגית (= אקוויוולנטיות), אזי ניתן להחליף כל אחת בשנייה בכל קונטקסט.[3]

קל לברר, ששתי נוסחאות A ו- B הן שקולות לוגית אם"ם הנוסחה $A \Leftrightarrow B$ היא אמת לוגית. לאמיתות לוגיות מסוג זה קוראים "שקילויות לוגיות".

טבלה מס' א.3 מרכזת רשימה של השקילויות הלוגיות החשובות ביותר. את כל אלה בהן הכמתים אינם מעורבים אפשר לאמת בקלות בעזרת טבלאות אמת: השקילות תקבל ערך t בכל פירוש של A, B ו- C על-ידי מתן ערכי אמת (במלים אחרות: שני אגפי השקילות יקבלו תמיד אותו ערך אמת). השאר ברורות מאליהן (אחרי שנבין לאשורו מה בדיוק כתוב בטבלה, דבר שנבהיר בהמשך לכל שקילות).

טבלה א.3: שקילויות לוגיות חשובות

קוניוקציה ודיסיונקציה

$(1)_a$ $A \wedge A \Leftrightarrow A$	$(1)_b$ $A \vee A \Leftrightarrow A$
$(2)_a$ $A \wedge B \Leftrightarrow B \wedge A$	$(2)_b$ $A \vee B \Leftrightarrow B \vee A$
$(3)_a$ $(A \wedge B) \wedge C \Leftrightarrow A \wedge (B \wedge C)$	$(3)_b$ $(A \vee B) \vee C \Leftrightarrow A \vee (B \vee C)$
$(4)_a$ $A \wedge (B \vee C) \Leftrightarrow (A \wedge B) \vee (A \wedge C)$	$(4)_b$ $A \vee (B \wedge C) \Leftrightarrow (A \vee B) \wedge (A \vee C)$

גרירה

$(5)_a$ $(A \Rightarrow B) \Leftrightarrow (\neg A \vee B)$	$(5)_b$ $(A \vee B) \Leftrightarrow (\neg A \Rightarrow B)$
(6) $(A \Rightarrow B) \Leftrightarrow (\neg B \Rightarrow \neg A)$	
(7) $(A \Rightarrow (B \Rightarrow C)) \Leftrightarrow (A \wedge B \Rightarrow C)$	
$(8)_a$ $(A \Rightarrow B \wedge C) \Leftrightarrow (A \Rightarrow B) \wedge (A \Rightarrow C)$	$(8)_b$ $(A \vee B \Rightarrow C) \Leftrightarrow (A \Rightarrow C) \wedge (B \Rightarrow C)$

[3] יש חוסר דיוק מסוים במלים "*בכל* קונטקסט", אך נוכל להתעלם כאן ממנו. נוסח מדויק ניתן בקורס בלוגיקה.

שלילה

		(9)	$\neg\neg A \Leftrightarrow A$
$(10)_a$	$\neg(A \wedge B) \Leftrightarrow \neg A \vee \neg B$	$(10)_b$	$\neg(A \vee B) \Leftrightarrow \neg A \wedge \neg B$
		(11)	$\neg(A \Rightarrow B) \Leftrightarrow A \wedge \neg B$
$(12)_a$	$\neg\forall xA \Leftrightarrow \exists x\,\neg A$	$(12)_b$	$\neg\exists xA \Leftrightarrow \forall x\,\neg A$

כמתים

$(13)_a$	$\forall x(\forall yA) \Leftrightarrow \forall y(\forall xA)$	$(13)_b$	$\exists x(\exists yA) \Leftrightarrow \exists y(\exists xA)$
$(14)_a$	$\forall xA \Leftrightarrow \forall yA(y/x)$	$(14)_b$	$\exists xA \Leftrightarrow \exists yA(y/x)$
$(15)_a$	$\forall x(A \wedge B) \Leftrightarrow \forall xA \wedge \forall xB$	$(15)_b$	$\exists x(A \vee B) \Leftrightarrow \exists xA \vee \exists xB$

שמונה השקילויות ($(1)_a$ – $(4)_b$) הן פשוטות מאוד, והשימוש בהן נעשה באופן אוטומטי. לכולן מקבילות (כמו שנראה) זהויות הקשורות בקבוצות ובאופרציות על קבוצות, וזה עיקר חשיבותן בקורס זה. נציין, עם זאת, שהשקילויות ב- (3) מאפשרות לנו לכתוב (למשל) $A_1 \wedge A_2 \wedge \ldots \wedge A_n$, בלי לציין במפורש את מקום הסוגריים השונים.

לקבוצת השקילויות הבאה (של הגרירה) יש חשיבות רבה בהוכחות, כיוון שהיא משמשת בסיס לגישות אפשריות להוכחה. $(5)_b$, לדוגמה, משמשת בסיס לשיטה הבאה להוכחת דיסיונקציה $A \vee B$: להוכיח ש- B נובע מההנחה, ש- A אינו נכון. (6) מאפשרת להוכיח משפטים בעזרת שיטה הידועה כ"קונטרה-פוזיציה": במקום להוכיח ש- B נובע מ- A, מראים, שאם B לא נכון אז גם A לא נכון (לדוגמה, כדי להוכיח, שאם במשולש זווית $\prec B$ שווה לזווית $\prec C$, אז $AB = AC$, משתמשים ספרי לימוד רבים ב"דרך השלילה": הם מראים, שאם $AB \neq AC$, אז או שזווית $\prec B$ גדולה מ- $\prec C$ או להיפך, ובכל מקרה לא ייתכן אז, ש- $\prec B = \prec C$). לפי $(8)_a$, כדי להוכיח שקוניוקציה $B \wedge C$ נובעת מ- A, יש לבצע שני דברים באופן נפרד: להוכיח ש- B נובע מ- A, ולהוכיח ש- C נובע מ- A. לפי $(8)_b$, לעומת זאת, אם ברצוננו להראות, שטענה C נובעת מדיסיונקציה $A \vee B$, אז עלינו להראות, שהיא נובעת בנפרד מכל אחד משני חלקי הדיסיונקציה. לבסוף, שקילות (7) מאפשרת להחליף גרירה מקוננת (nested) בגרירה פשוטה יותר. בדרך-כלל עושים אנו שימוש בה באופן אוטומטי, מבלי לתת את הדעת, על מה אנו מסתמכים. ניקח לדוגמה את המשפט: "במשולש ישר זווית, אם אחת הזוויות שווה 30^0, אז הצלע מולה שווה לחצי היתר". מה שנטען לפי ***ניסוח*** המשפט הוא:

$$\forall A \forall B \forall C(\prec ACB = 90° \Rightarrow (\prec ABC = 30° \Rightarrow 2\overline{AC} = \overline{AB}))$$

ברם, בניסוח ה"נתון – צריך להוכיח" כותבים מיד:

נתון: $\prec ACB = 90$
$\prec ABC = 30$

צ"ל: $\overline{AC} = 2\overline{AB}$

במלים אחרות, מה שמוכיחים הוא בעצם:

$$\forall A \forall B \forall C(\prec ACB = 90° \wedge \prec ABC = 30° \Rightarrow \overline{AC} = 2\overline{AB}))$$

לקבוצה הבאה של שקילויות, אלה הקשורות בשלילה, חשיבות מיוחדת בפישוט טענות ובהוכחות. כללית, ככל שהנוסחאות הנשללות בעזרת קשר השלילה הן פשוטות יותר – כן עדיף. אידיאלית, כדאי שהוא יופיע (אם בכלל) רק לפני הנוסחאות האטומיות (ואז אפשר לעתים קרובות להיפטר ממנו כליל. למשל, במקום $\neg(a < b)$ אפשר לכתוב $a \geq b$, אם עוסקים במספרים הטבעיים או הממשיים). השקילויות (12)-(9) מאפשרות להשיג מטרה זו תמיד. כדוגמה פשוטה נחזור לדוגמה המרכזית מהפרק הראשון. שם הגענו, כזכור, לניסוח הקצר הבא של הטענה "למספר שלילי אין שורש":

(8) $\forall a < 0 \neg \exists b. b^2 = a$

עתה, לפי כלל $(12)_b$, $\neg \exists b. b^2 = a$ שקול ל- $\forall b. \neg b^2 = a$. במקרה זה, במקום $\neg b^2 = a$ אפשר לכתוב $b^2 \neq a$.[4] וכך נקבל:

(9) $\forall a < 0 \; \forall b. b^2 \neq a$

נוכל עתה לחזור, סוף-סוף, לבעיה של שלילת הטענה "למספר שלילי אין שורש", בה התחלנו את דיוננו בפרק הראשון. בעזרת ההצרנה ושקילויות השלילה נוכל לפתור בעיה זו בקלות ובאופן מכני: אם נתחיל משלילת (9), נקבל:

		$\neg \forall a < 0 \; \forall b. \neg(b^2 = a)$
לפי $(12)_a$	$\Leftrightarrow$	$\exists a < 0 \; \neg \forall b. \neg(b^2 = a)$
לפי $(12)_a$	$\Leftrightarrow$	$\exists a < 0 \; \exists b. \neg\neg(b^2 = a)$
לפי שקילות מס' (9)	$\Leftrightarrow$	$\exists a < 0 \; \exists b. \; b^2 = a$

ובעברית: קיים מספר שלילי, שיש לו שורש.

הערה: במקרה זה היינו מגיעים לזה ביתר קלות, אם היינו מתחילים ב- (8).

---

[4] זה נוהג מקובל למדי לכתוב aRb במקום $\neg(aRb)$ (כש- R סימן יחס בינארי כמו $=$, $\in$ ועוד).

קוראים זהירים ישימו לב, שבדוגמה התיימרנו להשתמש (פעמיים) ב- $(12)_a$, אך למעשה $(12)_a$ מתייחס רק לנוסחאות מהצורה $\forall a...$ ולא לקיצורים כמו: $\forall a < 0...$. האם הדבר אכן לגיטימי? נבדוק זאת על-ידי שנפעיל את הכללים כמות שהם ל- (9), כשהוא כתוב בצורה מלאה. (9) הינו צורה מקוצרת ל-

$$\forall a(a < 0 \Rightarrow \forall b.\neg(b^2 = a))$$

לכן הפישוט המדויק נראה כך:

$\neg\forall a(a < 0 \Rightarrow \forall b.\neg(b^2 = a))$		
$\exists a.\neg\,(a < 0 \;\Rightarrow \forall b.\neg(b^2 = a))$	$\Leftrightarrow$	לפי $(12)_a$
$\exists a\,(a < 0 \;\wedge \neg\forall b.\neg(b^2 = a))$	$\Leftrightarrow$	לפי (11)
$\exists a\,(a < 0 \;\wedge \exists b.\,\neg\neg(b^2 = a))$	$\Leftrightarrow$	לפי $(12)_a$
$\exists a\,(a < 0 \;\wedge \exists b.\, b^2 = a)$	$\Leftrightarrow$	לפי (9)

אבל $\exists a < 0\;\; \exists b.\, b^2 = a$, שקיבלנו למעלה, אינו אלא כתיבה מקוצרת של מה שקיבלנו כאן! אנו רואים (וקל להוכיח זאת באופן כללי), שהקיצורים שהנהגנו (ודומים שנעשה בעתיד) של $\forall a(a < c \Rightarrow ...)$ ל- $\forall a < c(...)$ ושל $\exists b(b < c \wedge ...)$ ל- $\exists b < c\,(...)$, הם לא רק נוחים, אלא גם קוהרנטיים עם השקילויות הנוגעות לשלילה.

הקבוצה האחרונה של השקילויות בטבלת השקילויות עוסקת בכמתים. נקדיש לה עתה דיון קצר, כי אבחנות הקשורות במשתנים ובכמתים הן חשובות ביותר, והתעלמות מהן מהווה מקור בלתי נדלה של טעויות לוגיות.

ראשית נעיר, כי במקום "x" ו- "y" יכולים להופיע בכללים השונים משתנים אחרים כלשהם. כך, למשל, $\forall a(\forall z_1 A)$ ו- $\forall z_1(\forall aA)$ שקולים לפי כלל $(13)_a$. מעתה, אגב, נכתוב פשוט $\forall x\forall yA$ (או אפילו $\forall x,yA$) במקום $\forall x(\forall yA)$, כמו שאכן עשינו כבר באי-אלו דוגמאות קודמות.

שנית, בשקילויות (14) מופיע סימן הדורש הבהרה: $A(y/x)$ הכוונה היא לנוסחה המתקבלת מ- A על-ידי החלפת משתנה x במשתנה y.[5] לדוגמה, אם במקום x ו- y נשתמש ב- a וב- b, ובתור A ניקח את הנוסחה $(a + 1)^2 = a^2 + 2a + 1$, אזי לפי $(14)_a$ $\forall a((a + 1)^2 = a^2 + 2a + 1)$ הינה שקולה לוגית ל- $\forall b((b + 1)^2 = b^2 + 2b + 1)$.

---

[5] למעשה, לא בכל מקום, אלא רק בכל מקום ב- A בו x מופיע *חופשי*. עניין זה יובהר בפרק א.5.

עקרון זה (ללא הכמתים בהתחלה) ידוע ומוכר לנו משכבר. למעשה, בתיכון אנו לא אומרים, ששתי הנוסחאות שקולות לוגית, אלא שמדובר ב*אותה נוסחה*. האמת היא, שאכן (14) מבטא עקרון כללי ביותר של זיהוי, שאינו חל רק על נוסחאות לוגיות. נלמד אותו בפרק א.5 (שם נקרא לו עקרון α).

חשוב לציין, שעל (14) חלה הגבלה, ש- y חייב להיות משתנה ***חדש***, שאינו מופיע כבר בנוסחה A. לדוגמה, אם ניקח את $\exists i\exists j.i+j=3$ ונחליף בה את i ב- j, נקבל את $\exists j\exists j.j+j=3$, שאינו אלא $\exists j.j+j=3$. הטענה המקורית נכונה במספרים הטבעיים, השנייה – שקרית. ברור לכן, שהן אינן שקולות לוגיות.

בנוגע לשקילויות האחרות בטבלה, חשוב אולי יותר לשים לב ולזכור, מה ***לא*** מופיע בה מאשר מה כן. נתבונן, לדוגמה, בכללים (15). דוגמה ל- $(15)_a$ היא העובדה הברורה, שהטענה "כל ילד אוהב לאכול גלידה וגם כל ילד אוהב לאכול שוקולד" שקולה לחלוטין לטענה ש"כל ילד אוהב לאכול גלידה וגם אוהב לאכול שוקולד". לעומת זאת, הטענה ש"כל סטודנט יצליח או ייכשל בבחינה" אינה שקולה לטענה, ש"כל סטודנט יצליח בבחינה, או כל סטודנט ייכשל בה". $\forall x(A\vee B)$ אינה שקולה לנוסחה $\forall xA\vee\forall xB$! (עם $\exists$ המצב הוא הפוך, כמובן).

אבחנה חשובה עוד יותר קשורה בשקילויות (13), שעוסקות ב***סדר*** של הכמתים. תמציתן היא, שלגבי כמתים ***זהים*** אין חשיבות לסדר. הטענות $\exists x\exists y.x>y$ ו- $\exists y\exists x.x>y$ שקולות לחלוטין זו לזו. בדומה, הטענה "כל אם וכל בת מדליקות נרות שבת" שקולה לחלוטין (מבחינה לוגית, לפחות) לטענה "כל בת וכל אם מדליקות נרות שבת".

המצב שונה באופן קיצוני, כשעוסקים בכמתים מסוגים שונים. כאן הסדר משנה ועוד איך. $\forall x\exists yA$ הוא דבר שונה (וחלש יותר) מ- $\exists y\forall xA$. לדוגמה, $\forall x\exists y.x<y$ נכון בטבעיים. $\exists y\forall x.x<y$ לא.

אי הבנת ההבדל בין $\forall x\exists yA$ ובין $\exists y\forall xA$ היוותה ומהווה גורם ראשון במעלה לשגיאות לוגיות, שאפילו גדולי המתמטיקאים נכשלו בהן בזמנם. מקור הקשיים בכך, שהשפה הטבעית אינה מבדילה כל-כך בין שני הניסוחים. אין, למשל, הבדל במשמעות בין "לכל אדם כוכב יש בשמיים" ובין "יש כוכב בשמיים לכל אדם". ההבנה, למה הכוונה, נעשית במסגרת השפה הטבעית לפי הקונטקסט. לדוגמה, פירוש משפט כמו "יש מישהו שיצר את כל הפסלים הללו" יכול להיות: "לכל הפסלים הללו יש אותו יוצר". בקונטקסט אחר (כמו, למשל, עת מבקרים באיי הפסחא ורואים את עשרות

הפסלים המסתוריים, שנמצאים שם) הפירוש יכול להיות, שהפסלים לא נוצרו יש מאין, אלא יש בני-אדם שיצרו אותם (אבל ייתכן, שלא כולם נוצרו על-ידי אותו פסל). מבחינה לוגית הפירוש הראשון גורר את השני – אך לא להיפך. במלים אחרות, $\exists y\forall xA \Rightarrow \forall x\exists yA$ נכון תמיד, אך $\forall x\exists yA \Rightarrow \exists y\forall xA$ בדרך-כלל לא.

%%

ההבדל בין $\exists y\forall xA$ ובין $\forall x\exists yA$ עומד ביסוד כמה מהמושגים החשובים והאבחנות העדינות של המתמטיקה. כל מושג באנליזה (חשבון דיפרנציאלי ואינטגרלי) הכולל את צירוף המלים "במידה שווה" קשור בו. לדוגמה, ההבדל באנליזה בין רציפות סתם ובין רציפות במידה-שווה הינו, בעיקרו של דבר, הבא:[6]

להגיד ש-f רציפה בקטע I פירושו:

$$\forall x\, \forall \varepsilon > 0\, \exists \delta > 0\, \forall y(|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon)$$

מה ששקול לפי $(13)_a$ ל-

$$\forall \varepsilon > 0\, \underline{\forall x\, \exists \delta > 0}\, \forall y(|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon)$$

לומר ש-f רציפה במידה-שווה ב-I פירושו, לעומת זאת:

$$\forall \varepsilon > 0\, \underline{\exists \delta > 0\, \forall x}\, \forall y(|x-y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon)$$

אנו רואים ש"כל" ההבדל (הבדל מהותי ביותר – שלא תהיה מקום לטעות!) הוא בסדר הכמתים. ליתר דיוק: בהחלפה, במקום אחד, של "$\forall x\exists\delta$" ב- "$\exists\delta\forall x$".

מוסר השכל הוא: בקוראנו ובנסחנו משפטים מתמטיים יש לשים לב לכל מלה ומלה, וגם לסדר המלים!

%%

6 ב- "$\forall x$" בשורות הבאות הכוונה "לכל x בקטע I". הסימון, המאפשר לומר זאת עם התייחסות לקטע I, יילמד בהמשך.

*א.4. כיצד מוכיחין?

למרות שלשיטת טבלאות האמת ולשימוש בשקילויות יש חשיבות, הרי הוכחות בטקסטים מתמטיים לעתים רחוקות נעזרות בהם. למעשה, הוכחות מתמטיות משתמשות במספר לא גדול של *כללי היסק*. כללי היסק אלו מסוכמים בטבלה מס' א.4. מטרתנו בפרק זה היא להבין, מה בדיוק כתוב בטבלה זו (ולהביא דוגמאות).

טבלה א.4: כללי ההיסק הבסיסיים

(1) $\dfrac{\mathbf{T}, A \vdash B}{\mathbf{T} \vdash A \Rightarrow B}$	(2) $\dfrac{\mathbf{T} \vdash A \quad T \vdash A \Rightarrow B}{\mathbf{T} \vdash B}$
(3) $\dfrac{\mathbf{T} \vdash A \quad \mathbf{T} \vdash B}{\mathbf{T} \vdash A \wedge B}$	(4) $\dfrac{\mathbf{T} \vdash A \wedge B}{\mathbf{T} \vdash A} \quad \dfrac{\mathbf{T} \vdash A \wedge B}{\mathbf{T} \vdash B}$
(5) $\dfrac{\mathbf{T} \vdash A}{\mathbf{T} \vdash A \vee B} \quad \dfrac{\mathbf{T} \vdash B}{\mathbf{T} \vdash A \vee B}$	(6) $\dfrac{\mathbf{T} \vdash A \vee B \quad \mathbf{T}, A \vdash C \quad \mathbf{T}, B \vdash C}{\mathbf{T} \vdash C}$
(7) $\vdash \neg A \vee A$	(8) $\dfrac{\mathbf{T} \vdash \neg A \quad \mathbf{T} \vdash A \vee B}{\mathbf{T} \vdash B}$
(9) $\dfrac{\mathbf{T}, A \vdash B \quad \mathbf{T}, A \vdash \neg B}{\mathbf{T} \vdash \neg A}$	(10) $\dfrac{\mathbf{T}, \neg A \vdash B \quad \mathbf{T}, \neg A \vdash \neg B}{\mathbf{T} \vdash A}$
(11) $\dfrac{\mathbf{T} \vdash A \Rightarrow B \quad \mathbf{T} \vdash B \Rightarrow A}{\mathbf{T} \vdash A \Leftrightarrow B}$	(12) $\dfrac{\mathbf{T} \vdash A \Leftrightarrow B}{\mathbf{T} \vdash A \Rightarrow B} \quad \dfrac{\mathbf{T} \vdash A \Leftrightarrow B}{\mathbf{T} \vdash B \Rightarrow A}$
(13)* $\dfrac{\mathbf{T} \vdash A(y/x)}{\mathbf{T} \vdash \forall x A}$	(14)** $\dfrac{\mathbf{T} \vdash \forall x A}{\mathbf{T} \vdash A(t/x)}$
(15)** $\dfrac{\mathbf{T} \vdash A(t/x)}{\mathbf{T} \vdash \exists x A}$	(16)* $\dfrac{\mathbf{T} \vdash \exists x A \quad \mathbf{T}, A(y/x) \vdash B}{\mathbf{T} \vdash B}$

כשאנו קוראים הוכחה מתמטית, אנו קוראים סדרה של טענות. כל אחת ואחת מהן מוכחת (בשלב בו היא נטענת) על סמך קבוצה כלשהי של הנחות.[7] קבוצה זו מורכבת מהאקסיומות הכלליות של התחום הנדון (כמו אקסיומות הגיאומטריה, כשמדובר בהוכחה גיאומטרית), הנתונים המיוחדים של המשפט המוכח והנחות זמניות, שאנו עושים במהלך ההוכחה. לכן מה שמוכח בכל שלב אינו בעצם הטענה הכתובה, אלא העובדה, שהיא נובעת מקבוצה מסוימת של הנחות, הנמצאות ברקע שלה. כללי ההיסק, בעזרתם אנו עוברים מטענות שכבר הוכחו לטענה חדשה, אומרים בעצם משהו מהצורה הבאה: "אם הטענה A_1 נובעת מקבוצת ההנחות $\mathbf{T}_1$, A_2 נובעת מקבוצת ההנחות $\mathbf{T}_2$, ..., A_n נובעת מקבוצת ההנחות $\mathbf{T}_n$, אז B נובעת מקבוצת ההנחות $\mathbf{T}$". כללי ההיסק בטבלה א.4 מבטאים זאת בדיוק. מה שמעל "קו השבר" בכל אחד מהם מבטא את חלק ה"אם" האמור לעיל ("אם טענה A_1..."), בעוד מה שמתחתיו את חלק ה"אז" ("אז B נובעת ..."). הדוגמאות, שתבואנה מיד, תבהרנה, למה בדיוק הכוונה.

בעת שמשתמשים בכללי ההיסק יש לזכור, ראשית, שכל טענה נובעת מעצמה. כמו כן, אם טענה נובעת מקבוצה מסוימת של הנחות, אזי היא נובעת מכל קבוצה רחבה יותר של הנחות. הסימון "$\mathbf{T}, A \vdash B$" משמעותו, שהטענה B נובעת מההנחות ב- $\mathbf{T}$ בתוספת עם ההנחה A.

כללי ההיסק הנהוגים בהוכחות מתחלקים לשני סוגים:

א. כללים, שבעזרתם מוכיחים טענות מהסוג $A \Rightarrow B$, $A \wedge B$ וכו'. לכללים כאלה קוראים "כללי הכנסה".

ב. כללים, שבעזרתם משתמשים בידע מהצורה $A \wedge B$, $A \Rightarrow B$ וכו'. לכללים כאלה קוראים "כללי סילוק".

נעבור עתה על הכללים בטבלה, ונסביר אותם.

(I) כללי הגרירה

הכלל הברור והמובן מאליו הקשור בגרירה (ובכלל) הוא כלל (2) הידוע בשם "מודוס פוננס" (MP). הוא אומר פשוט, שאם ידוע ש- B נובע מ- A, וידוע A,

[7] אפשרי, אם כי נדיר ביותר, שטענה מוכחת ללא כל הנחות ברקע. במקרה זה נאמר בהמשך, שקבוצת ההנחות $\mathbf{T}$ הינה "ריקה".

אז ניתן להסיק את B. זוהי דוגמה לכלל סילוק: זהו כלל, בעזרתו אנו משתמשים בידע מהצורה $A \Rightarrow B$.

ואיך מוכיחים גרירה מהצורה $A \Rightarrow B$ על סמך קבוצת הנחות $\mathbf{T}$? ובכן, לפי כלל מס' (1), כל שעלינו לעשות לצורך זאת, הוא לצרף את A למערכת ההנחות $\mathbf{T}$ ולהוכיח את B על סמך אוסף ההנחות המורחב. אם נצליח (כלומר, אם נראה ש- $\mathbf{T},A \vdash B$) נוכל להסיק מזה ש- $A \Rightarrow B$ נובע מ- $\mathbf{T}$ ($\mathbf{T} \vdash A \Rightarrow B$). לדוגמה, נניח שברצוננו להראות, שמאקסיומות הגיאומטריה נובע, שאם במשולש יש שתי זוויות שוות, אז הוא שווה-שוקיים. ברצוננו להראות, אפוא, ש- $\mathbf{T} \vdash A \Rightarrow B$, כאשר $\mathbf{T}$ היא קבוצת האקסיומות של הגיאומטריה, A היא הטענה, שבמשולש ABC יש שתי זוויות שוות, ו- B – שמשולש זה הוא שווה-שוקיים. כיצד נעשה דבר זה בפועל בבית-הספר התיכון? ובכן, ראשית חכמה רושמים:

נתון: $\measuredangle ABC = \measuredangle ACB$

צ"ל: $AB = AC$

כלומר: מוסיפים את ההנחה ש- $\measuredangle ABC = \measuredangle ACB$ לאקסיומות של הגיאומטריה, ומשתדלים להוכיח ש- $AB = AC$ על סמך האוסף המתקבל. במלים אחרות: מוכיחים ש- $\mathbf{T}, A \vdash B$ (עם אותם $\mathbf{T}$, A, B). זהו שימוש מובהק, הנעשה באופן אוטומטי, בכלל (1) – כלל ההכנסה של הגרירה.

בעזרת שני הכללים הבסיסיים של הגרירה אפשר לגזור עוד הרבה כללים שימושיים אחרים. נראה לדוגמה כלל חשוב במיוחד: שאם A גורר את B ו- B גורר את C אז A גורר את C:

(i) $A \Rightarrow B\,,\ B \Rightarrow C\,,\ A \vdash A \Rightarrow B$ (עיקרון בסיסי)

(ii) $A \Rightarrow B\,,\ B \Rightarrow C\,,\ A \vdash B \Rightarrow C$ (עיקרון בסיסי)

(iii) $A \Rightarrow B\,,\ B \Rightarrow C\,,\ A \vdash A$ (עיקרון בסיסי)

(iv) $A \Rightarrow B\,,\ B \Rightarrow C\,,\ A \vdash B$ ((i), (iii) וכלל (2))

(v) $A \Rightarrow B\,,\ B \Rightarrow C\,,\ A \vdash C$ ((ii), (iv) וכלל (2))

(vi) $A \Rightarrow B\,,\ B \Rightarrow C \vdash A \Rightarrow C$ ((v) וכלל (1))

חשוב לציין, שבטקסטים אמיתיים לעולם לא יהיה פירוט כזה. ראשית, אין מציינים בכל שורה ושורה את מערכת ההנחות, ממנה הוסקה הטענה, הנמצאת באותה שורה. על הקורא לעקוב אחרי זה בעצמו. שנית, לא מציינים, איזה

כללים לוגיים מפעילים והיכן. פשוט משתמשים בהם. הוכחה כנ"ל בטקסט מתמטי אמיתי דומה יותר לדבר הבא:

(i) $A \Rightarrow B$ (הנחה)

(ii) $B \Rightarrow C$ (הנחה)

(iii) A (הנחה)

(iv) B (מ- (i) ו- (iii))

(v) C (מ- (ii) ו- (iv))

כדאי לשים לב, שאין טורחים כלל לעשות את המעבר המפורש מ- (v) ל- (vi). הקורא אמור להבין בעצמו, שכאן הוכח $A \Rightarrow C$ על סמך ההנחות $A \Rightarrow B$ ו- $B \Rightarrow C$!

(II) כללי הקוניוקציה

אלה הכללים הפשוטים ביותר, ולא נבזבז עליהם מלים מיותרות. כלל (3) קובע, שכדי להוכיח קוניוקציה $A \wedge B$, עלינו לספק שתי הוכחות: אחת ל- A, השנייה ל- B. הכללים ב- (4), לעומתם, אומרים, שפיסת ידע מהצורה $A \wedge B$ כמוה כשתי פיסות ידע, A ו- B.

◆ **דוגמה:**

נתבונן במשפט "חוצה זווית הראש במשולש שווה-שוקיים הוא גם גובה וגם תיכון". זוהי טענה מהצורה $\underline{A} \wedge \underline{B} \Rightarrow \underline{C} \wedge \underline{D}$, כאשר $\underline{A}$ היא הטענה, שמשולש ABC הוא שווה-שוקיים, $\underline{B}$ – הטענה ש- AD הוא חוצה זווית הראש, $\underline{C}$ קובעת ש- AD מאונך לבסיס, ו- $\underline{D}$ – ש- AD הוא תיכון. כמו בכל גרירה, מניחים בהוכחה את $\underline{A} \wedge \underline{B}$ ומוכיחים את $\underline{C} \wedge \underline{D}$. עתה את הנתון $\underline{A} \wedge \underline{B}$ מפרקים אוטומטית, לפי (4), לנתונים $\underline{A}$ ו- $\underline{B}$. גם את מה שצריך להוכיח מפרקים אוטומטית, הפעם לפי (3), לשתי מטלות: להוכיח את $\underline{C}$ ולהוכיח את $\underline{D}$. מה שכותבים הוא לכן:

נתון: (1) $AB = AC$ (2) $\sphericalangle BAD = \sphericalangle CAD$

צ"ל: (1) $AD \perp \mathrm{BC}$ (2) $BD = DC$

(III) כללי הדיסיונקציה

השיטה הישירה, הקונסטרוקטיבית, להראות ש- $A \vee B$ נכון, ניתנת על-ידי כלל (5): פשוט מוכיחים, שאחת משתי האלטרנטיבות (A ו- B) נכונה היא. ברם: אין

זו תמיד הדרך היחידה. לפעמים אפשר לדעת $A \vee B$ בלי לדעת, מי משתי הברירות היא הנכונה. מקרה בסיסי כזה ניתן על-ידי כלל (7), הידוע בשם "חוק השלישי הנמנע". לפי כלל זה, כל טענה היא או נכונה או לא. כך כל סטודנט יודע, שאו שהוא יצליח בבחינה, או שלא. לא כל סטודנט בטוח, עם זאת, מי משתי האלטרנטיבות תתגשם!

ואיך משתמשים בידע מהצורה $A \vee B$? יש שתי דרכים עיקריות לכך. האחת, הקונסטרוקטיבית, ניתנת על-ידי כלל (6) (הידוע כ"כלל הדילמה"): כדי להראות ש- C נכון, כשכל שידוע הוא $A \vee B$, עלינו להראות, שכל אחת משתי האלטרנטיבות לחוד מובילה למסקנה C. למשל, אם בגמר גביע המדינה בכדורגל מתמודדות הפועל תל-אביב ומכבי תל-אביב, אז בהרבה מוספי ספורט יכתבו לפני המשחק: "בכל מקרה הגביע יגיע לתל-אביב". ההגיון הוא, שאו שהפועל תל-אביב תזכה, או שמכבי תל-אביב תזכה. אם הפועל תל-אביב תזכה – הגביע יגיע לתל-אביב. אם מכבי תל-אביב תזכה – גם אז הגביע יגיע לתל-אביב. מכאן, שבכל מקרה הגביע יגיע לתל-אביב.

השיטה השנייה לשימוש בדיסיונקציה היא השיטה, הניתנת על-ידי כלל (8) (הנקרא, כזכור, "הסילוגיזם הדיסיונקטיבי"). היא מקובלת, בין היתר, על בלשים, שעורכים רשימת חשודים ואחר-כך מתחילים לצמצם אותה ("או ששמעון הוא הרוצח, או שזה לוי. לשמעון יש אליבי מוצק. מכאן שהרוצח הוא לוי").

(IV) כללי השלילה

כלל (9) נותן את הדרך העיקרית להוכחת שלילות: כדי להראות ש- A אינו נכון, מראים, שמ- A נובעת סתירה. אפשרות אחרת, כמובן, היא להוכיח משהו, שהינו שקול לשלילת A על-סמך השקילויות, שנלמדו בפרק הקודם.

ואיך משתמשים בידע מהצורה $\neg A$? דרך מרכזית, שכבר ראינו אותה, ניתנה על-ידי כלל מס' (8). אפשרות אחרת היא, שוב, להשתמש בשקילויות של הפרק הקודם.

כדאי לציין, שאת כל השקילויות, שלמדנו בפרק הקודם ניתן להוכיח בעזרת הכללים של טבלה א.4. כדוגמה חשובה במיוחד ניקח את ההסקה של A מ- $\neg\neg A$. גזירתה פשוטה בתכלית: מ- $\neg\neg A$ ו- $A \vee \neg A$ (כלל (7)) נובע A לפי כלל (8)! עתה, האפשרות של הסקת A מ- $\neg\neg A$ היא חשובה במיוחד, כי היא משמשת בסיס להוכחות בדרך השלילה. בהוכחות כאלה, כאשר רוצים להוכיח A, מניחים $\neg A$ ומגיעים לסתירה. מזה נובע, לפי כלל (9), $\neg\neg A$, וממנו נובע, כאמור, A.

התהליך כולו מסוכם בטבלה א.4 ככלל בפני עצמו: כלל (10) (למרות שכלל זה נובע, כאמור, משאר הכללים בטבלה, בחרנו לכלול אותו בה בגלל חשיבותו).

(V) כללי השקילות

$A \Leftrightarrow B$ שקול ל- $(A \Rightarrow B) \wedge (B \Rightarrow A)$, ולכן הכללים הקשורים ב- $\Leftrightarrow$ ב- (11) וב-(12) נגזרים ישירות מכללי הגרירה והקוניוקציה. נזכיר עוד את כלל הצבת האקויולנטים, שהוא דרך מרכזית להשתמש בשקילויות. כמו כן כדאי להדגיש, שלפי כלל (11), על מנת להוכיח את שקילותם של A ו- B, יש להראות ***שני*** דברים: ש- A גורר את B, וש- B גורר את A.

(VI) כללי $\forall$

הדרך הרגילה להוכיח טענה אוניברסלית מהסוג $\forall xA$ היא לקחת איבר "כלשהו" x בתחום הנדון, ששום דבר לא ידוע עליו, וגם לא הנחנו עליו דבר עד אותו רגע, ולהוכיח, שיש לו התכונה A. אין שום הכרח, אגב, לקרוא לו x דווקא. כל שם אחר אפשרי, ובלבד שלשם זה אין כבר פירוש, ועל העצם, שהשם מתייחס אליו, לא הונח דבר עד אותו רגע. צורת הוכחה זו היא המתוארת בכלל מספר (13) (ה"כוכבית" הצמודה למספר "13" באה לציין קיום הגבלות על y. ניסוחן המדויק של ההגבלות הללו יינתן בפרק הבא).

הבה נביא דוגמאות.

דוגמה ראשונה מתחום הגיאומטריה: נניח שאנו רוצים להוכיח, כי בכל משולש שווה-שוקיים שוות זוויות הבסיס. משפט זה מדבר על ***כל*** משולש. למרות זאת, בהוכחה עצמה אנו מתייחסים למשולש ABC, שהוא, כביכול, ספציפי, ואפילו משרטטים אותו בדרך-כלל. כמובן, ידוע לנו היטב, שבמהלך ההוכחה אין להשתמש בשום תכונה, שיש למשולש הספציפי ששרטטנו (כמו היות זווית הראש שלו חדה, או אולי קהה). משולש זה אמור לייצג משולש ***כלשהו***. כיוון שכך, ההוכחה לגביו תופסת לגבי כל משולש שהוא. ניתן לתאר זאת בצורה טובה שונה באופן הבא: כיוון שאין אנו מתייחסים בהוכחתנו לשום תכונה של המשולש, אתו אנו עובדים (פרט להיותו שווה-שוקיים), ניתן לשחזר את ההוכחה לגביו עבור כל משולש שווה-שוקיים אחר. לכן ההוכחה לגביו כמוה כהוכחה לכל משולש שווה-שוקיים.

דוגמה שנייה: נניח שרוצים להוכיח, שאם נעלה מספר טבעי אי-זוגי בריבוע ונחלק את התוצאה ב- 4, נקבל שארית 1. המלה "לכל" אינה מופיעה אמנם

בניסוח, אך הטענה היא בכל זאת אוניברסלית. ניסוח סטנדרטי של ההוכחה ייראה כך: יהי n מספר אי-זוגי. אז יש k כך ש- $n = 2k + 1$. לכן:

$$n^2 = (2k + 1)^2 = 4k^2 + 4k + 1 = 4(k^2 + k) + 1$$

מכאן, שאם נחלק את n^2 ב- 4 נקבל מנה $k^2 + k$ ושארית 1. מש"ל.

השימוש בכלל (13) בדוגמה זו מסתתר במלים, בהן נפתחת ההוכחה: "יהי n מספר אי-זוגי". הרעיון: אנו רוצים להראות משהו עבור *כל* מספר אי-זוגי. לכן אנו לוקחים מספר כלשהו כזה, n (שכל מה שאנו מניחים לגביו הוא, שהינו אי-זוגי), ומוכיחים את הטענה לגביו. כיוון ש- n היה כלשהו, מסקנתנו (שאיננו טורחים אפילו לכתוב אותה במפורש) היא, שהטענה נכונה עבור כל מספר אי-זוגי. זהו בדיוק שימוש בכלל (13).

כדאי לשים לב, שהפתיחה "יהי x..." היא פתיחה אופיינית להוכחת משפטים אוניברסליים מהצורה $\forall xA$, והיא סימן לכך, שאנו עומדים להשתמש בכלל (13).

לבסוף, דוגמה לשיטת הוכחה של טענות מהצורה: "לכל מספר טבעי..." (כלומר: טענות מהצורה $\forall nA(n)$), בה אין משתמשים בכלל (13) באופן ישיר: האינדוקציה המתמטית. בשיטה זו מסתמכים על האקסיומה הבאה:

$$[A(0/n) \wedge \forall n(A(n) \Rightarrow A(n + 1/n))] \Rightarrow \forall nA(n)$$

לפי אקסיומה זו (הידועה כ"אקסיומת האינדוקציה"), מספיק, לפי כלל (2), להוכיח את הרישא: $A(0/n) \wedge \forall n(A(n) \Rightarrow A(n + 1/n))$. עבור זה צריך, לפי כלל (3), להוכיח את $A(0/n)$ (מה שנקרא "בסיס האינדוקציה") ואת $\forall n(A(n) \Rightarrow A(n + 1/n))$ (מה שנקרא "שלב המעבר"). עתה, את שלב המעבר *כן* מוכיחים לפי כלל (13)! במלים אחרות: לוקחים n כלשהו ומוכיחים $A(n) \Rightarrow A(n + 1/n)$. כרגיל, עם גרירות, בשביל זה אנו מניחים ש- $A(n)$ נכון, ומוכיחים על סמך הנחה זו, ש- $A(n + 1/n)$ נכון. חשוב לציין, שבניגוד למה שנדמה לפעמים, *אין* אנו מניחים פה את מה שאנו רוצים להוכיח. מה שאנו רוצים להוכיח הוא $\forall nA(n)$, ואילו מה שאנו מניחים באותו שלב בהוכחה, הוא רק $A(n)$ לאיזה n לא ידוע. זה שונה מאוד מלהניח, *שלכל* n יש התכונה A!

וכיצד משתמשים בידע מהצורה $\forall xA$? התשובה: על-ידי יישומו למקרים פרטיים. זהו תוכן כלל (14). הוא קובע, שאם ידוע $\forall xA$, אז ניתן להסיק, בהינתן עצם ספציפי t, שיש לו התכונה A. במלים אחרות: ניתן להציב ב- A את t במקום x, ומה שמתקבל הוא מסקנה מהידוע. לדוגמה: אם ידוע ש- $\forall a((a + 1)^2 = a^2 + 2a + 1)$, אז ניתן להסיק מזה ש- $(7 + 1)^2 = 7^2 + 2 \cdot 7 + 1$ (כשמציבים $t = 7$) וגם ש- $(2a + 1)^2 = (2a)^2 + 2 \cdot (2a) + 1$ (כשמציבים $t = 2a$).

גם כאן מרמזות שתי הכוכביות ליד המספר (14), שיש הגבלה בנוגע לשמות העצם t, שניתן להציב כאן במקום x. הסבר הגבלה זו יינתן בפרק הבא.

(VII) כללי $\exists$

הוכחה ישירה, קונסטרוקטיבית, לכך, שמשהו בעל תכונה מסוימת קיים ($\exists xA$) היא על-ידי מתן דוגמה קונקרטית. למשל: מהעובדה ש- $1^3 - 3 \cdot 1 + 2 = 0$ נובע, שלמשוואה $x^3 - 3x + 2 = 0$ יש פתרון (כלומר: $\exists x.x^3 - 3x + 2 = 0$). זהו תוכנו של כלל (15). שתי הכוכביות לידו מרמזות שוב על הגבלה, שתוסבר בפרק הבא.

חשוב לציין, שלמרות ששימוש בכלל (15) להוכחת קיום משהו היא הדרך הישרה (והעדיפה, אם אפשר), אין זו הדרך היחידה. אפשרות אחרת היא ללכת בדרך השלילה, דהיינו: להניח, שלא קיים עצם כמבוקש (כלומר: $\neg\exists xA$, מה ששקול ל- $\forall x\neg A$, כלומר: שלכל x אין התכונה A) ולהגיע לסתירה. דרך זו עלולה שלא לתת רמז, איך למצוא בפועל x בעל התכונה A.

ואיך משתמשים בידע מהצורה $\exists xA$? ובכן, בדרך-כלל מה שעושים, אחרי שהראינו זאת, הוא לקחת משתנה y (לעתים קרובות, אך לא תמיד, זהו x עצמו), עליו לא הונח דבר עד אותו רגע, ולהוסיף לרשימת ההנחות את ההנחה, ש- y מייצג משהו בעל התכונה A ($A(y/x)$). כלל (16) קובע אז, שאם B נובע מאוסף ההנחות המורחב, אז הוא נובע גם מהאוסף המקורי. גם כאן יש הגבלות מסוימות, ואלה תוסברנה בסעיף הבא.

השימוש בכלל (16) נעשה בדרך-כלל באופן מוסווה למדי, ובאופן כה טבעי, עד כי קשה להרגיש, שאכן נעשה בו שימוש. (16) הוא מסוג הכללים, שבדרך-כלל קל הרבה יותר להשתמש בהם, מאשר להבין את ניסוחם!

נביא עתה דוגמה מפורטת להוכחה מתמטית אופיינית. בהוכחה זו נעשה שימוש בחלק ניכר מהכללים, כולל כלל (16). אני מקווה, שהכלל והשימוש בו יובנו אז טוב יותר.

%%

♦ דוגמה:

נניח שידוע, שאת המספר π (שאינו רציונלי) אפשר לקרב קירוב טוב כרצוננו על-ידי מספרים רציונליים, כלומר: שלכל ε חיובי אפשר למצוא מספר רציונלי, שמרחקו מ- π קטן מ- ε. נניח עוד, שגם ל- $\sqrt{2}$ (שאף הוא אינו רציונלי) יש אותה התכונה. צריך להוכיח על סמך זה, שגם לסכומם, $\pi + \sqrt{2}$, יש אותה התכונה.

הוכחת טענה זו בטקסט מתמטי אופייני תיראה כך: "יהי $\varepsilon > 0$. מהנתון על π נובע, שיש a_1 רציונלי כך ש- $|\pi - a_1| < \frac{\varepsilon}{2}$. מהנתון על $\sqrt{2}$ נובע, שיש a_2 רציונלי כך ש- $\left|\sqrt{2} - a_2\right| < \frac{\varepsilon}{2}$. עתה $a_1 + a_2$ הינו רציונלי, ומתקיים:

$$\left|(\pi + \sqrt{2}) - (a_1 + a_2)\right| = \left|(\pi - a_1) + (\sqrt{2} - a_2)\right| \leq |\pi - a_1| + \left|\sqrt{2} - a_2\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

מכאן ש- $a_1 + a_2$ מספר כמבוקש".

כדאי לשים לב, שכדי להבין את ההוכחה אין שום צורך לדעת מה זה, בעצם, "מספר רציונלי". כל מה שצריך לדעת הוא, שאם נחבר שני מספרים רציונליים, נקבל מספר רציונלי. מלבד זאת דרוש רק ידע בסיסי ביותר על חיבור, חיסור, $\leq$ וערך מוחלט (בעיקר "אי שוויון המשולש").

נעבור עתה לניתוח לעומק של ההוכחה הקצרה למעלה ונראה, באילו כללים השתמשנו בה והיכן. נתחיל בניסוח פורמלי מדויק של מה שצריך להוכיח ושל הנתונים (פרט לאלו הקשורים בידע האריתמטי הבסיסי הנחוץ, ומהווים את האקסיומות שברקע).

נתון:

(1) $\forall \varepsilon > 0 \exists a(\text{Rational}(a) \wedge |\pi - a| < \varepsilon)$

(2) $\forall \varepsilon > 0 \exists a(\text{Rational}(a) \wedge |\sqrt{2} - a| < \varepsilon)$

(3) $\forall a \forall b(\text{Rational}(a) \wedge \text{Rational}(b) \Rightarrow \text{Rational}(a + b))$

צ"ל:

$$\forall \varepsilon > 0 \exists a(\text{Rational}(a) \wedge |(\pi + \sqrt{2}) - a| < \varepsilon)$$

אם נתבונן במה שצריך להוכיח, נראה, שמדובר במשפט אוניברסלי, המתחיל ב- "$\forall \varepsilon > 0 ...$". משפט אוניברסלי, כאמור למעלה, מוכיחים בדרך-כלל בעזרת כלל (13) (כלל ההכנסה של "$\forall$"). המלים המאפיינות שימוש בכלל זה הן "יהי $\varepsilon > 0$". פירושן הוא, שאנו לוקחים ε חיובי כלשהו, ומוכיחים לגביו ש-

$$\exists a(\text{Rational}(a) \wedge |(\pi + \sqrt{2}) - a| < \varepsilon).$$

אם נצליח, הרי נוכל להסיק מזאת את המבוקש על-ידי הפעלת כלל (13) *בסוף* ההוכחה.

למעשה, הפתיחה "יהי $\varepsilon > 0$" טומנת בחובה רמז לשימוש *בשני* כללים בסוף ההוכחה. "$\forall \varepsilon > 0 ...$" הינו, כזכור, קיצור של "$\forall \varepsilon(\varepsilon > 0 \Rightarrow \cdots)$". כדי להסיק את מה שצריך להוכיח על סמך כלל (13), יש, למעשה, לקחת ε "כלשהו" ולהסיק עבורו גרירה: שאם

ε זה גדול מאפס, אז ... . הוכחת גרירה, מצדה, נעשית כמעט תמיד על-ידי שימוש בכלל (1). בהוכחה לעיל משמעותו היא, שאנו מצרפים לנתונים (3)-(1) את ההנחה (הזמנית) הבאה:

$$(4) \quad \varepsilon > 0$$

בהמשך אנו מוכיחים את $\exists a(\text{Rational }(a) \wedge |(\pi + \sqrt{2}) - a| < \varepsilon)$ על סמך (4)-(1). מזה נקבל, על סמך כלל (1), ש- $\varepsilon > 0 \Rightarrow \exists a(\ldots)$ נובע מ- (3)-(1) (בלבד!), ומזה נסיק, בעזרת כלל (13), שמה שצריך להוכיח אכן נובע מהנתונים (3)-(1).

אם נסכם: הפתיחה "יהי $\varepsilon > 0$" מציינת שימוש בכללים (1) ו- (13) (בסדר זה) ***בסוף***. העובדה, שאנו משתמשים בסוף בכללים אלה, לא מתוארת במפורש, אלא על הקורא להבין זאת בעצמו!

בשורה הבאה, בהוכחה בעברית למעלה, מצוין, כי אנו מסיקים משהו ממה שנתון על π, כלומר מנתון (1). נתון זה הוא שוב משפט אוניברסלי מהצורה $\forall\varepsilon(\ldots)$. שימוש בנתון כזה נעשה בדרך-כלל על סמך כלל (14) – כלל הסילוק של $\forall$. במקרה שלפנינו מופיע המשתנה "ε" במקום המשתנה "x", שמופיע בניסוח כלל (14), ושם העצם t, לגביו אנו מיישמים את הכלל, הוא $t = \frac{\varepsilon}{2}$. אנו מקבלים אז מהנתון (1), הלא הוא:

$$\forall\varepsilon(\varepsilon > 0 \Rightarrow \exists a(\text{Rational }(a) \wedge |\pi - a| < \varepsilon))$$

את:

$$\frac{\varepsilon}{2} > 0 \Rightarrow \exists a(\text{Rational }(a) \wedge |\pi - a| < \frac{\varepsilon}{2})$$

בשלב זה יש מספר צעדים הנעשים באופן אוטומטי, וכלל לא מוזכרים בהוכחה. כך ידוע באופן כללי, ש- $\forall x(x > 0 \Rightarrow \frac{x}{2} > 0)$. מזה ומכלל (14) (הפעם עם $t = \varepsilon$) נקבל $\varepsilon > 0 \Rightarrow \frac{\varepsilon}{2} > 0$. מזה ומההנחה (4) נובע $\frac{\varepsilon}{2} > 0$, ויחד עם המסקנה הקודמת מקבלים (בעזרת MP):

$$(\text{I}) \qquad \exists a(\text{Rational }(a) \wedge |\pi - a| < \frac{\varepsilon}{2})$$

המסקנה, שהגענו אליה זה עתה, היא טענה אקזיסטנציאלית ($\exists a(\ldots)$). שימוש במסקנה זו נעשה כמעט תמיד בעזרת כלל (16) – כלל הסילוק של $\exists$: ברגע שהוכחנו, שקיים "a" בעל תכונה מסוימת, אנו נותנים ***שם*** ל- "a" כזה. השם צריך להיות שונה מכל שם, שנעשה בו שימוש עד שלב זה. בהוכחה כאן הוא לא יכול לכן להיות π או ε. שמות אלו תפוסים. הוא כן יכול להיות a עצמו (שעדיין לא משמש כשם לשום דבר

ספציפי!) – ובדרך-כלל זה אכן מה שעושים. בניסוח ההוכחה בעברית למעלה נבחר אבל שם אחר: a_1 (הסיבה תובהר בהמשך). אחרי בחירת השם מתווספת לרשימת ההנחות הנחה נוספת, המתקבלת מ"זריקת" $\exists a$ שבראש (I) למעלה, והצבת a_1 במקום a במה שנותר:

$$\text{(5)} \qquad \text{Rational}\ (a_1) \wedge |\pi - a_1| < \frac{\varepsilon}{2}$$

חשוב לציין, ש- (5) ***אינו*** מסקנה לוגית של (I). הוא ***הנחה*** חדשה. כלל (16) מבטיח אבל, כי טענה, שנוכיח בעזרת הנחה זו, נובעת גם מ- (I) – בתנאי שאינה מזכירה את a_1. אנו נשתדל לכן מעתה להוכיח את $\exists a(\text{Rational}\ (a) \wedge |(\pi + \sqrt{2}) - a)| < \varepsilon)$ על סמך (5)-(1). אם נצליח, הרי מכלל (16) ינבע, שטענה זו נובעת בעצם מ- (4)-(1) בלבד, כמבוקש. גם כאן אין כל זאת מוסבר בגוף ההוכחה. כמו במקרים קודמים, העובדה, שנעשה שימוש בכלל (16) בשלב מאוחר, מובלעת במלים "קיים a_1 כך ש...".

בשלב הבא בהוכחה המקורית (המשפט השלישי בה!) חוזרים על אותה צורת חשיבה כמו בשלב הקודם, אלא שבמקום בנתון (1) משתמשים בנתון (2). הפעם מקבלים:

$$\text{(II)} \qquad \exists a(\text{Rational}\ (a) \wedge |\sqrt{2} - a| < \frac{\varepsilon}{2})$$

בנקודה זו עושים שוב הכנה לשימוש בכלל (16) על-ידי בחירת שם מתאים עבור אותו "a", שקיומו מובטח ב- (II). השם "a_1" תפוס עתה – כבר נעשה בו שימוש. אי לכך בוחרים שם חדש – "a_2" במקרה זה. עתה מוסיפים לרשימת ההנחות את ההנחה של-a_2 התכונה המתוארת ב- (II):

$$\text{(6)} \qquad \text{Rational}\ (a_2) \wedge |\sqrt{2} - a_2| < \frac{\varepsilon}{2}$$

את המסקנה המבוקשת ננסה עתה להוכיח על סמך (6)-(1).

הערה:
אילו בשלב הקודם היינו עושים שימוש בשם "a" עצמו ולא ב- "a_1", היינו חייבים בכל מקרה לבחור כאן שם אחר, כמו b, או a'. בחירת השמות a_1 ו- a_2 מרמזת על קשר לנתונים (1) ו- (2) (בהתאמה), ויש לה כאן לכן עדיפות ***פסיכולוגית***.

המשפט הבא, הרביעי במספר, בהוכחה בעברית, הוא: "עתה, $a_1 + a_2$ הינו רציונלי". עובדה זו מוסקת מ- (3), (5) ו- (6) באופן הבא: מנתון (3) מסיקים, על-ידי שני שימושים בכלל (14):

$$\text{(III)} \qquad \text{Rational}\ (a_1) \wedge \text{Rational}\ (a_2) \Rightarrow \text{Rational}\ (a_1 + a_2)$$

מצד שני, מ- (5) נובע, בעזרת כלל (4), $\text{Rational}\ (a_1)$, בעוד מ- (6) נובע, שוב בעזרת כלל (4), $\text{Rational}\ (a_2)$. משני אלה נקבל באמצעות כלל (3):

$$\text{(IV)} \qquad \text{Rational}\ (a_1) \wedge \text{Rational}\ (a_2)$$

מ- (III) ו- (IV) אפשר להסיק מיידית בעזרת MP (כלל (2)) את $\text{Rational}\ (a_1 + a_2)$. זהו בדיוק תוכן המשפט הרביעי בהוכחה בעברית. ראוי לציין, שכדי לקבל אותו יש צורך בשישה צעדים לוגיים! זהו מצב אופייני. פירוט מלא של כל הצעדים הלוגיים היה הופך כל הוכחה פשוטה לארוכה באופן בלתי נסבל (לבני אדם, לפחות), ולכן "קפיצות" נעשות כל הזמן.

השורה הבאה בהוכחה בעברית למעלה מכילה חישוב, המתבסס על תכונות של $=$, $\leq$, $+$ וערך מוחלט (כמו למשל: $\forall x \forall y.|x + y| \leq |x| + |y|$). לא נפרט אותו כאן, כי שלב זה בהוכחה אינו עושה בעיות בהבנה. נציין רק, שגם בשלב זה נעשה, כמובן, שימוש בכללים לוגיים, בעיקר בכלל (14) (איפה?). המסקנה הסופית מחישוב זה היא, על כל פנים:

$$\text{(V)} \qquad |(\pi + \sqrt{2}) - (a_1 + a_2)| \leq \varepsilon$$

מזה ומהמסקנה $\text{Rational}\ (a_1 + a_2)$ של השלב הקודם, אנו מקבלים, לפי כלל (3):

$$\text{Rational}\ (a_1 + a_2) \wedge |(\pi + \sqrt{2}) - (a_1 + a_2)| < \varepsilon$$

מעובדה זו אנו מסיקים, בעזרת כלל (15) (כלל ההכנסה של $\exists$):

$$\text{(VI)} \qquad \exists a(\text{Rational}\ (a) \wedge |(\pi + \sqrt{2}) - a| < \varepsilon)$$

השימוש בכלל (15) נעשה עם המשתנה a במקום המשתנה x, ועם $t = a_1 + a_2$. המלים "$a_1 + a_2$ הוא מספר כמבוקש" מציינות שזהו אכן t, ושאכן נעשה פה שימוש בכלל (15).

ההוכחה בעברית נסתיימה בשלב זה. למעשה הוכח עד כה רק ש- (VI) נובע מ- (6)-(1). נזכיר לקוראים, שעתה באים למעשה שני שימושים בכלל (16), בעזרתם מקבלים, ש- (VI) נובע בעצם מ- (4)-(1) בלבד. לבסוף באים שימוש בכלל (1) ואחר-כך שימוש בכלל (13). רק אז מקבלים, שמה שרצינו להוכיח, נובע אכן מהנתונים (3)-(1).

%%

א.5. על השימוש במשתנים במתמטיקה

בכל שפה יש מספר קטגוריות דקדוקיות. בשפה העברית, למשל, יש קטגוריות כמו "פועל", "שם עצם", "תואר השם", "משפט" ועוד. בשפה המתמטית מספר הקטגוריות הבסיסיות הוא 2: שמות עצם (או ביטויים – terms) ונוסחאות (או פסוקים, או טענות). דוגמה טובה לסוג הראשון הם ביטויים אלגבריים כמו $\frac{a}{a+2}$. משוואות וזהויות אלגבריות הן דוגמה טובה לסוג השני. בלא שנספק הגדרות מדויקות, ביטויים הם דבר לו אנו מייחסים ערך (value), בעוד נוסחאות הן דבר אותו אנו רואים כנכון או לא נכון. לדוגמה: ערך הביטוי $1 + 2$ הוא 3, $1 < 2$ היא נוסחה נכונה, ואילו $1 = 2$ היא נוסחה שקרית.[8]

להוציא מקרים פשוטים במיוחד, כמו הדוגמאות שהבאנו לפני רגע, הן ביטויים והן נוסחאות מכילים בדרך-כלל ***משתנים*** (variables). משתנים הם יצירים סינטקטיים בעלי שימוש רב הן במתמטיקה והן בשפות מחשב. בסעיף זה נדון לעומק בצורת השימוש במשתנים במתמטיקה.

נתחיל באבחנה יסודית בין שתי הצורות, שבהן יכול משתנה להופיע בביטוי או נוסחה מתמטיים: כמשתנה חופשי או כמשתנה קשור. כדי להבין למה הכוונה נפתח בדוגמאות.

נחזור לדוגמה של הביטוי $\frac{a}{a+2}$. זהו ביטוי, ולכן אמור להיות לו ערך. אבל מהו ערך זה? אין לכך, כמובן, תשובה מוגדרת. הדבר תלוי בערכו של המשתנה a. כאשר $a = 1$, ערך הביטוי הוא $\frac{1}{3}$. כש- $a = 0$, ערכו הוא 0. המשתנה משמש כאן לא לצורך יצירת ביטוי בעל ערך קונקרטי, אלא לצורך יצירת **תבנית** של ביטויים בעלי ערך קונקרטי. ביטויים כאלה מתקבלים, כאשר מציבים ערכים קונקרטיים במקום המשתנה a. כאשר משתמשים במשתנה בצורה כזו, דהיינו: לצורך יצירת תבניות של ביטויים או תבניות של נוסחאות, אנו קוראים לו משתנה ***חופשי*** (בביטוי או בנוסחה) או ***פרמטר***.

---

[8] יש לשים לב היטב, שמובן המונח "נוסחה" כאן הוא רחב בהרבה ממה שהקוראים רגילים לו!

הבה נשווה עתה את השימוש ב- a בביטוי $\frac{a}{a+2}$ לשימוש ב- x בביטוי $\int_0^1 x^2dx$. הביטוי השני מכיל משתנה (x), ובכל זאת יש לו ערך קונקרטי ($\frac{1}{3}$ במקרה זה). אין כאן צורך להציב ערכים במקום x. למעלה מזאת: אי אפשר כלל להציב ערכים במקום x! אם, למשל, נציב 2 במקום x, נקבל את $\int_0^1 2^2d2$, וזהו צירוף סימנים חסר כל מובן.
ברור אפוא, ש- x משמש כאן לצורך שונה מזה שבמקרים הקודמים. מבלי לנסות להגדיר זאת במדויק, נאמר, שלפנינו שימוש ב- x כמשתנה *קשור*. סימן מובהק להיות משתנה קשור הוא, שאין מובן להצבת ערכים קונקרטיים במקומו.

נעבור לדוגמה נוספת. בנוסחה $x^2 = x$ x הוא משתנה חופשי. אם נציב במקום x 1, נקבל טענה נכונה, ואם נציב במקומו 2, נקבל טענה לא נכונה. לעומת זאת, ב- $\forall x(x^2 = x)$ וב- $\exists x(x^2 = x)$ x הוא משתנה קשור. לשתי הנוסחאות יש ערך אמת ברור במספרים הטבעיים (הראשונה שקרית, השנייה אמיתית). ערך זה אינו תלוי בשום הצבה עבור המשתנה x. נהפוך הוא. ניסיון להציב 7, למשל, במקום x ייתן את $\forall 7(7^2 = 7)$ ו- $\exists 7(7^2 = 7)$: שתי מחרוזות של סימנים, שהן חסרות מובן ולא דקדוקיות!

דוגמה שלישית: נתבונן בנוסחה $\forall x(x \geq y)$. מופיעים כאן שני משתנים: x ו- y. y הוא משתנה חופשי. בגלל נוכחותו אין הנוסחה מביעה טענה בעלת ערך-אמת מוגדר. נוכל אבל להציב במקום y 2 ולקבל את הטענה השקרית $\forall x(x \geq 2)$. נוכל גם להציב במקום y 0 ולקבל את הטענה $\forall x(x \geq 0)$. טענה זו היא נכונה בתחום המספרים הטבעיים, ושקרית בתחום המספרים הממשיים. המשתנה x, לעומת זאת, הוא קשור כאן, ואין לערכו שום קשר לנכונות או אי-נכונות הנוסחה. הצבת 2 במקומו תיתן $\forall 2(2 \geq y)$, וזוהי מחרוזת סימנים חסרת-שחר (בלא כל קשר לתחום).

מהו הדבר ההופך משתנה חופשי למשתנה קשור? התשובה היא: הכנסתו לתוך הטווח (scope) של מה שנקרא ***אופרטור קשירה*** (binding operator). למושג זה לא ניתן כאן הגדרה מדויקת. בסופו של דבר, אופרטור קושר משתנים אם הוא מוכרז בתור שכזה. טבלה מס' א.5 כוללת דוגמאות חשובות לאופרטורי קשירה במתמטיקה (ובשפות תכנות). אנו נסקור את מרבית הדוגמאות, ונעמוד על מספר מאפיינים המשותפים להם (שני האופרטורים האחרונים בטבלה הם חשובים במיוחד, אך יילמדו רק בהמשך הקורס). זה יספיק לצרכינו. בכל הדוגמאות כולל תיאור אופרטור הקשירה את שלושת המרכיבים הבאים:

(א) סימבול מיוחד המשמש כשם האופרטור.

(ב) אמצעי לציון אילו משתנים האופרטור קושר בכל מקרה של שימוש בו.

(ג) אמצעי לציון הטווח של האופרטור, דהיינו: התחום שבו הוא קושר את המופעים (occurrence) ה*חופשיים* של המשתנה הנקשר (משפט זה יובהר בהמשך).

טבלה א.5: דוגמאות לאופרטורים קושרים

	אופרטור	**דוגמה**
(1	$\forall x(\ldots)$ $\quad$ $\exists x(\ldots)$	$\forall x((x+1)^2 > x)$ $\quad$ $\exists x((x+1)^2 > x)$
(2	$\int(\ldots)dx$	$\int_0^1 (x^2+x)dx$
(3	$\sum_{j=\ldots}(\quad)$	$\sum_{j=1}^{10} (j^2+j)$
(4	$\lim_{y\to\ldots}(\quad)$	$\lim_{y\to 0}(2^y-1)$
(5	**function** ______ (z: real): real; **begin** ... **end**	**function** f(x: real): real; **begin** f := sqr(x + 1) **end**
(6	$\mu k. \ldots$	$\mu k.\exists i(k = 3\cdot i) \wedge \exists j(k = 2j)$
(7	$\iota x. \ldots$	$\iota x. x > 0 \wedge (x+1)^2 = 4$
(8	$\{x \mid \ldots\}$	$\{x \mid (x+1)^2 > x\}$
(9	$\lambda x. \ldots$	$\lambda x. (x+1)^2$

נעבור לפירוט הדוגמאות.

1. בלוגיקה, הכמתים (שסימוניהם, כזכור, $\forall$, $\exists$) הם אופרטורי קשירה. הם קושרים את המשתנה הנכתב מיד אחריהם, בעוד טווח הקשירה הוא התחום בתוך הסוגריים, הבאים מיד אחר-כך. הבה נתבונן, למשל, בנוסחה הבאה:

$$(*) \qquad \forall y[(\exists x(y^2 = x)) \vee x \cdot y = 0]$$

ב- (*) יש שלושה מופעים של המשתנה y. כולם הינם קשורים על-ידי הכמת האוניברסלי $\forall$, שמופיע בהתחלה. העובדה ש- $\forall$ זה קושר דווקא את y, נקבעת על-ידי זה, ש- y הוא המשתנה המופיע מיד אחריו. טווח הקשירה של $\forall$ זה הוא מה שמצוי בתוך הסוגריים המרובעות, דהיינו: הנוסחה $\exists x(y^2 = x) \vee x \cdot y = 0$. בנוסחה זו, ***כשלעצמה*** (כאשר מנתקים אותה מהקונטקסט (*), בו היא מופיעה כאן) יש ל- y שני מופעים חופשיים. בנוסחה (*) המלאה, שני מופעים חופשיים אלה נקשרים על-ידי $\forall y$. בנוסחה (*) מופיע גם הכמת $\exists$. המשתנה, שהוא קושר בה, הוא x. טווח הקשירה שלו הוא הנוסחה $y^2 = x$, הנמצאת בסוגריים שאחרי $\exists x$. לכן המופע השני של x ב- (*) (זה שאחרי סימן השוויון) הינו קשור. לעומת זאת, המופע השלישי והאחרון של x ב- (*) (זה שב- $x \cdot y = 0$) נמצא מחוץ לטווח הקשירה של $\exists x$, ולכן הינו ***חופשי***. נוכל להציב במקומו 0 ולקבל את הפסוק הנכון $\forall y[(\exists x(y^2 = x) \vee 0 \cdot y = 0]$. נוכל גם להציב במקומו את 1 ולקבל את הפסוק השקרי (בממשיים) $\forall y[(\exists x(x^2 = y) \vee 1 \cdot y = 0]$ (הערה: ***"פסוק"*** הוא שם מקובל בלוגיקה לנוסחה, שאין בה משתנים חופשיים. לפסוק חייב להיות ערך אמת מוגדר: אמת או שקר, בהתאם לתחום בו מדובר).

2. הדוגמה המפורשת הראשונה של אופרטור קשירה, שסטודנטים נתקלים בה בחייהם, היא בדרך-כלל אופרטור האינטגרציה. הסימן המקובל הוא, כידוע: $\int$. את המשתנה הנקשר כותבים דווקא בסוף הביטוי, אחרי האות d, ואילו טווח הקשירה הוא כל מה שנמצא בין סימן האינטגרל $\int$ ובין אותו d. כך, למשל, בביטוי $\int_0^1 (ax^2 + 2x)dx$ המשתנה x הינו קשור בכל מקום בו הוא מופיע (כפי שמלמד ה- "dx"), בעוד a הינו פרמטר, דהיינו משתנה חופשי.

אופרטור האינטגרציה הינו אופרטור מסובך למדי מבחינה תחבירית, כי הוא קושר באופן חלקי. כאשר הוא קושר משתנה x, הוא קושר את המופעים שלו בטווח שציינו (בין ה- $\int$ וה- dx), אך לא שום דבר המופיע בגבול העליון או בגבול התחתון של סימן האינטגרל. כך בביטוי $\int_0^y xdx$ ה- y הינו משתנה חופשי. למעלה מזאת: גם

בביטוי $\int_0^x xdx$ המופע של x, הנמצא מעל סימן האינטגרל, הינו חופשי, ואין כל קשר בינו ובין המופעים של x הנמצאים בטווח הקשירה. בביטוי $\int_0^x xdx$ אפשר להציב $x = 1$, למשל, והתוצאה תהיה $\int_0^1 xdx$ (ולא $\int_0^1 1d1$, שהוא קשקוש).

3. אופרטור הסכום Σ (ואופרטור המכפלה $\prod$) גם הם אופרטורי קשירה. במקרה זה מצוין המשתנה הנקשר משמאלו של סימן השוויון המופיע מתחת לסימן Σ. טווח הקשירה הוא האזור שמימין לסימן Σ (כשלעתים משתמשים בסוגריים כדי לסמן היכן הוא מסתיים). כך, למשל, בביטויים $\sum_{i=k}^{n} ni^2$ ו- $\sum_{i=k}^{n} (ni^2 + k)$ i הינו משתנה קשור, בעוד k ו- n הינם משתנים חופשיים. ערך הביטוי הראשון כש- $k = 1$ ו- $n = 3$, הינו $\sum_{i=1}^{3} 3i^2$ (כלומר $42 = 3 \cdot 3^2 + 3 \cdot 2^2 + 3 \cdot 1^2$). אין, לעומת זאת, שום משמעות, כמובן, להציב כאן $i = 2$ (נאמר). ל"ביטוי" $\sum_{2=k}^{n} n \cdot 2^2$ אין כל משמעות!

4. בחשבון דיפרנציאלי יש שימוש רחב לאופרטור הגבול lim, שאף הוא אופרטור קשירה. המקום בו מצוין המשתנה הנקשר דומה כאן למקרה של Σ (אלא שבמקום סימן השוויון "=" משתמשים כאן ב- "$\to$"). כך גם מיקום טווח הקשירה. למשל: בביטוי $\lim_{n\to\infty} \frac{k}{n}$ n הינו קשור, בעוד k חופשי. בדומה, ב- $\lim_{x\to a} (x^2 - x)$ x הינו משתנה קשור, בעוד a חופשי (שוב, בביטוי השני אפשר להציב $a = 1$ ולקבל $\lim_{x\to 1} (x^2 - x)$, שזה 0, אך אי-אפשר להציב $x = 1$: אין משמעות ל- $\lim_{1\to a} (1^2 - 1)$).

5. הדוגמה החמישית בטבלה א.5 היא דוגמה לאופרטור קשירה בשפת התכנות פסקל. המשתנה הנקשר מוצהר כאן בתוך סוגריים, המופיעים מיד אחרי השם הניתן לפונקציה המוגדרת, בעוד טווח הקשירה הוא מה שנמצא בין המלים "begin" ו- "end".

בטרם נמשיך בדוגמאות, הבה נהרהר רגע בשאלה, שדומני כי היא מתבקשת: האם באמת הכרחי הדבר, שיהיו כל-כך הרבה צורות שונות ומשונות על מנת לציין, מיהו

המשתנה הנקשר, ומהו טווח הקשירה? האם אין זה אפשרי ועדיף לעשות זאת בצורה אחידה ופשוטה, כמו במקרה של $\forall$ ו- $\exists$? התשובה היא, שזה בהחלט אפשרי ובהחלט עדיף. עם זאת, במקרה של דוגמאות 2-4 אנו כבולים עדיין למורשת היסטורית של סימונים, שנוצרה בתקופה, שבה נושא השימוש במשתנים בכלל (וקשירתם בפרט) היה מאוד לא ברור. מסורות, לרוע המזל, קשה מאוד (מאוד!) לשנות. אף-על-פי-כן, כבר כיום נהוג סימון אחיד באי-אלו מערכות ממוחשבות, שנועדו לבדיקה אוטומטית של הוכחות מתמטיות ואף לגילוין. ייתכן מאוד, שהשפעת המחשב תביא בסופו של דבר לשינוי גם כאן.

שתי הדוגמאות הבאות לאופרטורי קשירה אינן נמצאות כל-כך בשימוש במתמטיקה, אלא בלוגיקה מתקדמת ובמדעי המחשב התיאורטיים. הם הוכנסו לשימוש במאה העשרים, וצורת הסימון היא לכן מודרנית יותר:

6. בתחום תורת הפונקציות החשיבות (computable) מקובל השימוש באופרטור הקשירה μ ("מיו"). המשתנה הנקשר נכתב כאן (כמו במקרה של $\forall$ ו- $\exists$) מיד אחרי האופרטור. טווח הקשירה נכתב או בסוגריים מיד אחר-כך, או מתחיל בנקודה הנכתבת מיד לאחר ציון המשתנה הנקשר, ונגמר או בסוף הטקסט, או באיזה סוגר ימני ")", שהסוגר השמאלי "(" המתאים לו נכתב עוד לפני ה- μ (כמו, למשל, בביטוי $7 \cdot (1 + \mu n.\underline{n \geq n^2})$. טווח הקשירה כאן הוא הנוסחה $n \geq n^2$).
משמעות הביטוי $\mu n.$____ היא: "ה- n הטבעי הראשון כך ש- ______". למשל: הביטוי $\mu k.\exists a(k = 3 \cdot a) \wedge \exists b(k = 2 \cdot b)$ מציין את המספר הטבעי הראשון, המתחלק גם ב- 3 וגם ב- 2, דהיינו: 6 (או 0, כאשר 0 נחשב למספר טבעי).
יש לציין, שלא תמיד לביטוי עם μ יש ערך, אפילו אם הוא תקין מבחינה דקדוקית. כך אין שום ערך לביטוי $\mu n.\, n > n + 1$. עניין זה אין בו כל חידוש, כמובן. גם הביטויים $\lim_{n \to \infty} (-1)^n$ ו- $1/0$ הם כשרים מבחינה דקדוקית, אך אין להם ערך מוגדר.

7. סימון אחר, שנשתמש בו בהמשך (בצורה מוגבלת) ואינו מוכר למרבית המתמטיקאים, הוא האופרטור ι ("איוטה"), המקביל להא-הידוע בעברית. הכללים כאן עבור ציון המשתנה הנקשר וטווח הקשירה זהים לאלה של μ. המשמעות אבל שונה: $\iota x.$____ מציין את ה- x היחיד, שיש לו התכונה ______. למשל: ערך הביטוי $\iota x.2x + 1 = 7$ הוא 3. כמו כן, $(\iota x.x < 0 \wedge x^2 = 1) = -1$, כיוון ש- -1 הוא המספר היחיד, שהוא גם שלילי וגם ריבועו הוא 1. ברור שהביטוי $\iota x.$___ הוא בעל משמעות רק כ*שיש* אכן x *יחיד* בעל התכונה ____.

כך הביטוי $\iota x\, x^2 = -1$ אינו מוגדר בממשיים, כיוון שאין בכלל מספר ממשי שריבועו -1. אותו הביטוי עצמו אינו מוגדר גם במרוכבים, אבל מסיבה אחרת: שם יש שני מספרים, שריבועם -1. הדבר זהה לחלוטין לשימושה של הא-הידוע בעברית. לשם העצם "המלכה הנוכחית של אנגליה" יש ערך מוגדר בזמן כתיבת שורות אלו. לשמות העצם "המלכה הנוכחית של צרפת" ו"הרב הראשי של ישראל" אין, אך מסיבות שונות.

הערות

%%

א. כל הדוגמאות בטבלה הן של אופרטורי קשירה, הקושרים בדיוק משתנה אחד. אופרטורים הקושרים בבת-אחת יותר ממשתנה אחד נדירים יותר, אך קיימים. כך מקובל בחשבון דיפרנציאלי ואינטגרלי מתקדם להשתמש באופרטור הגבול (lim) לצורך קשירת מספר משתנים בבת-אחת. במקרה של lim, למשל, מציין הביטוי $\lim\limits_{\substack{x\to 0\\ y\to 0}} \frac{xy}{x^2+y^2}$ "גבול כפול", וה- lim קושר פה בבת-אחת גם את x וגם את y. יש להבדיל בין ביטוי זה ובין הביטוי $\lim\limits_{y\to 0}\lim\limits_{x\to 0} \frac{xy}{x^2+y^2}$, שגם בו שני המשתנים קשורים, אך כל אחד מהם נקשר על-ידי אופרטור lim אחר. שני הביטויים שונים קודם כל מבחינה *דקדוקית*, ובמקרה פרטי זה גם בערכם (לראשון *אין* כאן, למעשה, ערך). אחת השאלות החשובות בחשבון דיפרנציאלי ואינטגרלי מתקדם היא: מתי ערך ביטוי המוגדר בעזרת קשירה "מקבילה" שווה לערך הביטוי המוגדר על-ידי קשירה "סדרתית"[9] (כלומר, שהמשתנים נקשרים בו זה אחר זה, לא בבת-אחת). שאלה דומה מתעוררת בקשר לאופרטור של אינטגרל כפול: $\iint$. אופרטור זה גם הוא קושר בבת-אחת שני משתנים, כמו בביטוי: $\iint\limits_{x^2+y^2\le 1} (x+2y)dxdy$. כל זאת שייך אבל לקורס שני בחשבון דיפרנציאלי ואינטגרלי. בקורס *הנוכחי* נשתמש מדי פעם באופרטורים של דוגמאות 8, 9 לצורך קשירת מספר משתנים בבת-אחת, אולם עד שנגיע לזה – עוד חזון למועד.

%%

ב. השימוש באופרטורים השונים כרוך בכללים תחביריים נוספים פרט לאלה שציינו. במיוחד חשוב לשים לב ל"קלט" ול"פלט" המצופים כאן. כך לגבי $\forall$ ו- $\exists$ חייבים אנו להציב *נוסחאות* בטווח שלהם, ומה שנקבל יהיה אף הוא נוסחה. אם ננסה להציב ביטוי בטווח של $\forall$ ו- $\exists$ נקבל משהו, שאינו תקין מבחינה תחבירית.

---

[9] הטרמינולוגיה שאולה כאן ממדעי המחשב.

בדוגמאות 2-4 יש להציב ***ביטויים*** בטווח והתוצאה תהיה אף היא ***ביטוי***. הפרת כללים אלה תיצור דברים חסרי שחר מבחינה תחבירית.

ג. כמו שראינו בדוגמאות השונות, מה שאופרטורי הקשירה קושרים אינו בעצם משתנים, אלא ***מופעים*** של משתנים. אותו משתנה עצמו יכול לכן להופיע בנוסחה או בביטוי מסוימים הן כמשתנה חופשי והן כמשתנה קשור (במקומות שונים, כמובן). יתר על כן, מופעים שונים של אותו משתנה יכולים להיקשר על-ידי אופרטורים שונים, כמו בדוגמה הבאה: $\forall x(x^2 \geq 0) \vee \forall x(x^2 < 0)$. שני המופעים הראשונים של x נקשרים על-ידי ה- $\forall$ הראשון, שני המופעים האחרונים על-ידי ה- $\forall$ השני. אין כל קשר לכן בין שני המופעים הראשונים ובין שני המופעים האחרונים. מכל בחינה פרקטית מדובר במופעים של משתנים שונים. בדומה, אין כל קשר בין מופעים חופשיים ומופעים קשורים של אותו משתנה. כבר ראינו, שהצבת $x = 2$ בביטוי כמו $\int_0^x x^2dx$ נותנת $\int_0^2 x^2dx$ $(= 8/3)$ ולא $\int_0^2 2^2d2$. כלומר, אין כל קשר בין ה- x-ים בטווח הקשירה ובין ה- x-ים בגבול העליון (או התחתון) של האינטגרל.

ד. מה שאופרטור קשירה קושר אינו כל המופעים של המשתנה הנקשר הנמצאים בטווח הקשירה, אלא רק אותם מופעים, שהם עדיין חופשיים שם. לדוגמה, נניח שלנוסחה (*) למעלה (בדוגמה של $\forall$ ו- $\exists$) נוסיף $\exists x$ בהתחלה. נקבל:

$$\exists x\{\forall y[(\exists x(y^2 = x)) \vee x \cdot y = 0]\} \qquad (**)$$

טווח הקשירה של ה- $\exists x$ החדש הוא מה שנמצא בסוגריים המסולסלים, כלומר הנוסחה (*) המקורית. הוא קושר שם אבל רק את המופע האחרון של x (זה שב- $(xy = 0)$). מופע זה, כזכור, הינו חופשי ב- (*) (אבל ב- (**) הוא קשור, כמובן). לעומת זאת, המופעים האחרים של x בתוך הסוגריים המסולסלות הינם קשורים כבר על-ידי ה- $\exists x$ הפנימי, ואי-אפשר לקשור אותם מחדש!

ה. קורה לא פעם, שמסיבות של נוחיות (או מסיבות היסטוריות) נחשבים מופעים של משתנים בקונטקסט מסוים לקשורים, למרות ששום אופרטור קשירה מפורש לא נראה בשטח. אופרטור קשירה, שמושמט לעתים קרובות מאוד, הוא האופרטור λ, עליו נלמד בהמשך. עוד יותר נפוצה השמטתו של הכמת $\forall$, כאשר הוא מופיע בראשית נוסחה. ודאי זכורה לקוראים ההבדלה, הנעשית בבית-הספר

התיכון, בין "זהויות" ובין שוויונות סתם, וחשיבות הבדלה זו: הטיפול הלוגי הוא שונה מאוד כאשר מדובר ב"זהות". עם זאת, אין בבית-הספר התיכון כל הבדל בין הצורה, שבה נכתבת "זהות", ובין הצורה שבה נכתב שוויון סתם. האמת היא אבל, ש"זהות" אינה אלא נוסחה עם כמתים אוניברסליים בהתחלה, שפשוט לא טורחים לכתוב אותם. כאשר אנו אומרים, למשל, שהנוסחה $(a+b)^2 = a^2 + 2ab + b^2$ הינה זהות, כוונתנו היא, בעצם, שהנוסחה הבאה היא נכונה:

$$\forall a \forall b . (a+b)^2 = a^2 + 2ab + b^2$$

בדומה, כשאנו מוצאים בספרי לימוד את המשפטים הבאים:

$$1 + 2 + \ldots + n = \frac{n(n+1)}{2}$$

$$\int_0^x x^a dx = \frac{x^{a+1}}{a+1} \qquad (a \neq -1)$$

הכוונה היא, בעצם, לטענות הבאות:

$$\forall n \, . \, 1 + 2 + \ldots + n = \frac{n(n+1)}{2}$$

$$\forall x \forall a \left(a \neq -1 \Rightarrow \int_0^x x^a dx = \frac{x^{a+1}}{a+1} \right)$$

(כדאי לשים לב לכך, שבדוגמה האחרונה יש מופעים של x הנקשרים על-ידי $\forall x$, בעוד מופעים אחרים נקשרים על-ידי $\int \ldots dx$).

אחת הדרכים להיווכח, שבזהות כל המשתנים הינם קשורים, הינה, שעל זהות אנו שואלים אם היא נכונה או לא נכונה באופן אבסולוטי, ללא תלות בערכם של המשתנים החופשיים, שכביכול נמצאים בה. בדומה, אנו אומרים לעתים קרובות, שאין זה נכון ש- $(a+b)^2 = a^2 + b^2$. זאת, למרות שיש הצבות (כש- $a = 0$, למשל), שבהן נקבל מנוסחה זו עובדות נכונות דווקא. מה שמתכוון מי שטוען טענה זו הוא, כמובן, שהנוסחה $\forall a \forall b . (a+b)^2 = a^2 + b^2$ אינה נכונה.

זיהוי העובדה, שמשתנים מסוימים הם בעצם קשורים, ומי (והיכן) הוא האופרטור הקושר אותם, הוא חשוב ביותר, ובמיוחד כשאופרטור זה מושמט. זאת, כיוון שהטיפול בהם נעשה כאילו האופרטור היה כתוב במפורש!

הקדשנו דיון ארוך מאוד להבחנה בין משתנים קשורים וחופשיים – דיון שעלול להיראות בקריאה ראשונה מסובך למדי. מן הסתם, התגנבה בשלב כלשהו ללב הקוראים התהייה: בשביל מה כל זה? התשובה היא, שהטיפול הלוגי בשני סוגי המשתנים הינו שונה לחלוטין. למעשה, מקור מרבית אי ההבנות, הטעויות והבלבול, שיש לסטודנטים בתחומים רבים של המתמטיקה, הוא באי הבחנה, באיזו צורה משתנים מתפקדים בקונטקסט מסוים. כדי שהדיון שעשינו יוכל לעזור לנו להימנע מטעויות כאלה בעתיד, עלינו להכיר את הכללים הלוגיים העיקריים הקשורים למשתנים. זו תהיה מטרתנו הבאה.

כלל α

כבר ציינו מספר פעמים, שאי אפשר להחליף משתנים הקשורים בקונטקסט מסוים בערכים קונקרטיים ולקבל משהו בעל משמעות (אלא אם כן נעשה שינוי גם בקונטקסט – דוגמה חשובה במיוחד לכך נראה בהמשך). בדומה, אין גם כל מובן להצבת ביטוי מורכב במקום משתנה קשור.[10] כך אין כל משמעות למחרוזת הסימנים: $\forall(x+1).\ x+1 \geq y$, המתקבלת על-ידי הצבת $x+1$ במקום z ב- $\forall z.\ z \geq y$ (יש אבל, כמובן, בהחלט מובן להצבת $x+1$ במקום y בנוסחה זו, כיוון ש- y חופשי בה). עם זאת, יש סוג אחד של הצבה, שאפשר לעשות במקום משתנים קשורים, והתוצאה לא זו בלבד שהיא בעלת מובן, אלא שמובן זה *זהה* למובן של הביטוי או הנוסחה המקוריים (למעט מקרים יוצאים מן הכלל, אותם נפרט בהמשך). הכוונה היא להחלפתו של משתנה קשור אחד במשתנה אחר. ליתר דיוק: אפשר להחליף את המשתנה הנקשר על-ידי אופרטור קשירה מסוים (ביחד עם כל המופעים של אותו משתנה, הנקשרים על-ידי אותו אופרטור) במשתנה אחר. אפשרות זו ידועה בשם כלל (α) . הפעלת כלל α על ביטוי תיתן ביטוי השווה לביטוי המקורי (עבור כל ערך של המשתנים החופשיים שלו). הפעלת כלל α על נוסחה תיתן נוסחה השקולה לוגית לנוסחה המקורית.

דוגמאות:

$\forall x.\ x=x$ **שקול לוגית ל-** $\forall y.\ y=y$.

$\forall x \exists y.\ x>y$ **שקול לוגית ל-** $\forall x \exists a.\ x>a$ **ול-** $\forall b \exists a.\ b>a$

$$\int_0^a x^2 dx = \int_0^a y^2 dy$$

[10] בכל המקומות המעטים בטקסטים, בהם כביכול נעשה הדבר, אין הכתוב אלא צורת כתיבה וולגרית מקוצרת למשהו, שבעצם היה צריך להיכתב אחרת.

$$\sum_{k=1}^{n} k^2 = \sum_{j=1}^{n} j^2$$

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \lim_{k\to\infty}\left(1+\frac{1}{k}\right)^k$$

וכן הלאה.

עם זאת, יש צורך, כפי שרמזנו, בזהירות מסוימת פה. לא נוכל להחליף משתנה קשור מסוים *בכל* משתנה אחר. כך, למשל, בביטוי $\int_0^1 yxdx$ לא נוכל להחליף את המשתנה הקשור x במשתנה y. אם נעשה זאת נקבל $\int_0^1 y^2dy$, שזהו משהו שונה לחלוטין. למעלה ראינו גם, שבנוסחה $\exists i\exists j.i+j=3$ לא נוכל להחליף את j ב- i (או להיפך).

התנאים המדויקים לתקפות כלל α הם כדלקמן: ניתן להחליף את המשתנה x, הנקשר על-ידי אופרטור מסוים, במשתנה אחר y, אם מתקיימים *שני* התנאים הבאים:

(1) y אינו מופיע חופשי בטווח של האופרטור בו מדובר.
(2) בשום מקום בטווח הנ"ל אין x מופיע חופשי בטווח של אופרטור, הקושר את y.

לדוגמה, בביטוי $\int_0^1 yxdx$ לא נוכל להחליף את המשתנה הקשור x במשתנה y, כיוון שתנאי (1) יופר אז (y מופיע חופשי בטווח הקשירה של x). לעומת זאת, בנוסחה $\exists i\exists j.i+j=3$ לא נוכח להחליף את המשתנה הקשור i במשתנה j, כיוון שאז יופר תנאי (2): בטווח הקשירה של $\exists i$ (הנוסחה $\exists j.i+j=3$) יש ל- i מופע חופשי בתוך הטווח של $\exists j$, שהוא אופרטור הקושר את j.

התנאים (1) ו- (2) קשים מעט לזיכרון. מומלץ לכן להפעיל את כלל α באופן הפשוט הבא: כדאי להחליף (כשמתעורר הצורך) משתנה קשור x במשתנה y, שאינו מופיע *בכלל* בטווח הקשירה של x. דהיינו, כדאי ש- y יהיה משתנה חדש וטרי מהתנור. זה מבטיח את קיום התנאים (1) ו- (2), קל לזיכרון, ומספיק לכל הצרכים המעשיים.

הערות:

(א) כדי למנוע כל אי-הבנה, נציין שיש לנו מאגר בלתי נדלה של משתנים, שנוכל להשתמש בהם, לא רק 52 האותיות של הא"ב הלטיני. מתמטיקאים נוהגים להשתמש בעיקר באותיות לטיניות ויווניות (כמו α, β, ε, δ), בלי או עם

אינדקסים: x, x_1, y_{27}, α_{101}. לפעמים הם גם מעטרים את המשתנים בתגים וכוכבים למיניהם: z^*, x''. בשפות תכנות מקובלת בדרך-כלל כמשתנה כל מחרוזת של סימנים, המתחילה באות לטינית. למשל: המחרוזת

this_is_a_very_long_variable

היא דוגמה למשתנה ארוך.

(ב) בכלל α נוהגים להשתמש כשיש עודף של שימוש במשתנה מסוים, העלול לגרום לטעויות. כך, למשל, נוהגים לעתים קרובות בראשית לימוד האינטגרלים להיעזר בכלל α ולכתוב $\int_0^x t^2dt$ במקום $\int_0^x x^2dx$ (החלפת המשתנה *הקשור* x במשתנה t).

אין, לפי כלל α, כל הבדל בין שני הביטויים (פרט להבדל פסיכולוגי), ומאוחר יותר לא טורחים לעשות שינוי זה (עם זאת, קשה להגיד, שמתמטיקאים הינם עקביים בנקודה זאת. בעוד כתיבת ביטוי כמו $\int_0^x x^2dx$ היא דבר מקובל מאוד, לא תמצאו בשום ספר לימוד ביטוי כמו: $\sum_{i=0}^{i} i^2$, למרות שגם כאן אין שום קשר בין ה- i למעלה, שהוא חופשי, ובין שאר ה- i-ים. למעשה, אפילו בביטוי $\sum_{i=i}^{3} i^2$ אין שום קשר בין שני ה- i-ים הכתובים ב"שוויון" $i = i$ למטה. הסימן "=" כלל אינו מביע שוויון כאן, אלא רק מפריד בין ה- i משמאל, שהוא קשור, וה- i מימין, שהוא חופשי. ערך הביטוי $\sum_{i=i}^{3} i^2$, כש- $i = 1$, הוא $\sum_{i=1}^{3} i^2$ (כלומר: $1^2 + 2^2 + 3^2 = 14$).

אין כל הבדל עקרוני בין ביטוי כמו $\sum_{i=i}^{3} i^2$ ובין ביטוי כמו $\int_x^3 x^2dx$. בכל זאת, בביטוי הראשון שום מתמטיקאי אינו משתמש, בעוד בשני – כל מתמטיקאי משתמש. למה? ככה!).

בדומה, מקובל לעתים בבית-הספר התיכון (בעיקר ברמת "3 נקודות"), בתרגילים המתבססים על זהויות כמו $(a + b)(a - b) = a^2 - b^2$, להשתמש בכלל α ולעבור לשימוש בזהות $(x + y)(x - y) = x^2 - y^2$ (שזו אותה זהות, כמובן). זה קורה כשהמטלה היא לפשט ביטויים המכילים את a או b כמו $(a^2 + b^2)(a^2 - b^2)$. המטרה של המורה העושה זאת היא, שוב, מניעת בלבול (כזכור, בזהויות המשתנים הינם משתנים הקשורים על-ידי כמתי ∀ סמויים. לכן כלל α יָשים).

ברמה של 5-4 נקודות משתמשים לעתים בכלל α בעת לימוד נושא האינדוקציה. כזכור, אקסיומת האינדוקציה, כאשר A היא נוסחה שבה n מופיע חופשי, הינה:

$$[A(0/n) \wedge \forall n(A \Rightarrow A(n+1/n))] \Rightarrow \forall nA$$

באקסיומה זו נקשר המשתנה n במקומות שונים על-ידי אופרטורי $\forall$ שונים. כדי למנוע בלבול, ספרי לימוד רבים משתמשים לכן בכלל α ומנסחים (למעשה) את האקסיומה כך: $[A(0/n) \wedge \forall k(A(k/n) \Rightarrow A(k+1/n))] \Rightarrow \forall nA$. זה מתבטא בכך, שבשלב "המעבר" מניחים נכונות ל- "$n = k$" ומוכיחים ל- "$n = k + 1$". מובן, שזה אפשרי רק כאשר התנאים לתקפות כלל α מתקיימים. אם, למשל, k מופיע כפרמטר ב- A, צריך (לפי שיטה זו) להשתמש במשתנה אחר.

(ג) תקפות כלל α יכולה לשמש כמבחן מועיל לזיהוי משתנים קשורים. "זהויות" הן דוגמה טובה לכך.

כללי הכמתים

הכללים בטבלה א.4 (עמ' 26), העוסקים בכמתים (כללים 13-16), הם המקום העיקרי בו האבחנה בין משתנים קשורים וחופשיים היא קריטית. אם נתבונן בכללים אלו נראה, שליד המספרים (13, 14 וכו') מופיעות כוכביות. כוכביות אלו באות לציין, שיש הגבלות על השימוש בכללים אלה. ההגבלה החשובה ביותר היא זו על כלל 14. כלל זה הוא פשוט ביותר, ואנו משתמשים בו בלי הרף מאז הבנו את המלה "כל". הכלל קובע, כזכור, שאם לכל העצמים בתחום מסוים יש תכונה מסוימת, אז אם ניקח עצם מסוים בתחום, מובטח לנו, שתהיה לו תכונה מסוימת זו. פורמלית הכלל אומר, שאם הסקנו את $\forall xA$ בצורה כלשהי מקבוצת הנחות **T**, אז נוכל להסיק מאותה קבוצה (כמעט) כל נוסחה מהצורה $A(t/x)$, דהיינו (כמעט) כל נוסחה המתקבלת מהצבת הביטוי t במקום המופעים החופשיים של x ב- A (ב- $\forall xA$ אין, כמובן, ל- x מופעים חופשיים, אך ב- A עצמה קרוב לוודאי שיש לו). למשל, אם הגננת מודיעה, ש"כל ילד חייב לשטוף ידיים לפני הארוחה" (המלה "ילד" מתפקדת כמשתנה במשפט זה!), אזי מסיק אורי הקטן, ש"אורי חייב לשטוף ידיים לפני הארוחה", למרות שהגננת לא אמרה זאת במפורש. אורי נעזר, בלי דעת, בכלל 14, כאשר x, במקרה זה, הוא "ילד" ואילו t – השם "אורי".

דוגמאות נוספות: מהנוסחה $\forall x\exists y.\ x < y$ ניתן להסיק ש- $\exists y.\ 3 < y$ (הצבת $t = 3$ במקום x). וכן גם $\exists y.\ (x + z) + 1 < y$ (הצבת $t = (x + z) + 1$ במקום x).

בדומה, מהזהות הידועה $1 + 2 + \ldots + n = \frac{n(n + 1)}{2}$ נובעת, לפי כלל 14, הזהות $1 + 2 + \ldots + (m^2 - 1) = \frac{(m^2 - 1)m^2}{2}$, על-ידי ההצבה $t = m^2 - 1$ במקום n (n, כזכור, קשור כאן על-ידי כמת $\forall n$ סמוי הנמצא בהתחלה. לכן כלל 14 ישים. העובדה, שמה שמתקבל הוא זהות גם כן, דהיינו: אפשר להוסיף לו $\forall m$ בהתחלה, נובעת מכלל 13).

יש אבל, כמו שרמזנו, הגבלה חשובה ביותר על הביטוי t, שאפשר להציב במקום x בכלל 14, ***ויש לזכור אותה תמיד כדי להימנע משטויות***: ניסוחה הטכני הוא, ש- t חייב להיות ***חופשי להצבה*** ב- A במקום x. פירוש הדבר הוא, ששום מופע חופשי ב- t של משתנה אסור לו שיהפוך לקשור בעקבות ההצבה!!

דוגמאות:

1. מ- $\forall x\exists y.\ x < y$ (פסוק נכון במספרים הממשיים) אי אפשר, למרבה המזל, להסיק ש- $\exists y.\ y + 1 < y$ (מה שהיינו מקבלים מכלל 14, לו $t = y + 1$ היה חופשי להצבה במקום x בנוסחה $\exists y.\ x < y$). הסיבה היא, שהביטוי $y + 1$ מכיל את המשתנה y, שהינו חופשי בו, אך נקשר (על-ידי $\exists y$) בעקבות ההצבה.

2. מהזהות $\int_0^x ax^2dx = \frac{ax^3}{3}$ (דהיינו $\forall a\forall x\left[\int_0^x ax^2dx = \frac{ax^3}{3}\right]$) ***לא*** נובע (על-ידי הצבת $t = x$ במקום a), ש- $\forall x\left[\int_0^x x \cdot x^2dx = \frac{x \cdot x^3}{3}\right]$. האמת, כמובן, הינה, ש- $\forall x\left[\int_0^x x^3dx = \frac{x^4}{4}\right]$!

3. אחת הזהויות החשובות ביותר הקשורות בנגזרות הינה:

$$(x^a)' = ax^{a-1}$$

כמעט כל תלמיד תוהה, בשלב כלשהו, מדוע לא נוכל להציב בנוסחה זו x במקום a ולקבל ש- $(x^x)' = x \cdot x^{x-1}$ (דהיינו: $(x^x)' = x^x$). הרי הזהות הזו נכונה ***לכל*** a, לא? היא נכונה עבור $a = 7$ ו- $a = -12$ ו- $a = \sqrt{2}$, לא? אז למה לא עבור $a = x$?? התשובה המעורפלת, שהתלמיד מקבל בדרך-כלל, היא, ש- a כאן הוא "קבוע".

הכיצד קבוע הוא, אם מותר להציב במקומו ערכים שונים, וזוהי אפילו עיקר השימושיות של הנוסחה? ובמה קבוע הוא כאן יותר מ- x ??

הבה ננתח מה ***באמת*** קורה כאן. ראשית, ה"זהות" הנ"ל הינה בעצם הטענה:

$$\forall a \forall x[(x^a)' = ax^{a-1}] \qquad (*)$$

כאשר $(x^a)'$ היא צורה מקוצרת לכתיבת $\lim\limits_{h\to 0}\dfrac{(x+h)^a - x^a}{h}$.

עתה, ההצבה $a = 7$, למשל, הנותנת $\forall x[(x^7)' = 7x^6]$ (ובקיצור יותר: הזהות $(x^7)' = 7x^6$), אינה אלא הפעלת כלל 14 על (*) עם $t = 7$ (ו- a במקום ה- x, שמופיע בניסוח הכלל). כלל 14 מאפשר אף להציב y ו- $y-1$ במקום a כאן (ומשתמשים בזה ב"חדו"א II"). מה שכלל 14 ***אינו*** מאפשר כאן הוא הצבת x (או כל ביטוי אחר בו x הוא חופשי, כמו $x-1$ או $\int\limits_0^x x^2dx$) במקום a, כי ביטויים המכילים מופעים חופשיים של x אינם חופשיים להצבה במקום a בנוסחה $\forall x[(x^a)' = ax^{a-1}]$! זהו כל הסיפור כולו, ואין בו שום שוני ממה שקורה בשתי הדוגמאות הקודמות.

הערה:
פירשנו כאן את הזהות $(x^a)' = ax^{a-1}$ כמציינת שוויון בין שני ***מספרים***. ערך שני מספרים אלו תלוי בערך המשתנים x ו- a. בפירוש זה אין כל הבדל ***לוגי*** בין תפקידי המשתנים x ו- a, פרט להסכם הג'נטלמני ש- $(x^a)'$ הוא קיצור של $\lim\limits_{h\to 0}\dfrac{(x+h)^a - x^a}{h}$ דווקא, ולא של $\lim\limits_{h\to 0}\dfrac{x^{a+h} - x^a}{h}$. דבר זה יתברר אם נכתוב את הזהות למעלה בצורתה המלאה:

$$\lim_{h\to 0}\frac{(x+h)^a - x^a}{h} = ax^{a-1}$$

ממש כשם שבזהות זו ניתן להציב במקום a כל ביטוי, שאינו מכיל מופע חופשי של x, ולקבל כך זהות חדשה, כך ניתן להציב בו במקום x כל ביטוי, שאינו מכיל מופע חופשי של a, ולקבל כך זהות חדשה! קיימת אבל אינטרפרטציה אחרת של הזהות $(x^a)' = ax^{a-1}$, לפיה זהות זו מביעה לא שוויון של מספרים, אלא שוויון של ***פונקציות*** (יש, כמובן, קשר הדוק בין שתי האינטרפרטציות). באינטרפרטציה זו נטפל בפרק ב.4, ובה יש אכן הבדל מסוים בתפקוד של x ושל a (אם כי עדיין שניהם יהיו קשורים). על כל פנים, בשתי האינטרפרטציות הצבת $a = 7$ או

$a = y + 1$ אפשרית לפי כלל 14, ואי האפשרות להציב x במקום a נובעת מה*הגבלה* על כלל זה.

4. דוגמה נוספת למצב נפוץ, בו נעשית קשירה של משתנים, ללא שהדבר נראה ישירות, היא בהגדרת פונקציות. כשכותבים משהו כמו: "נגדיר פונקציה f על-ידי הכלל $f(x) = x^2$", המשתנה x המופיע שם הינו קשור (תקפות כלל α היא מבחן מצוין לכך. אם נכתוב: "נגדיר פונקציה f על-ידי הכלל $f(z) = z^2$", הרי f ה"חדשה" תהיה בדיוק ה- f הקודמת!). למעשה, גם כאן יש להתייחס לכָתוב כאילו המשתנה x נקשר על-ידי $\forall$. במלים אחרות, מה שכתוב הוא בעצם:

$$\forall x.\, f(x) = x^2$$

מזה נובע, לפי כלל 14, ש- $f(2) = 2^2$ ו- $f(a-1) = (a-1)^2$.
מובן, לאור זאת, שההגבלה על השימוש בכלל 14 תחול גם על המקרה של הגדרת פונקציות. כך, אם נגדיר פונקציה g על-ידי:

$$g(x) = \int_0^x xy^2dy \qquad (= \frac{x^4}{3})$$

אז
$$g(2) = \int_0^2 2y^2dy = \frac{16}{3}$$

ו-
$$g(x+z) = \int_0^{x+z} (x+z)\cdot y^2dy = \frac{(x+z)^4}{3}$$

בניגוד לכך, ***לא*** נוכל להציב בהגדרת g את הביטוי y (או $y - 1$) במקום x, כי ביטויים אלו אינם חופשיים להצבה: יהיו מופעים של y (לא כולם!) שיקָשרו בעקבות ההצבה! אם נעשה הצבה אסורה זאת "נקבל":

$$g(y) = \int_0^y y\cdot y^2dy = \frac{y^4}{4}$$

בעוד שלאמיתו של דבר, $g(y) = \frac{y^4}{3}$! (כדי למצוא זאת מ*ההגדרה* של g עלינו להחליף תחילה, לפי כלל α, את המשתנה הקשור y שבהגדרת g במשתנה אחר, כמו z, ***ואז*** להציב y).

לסיכום, כמעט כל השגיאות הנעשות בעת הצבות (והן מרובות מאוד) קשורות בהפרת ההגבלה על כלל 14. חיוני אפוא, לפני שמבצעים הצבה, לבדוק היטב, מי הם המשתנים הקשורים ועל-ידי מה הם קשורים (כולל קשירה על-ידי אופרטורי קשירה סמויים – במיוחד $\forall$), ומי והיכן הם המשתנים החופשיים. בדיקה זו תאפשר לנו לוודא, שאכן הביטויים שאנו מציבים הינם אכן חופשיים להצבה במקומות, בהם אנו מציבים אותם.

%%

נפרט עתה, לצורך השלמות, את ההגבלות על שאר הכללים של הכמתים:

כלל 13: הכלל תקף בתנאי של- y אין מופע חופשי באף אחת מההנחות בקבוצת ההנחות, מהן הוסקה A (דהיינו: **T**). אם y איננו x עצמו, אז אסור גם שיהיה לו מופע חופשי ב- A.

כלל 15: אותה הגבלה כמו בכלל 14: t חייב להיות חופשי להצבה במקום x ב- A.

כלל 16: הכלל תקף בתנאי, שלמשתנה y אין שום מופע חופשי ב- **T** או ב- B. אם y איננו x עצמו, אז אסור גם שיהיה לו מופע חופשי ב- A.

%%

ב. יסודות תורת הקבוצות

ב.1 מושגי יסוד

מושג הקבוצה הינו אחד המושגים היסודיים של המתמטיקה. למעשה זהו *המושג היסודי* של המתמטיקה המודרנית. בהתאם, הלשון של תורת הקבוצות היא הלשון בה מוצגים כיום ענפי המתמטיקה השונים, ומשתמשים בה רבות גם במדעי המחשב. לימוד לשון זו, יחד עם לימוד התוצאות היסודיות של תורת הקבוצות, הינם המטרות של חלק זה.

מהי, אם כן, קבוצה? ובכן, קבוצה (set) הינה אוסף של עצמים, המהווה עצם בעצמו. לעצמים, מהם מורכבת קבוצה, קוראים *איברי* הקבוצה, ועל כל אחד מהם אומרים, שהוא *שייך* לקבוצה, או שהקבוצה *כוללת* אותו.

את מה שנכתב בפסקה הקודמת אין לראות בגדר *הגדרה* של מושג הקבוצה. מסופקתני, אם המושג "אוסף", המופיע ב"הגדרה" זו, הינו ברור יותר מאשר מושג הקבוצה עצמו. בסופו של דבר, המושגים של "קבוצה", "איבר", ו"שייכות" הינם מושגים יסודיים של תורת הקבוצות, ממש כמו ש"נקודה", "ישר" ו"המצאות נקודה על ישר" הם מושגים יסודיים בגיאומטריה[1] . יש לראות אפוא בפסקה האחרונה הסבר מסייע בלבד. אשר לתוספת "המהווה עצם בעצמו" המופיעה שם - כרגע היא נראית סתומה, מן הסתם. משמעותה והצורך בה יובהרו בהמשך.

הנקודות הבאות חשובות להבנת מושג הקבוצה:

(א) אין הגבלה מראש על מה שיכול לשמש כאיבר בקבוצה. כל דבר, אותו רואים אנו כעצם, כשיר לצורך כך. כך מורכבת קבוצת חברי כנסת מבני-אדם, בעוד הקטע [0,1] (קבוצת המספרים בין 0 ל- 1, כולל 0 ו- 1) הינו קבוצה של מספרים. השאלה, מה יכול לשמש איבר בקבוצה, זהה למעשה לשאלה, מה מוכנים אנו לראות כעצם לגיטימי. התשובה עשויה להיות תלויית קונטקסט. כשעוסקים, למשל, במיתולוגיות, הרי אפילו אלי האולימפוס יהיו (ביחד) קבוצה. בקורס זה, באופן טבעי, נתעניין בעיקר בעצמים מתמטיים. העקרונות שנלמד יהיו נכונים אבל באופן כללי.

[1] אוקלידס, בספרו המקורי, "הגדיר" כל מושג, כולל המושגים היסודיים, אולם בנוגע למושגים היסודיים, כמו "נקודה", אין לראות ב"הגדרות" אלו יותר מהסבר, המסייע לקורא להבין במה מדובר.

(ב) קבוצות אינן חייבות להיות הומוגניות כמו בדוגמאות, שהבאנו בהערה הקודמת. כך הקבוצה, המורכבת מחברי כנסת יחד עם המספרים הטבעיים בין 91 למאה, היא קבוצה לכל דבר, ויש בה 130 איברים.

(ג) קבוצות יכולות להיות סופיות (כמו בדוגמא של הכנסת) או אין-סופיות (כמו בדוגמא של הקטע $[0,1]$).

(ד) כמו שהכרזנו ב"הגדרת" מושג הקבוצה, קבוצות נחשבות בעצמן כעצמים לגיטימיים. אי לכך כל קבוצה יכולה לשמש כאיבר של קבוצה אחרת. דוגמא מהחיים: כיתה בבית הספר הינה קבוצה של תלמידים. ממילא קבוצת הכיתות של אותו בית ספר הינה קבוצה, שכל איבריה הם קבוצות בעצמם.

(ה) אנו רואים קבוצה כנקבעת לחלוטין על-ידי איבריה ולא על-ידי שום דבר אחר (כמו למשל, הסדר, שבו הם מוצגים). כך, אם נסדר יום אחד את התלמידים בכיתה מסוימת לפי סדר אלף-ביתי של שמם, וביום אחר לפי מספרי הזהות שלהם, הרי שבשני המקרים מדובר עדיין באותה קבוצה (בהנחה ששום תלמיד לא נגרע בינתיים, ושום תלמיד חדש לא התווסף). הניסוח המדויק של רעיון זה מתבטא בעיקרון הבא:

עקרון האקסטנציונליות:

שתי קבוצות הן שוות, אם ורק אם יש להן בדיוק אותם איברים.

כיוון שמושג ההשתייכות של עצם לקבוצה הינו אחד המושגים הבסיסיים של תורת הקבוצות, נכניס לו סימון מיוחד $\in$ (או ε)[2]. מכאן ואילך, כשנרצה להגיד שהעצם t הינו איבר של הקבוצה A נכתוב בקיצור $t \in A$ (לדוגמא: $\frac{1}{2} \in [0, 1]$). במקום $\neg(t \in A)$ נכתוב בדרך כלל $t \notin A$ (לדוגמה $2 \notin [0, 1]$). קיצורים מקובלים אחרים, דומים לאלו הקשורים ב- $\leq$, הינם:

לכתוב $\forall x \in A\ (...)$ – במקום $\forall x(x \in A \Rightarrow \ldots)$

לכתוב $\exists x \in A\ (...)$ – במקום $\exists x(x \in A \wedge \ldots)$

בעזרת הסימן החדש נוכל עכשיו לבטא את עיקרון האקסטנציונליות כך:

$$\textbf{(Ext1):} \quad \forall A\ \forall B\ (A = B \Leftrightarrow (\forall x.\ x \in A \Leftrightarrow x \in B))$$

---

[2] $\in$ ו- ε הם שתי צורות מקובלות של האות היוונית "אפסילון", ובשתיהן ניתן להיתקל בספרות.

אם נשתמש בשקילויות לוגיות (הגדרת $\Leftrightarrow$ ושקילות (15a) מטבלה א.3), נקבל שזה שקול ל:

$$\textbf{(Ext2)}: \quad \forall A \, \forall B \, (A = B \Leftrightarrow [(\forall x.\, x \in A \Rightarrow x \in B) \wedge (\forall x.\, x \in B \Rightarrow x \in A)])$$

הניסוח האחרון (במיוחד שני הקוניוקטים בתוך הסוגריים המרובעים) מביא אותנו באופן טבעי להכנסת מושג בסיסי חדש של תורת הקבוצות:

הגדרה:

נאמר שקבוצה A היא ***תת-קבוצה*** של קבוצה B (או ש- A ***חלקית*** ל- B, או ש- B ***מכילה*** את A) אם כל איבר של A הינו איבר של B. A נקראת ***חלקית ממש*** ל- B אם A חלקית ל- B אך לא שווה לה.

סימונים:

$A \subseteq B$ פירושו ש- A חלקית ל- B

$A \subset B$ פירושו ש- A חלקית ממש ל- B

בשפה הפורמלית, לכן, הגדרת $\subseteq$ ו- $\subset$ הינה:

$$A \subseteq B =_{Df} \forall x\, (x \in A \Rightarrow x \in B)$$
$$A \subset B =_{Df} A \subseteq B \wedge A \neq B$$

(הסימון $=_{Df}$ פירושו, כזכור, שאגף ימין מהווה הגדרה של אגף שמאל).

אם נתבונן עתה ב- (Ext2), הניסוח השני של עקרון האקסטנציונליות, נראה שבעזרת הסימונים החדשים ניתן לנסחו כך:

$$\forall A \, \forall B \, (A = B \Leftrightarrow (A \subseteq B \wedge B \subseteq A))$$

במלים: שתי קבוצות הינן שוות, אם כל אחת מהן הינה חלקית לשנייה. מזה נגזרת הצורה הנפוצה ביותר להוכחת שוויון של שתי קבוצות A ו- B. ההוכחה בצורה זו מתחלקת לשני חלקים: מראים לחוד ש- A חלקית ל- B ולחוד ש- B חלקית ל- A (דוגמאות נראה בפרק הבא).

הערות על סימונים:

1. כמקובל במקרים אחרים, נכתוב לעתים קרובות $A \not\subset B$, $A \not\subseteq B$ במקום $\neg(A \subseteq B)$, $\neg(A \subset B)$, בהתאמה.

2. לרוע המזל, השימוש בסימנים $\subseteq$ ו- $\subset$ ובטרמינולוגיה למעלה אינו אחיד בספרות. יש האומרים "חלקי או שווה" היכן שאנו אומרים "חלקי" , ו"חלקי" במקום בו אנו אומרים "חלקי ממש". בדומה, יש כאלה המשתמשים בסימן $\subset$ היכן שאנו משתמשים ב $\subseteq$, ו- $\subsetneq$ במקום שאנו כותבים $\subset$.

3. ממש כמו לגבי $\leq$ ו- $\in$ גם לגבי $\subseteq$ ו- $\subset$ מקובלים הקיצורים

$$\forall x \subseteq A(...) =_{Df} \forall x(x \subseteq A \Rightarrow ...)$$
$$\exists x \subseteq A(...) =_{Df} \exists x(x \subseteq A \wedge ...)$$

וכנ"ל לגבי $\subset$.

דוגמאות:

(1) קבוצת המספרים הזוגיים חלקית לקבוצת המספרים הטבעיים. יתר על כן: היא חלקית ממש לקבוצה זו.

(2) קבוצת הנקודות בפנים מעגל מסוים חלקית (ממש) לקבוצת הנקודות במישור.

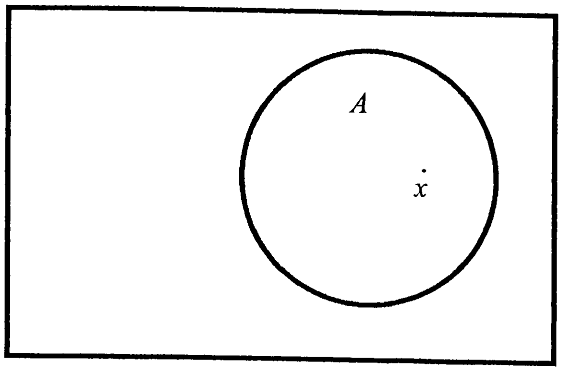
(3) כל קבוצה הינה חלקית לעצמה. ואכן, לפי הגדרה $A \subseteq A$ אם ורק אם $\forall x\ (x \in A \Rightarrow x \in A)$, וזוהי אמת לוגית. לעומת זאת $A \not\subset A$.

אזהרה:

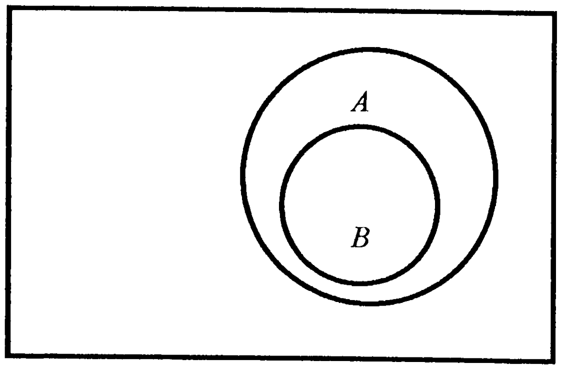
יש להיזהר מאד מלבלבל בין מושג השייכות ($\in$) ובין מושג החלקיות ($\subseteq$) . קבוצת המספרים הזוגיים היא , כאמור , חלקית לקבוצת המספרים הטבעיים, אך היא אינה שייכת לה, כיוון שהיא עצמה איננה מספר טבעי. אזהרה זו הינה חשובה, בגלל שהשפות הטבעיות מטשטשות לעתים קרובות את ההבדל בין שני היחסים. לדוגמה: כשאנו אומרים ש"היהודים הם בני אדם", כוונתנו לכך שקבוצת היהודים הינה ***חלקית*** לקבוצת בני האדם. לעומת זאת באומרנו כי "היהודים הם עם ככל העמים" כוונתנו היא, שקבוצת היהודים ***שייכת*** למשפחת העמים (שהיא קבוצה, שאבריה הם בעצמם קבוצות). מכאן, שמשמעות מלה כמו "הם" יכולה לפעמים להיות "$\in$" ולפעמים "$\subseteq$" , בהתאם לקונטקסט!

כדי *להמחיש* את המושגים היסודיים של תורת הקבוצות משתמשים לעתים באמצעי הידוע *כדיאגרמת וון*. בדיאגרמות אלו מיוצג עולם העצמים, שבו יש לנו עניין, על-ידי מלבן במישור. הקבוצות השונות של עולם זה מיוצגות על ידי עיגולים או אליפסות, הפנימיים למלבן זה (ולעתים על ידי חלקים שלהם, או קבוצה סופית של חלקים

כאלה). העצמים בעולם זה מיוצגים על-ידי הנקודות בפנים המלבן. כך למשל מומחשות העובדות ש- $x \in A$ וש- $B \subseteq A$ באופן הבא:



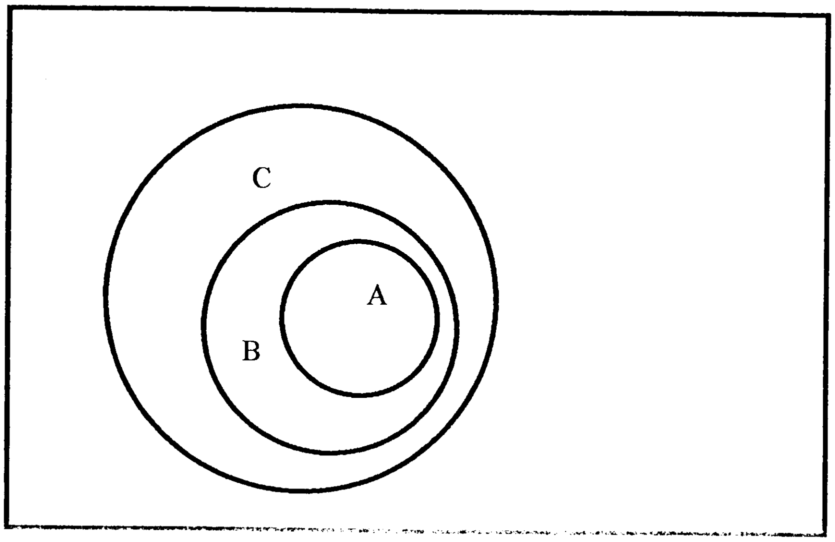
$x \in A$



$B \subseteq A$

דוגמה ***לאינטואיציה***, שהשימוש בדיאגרמות וון יכול לספק נותנת לנו הדיאגרמה הבאה:

ציור ב.1: $A \subseteq B \wedge B \subseteq C$



הדיאגרמה ממחישה מצב שבו A, B ו- C הינן קבוצות כך ש- $A \subseteq B$ וגם $B \subseteq C$. התבוננות בה מגלה (אם נמחק בדמיוננו את קו השפה של B), שבמקרה זה $A \subseteq C$. זהו אכן עקרון פשוט וחשוב, הידוע בשם:

טרנזיטיביות יחס ההכלה:

אם $A \subseteq B$ ו- $B \subseteq C$ אז $A \subseteq C$
(או באופן פורמלי: $\forall A \forall B \forall C (A \subseteq B \wedge B \subseteq C \Rightarrow A \subseteq C)$).

חשוב להבין, שלמרות שהשימוש בדיאגרמה הוביל אותנו לאינטואיציה בדבר עיקרון זה ואף נותן אינטואיציה למה הוא נכון, ***אין*** הוא מהווה ***הוכחה***, כי הוא מטפל בסוג מיוחד מאוד של קבוצות. ניתן, אפוא, הוכחה אמיתית:

הוכחת טרנזיטיביות ההכלה:

עלינו להראות שלכל x, אם $x \in A$ אז $x \in C$ יהי לכן $x \in A$ כיוון ש- $A \subseteq B$ נובע מזה ש- $x \in B$ מזה ומהעובדה ש- $B \subseteq C$ נובע ש- $x \in C$ מש"ל.

%%

למען אלו שלא קראו את פרק א.4, ולמען אלה שרוצים לקראו שנית, אך עם דוגמה אחרת, נביא עתה ניתוח לוגי מפורט של ההוכחה שהבאנו עתה, ומה בדיוק נעשה בה[3] (בעתיד לא נעשה עוד ניתוחים כאלה!).

ובכן, הטענה שאנו רוצים להוכיח היא (בסגנון הגיאומטריה בתיכון):

(I) נתון: (1) $A \subseteq B$
(2) $B \subseteq C$
צ"ל: $A \subseteq C$

הצעד הראשון בהוכחה שהבאנו הוא צעד, שנעשה באופן ***לא-מפורש***. כותב ההוכחה מניח במקרים כאלו, שגם הקורא עושה צעד זה לעצמו באופן אוטומטי. הצעד הינו החלפת הסימנים והמושגים, המופיעים בנתונים ובמה שצריך להוכיח, במשמעותם המקורית. במקרה שלפנינו משמעות $A \subseteq B$, כפי שניתנה בהגדרה למעלה, היא, כזכור: $\forall x\, (x \in A \Rightarrow x \in B)$. מכאן, שבעצם הבעיה היא הבעיה הבאה:

(II) נתון: (1) $\forall x\, (x \in A \Rightarrow x \in B)$
(2) $\forall x\, (x \in B \Rightarrow x \in C)$
צ"ל: $\forall x\, (x \in A \Rightarrow x \in C)$

ההוכחה למעלה מתחילה למעשה בנקודה זו. הבה ננתחה משפט אחר משפט.

---

[3] הסברים *למה* הכללים הלוגיים אכן תקפים ניתן למצוא בפרק א.4. בדוגמא כאן נתרכז בהבנת צורת השימוש בהם.

"עלינו להראות שלכל x, אם $x \in A$ אז $x \in C$"
לכאורה אין כאן אלא ניסוח ב"עברית מתמטית" של מה שצריך להוכיח. למעשה, יש כאן, כמו שנראה מייד, גם רמז, שאנו עומדים להשתמש בכלל 13 (וגם 1) מטבלה א.4, עמוד 26.

"יהי לכן $x \in A$"
משפט קצר ותמים-למראה זה טומן בחובו הפעלה של שני כללים! ראשית, מה שאנו רוצים להוכיח הוא משפט אוניברסלי מהצורה $\forall x\varphi$ (כש- φ היא כאן הנוסחה $x \in A \Rightarrow x \in C$). משפט כזה מוכיחים בדרך כלל בעזרת כלל 13 (כל המספרים – לפי טבלה א.4). אם נקרא כלל זה ***מלמטה למעלה***, הוא אומר, שכדי להוכיח משהו מהצורה $\forall x\varphi$, עלינו להוכיח פשוט את φ – ובלבד שהמשתנה x בו מדובר אינו מופיע חופשי בהנחות. במקרה שלנו המשתנה x אומנם ***מופיע*** בהנחות (1) ו- (2), אך ***אינו*** מופיע בהם ***חופשי***. לכן הפעלת כלל 13 הינה אפשרית[4]. מספיק לכן להוכיח את הנוסחה $x \in A \Rightarrow x \in C$ מהנתונים (1) ו- (2). נוסחה זו היא בעלת צורה של גרירה, וגרירה מוכיחים בדרך כלל בעזרת כלל 1. כלל זה קובע (אם נקרא גם אותו מלמטה למעלה), שכדי להוכיח נוסחה מהצורה $\varphi \Rightarrow \psi$ עלינו להוסיף את φ לרשימת הנתונים, ולנסות להוכיח את ψ מרשימת הנתונים המורחבת. במקרה שלנו צריכים אנו לכן להוסיף את $x \in A$ לרשימת הנתונים ולנסות להוכיח את $x \in C$ מרשימת הנתונים החדשה.

נסכם אפוא: כדי להוכיח את (II) למעלה, די אם נוכיח את הטענה הבאה:

(III)	נתון:	(1)	$\forall x\,(x \in A \Rightarrow x \in B)$
		(2)	$\forall x\,(x \in B \Rightarrow x \in C)$
		(3)	$x \in A$
	צ"ל:		$x \in C$

אם נצליח, אז בעזרת כלל (1) נוכל להסיק, ש- $x \in A \Rightarrow x \in C$ נובע מהנתונים (1) ו-(2) (בלבד!), ואז לפי כלל (13), ש- $\forall x(x \in A \Rightarrow x \in C)$ נובע גם הוא מנתונים אלו. כך נקבל הוכחה של (II).

---

[4] לו היה x מופיע חופשי באחת ההנחות, היה עלינו להשתמש תחילה בכלל (α) ולהחליף את $\forall xA$ ב- $\forall y(A(y/x))$, כש-y משתנה חדש בתכלית – ומכאן להתקדם באותה צורה.

כדאי לשים לב לשני הדברים הבאים:

(א) הכללים (1) ו- (13) הם כללים, שבאופן מעשי משתמשים בהם כדי לעשות *רדוקציה* של הבעיה לבעיה פשוטה יותר. במקום להוכיח טענה מסוימת אנו מוכיחים משהו אחר, שהוא *מספיק*, כי אם נצליח, אז כללים (1) ו- (13) יסיימו את המלאכה. לכן, למרות שבהוכחה מלאה, פורמלית, כללים אלו מופעלים בדרך כלל *בסוף*, *ההתיחסות* אליהם בהוכחות לא פורמליות נעשית (ברמז בלבד!) *בהתחלה*! רמז לכך במקרה שלנו ניתן במלים "עלינו להראות", בהן נפתח המשפט הראשון (בעברית) בהוכחה. יש בהן רמז, שאנו מתחילים בהוכחת משהו, *שמספיק* לצורך הוכחת מה שאנו רוצים להוכיח באמת (ולכן, בדרך כלל, זהו רמז לשימוש בכללים (13) ו- (1)).

(ב) ביטויים כמו *"יהי* $x \in A$", "יהי $\varepsilon > 0$" וכדומה מרמזים *תמיד* על שימוש בכלל 13, וכמעט תמיד – בצירוף עם כלל (1).

נמשיך עתה בניתוח הלוגי של ההוכחה.

"כיוון ש- $A \subseteq B$, נובע מזה ש- $x \in B$"
גם משפט זה טומן בחובו שימוש בשני כללים. תחילה אנו מסיקים, בעזרת כלל 14, מהנתון (1) ב- III (שאינו אלא $A \subseteq B$, כזכור) ש- $x \in A \Rightarrow x \in B$. מנוסחה אחרונה זו והנתון החדש (3) ($x \in A$) נובע $x \in B$ בעזרת כלל (2).

"מזה ומהעובדה ש- $B \subseteq C$ נובע $x \in C$"
הסיפור כאן דומה. מ"זה" (כלומר: מ- $x \in B$) ומנתון (2) ($B \subseteq C$) נובע, בעזרת כלל 14 ובעקבות הפעלה של כלל (2), $x \in C$.

"מש"ל"
ה- "מש"ל" אומר, בעצם, שסיימנו הוכחת III. מכל מה שהסברנו פירוש הדבר, שאנו יכולים להוכיח גם את II וממילא את I. אין התייחסות מפורשת לכך, ולעולם גם לא תהיה!

הערה
המעמיקים ישימו לב כאן, שלמעשה, אפילו כדי לקבל את הצורה (I) מהניסוח הפורמלי של המשפט מספר שורות קודם לכן, הופעלו עוד כללים לוגיים רבים: תחילה אנו אומרים, שבמקום להוכיח

$$\forall A \forall B \forall C (A \subseteq B \wedge B \subseteq C \Rightarrow A \subseteq C)$$

די, בגלל כלל 13, להוכיח את $A \subseteq B \wedge B \subseteq C \Rightarrow A \subseteq C$. בשביל זה די, לאור כלל (1), להניח את $A \subseteq B \wedge B \subseteq C$ ולהוכיח את $A \subseteq C$. את הנתון $A \subseteq B \wedge B \subseteq C$ אנו מפרקים מיד, לפי כלל (4), לשני נתונים: $A \subseteq B$ ו- $B \subseteq C$. רק אז מקבלים אנו, שמספיק להוכיח את (I) !

%%

לסיום הפרק, שתי הערות עדינות-משהו:

(1) קוראה חדת-עין הבחינה אולי בבעייתיות מסוימת בצורה, בה ניסחנו את עקרון האקסטנציונליות למעלה (ב- **Ext1**, למשל). כאשר כתוב שם ...$\forall x$ הכוונה היא אכן *לכל* עצם אפשרי. לעומת זאת כאשר כתוב , באותה שורה, ...$\forall A \forall B$ הכוונה היא רק לכל *קבוצה* A ולכל *קבוצה* B. האמת היא, שבתורת הקבוצות המתמטית הטהורה אין הבדל, כי שם כל עצם מתמטי הינו קבוצה (כן, אפילו המספרים השונים מוגדרים שם כקבוצות מסוימות!) . עם זאת, בקורס הנוכחי אנו מניחים את אפשרות קיומם של עצמים, שאינם קבוצות. למעלה מזאת: אנו אכן נתייחס למספרים, נקודות במישור ודברים אחרים כאל עצמים "אטומיים", שאינם קבוצות. כדי לנסח לכן את **Ext1** ומשפטים אחרים בצורה מדויקת, היה עלינו להכניס יחס (או "פרדיקט") מיוחד, כמו $\text{set}(A)$, שמובנו: "A הינה קבוצה". היה עלינו אז לנסח את **Ext1** כך:

$$\forall A \forall B\ (\text{set}(A) \wedge \text{set}(B) \Rightarrow [\text{A} = B \Leftrightarrow (\forall x.\ x \in A \Leftrightarrow x \in B)])$$

כדי למנוע סרבול זה, לא נשתמש בפרדיקט $\text{set}(X)$. במקום זאת נשתמש בשיטה של משתנים מיוחדים עבור קבוצות. מכאן ואילך (ולמעשה כבר לכל אורך פרק זה) אותיות לטיניות גדולות A, B, C ...X, Y, Z ישמשו כמשתנים עבור קבוצות בלבד. רשמית לכן:

כאשר נכתוב $\forall A(...)$ הכוונה היא ל: $\forall A(\text{set}(A) \Rightarrow ...)$

כאשר נכתוב $\exists A(...)$ הכוונה היא ל: $\exists A(\text{set}(A) \wedge ...)$

(2) כמו כן נלך גם אנו לעתים קרובות בעקבות הנוהג המקובל ונשמיט כמתים אוניברסליים בהתחלת טענה. משהו כמו **Ext1** ייכתב לכן, בדרך כלל, פשוט כך:

$$A = B \Leftrightarrow \forall x.\ x \in A \Leftrightarrow x \in B$$

ב.2 : הגדרת קבוצות וסימונן

בפרק הקודם, עת רצינו להתייחס לקבוצה מסוימת, השתמשנו בלשון הבאה: "קבוצת כל המספרים הטבעיים הזוגיים", "קבוצת המספרים הממשיים בין 0 ל- 1" וכו'. קבוצות הוגדרו על-ידי *תכונה* מסוימת, המשותפת לכל איבריהן, ולאיבריהן בלבד יש תכונה זו. כיוון שזו הצורה הנפוצה ביותר להגדרת קבוצות, הוכנס עבורה סימן מיוחד. במקום "קבוצת כל העצמים x, שיש להם תכונה ψ", כותבים : "$\{x\,|\psi\}$" . בביטוי זה x חייב להיות משתנה, ו- ψ היא נוסחה, ש- x הוא אחד ממשתניה החופשיים (לא בהכרח היחיד). למשל:

(א) $\{n \mid (\text{הינו טבעי } n) \wedge \exists k. (\text{הינו טבעי } k) \wedge (n = 2\cdot k)\}$ הינה קבוצת כל המספרים הטבעיים הזוגיים.

(ב) $\{x \mid (\text{הינו ממשי } x) \wedge 0 \le x \le 1\}$ הינה קבוצת המספרים הממשיים בין 0 ל- 1.

את הנוסחאות בעברית "n הינו טבעי", "x הינו ממשי" נחליף בהמשך בסימונים קצרים יותר. איתם או בלעדיהם מקובל לקצר ביטויים כמו השניים האחרונים על ידי שימוש במשתנים מסוג מיוחד. כך מקובל מאוד, שמשתנים המתחילים באותיות i,j,k,l,m,n הינם משתנים עבור המספרים הטבעיים. לכן נכתוב לעתים קרובות $\{n \mid \exists k.\ n = 2k\}$ במקום הביטוי הארוך יותר למעלה (עוד נשוב לנקודה זו).

האופרטור $\{\ldots \mid \_\_\}$ הוא אופרטור קשירה. את המשתנה הנקשר כותבים משמאל לקו ההפרדה " | " שבסוגריים המסולסלים. טווח הקשירה הוא הצד הימני של קו הפרדה זה. כיוון שכך, חלים כאן כל העקרונות שעליהם דיברנו בפרק הקודם. כך למשל כלל (α) ישים, ונובע ממנו ש:

$$\{x \mid \psi\} = \{y \mid \psi(y/x)\}$$

ובלבד ש- y כשיר להפעלת כלל (α) כאן. כך לדוגמה:

$$\{n \mid \exists k.\ n = 2k\} = \{j \mid \exists k.\ j = 2k\} = \{j \mid \exists i.\ j = 2i\}$$

$$\{x \mid 0 \le x \le 1\} = \{z \mid 0 \le z \le 1\}$$

התכונה היסודית של $\{x \mid \psi\}$ ניתנת על ידי:

$$\forall x(x \in \{x \mid \psi)\} \Leftrightarrow \psi)$$

מה ששקול, לפי כלל (α) עבור $\forall$, ל :

$$\forall y(y \in \{x \mid \psi\} \Leftrightarrow \psi(y/x))$$

מזה נובע, לפי הכלל הבסיסי של $\forall$ (כלל 14), שלכל ביטוי t, החופשי להצבה במקום x ב- ψ מתקיים:

כלל (β) : $\boxed{t \in \{x \mid \psi\} \Leftrightarrow \psi(t/x)}$

דוגמאות

1. $\frac{1}{4} \in \{x \mid 0 \le x \le 1\} \Leftrightarrow 0 \le \frac{1}{4} \le 1$

כיוון שצד ימין של האקויולנציה נכון, דהיינו: אכן $0 \le \frac{1}{4} \le 1$, אז גם צד שמאל שלה הוא נכון, כלומר $\frac{1}{4} \in \{x \mid 0 \le x \le 1\}$.

2. $2 \in \{x \mid 0 \le x \le 1\} \Leftrightarrow 0 \le 2 \le 1$

כיוון שכאן צד ימין של האקויולנציה *אינו* נכון, גם צד שמאל אינו נכון, כלומר: $2 \notin \{x \mid 0 \le x \le 1\}$.

3. **נסמן את קבוצת המספרים הטבעיים ב- $\mathbf{N}$. אזי[5]:**

$$8 \in \{n \mid n \in \mathbf{N} \wedge \exists k \in \mathbf{N}.\ n = 2k\} \Leftrightarrow 8 \in \mathbf{N} \wedge \exists k \in \mathbf{N}.\ 8 = 2k$$

עתה כאן אגף ימין של השקילות הוא נכון. זאת, משום ששני הקוניוקטים בו הם נכונים. $8 \in \mathbf{N}$ (כלומר 8 אכן מספר טבעי) וכמו כן $4 \in \mathbf{N}$ ו- $8 = 2 \cdot 4$, ולכן $\exists k \in \mathbf{N}. 8 = 2k$. מכאן שאכן $8 \in \{n \mid n \in \mathbf{N} \wedge \exists k \in \mathbf{N}.\ n = 2k\}$. במלים אחרות: 8 שייך לקבוצת המספרים הזוגיים (דהיינו: 8 הינו מספר זוגי).

הערה: במקום לכתוב "n הינו טבעי" כמו למעלה, כתבנו פה "$n \in \mathbf{N}$".

4. **הבה נבדוק האם**

$$(*) \qquad 1.5 \in \{x \mid \forall y.\ y \in \{x \mid x^2 < 2\} \Rightarrow x \ge y\}$$

לפי הכלל הבסיסי (β) זה נכון אם ורק אם

$$\forall y.\ y \in \{x \mid x^2 < 2\} \Rightarrow 1.5 \ge y$$

[5] נזכיר ש: $\exists k \in \mathbf{N}\ldots$ פירושו: $\exists k(k \in \mathbf{N} \wedge \ldots)$

(כאן ψ היא הנוסחה "$\forall y.\ y \in \{x | x^2 < 2\} \Rightarrow x \geq y$", והצבנו את 1.5 במקום המופע החופשי היחיד שיש ל-x בה, שהוא המופע האחרון בה של x). שוב לפי הכלל הבסיסי (β), מופעל כאן עתה על $y \in \{x | x^2 < 2\}$, מה שקיבלנו הוא נכון אם ורק אם:

$$(**)\quad \forall y.\ y^2 < 2 \Rightarrow 1.5 \geq y$$

טענה אחרונה זו הינה נכונה: 1.5 אכן גדול מכל מספר, שריבועו קטן מ- 2. מכאן שגם הטענה (*), בה התחלנו, הינה נכונה.

הערה:

בתרגום לעברית, (*) אומר, ש- 1.5 שייך לקבוצת החסמים מלעיל של קבוצת המספרים, שריבועם קטן מ- 2 (כלומר 1.5 הוא חסם מלעיל של קבוצת המספרים, שריבועם קטן מ- 2).

שאלה:

בקבוצה הנ"ל יש איבר קטן ביותר. מיהו?

הערה נוספת:

ב- (*) המשתנה x נקשר במספר מקומות על ידי אופרטורי קשירה שונים. כרגיל, מי שחושש להתבלבל יכול להפעיל את כלל (α) ולהימנע ממצב עניינים זה. אפשר להעביר תחילה בעזרתו את (*) ל:

$$1.5 \in \{z | \forall y.\ y \in \{x | x^2 < 2\} \Rightarrow z \geq y\}$$

או ל:

$$1.5 \in \{x | \forall y.\ y \in \{z | z^2 < 2\} \Rightarrow x \geq y\}$$

את הפישוטים בעזרת כלל (β) ניתן להתחיל מאחת הצורות הללו. התוצאה הסופית, (**) לא תהיה שונה (בדקו!).

הערה שלישית ואחרונה:

אין ספק שכשלעצמו, (*) קשה מאד לקריאה והבנה. מבחינה זו הניסוח בעברית עדיף. לעומת זאת, כשאנו באים לברר את שאלת *הנכונות*, הרי הרבה יותר קל להשיב נכונה על השאלה אם (**) נכון, מאשר אם הטענה, כמו שהיא מנוסחת בעברית, היא נכונה (ורמת הסיבוכיות כאן אינה גבוהה במיוחד!). עתה, הנקודה המרכזית כאן היא, שהמעבר מ- (*) ל- (**) נעשה לפי כללים מדויקים, שהפעלתם ***אינה דורשת כל מחשבה ואינטליגנציה, אלא רק רכישת מיומנות***. פישוט (*) ל- (**) דומה לפישוט ביטויים אלגבריים על פי נוסחאות, הנעשה בחטיבת הביניים (למעשה הוא מסובך

פחות!) , או לגזירת פונקציות בסוף התיכון. אפשר ללמוד לגזור פונקציות נכונה בלי להבין כלל את מושג הנגזרת. בתהליך הלימוד נעשות שגיאות לא מעטות (למשל: שגיאות הנובעות מאי הפעלה נכונה של הכלל לגזירת פונקציה מורכבת), אבל בסופו של דבר כולם לומדים לגזור נכון. זו רק שאלה של זמן ותרגול. לימוד הפעלה נכונה של כללים כמו (α) ו (β) הוא דומה, אך הוא פשוט לאין ערוך!

אנו משאירים לבדיקת הקוראים את נביעת העובדות הפשוטות (אך חשובות) הבאות מן ההגדרות:

(i) $\{x|\psi\}\subseteq\{x|\varphi\}\Leftrightarrow\forall x(\psi\Rightarrow\varphi)$

(ii) $\{x|\psi\}=\{x|\varphi\}\Leftrightarrow\forall x(\psi\Leftrightarrow\varphi)\Leftrightarrow\forall x(\psi\Rightarrow\varphi)\wedge\forall x(\varphi\Rightarrow\psi)$

(iii) $[\forall x\in\{y|\varphi\}.\ \psi]\Leftrightarrow[\forall x.\ \varphi(x/y)\Rightarrow\psi]$

(iv) $[\exists x\in\{y|\varphi\}.\ \psi]\Leftrightarrow[\exists x.\ \varphi(x/y)\wedge\psi]$

ב- (iii) וב- (iv) x חייב, כמובן, להיות חופשי להצבה במקום y בנוסחה φ.

דוגמאות:

$$[\forall x\in\{z|0\le z\le 1\}.\ x^2\le 1]\Leftrightarrow[\forall x.\ 0\le x\le 1\Rightarrow x^2\le 1]$$

$$[\exists a\in\{z|0\le z\le 1\}.\ a=a^2]\Leftrightarrow[\exists a.\ 0\le a\le 1\wedge a=a^2]$$

הקדשנו לא מעט זמן לסימון $\{x|\psi\}$ בגלל חשיבותו המרובה. נקודה שלא התייחסנו אליה עד כה היא: אלו נוסחאות ψ קבילות לצורך הגדרת קבוצות בצורה $\{x|\psi\}$? כדי לראות, שהשאלה חשובה היא, ושהגבלות כלשהן הינן נחוצות, נתבונן בדוגמאות הבאות.

פרדוקס ברי (Berry):

תהי B ה"קבוצה" הבאה:

$B=\{n|$ n הוא מספר טבעי, שאי-אפשר להגדיר בעברית בפחות מ- 20 מלים $\}$

בתוך הסוגריים המסולסלות, מימין ל- " | ", רשומה נוסחה בשפה העברית. בכך אין פסול: בשום מקום עד כה לא התייחסנו לשפה, בה רשומה הנוסחה ψ בביטוי כמו$\{x|\psi\}$, ולגבינו לכן כל נוסחה בעלת מובן, בכל שפה, הינה באה בחשבון. כמו כן נוסחה זו מגדירה אכן, כמדומה, תכונה של עצמים. לכל העצמים, שאינם מספרים

טבעיים, אין ,כמובן, תכונה זו. ברור לעומת זאת, שלרוב המספרים הטבעיים *יש* תכונה זו. בשפה העברית יש מספר סופי של מלים, ולכן רק מספר סופי של ביטויים כשירים דקדוקית, המכילים פחות מעשרים מלים. מתוך אלה רק חלק מגדיר מספרים טבעיים (למשל, הביטוי: " המספר הזוגי החיובי הקטן ביותר", המגדיר את המספר הטבעי 2, ויש בו חמש מלים בלבד). מכאן, שיש מספר סופי של מספרים, להם אין התכונה, ולכל השאר (אין-סוף מספרים) יש. עתה, לפי אחד העקרונות הבסיסיים של המתמטיקה, בכל קבוצה לא-ריקה של מספרים טבעיים יש מספר מינימלי. בפרט נכון הדבר ל-B, אם אכן קבוצה היא. נניח אפוא, ש- b הוא המספר המינימלי ב- B. b הוא לכן ***המספר הטבעי הקטן ביותר שאי אפשר להגדיר בפחות מעשרים מלים*** - והרגע הגדרנו אותו בעשר מלים (המלים המודגשות במשפט האחרון). עשר, כידוע, קטן מעשרים. קיבלנו לכן, שאת b אי אפשר להגדיר בפחות מעשרים מלים, אך אפשר להגדיר בעשר מלים. הכיצד??

למצב עניינים, שאינו מתקבל על הדעת, קוראים פרדוקס[6] , ופרדוקסים הינם דבר מטריד, הדורש התייחסות. ניתוח לעומק של פרדוקס ברי מראה, שמקור הבעיה הוא בכך, שניסוחים בשפה העברית (או כל שפה טבעית אחרת), שהינם תקינים דקדוקית, הם לא תמיד גם בעלי מובן, אפילו אם נדמה כך. במקרה שלפנינו הביטוי "מספר טבעי שאי אפשר להגדיר בפחות מעשרים מלה" אינו צירוף מלים, המגדיר תכונה ברורה, שלכל מספר יש או אין. השפה העברית פשוט עשירה ומורכבת מידי. בתורת הקבוצות האקסיומטית משתמשים לכן בשפה, שהיא גם מצומצמת הרבה יותר וגם פורמלית. בעיות מסוג פרדוקס ברי אינן מתעוררות לגבי שפה זו. אנו לא נתאר כאן בצורה מדויקת את השפה של תורת הקבוצות האקסיומטית. נבטיח עם זאת לקורא, שלהגדרת קבוצות נשתמש אנו רק בשפה מתמטית ברורה, שלכל הנוסחאות בה יהיה מובן חד-משמעי.

בעיה קשה הרבה יותר מציב לפני תורת הקבוצות הפרדוקס הבא:

פרדוקס ראסל:

ותהיה S ה"קבוצה" $\{x \mid x \notin x\}$. אזי

$$S = \{x \mid x \notin x\}$$

---

[6] ברצוני לנצל הזדמנות זו על מנת להמליץ על ספרה של ענת בילצקי אודות פרדוקסים (בהוצאת האוניברסיטה המשודרת של גל"צ). ספרון מאלף ומהנה זה מתאר הן את הפרדוקסים של תורת הקבוצות והן פרדוקסים מפורסמים בתחומים אחרים. עלי להוסיף, אמנם, שדעות המחבר כאן ודעות המחברת שם חלוקות בתכלית בכל הנוגע לתוכן הפרק, המסיים את ספרה!

לכן: $\forall y.\ y\in S \Leftrightarrow y\in \{x \mid x\notin x\}$

מכאן (בעזרת כלל (β)) $\forall y.\ y\in S \Leftrightarrow y\notin y$

לכן (כלל 14, עם $t = S$) $S\in S \Leftrightarrow S\notin S$

הטענה האחרונה אינה יכולה אבל להיות נכונה. ל- $S\in S$ ול- $S\notin S$ יש תמיד ערך אמת שונה, ולכן הן אינן יכולות להיות שקולות. דבר זה ניתן לברר בקלות בעזרת טבלת האמת של $\Leftrightarrow$ (גם בלעדיה ברור, שמהטענה האחרונה נובע, שאם S שייך ל- S אז S לא שייך ל- S ולהפך – מצב עניינים אבסורדי). הכיצד??

הערה:

השימוש בשם S בתיאור הפרדוקס נועד לצרכים פסיכולוגיים בלבד, ואין הוא למעשה חיוני. בלעדיו היה הפרדוקס נראה כך:

(β): $\forall y.\ y\in \{x \mid x\notin x\} \Leftrightarrow y\notin y$

לכן: $\{x \mid x\notin x\} \in \{x \mid x\notin x\} \Leftrightarrow \{x \mid x\notin x\} \notin \{x \mid x\notin x\}$

(זאת לפי כלל 14, כשלוקחים $t = \{x \mid x\notin x\}$). המשפט האחרון הוא שוב סתירה לוגית, כי שום טענה אינה יכולה להיות שקולה לשלילתה.

הבעיה שיוצר פרדוקס ראסל קשה שבעתיים מזו שמציג פרדוקס ברי. מצד אחד ברור ממנו, שהביטוי $\{x \mid x\notin x\}$ אינו יכול להגדיר קבוצה. מצד שני הנוסחה $x\notin x$ (כלומר $\neg(x\in x)$) הינה תקינה בהחלט בשפה שנשתמש בה, וגם בשפת תורת הקבוצות האקסיומטית. כיוון שכך, עלינו להשלים עם העובדה, שלא עבור כל נוסחה תקינה ψ מגדיר הביטוי $\{x \mid \psi\}$ ***קבוצה*** (כלומר אוסף, שהוא בעצמו ***עצם***, וממילא הכללים (β) והכלל הבסיסי של $\forall$ חלים עליו). נשאלת אפוא השאלה, כיצד נחליט, איזו ψ "כשרה" היא ואיזו לא? תשובה מקובלת ניתנת במסגרת תורת הקבוצות האקסיומטית הידועה בשם ZF (Zermelo-Fraenkel). ZF כוללת רשימה של אקסיומות, שמרביתן פשוט מבטיחות, לגבי נוסחאות רבות, שהן אכן כשרות לשימוש. אקסיומטיקה זו, כשלעצמה, אינה מטרת קורס זה. ברם, כמעט לכל אקסיומה של ZF מתאים ***סימון*** נפוץ עבור קבוצות. סימונים אלו וכללים גם וכללים המסגרת מה שכל מי שמשתמש בשפה של תורת הקבוצות חייב לדעת. נביא אפוא עתה את העקרונות של ZF עם הסימונים התואמים. עיקר ענייננו, נדגיש שוב, הוא ***בסימונים***: אנו נכניס לשימוש סוגים חדשים של ביטויים ונתאר, מתי ביטויים אלה מייצגים קבוצות, ואילו קבוצות הם מייצגים.

התואמים. עיקר ענייננו, נדגיש שוב, הוא *בסימונים*: אנו נכנים לשימוש סוגים חדשים של ביטויים ונתאר, מתי ביטויים אלה מייצגים קבוצות, ואילו קבוצות הם מייצגים.

עיקרון I:

יש מספר אוספים חשובים שעליהם נניח *במפורש*, שאכן קבוצות הם, ואף נעניק להם שם וסימון מיוחד. בין קבוצות אלה נמנה את:

(א) קבוצת המספרים הטבעיים (כולל 0) אותה נסמן ב- **N**.

(ב) קבוצת המספרים הטבעיים החיוביים ממש (ללא אפס), אותה נסמן ב- $\mathbf{N}^+$.

(ג) קבוצת המספרים השלמים $\{...-2,-1,0,1,2...\}$. סימונה: **Z**.

(ד) קבוצת המספרים הרציונליים (שברים אמיתיים או מדומים, חיוביים או שליליים, כולל 0). סימונה: **Q**.

(ה) קבוצת המספרים הממשיים. סימונה: **R**.

(ו) קבוצת המספרים המרוכבים. סימונה: **C**.

הערות:

(1) כאקסיומה יש להניח, למעשה, רק את קיומה (כקבוצה) של **N**. את כל השאר אפשר לבנות (ולהוכיח את עובדת היותן קבוצות) על סמך אקסיומה זו (בעזרת שאר האקסיומות). לא נעשה זאת בקורס זה[7].

(2) מה שאנו קוראים פה "שמות מיוחדים" נקרא בשפות תכנות "reserved words".

קבוצה חשובה במיוחד נוספת, שאת קיומה נניח (ולמעשה אף *נוכיח* בשלב מאוחר יותר) היא הקבוצה *הריקה*, דהינו הקבוצה, שאין בה איברים כלל (לפי עיקרון האקסטנציונליות תתכן לכל היותר קבוצה אחת כזו, ולכן השימוש בהא הידוע ביחס אליה מוצדק). חשיבותה של קבוצה זו מקנה לה את הזכות לסימון מיוחד משלה. הסימון המקובל הוא $\varnothing$ ("פי", בפא לא דגושה).

את $\varnothing$ ניתן להגדיר כך:

$$\varnothing = \{x \mid x \neq x\}$$

תכונה מיוחדת מאד של $\varnothing$ הינה, שהיא קבוצה חלקית לכל קבוצה אחרת: $\varnothing \subseteq A$ לכל קבוצה A. דבר זה נובע מיד מההגדרות. $\varnothing \subseteq A$ פירושו

$$\forall x.\ x \in \varnothing \Rightarrow x \in A$$

---

[7] המתעניינים בבניית **R**, למשל, יוכלו לעיין בפרקים הראשונים בספרו של מייזלר: "חשבון אינפיניטסימלי".

אבל אם x עצם כלשהו, אז, מהגדרה, $x\notin\varnothing$. מעובדה זו נובע לפי טבלת האמת של $\Rightarrow$, ש- $x\in\varnothing \Rightarrow x\in A$ (כיוון שערך האמת של $x\in\varnothing$ הוא f, ערך האמת של הגרירה כולה הוא t). זה נכון *לכל* עצם x, ולכן $\forall x.\ x\in\varnothing \Rightarrow x\in A$.

הסבר *אינטואיטיבי* למה $\varnothing \subseteq A$, ניתן לתת כך: אם לוקחים קבוצה A ו"זורקים" ממנה איברים בזה אחר זה, נקבל בכל שלב קבוצות מצומצמות יותר ויותר, אך כולן חלקיות כמובן לקבוצה המקורית A. הדעת נותנת אפוא, שמה שמתקבל בסוף התהליך, עת "זרקנו" את כל האיברים ונותרנו עם הקבוצה הריקה, גם הוא קבוצה חלקית של הקבוצה המקורית[8]. (כדאי לשים לב, שגם פה האינטואיציה ש- $\varnothing \subseteq A$ תואמת, ואפילו מכתיבה, את טבלת האמת של $\Rightarrow$).

לסיום עניין השמות המיוחדים נציין, שפרט לשמות הקבועים שמנינו כאן, מקובל לתת מדי פעם שמות *זמניים* לקבוצות בעלות חשיבות בקונטקסט מסוים בו עוסקים (כמו בניסוח משפט או בהוכחה כלשהי). כותבים אז משהו כמו: "נסמן $A = \{x|\psi\}$", או, בניסוח שקול: "נגדיר קבוצה A על-ידי $\forall x(x\in A \Leftrightarrow \psi)$" (את ה- "$\forall x$" משמיטים בדרך כלל בניסוחים אלה. במקום "A" יכול להופיע כאן כל שם אחר. ψ חייבת, כמובן, להיות נוסחה כך ש- $\{x|\psi\}$ היא קבוצה). אנו עצמנו השתמשנו באמצעי זה, עת הענקנו ל"קבוצת" ראסל $\{x|x\notin x\}$ (שאומנם אינה קבוצה כלל) את השם הזמני S. הסיבה שם היתה הרצון להגדיל בהירות ולקצר ניסוחים, וזוהי תמיד הסיבה להכנסת שמות זמניים. לשמות זמניים קוראים בשפות תכנות "קונסטנטות". הכוונה שם היא לשמות, שתוקפם חל רק במסגרת תכנית מסוימת, והוכרזו ככאלה על-ידי כותב אותה תכנית.

עיקרון II:

אם $t_1,\ldots,t_n$ הם ביטויים המייצגים עצמים, אז $\{t_1,\ldots,t_n\}$ הוא ביטוי המייצג קבוצה. ביטוי זה הינו קיצור של $\{x|x = t_1 \vee x = t_2 \vee \ldots \vee x = t_n\}$, כאשר x הוא משתנה כלשהו, שאינו מופיע חופשי ב- $t_1,\ldots,t_n$. לכן עבור כל ביטוי s מתקיים:

$$\boxed{s \in \{t_1,\ldots,t_n\} \Leftrightarrow s = t_1 \vee \ldots \vee s = t_n}$$

[8] המדקדקים יוכלו לטעון, שטיעון זה תקף לקבוצות סופיות בלבד. כיוון שמדובר בהסבר אינטואיטיבי, שאיננו, וגם לא מתימר להיות, *הוכחה*, אין עניין זה מדאיג פה. יתר על כן: ברגע שהשתכנענו, כי $\varnothing$ הינה חלקית לכל קבוצה סופית, הרי טרנסיטיביות יחס ההכלה, אותה הראינו למעלה, כופה זאת גם על קבוצות אין-סופיות.

דוגמאות:

(i) הביטוי $\{1,2,3\}$ מייצג קבוצה בת שלושה איברים: המספרים 1,2,3.

(ii) הביטוי $\{\varnothing\}$ מייצגת קבוצה בת איבר אחד. איברה היחיד הוא הקבוצה $\varnothing$. (יש לשים לב, ש- $\{\varnothing\}$ עצמה *אינה* ריקה: יש בה איבר!).

(iii) הביטוי $\{\varnothing,\{\varnothing\}\}$ מיצג קבוצה בת שני איברים: איבריה הם $\varnothing$ ו- $\{\varnothing\}$ (כדאי לשים לב, ש- $\{\varnothing,\{\varnothing\}\}$ היא קבוצה, שאיבריה הם בעצמם קבוצות!).

לקבוצות כמו $\{\varnothing\}$ ו- $\{2\}$, דהינו: קבוצות בעלות איבר יחיד, נוהגים לקרוא ***סינגלטונים***. לעומת זאת, לקבוצה מהצורה $\{t,s\}$ נוהגים לקרוא ***הזוג הלא סדור*** של t ו- s. $\{1,3\}$ ו- $\{\varnothing, \{\varnothing\}\}$ הן דוגמאות לזוגות לא סדורים. כדאי לשים לב, שלפי עיקרון האקסטנציונליות:

(א) $\{t,s\} = \{s,t\}$ (ומכאן השם זוג לא סדור).

(ב) $\{t\} = \{t,t\}$ (ולכן, עקרונית, כל סינגלטון הוא "זוג" לא סדור).

עיקרון III: עיקרון הקומפרהנסיה המוגבל:

אם –

(1) S ביטוי המייצג קבוצה

(2) ψ נוסחה

(3) x משתנה, שאינו חופשי ב- S

אז: הביטוי $\{x\in S \mid \psi\}$ הוא ביטוי המייצג קבוצה. ביטוי זה הוא קיצור של $\{x \mid x\in S \wedge\psi\}$. לכן:

$$\boxed{t \in \{x\in S \mid \psi\} \Leftrightarrow t \in S \wedge\psi(t/x)}$$

עבור כל ביטוי t, החופשי להצבה במקום x ב- ψ.

דוגמאות:

(i) אוסף הפתרונות הממשיים של המשוואה $\sin ax = 0$ הוא הקבוצה

$$\{x \in \mathbf{R} \mid \sin ax = 0\}$$

(כאן ψ מכילה את הפרמטר a, והביטוי מגדיר קבוצה תחת ההנחה, שמציבים במקום a ערכים מ- $\mathbf{R}$).

(ii) קבוצת המספרים הטבעיים הזוגיים היא הקבוצה:

$$\{n \in \mathbf{N} \mid \exists k \in \mathbf{N}.\ n = 2k\}$$

(iii) $\{x \in \mathbf{N} \mid x \neq x\} = \varnothing$. בדומה, לכל A $\{x \in A \mid x \neq x\} = \varnothing$

(iv) $\{x \mid x \in A\} = \{x \mid x \in A \wedge x = x\} = \{x \in A \mid x = x\} = A$.

(השוויון האחרון נובע מעיקרון האקסטנציונליות).

הערות:

(1) ברור ש- $\{x \in A \mid \psi\} \subseteq A$, כלומר: $\{x \in A \mid \psi\}$ הינה תמיד קבוצה חלקית ל - A.

(2) עיקר דוגמה (iv) הוקדש כדי להראות שהביטוי $\{x \mid x \in A\}$ מייצג קבוצה (בתנאי ש- A מייצג קבוצה, כמובן). מ-(iv) נובע אבל גם עיקרון חשוב, אם כי פשוט ביותר, הידוע בשם עיקרון (η)[9]. עיקרון זה קובע, שאם S ביטוי המייצג קבוצה, אז:

$$\boxed{\{x \mid x \in S\} = S}$$

(בתנאי ש- x אינו חופשי בביטוי S).

(3) דוגמה מספר (iii) מראה, שהעובדה, שאכן $\varnothing$ הינה קבוצה, נובעת למעשה מעיקרון הקומפרהנסיה המוגבל, ואינה מצריכה אקסיומה מיוחדת. כמו כן היא מציגה מזווית ראיה שונה את העובדה ש- $\varnothing \subseteq A$ לכל A (על סמך הערה (1)).

(4) **%%** בניסוח עיקרון הקומפרהנסיה המוגבל השמטנו אי אלו תנאים, שעלולים היו לבלבל. גם תנאי (3) שם הוא אולי סתום מעט. כדי להבין אותם נדגיש שוב, שאנו עוסקים בתיאור ***שפה*** ובמשמעות ביטוייה. כשמיישמים את עיקרון הקומפרהנסיה המוגבל, הרי S יהיה ***ביטוי***, אולי קצר יותר, אולי ארוך – ואולי אחד שמכיל משתנים חופשיים. ψ תהיה תמיד ***נוסחה***, ואף היא עשויה להכיל פרמטרים (כלומר: משתנים חופשיים). הביטוי $\{x \in S \mid \psi\}$ עשוי אפוא להכיל פרמטרים. הוא יוכל, כמובן, לייצג קבוצה קונקרטית רק כאשר מציבים ערכים קונקרטיים במקום הפרמטרים – וגם אז ייתכן, שנקבל ביטוי, המייצג קבוצה רק עבור ערכים מסוימים של הפרמטרים. כדוגמה ניקח את דוגמה מס' (i) למעלה: בביטוי $\{x \in \mathbf{R} \mid \sin ax = 0\}$ מופיע הפרמטר a. הביטוי מייצג קבוצות חלקיות ל- $\mathbf{R}$ רק כאשר מציבים במקום a מספרים ממשיים. למשל כש- $a = \pi$ נקבל את $\{x \in \mathbf{R} \mid \sin \pi x = 0\}$, שהיא אכן קבוצה חלקית ל- $\mathbf{R}$ ($\mathbf{Z}$, למעשה).

---

[9] (η) היא האות היוונית אטה. השמות α, β ו η לעקרונות, שכינינו כאן בשם זה, שאולים מתחשיב λ, כיוון שמדובר בעקרונות מאד דומים לאלה שם. על סימון λ ועקרונות α, β ו- η שלו נלמד בהמשך.

כשאנו מבינים כל זאת, נעבור לתנאי (3) בניסוח עיקרון הקומפרהנסיה, ונדגים את נחיצותו. נעיין אפוא בדוגמה הבאה: $S=\{y\,|\,y=x\}$ ו- $\psi = x\notin x$. עבור כל ערך קונקרטי של המשתנה החופשי x בביטוי S, ייצג ביטוי זה קבוצה (הסינגלטון $\{x\}$). תנאים (1) ו (2) מתקיימים פה לכן. תנאי (3) לא. ואכן :

$$\{x\in S\,|\,\psi\} = \{x\in\{y\,|\,y=x\}\,|\,x\notin x\}$$
$$= \{x\,|\,x\in\{y\,|\,y=x\}\wedge x\notin x\}$$ (לפי הגדרת הסימון $\{x\in\ldots|\ \_\ \}$)
$$= \{x\,|\,x=x\wedge x\notin x\}$$ (לפי עיקרון (β), מופעל על $x\in\{y\,|\,y=x\}$)
$$= \{x\,|\,x\notin x\}$$ (כי $x=x$ נכון לכל x)

ברם, $\{x\,|\,x\notin x\}$ הינו ביטוי, ***שאינו*** מייצג קבוצה, כמו שראינו עת תיארנו את פרדוקס ראסל.

%%

תרגילים

(א) הוכח: $\{x\in\{\varnothing,\{\varnothing\}\}\,|\,x\neq\varnothing\} = \{\{\varnothing\}\}$

(ב) הוכח: $\forall p.\ p\in\mathbf{N}\wedge p\neq 0 \Rightarrow 9p\in\{x\in\{n\in\mathbf{N}\,|\,\exists k\in\mathbf{N}.\ n=pk\}\,|\,x\geq 9\}$

(ג) $\forall x\in\{x\in A\,|\,\varphi\}.\psi \Leftrightarrow \forall x\in A(\varphi\Rightarrow\psi)$

(ד) $\exists x\in\{x\in \mathrm{A}\,|\,\varphi\}.\psi \Leftrightarrow \exists x\in A(\varphi\wedge\psi)$

(ה) הביטוי $\{x\,|\,x=x\}$ אינו מיצג קבוצה (במלים אחרות : הוא אינו מוגדר).

עיקרון IV: עיקרון ההחלפה

אם –

(1) S ביטוי המייצג קבוצה

(2) x משתנה שאינו חופשי ב- S

(3) t ביטוי

אז: $\{t\,|\,x\in S\}$ הוא ביטוי המייצג קבוצה. ביטוי זה הוא קיצור של $\{y\,|\,\exists x\in S.\ y=t\}$, כאשר y הוא משתנה כלשהו, שאינו חופשי ב- S או ב- t. מכאן שאם t' ביטוי ש- x אינו מופיע בו חופשי, אז:

$$\boxed{t'\in\{t\,|\,x\in S\}\Leftrightarrow\exists x\in S.\ t'=t}$$

דוגמאות

(א) $\{2n \mid n\in\mathbf{N}\}$ היא קבוצת המספרים הזוגיים.

(ב) $\{\{n\} \mid n\in\mathbf{N}\}$ היא קבוצת כל הסינגלטונים של מספרים טבעיים (כלומר: קבוצת כל הקבוצות המהוות סינגלטון, שאיברו היחיד הוא מספר טבעי).

אם ננתח את (ב) לפי עיקרון ההחלפה, הרי הביטוי S הוא כאן השם המיוחד (והקבוע) $\mathbf{N}$ (שמייצג קבוצה!), בתפקיד x משחק כאן המשתנה n, ו- t הינו $\{n\}$. $\{\{n\} \mid n\in\mathbf{N}\}$ הוא, לכן, קיצור של $\{z \mid \exists n\in\mathbf{N}.z=\{n\}\}$ (למה דווקא z ? ובכן, למה לא?).

אינטואיטיבית:

$$\{\{n\} \mid n\in\mathbf{N}\}=\{\{0\},\{1\},\{2\},...\}$$

נדגיש, עם זאת, ששפתנו הרשמית אינה כוללת ביטויים כמו זה שבאגף ימין של "המשוואה" האחרונה, כיוון שאין כללים ברורים, מה 3 הנקודות מייצגות בכל מקרה ומקרה. עם זאת, כיוון שהקוראים בני אדם הם ולא מחשבים, סביר להניח, כי יוכלו לנחש, למה הכוונה כאן!

עיקרון V: עיקרון קבוצה החזקה

אם A קבוצה אז אוסף כל הקבוצות החלקיות שלה הוא גם קבוצה. אוסף זה נקרא *קבוצת החזקה* של A.

סימון : אם S ביטוי המייצג קבוצה (כולל המקרה ש- S הוא משתנה!) אז $P(S)$ מייצג את קבוצת החזקה של קבוצה זו. (סימון אחר: 2^S).

הקבוצה $P(A)$ מוגדרת אפוא בצורה הבאה:

$$P(A)=\{x \mid x\subseteq A\}$$

ולכן אם S ביטוי המייצג קבוצה, ו- t הינו ביטוי, אז:

$$\boxed{t\in P(S)\Leftrightarrow t\subseteq S}$$

הערות

(1) כרגיל, בכותבנו הגדרה כמו זו של $P(A)$ למעלה, יש להתייחס אליה (כשמפעילים אותה) כאילו היה כתוב: $\forall A.\ P(A) = \{x \mid x \subseteq A\}$. הערה דומה תחול על כל הגדרה באותו סגנון בעתיד.

(2) ברוב המכריע של ספרי הלימוד כותבים פשוט: "אם A קבוצה אז $P(A)$ מסמן את קבוצת החזקה של A" (ולא את הדבר המסורבל שאנו כתבנו). ניסוח זה איננו מדויק, למעשה. אם ניקח, למשל, את קבוצת חברי הכנסת, הרי לא נוכל לשים אותם בסוגריים ולכתוב את האות "P" לפני התוצאה. מה שנוכל לשים בסוגריים הוא ***ביטוי*** לשוני, המתאר או ***מייצג*** את קבוצת חברי הכנסת, לא את הקבוצה עצמה. עם זאת, יש להודות שהניסוח שלנו, המדבר על ביטויים, אכן ***מסורבל*** הוא. כיוון שכך, נתחיל גם אנו להשתמש יותר ויותר בנוסח המקובל, הפשוט יותר לקריאה. תקוותנו ואמונתנו, שבשלב זה יבינו תמיד הקוראים, מה היה צריך לכתוב ***באמת***.

עובדות ברורות:

(1) $A \in P(A)$

(2) $\varnothing \in P(A)$

(3) $\{x \in A \mid \varphi\} \in P(A)$

דוגמאות :

(1) $P(\varnothing) = \{\varnothing\}$

(2) $P(\{1\}) = \{\varnothing, \{1\}\}$

(3) $P(\{1,2\}) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$

הערה:
בהקשר של הדוגמה הראשונה כדאי לשים לב, ש- $\varnothing \subseteq \varnothing$, ולכן $\varnothing \in \{\varnothing\}$.
כמו כן, $\varnothing \subseteq \{\varnothing\}$. עם זאת, $\varnothing \notin \varnothing$.

הדוגמאות מעלות בצורה ברורה את ההשערה הבאה, שאת הוכחתה באינדוקציה נשאיר לקוראים (כמו כן עוד נחזור למשפט למטה, בחלק שיוקדש לקומבינטוריקה).

משפט: אם A קבוצה סופית בת n איברים אז ב- $P(A)$ יש 2^n איברים.

בכך סיימנו כמעט לסקור את העקרונות הבסיסיים, המאפשרים יצירת קבוצות באופן קונסטרוקטיבי (כולל סימון מתאים). נותר עוד עיקרון אחד, שייסקר (במסגרת נושאים נוספים) בפרק הבא.

לסיום נזכיר עוד את הסימונים המקובלים הבאים מחדו"א עבור קבוצות חלקיות חשובות של $\mathbf{R}$:

$$(-\infty,a) = \{x\in \mathbf{R} \mid x < a\} \qquad (a,b) = \{x\in \mathbf{R} \mid a < x < b\}$$
$$(-\infty,a] = \{x\in \mathbf{R} \mid x \le a\} \qquad (a,b] = \{x\in \mathbf{R} \mid a < x \le b\}$$
$$(a,\infty) = \{x\in \mathbf{R} \mid x > a\} \qquad [a,b) = \{x\in \mathbf{R} \mid a \le x < b\}$$
$$[a,\infty) = \{x\in \mathbf{R} \mid x \ge a\} \qquad [a,b] = \{x\in \mathbf{R} \mid a \le x \le b\}$$

נספח: על משוואות ושאילתות

במערכות בסיסי נתונים (DBMS) מסתכלים על $\{x|\varphi\}$ כעל ***שאילתה***: מצא את כל העצמים המקיימים את התנאי φ. בדרך-כלל רוצים, שלשאילתה יהיה מספר סופי של תשובות, כך שאפשר יהיה לסדרן בטבלה (וכן שאפשר יהיה למצוא תשובות אלו באופן אפקטיבי). אחת הדרכים המרכזיות להשיג זאת היא להגביל את השאילתות המותנות למה שנקרא "שאילתות בטוחות". צורת שאילתות אלו היא $\{x\in A|\varphi\}$, כאשר A היא קבוצה סופית.

בהסתכלות על $\{x\in A|\varphi\}$ כעל שאילתה אין, למעשה, כל חידוש עבורנו. כשאנו נדרשים לפתור בביה"ס התיכון משוואה כמו $\sin x = 0$, אנו נדרשים למצוא את ***קבוצת*** הפתרונות הממשיים של משוואה זו. תיאור מדויק של הקבוצה המבוקשת ניתן על ידי: $\{x\in \mathbf{R} \mid \sin x = 0\}$. ברם תאור זה, מדויק ככל שיהיה, לא כל כך מספק את המורה (ודי בצדק). ברור מעמיק יותר מגלה, שמה שאנו באמת נדרשים לעשות, הוא לתת תיאור פשוט יותר, בעל צורה מסוימת, ***קנונית***, לקבוצת הפתרונות. מהי הצורה הנחשבת לקנונית בתיכון הוא דבר המשתנה לפי סוג הבעיה:

(א) כשמספר הפתרונות הוא סופי אנו רוצים ייצוג מהצורה $\{a_1, \ldots, a_n\}$ למשל:

$$\{x\in \mathbf{R} \mid x^2 - 3x + 2 = 0\} = \{1,2\}$$

(ב) במשוואות טריגונומטריות רוצים "הצגה פרמטרית", כמו :

$$\{x\in \mathbf{R} \mid \sin x = 0\} = \{k\pi \mid k\in\mathbf{Z}\}$$

(או איחוד של קבוצות כאלו. על "איחוד" נדבר בפרק הבא).

(ג) ב"אי-שוויונים" רוצים הצגה כאיחוד של קטעים :

$$\{x\in \mathbf{R} \mid x^2 > 1\} = \{x\in \mathbf{R} \mid x < -1 \lor x > 1\} \quad [= (-\infty,-1) \cup (1,\infty)]$$

אנו רואים אפוא, ש"משוואות" ו"אי-שוויונים" אינם אלא ייצוגים מסוימים של קבוצות של מספרים, ו"פתרונם" הוא ייצוג אותן קבוצות בצורה "קנונית", נוחה יותר לשימוש. ה"נעלמים" במשוואות הינם לכן משתנים קשורים (על-ידי $\{\ldots \mid \_\,\}$). ב"משוואות עם פרמטרים" הפרמטרים אכן כשמם כן הם: משתנים חופשיים.

את המשפט המרכזי על פתרון משוואות ריבועיות נוכל לנסח עתה כך:

$$\forall a\in \mathbf{R}\ \forall b\in \mathbf{R}\ \forall c\in \mathbf{R}\ \Big[a \neq 0 \Rightarrow$$

$$\Big((b^2 - 4ac < 0 \Rightarrow$$

$$\{x\in \mathbf{R} \mid ax^2 + bx + c = 0\} = \varnothing) \land$$

$$(b^2 - 4ac \geq 0 \Rightarrow$$

$$\{x\in \mathbf{R} \mid ax^2 + bx + c = 0\} = \Big\{ \frac{-b+\sqrt{b^2-4ac}}{2a}, \frac{-b-\sqrt{b^2-4ac}}{2a} \Big\}) \Big) \Big]$$

ב.3. פעולות יסודיות על קבוצות

בפרק זה נכיר את הפעולות הבסיסיות הקשורות לקבוצות ונלמד את תכונותיהן.

(א) חיתוך, איחוד, משלים, הפרש

הגדרה

א. ***האיחוד*** של שתי קבוצות היא הקבוצה המכילה את כל העצמים השייכים לפחות לאחת משתי הקבוצות.
האיחוד של שתי קבוצות A ו- B מסומן ב- $A\cup B$.

ב. ***החיתוך*** של שתי קבוצות היא הקבוצה המכילה את כל העצמים השייכים לשתי הקבוצות גם יחד.
החיתוך של שתי קבוצות A ו- B מסומן ב- $A\cap B$.

ג. ***ההפרש*** של שתי קבוצות, A ו- B, הוא קבוצת העצמים השייכים ל- A אך לא שייכים ל- B.
ההפרש של A ו- B מסומן ב- $A-B$.

ד. ***ההפרש הסימטרי*** של שתי קבוצות הוא אוסף העצמים הנמצאים בדיוק באחת משתי הקבוצות (כלומר: נמצאים באחת אך לא נמצאים באחרת).
ההפרש הסימטרי של A ו- B מסומן ב- $A\,\Delta\,B$.

הגדרה קצרה יותר של הפעולות $\cup$, $\cap$, $-$, Δ ניתנת על-ידי:

1. $A\cup B=\{x\,|\,x\in A\vee x\in B\}$
2. $A\cap B=\{x\,|\,x\in A\wedge x\in B\}=\{x\in A\,|\,x\in B\}$
3. $A-B=\{x\,|\,x\in A\wedge x\notin B\}=\{x\in A\,|\,x\notin B\}$
4. $A\,\Delta\,B=\{x\,|\,(x\in A\wedge x\notin B)\vee(x\in B\wedge x\notin A)\}=(A-B)\cup(B-A)$

הערות:

1. נזכיר שוב, כי להגדרות מהסוג של 1-4 יש להתייחס כאילו הן טענות עם כמתים כוללים $\forall A\forall B$ בתחילתם (וממילא כלל α וכללי $\forall$ ישימים).

2. מהשוויון השני, הן ב- 2 והן ב- 3, נובע, שאם A ו- B הן קבוצות, אז גם $A \cap B$ ו- $A - B$ הן קבוצות (לפי עקרון הקומפרהנסיה המוגבל). בהנחה שאיחוד של שתי קבוצות נותן קבוצה, נובע מזה ומהשוויון השני ב- 4, ש- $A \Delta B$ קבוצה (כאשר A ו- B קבוצות). לעומת זאת אין העובדה, ש- $A \cup B$ היא קבוצה, נובעת משום עיקרון, שלמדנו עד כה. נחוץ לכך עקרון מיוחד:

עקרון האיחוד, נוסח חלש:

האיחוד של שתי קבוצות הוא קבוצה. (ממילא, כש- A ו- B ביטויים המגדירים קבוצות אז הביטוי $\{x | x \in A \vee x \in B\}$ מגדיר קבוצה).
ניסוח אחר לעקרון זה הוא: אם הביטויים $\{x \mid \varphi\}$ ו- $\{x \mid \psi\}$ מגדירים קבוצות, אז גם הביטוי $\{x \mid \varphi \vee \psi\}$ מגדיר קבוצה.

דוגמאות:

(א) נניח:

$$A = [0,2] = \{x \in \mathrm{R} | \ 0 \le x \le 2\}$$
$$B = [1,4] = \{x \in \mathrm{R} | \ 1 \le x \le 4\}$$

אזי:

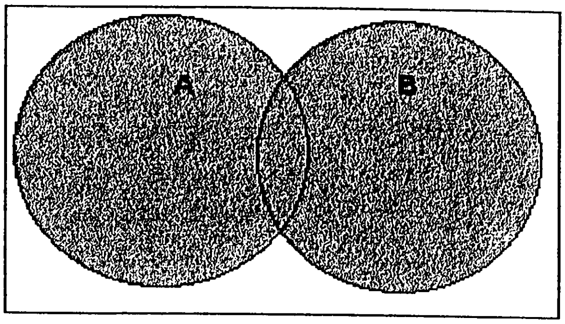
$$A \cup B = [0,4] = \{x \in \mathrm{R} | \ 0 \le x \le 4\}$$
$$A \cap B = [1,2] = \{x \in \mathrm{R} | \ 1 \le x \le 2\}$$
$$A - B = [0,1) = \{x \in \mathrm{R} | \ 0 \le x < 1\}$$
$$B - A = (2,4] = \{x \in \mathrm{R} | \ 2 < x \le 4\}$$
$$A \Delta B = \{x \in \mathrm{R} | \ 0 \le x < 1 \vee 2 < x \le 4\}$$

(ב)

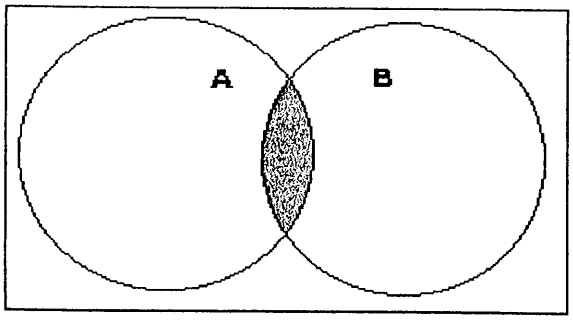
$$\mathrm{P(N)} \cap \mathrm{P}(\{\varnothing, 1\}) = \{ \varnothing, \{1\} \}$$
$$\mathrm{P}(\{\varnothing, 1\}) - \mathrm{P(N)} = \{ \{\varnothing\} , \{\varnothing, 1\} \}$$
$$\mathrm{P(N)} \cup \mathrm{P}(\{\varnothing, 1\}) = \{x \mid x \subseteq \mathrm{N} \vee x = \{\varnothing\} \vee x = \{\varnothing, 1\}\}$$

(ג) המחשה של האופרציות בעזרת דיאגרמות ואן אפשר למצוא בציור מס' ב.2 באיור הבא.

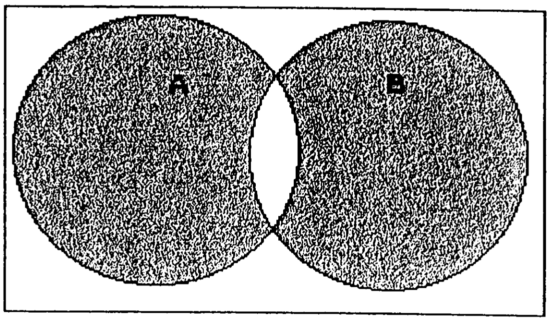
ציור ב.2: דיאגרמת וון עבור הפעולות



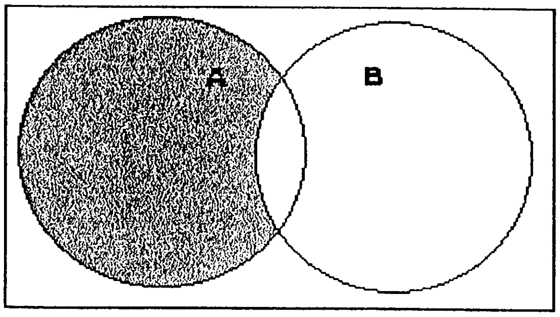
$A \cup B$



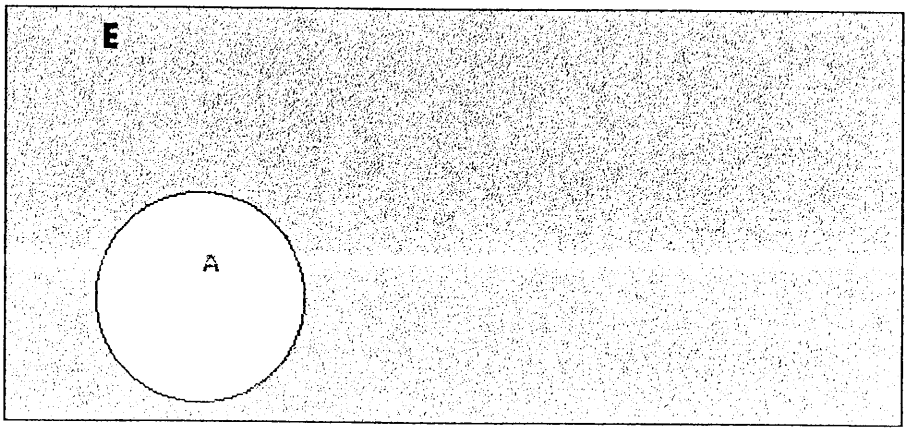
$A \cap B$



$A \Delta B$



$A - B$



$\overline{A}_E$

ההגדרות של חיתוך ואיחוד מראות על קשר הדוק בינן ובין הקשרים הלוגיים $\wedge$ ו- $\vee$. לעומת זאת, יש לשים לב, שאם A קבוצה, אז $\{x| x \notin A\}$ ***אינו*** ביטוי המגדיר קבוצה. ואכן, לו היה מגדיר קבוצה, אז לפי עקרון האיחוד גם $A \cup \{x| x \notin A\}$ היה מגדיר קבוצה. אבל:

$$A \cup \{x| x \notin A\} = \{x| x\in A\} \cup \{x| \neg (x \in A)\} =$$

$$= \{x| x\in A \vee \neg (x\in A)\} = \{x| x = x\}$$

וזאת כיון ש- $x\in A \vee \neg(x\in A)$ הינה אמת לוגית, הנכונה לכל x ו- A. ברם, ראינו למעלה, שהביטוי $\{x| x = x\}$ אינו ביטוי המגדיר קבוצה. מכאן, שגם הביטוי $\{x| x \notin A\}$ אינו מגדיר קבוצה. עם זאת, ברוב הסיטואציות בהם עוסקים בקבוצות, יש איזו "קבוצת גג" E, שכל הקבוצות המעניינות אותנו חלקיות לה. במקרה זה, אם $A \subseteq E$, נוהגים לקרוא ל- $E - A$ ***"המשלים של A ביחס ל- E"*** ומסמנים קבוצה זו ב- $\overline{A_E}$ או A_E^c. למעשה, בדרך כלל משמיטים את ה- "E" בה מדובר, מדברים פשוט על "המשלים של A" ומסמנים זאת ב- $\overline{A}$ או ב- A^c. יש לזכור אבל, שזהותה המדויקת של $\overline{A}$ תלויה בשאלה: מי היא הקבוצה E אותה לוקחים אנו כקבוצת הגג.

המחשה של $\overline{A_E}$ על ידי דיאגרמת וון ניתן למצוא בציור ב.1.

(ב) התכונות הבסיסיות של הפעולות הבסיסיות

בטבלה מספר ב.1 (עמ' 87) מרוכזות התכונות היסודיות של הפעולות, שהוגדרו בסעיף הקודם (למעט ההפרש הסימטרי, שחשיבותו פחותה). אין צורך להיבהל ממספרן הרב: מרביתן מובנות מאליהן. אנו אכן לא נטרח להוכיח את כל החוקים בטבלה. את מרביתן נשאיר כתרגילים לקוראים. כאן נסתפק רק במספר דוגמאות מייצגות.

♦ דוגמה 1:

נוכיח את 5b :

הוכחה א':

הוכחה זו מסתמכת על השקילויות בטבלה א.3 (עמודים 20-21) הקוראים מתבקשים למצוא באילו שקילויות (פרט להגדרת $\subseteq$ ו- $\cup$) אנו משתמשים בכל שלב.

$$
\begin{aligned}
A\cup B\subseteq C \quad &\Leftrightarrow \quad \forall x.\ x\in A\cup B \Rightarrow x\in C\\
&\Leftrightarrow \quad \forall x\,((x\in A \vee x\in B)\Rightarrow x\in C)\\
&\Leftrightarrow \quad \forall x\,((x\in A \Rightarrow x\in C)\wedge(x\in B\Rightarrow x\in C))\\
&\Leftrightarrow \quad \forall x\,(x\in A \Rightarrow x\in C)\wedge \forall x\ (x\in B\Rightarrow x\in C)\\
&\Leftrightarrow \quad A\subseteq C\wedge B\subseteq C
\end{aligned}
$$

הוכחה ב':

זוהי צורת ההוכחה הסטנדרטית. היא נעשית בעזרת כללי ההיסק הלוגים של טבלה א.4 (עמ' 26). עם זאת, כדי להבין אותה אין צורך להכיר טבלה זו!

ראשית, כיוון שמדובר בהוכחת שקילות, אנו מפרקים זאת לשני חלקים:

הוכחה ש: $A\cup B\subseteq C \;\Rightarrow\; A\subseteq C\wedge B\subseteq C$ (הכיוון $\Rightarrow$)

והוכחה ש: $A\cup B\subseteq C \;\Leftarrow\; A\subseteq C\wedge B\subseteq C$ (הכיוון $\Leftarrow$)

הוכחת ($\Rightarrow$) :

נניח ש- $A\cup B\subseteq C$. נוכיח ש- $A\subseteq C$ וש- $B\subseteq C$. נראה למשל, ש- $A\subseteq C$ (הוכחת $B\subseteq C$ היא דומה). יהי אפוא $x\in A$. כיוון ש- $A\subseteq A\cup B$, נובע מזה ש- $x\in A\cup B$. מזה ומהנתון $A\cup B\subseteq C$ נובע ש- $x\in C$.

הוכחת ($\Leftarrow$) :

נניח ש- $A\subseteq C$ וש- $B\subseteq C$. נראה ש- $A\cup B\subseteq C$. נניח לכן ש- $x\in A\cup B$ ונראה ש- $x\in C$. כיוון ש- $x\in A\cup B$, יש שתי אפשרויות :

אפשרות (i): $x\in A$. כיוון שנתון ש- $A\subseteq C$, נובע מזה ש- $x\in C$.

אפשרות (ii): $x\in B$. כיוון שנתון ש- $B\subseteq C$, נובע מזה ש- $x\in C$.

בשני המקרים $x\in C$, כמבוקש.

הערות:

1. במהלך ההוכחה השתמשנו בתכונה 2a מטבלת התכונות היסודיות. הנחנו, אם כך, שתכונה זו כבר הוכחה (ההוכחה היא אכן טריוויאלית במיוחד). למעשה, היה קצר יותר שלא להשתמש בה כלל ולכתוב כך:

   "יהי אפוא $x\in A$. מזה נובע $x\in A\vee x\in B$ ולכן $x\in A\cup B$."

   מקובל יותר אבל בטקסטים לכתוב בצורה שכתבנו למעלה.

2. "החלוקה למקרים" אינה אלא שימוש בכלל 6 מטבלה א.4.

טבלה ב.1: תכונות יסודיות של האופרציות הבסיסיות

1a $A \cap B \subseteq A$	1b $A \cap B \subseteq B$
2a $A \subseteq A \cup B$	2b $B \subseteq A \cup B$
3a $C - A \subseteq C$	3b $A \subseteq B \Rightarrow C - B \subseteq C - A$
4a $A \subseteq B \Leftrightarrow A \cap B = A$	4b $A \subseteq B \Leftrightarrow A \cup B = B$
5a $C \subseteq A \cap B \Leftrightarrow C \subseteq A \wedge C \subseteq B$	5b $A \cup B \subseteq C \Leftrightarrow A \subseteq C \wedge B \subseteq C$
6a $A \cap A = A$	6b $A \cup A = A$
7a $A \cap B = B \cap A$	7b $A \cup B = B \cup A$
8a $(A \cap B) \cap C = A \cap (B \cap C)$	8b $(A \cup B) \cup C = A \cup (B \cup C)$
9a $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$	9b $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
10a $A \cap \varnothing = \varnothing$	10b $A \cup \varnothing = A$
11a $A - A = \varnothing$	11b $A - \varnothing = A$

12. $C - (C - A) = C \cap A$

13a $C - (A \cap B) = (C - A) \cup (C - B)$	13b $C - (A \cup B) = (C - A) \cap (C - B)$
14a $A \cap (C - A) = \varnothing$	14b $A \cup (C - A) = C \cup A$

12 *. $\overline{\overline{A}} = A$

13a * $\overline{A \cap B} = \overline{A} \cup \overline{B}$	13b * $\overline{A \cup B} = \overline{A} \cap \overline{B}$
14a * $A \cap \overline{A} = \varnothing$	14b * $A \cup \overline{A_E} = E$

15. $A - B = A \cap \overline{B}$

הוכחה ג':

גם בהוכחה זו אנו מפרקים את מה שצריך להוכיח לשני חלקים, אך משתמשים בשביל כל אחד בתכונות פשוטות יותר מהטבלה (אותן צריך, כמובן, להוכיח קודם).

הוכחת ($\Rightarrow$) :

נניח ש- $A \cup B \subseteq C$, אז לפי 4b: $A \cup B \cup C = C$, כיוון ש- $A \subseteq A \cup B \cup C$ ו- $B \subseteq A \cup B \cup C$, נובע מזה ש- $A \subseteq C$ וש- $B \subseteq C$.

הוכחת ($\Leftarrow$) :

נניח ש- $A \subseteq C$ וש- $B \subseteq C$, אז לפי 4b: $A \cup C = C$ ו- $B \cup C = C$.
לפי כללים 6-8:

$$(A \cup B) \cup C = (A \cup B) \cup (C \cup C) = (A \cup C) \cup (B \cup C) = C \cup C = C$$

לכן, לפי 4b, $A \cup B \subseteq C$.

מוסר השכל: לטענות טריוויאליות אפשר למצוא הרבה הוכחות טריוויאליות ...

♦ דוגמה 2:

נוכיח: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ 9a.

הוכחה א':

$$\begin{aligned} x \in A \cap (B \cup C) &\Leftrightarrow x \in A \wedge x \in B \cup C \\ &\Leftrightarrow x \in A \wedge (x \in B \vee x \in C) \\ &\Leftrightarrow (x \in A \wedge x \in B) \vee (x \in A \wedge x \in C) \\ &\Leftrightarrow x \in A \cap B \vee x \in A \cap C \\ &\Leftrightarrow x \in (A \cap B) \cup (A \cap C) \end{aligned}$$

הוכחה ב':

($\subseteq$) :

נניח $x \in A \cap (B \cup C)$. אז $x \in A$ וגם $x \in B \cup C$, לכן $x \in B$ או $x \in C$.

אפשרות (i): $x \in B$.

במקרה זה, כיוון ש- $x \in A$, נקבל ש- $x \in A \cap B$.

כיוון ש- $A \cap B \subseteq (A \cap B) \cup (A \cap C)$, נקבל ש- $x \in (A \cap B) \cup (A \cap C)$.

אפשרות (ii): $x \in C$

במקרה זה , כיוון ש- $x \in A$**, נקבל ש-** $x \in A \cap C$.

לכן גם כאן נקבל ש- $x \in (A \cap B) \cup (A \cap C)$.

$(\supseteq)$:

לשם הגיוון, נוכיח כיוון זה בעזרת תכונות קודמות בדף.

ובכן, כיוון ש- $A \cap B \subseteq A$ **ו-** $A \cap C \subseteq A$**, נובע, לפי 5b, ש-**

$$(A \cap B) \cup (A \cap C) \subseteq A$$

כמו כן, $A \cap B \subseteq B \subseteq B \cup C$ **ו-** $A \cap C \subseteq C \subseteq B \cup C$.

לכן, שוב לפי 5b, מתקיים גם:

$$(A \cap B) \cup (A \cap C) \subseteq B \cup C$$

משתי העובדות שהוכחנו נובע, לפי 5a, ש:

$$(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$$

הערה:

בעתיד, כאשר נרצה להשתמש בתכונות יסודיות מהדף (כמו 5a ו- 5b), לא נציין בפירוש, באיזו תכונה אנו משתמשים, אלא פשוט נשתמש בה!

הערות:

1. לחלק מהחוקים יש שמות, שמן הראוי לזוכרם. הכללים ב- 6 נקראים החוקים האידמפוטנטיים, הכללים ב- 7: החוקים הקומוטטיביים (חוקי החילוף), הכללים ב- 8: החוקים האסוציאטיבים (חוקי הקיבוץ), הכללים ב- 9: החוקים הדיסטריבוטיבים (חוקי הפילוג), והכללים ב- 13: חוקי דה-מורגן. כדאי לציין, שלחלק מהשקילויות הלוגיות יש שמות דומים. דמיון זה אינו מקרי, כמובן, אלא מצביע על קשר אמיץ בין החוקים שם וכאן. כך, למשל, אם נבדוק את ההוכחה הראשונה שהבאנו לחוק הפילוג 9a, נראה, שפרט להגדרת חיתוך ואיחוד, הדבר היחיד, שהשתמשנו בו שם, הוא חוק פילוג לוגי מתאים (שקילות 4a בטבלה א.3, עמוד 20).

2. בגלל חוק הקיבוץ 8 הרשינו לעצמנו בהוכחות לכתוב $A \cup B \cup C$, בלי לפרט אם הכוונה ל- $(A \cup B) \cup C$ או ל- $A \cup (B \cup C)$. גם בהמשך ננהג כך (גם לגבי $\cap$ כמובן). כללית, נובע מחוקים 6-8, שבביטוי מהצורה $A_1 \cup A_2 \cup \ldots \cup A_n$ (או $A_1 \cap A_2 \cap \ldots \cap A_n$) אין חשיבות למיקום הסוגריים, לסדר האיברים או למספר הפעמים, שכל אחד מהם מופיע.

3. את חוקי פעולת ההפרש אין, למעשה, צורך לזכור. במקום זאת די לזכור את חוקי פעולת המשלים (להם יש כללים לוגים מקבילים). כל חוקי פעולת ההפרש ינבעו אז מהזהות (15):

$$A - B = A \cap \overline{B}$$

כאשר המשלים כאן, $\overline{B}$, יכול להילקח יחסית לכל קבוצה, המכילה גם את A וגם את B ($A \cup B$, למשל). לדוגמה:

$$C - (A \cup B) = C \cap (\overline{A \cup B}) = C \cap (\overline{A} \cap \overline{B}) =$$
$$= (C \cap \overline{A}) \cap (C \cap \overline{B}) = (C - A) \cap (C - B)$$

בחישוב זה פעולת המשלים נלקחה יחסית ל- $A \cup B \cup C$ (למשל), והשתמשנו בחוקי האיחוד, החיתוך והמשלים בלבד.
מאידך גיסא, כל החוקים של פעולת המשלים הם מקרים פרטיים של חוקי ההפרש, כאשר מניחים אנו ש- $A \subseteq E$ ו- $B \subseteq E$ (E – הקבוצה ביחס אליה נלקח המשלים). לדוגמה, *12 קובע, בעצם, שכאשר $A \subseteq E$ אז $E - (E - A) = A$. זה נובע מיד מ- 12 ומ- 4a.

4. הפעולות הבסיסיות על קבוצות ידועות גם בשם הפעולות הבוליאניות. הכללים שבטבלה ב.1 נכונים לא רק לקבוצות, אלא לכל מבנה מתמטי המהווה מה שנקרא "אלגברה בוליאנית". לא נוכל כאן, לצערנו, להרחיב את הדיבור על אלגבראות בוליאניות מעבר לרמז זה.

(ג) הכללת הפעולות של איחוד וחיתוך

עד כה הגדרנו איחוד של *שתי* קבוצות וחיתוך של *שתי* קבוצות. נכליל עתה פעולות אלו למקרה כללי הרבה יותר.

הגדרה:

תהי F קבוצה, שכל איבריה אף הן קבוצות (לקבוצה הזו קוראים לעתים קרובות, מסיבות בלתי ברורות, ***משפחה*** של קבוצות). ***האיחוד של F*** מוגדר כקבוצת כל העצמים, שנמצאים בלפחות אחת מהקבוצות ב- F. אם F אינה ריקה, אז ***החיתוך של F*** מוגדר כקבוצת כל העצמים, הנמצאים בכל איברי F.

סימונים:

אם A ביטוי המיצג קבוצה, אז "$\cup (S)$" המייצג את האיחוד של קבוצה זו, בעוד "$\cap (S)$" ביטוי המייצג את החיתוך שלה.

את ההגדרות אפשר לסכם כך:

$$\cup (A) = \{ x \mid \exists\ z \in A \,.\, x \in z\}$$

$$\cap (A) = \{ x \mid \forall z \in A \,.\, x \in z\}$$

יש לשים לב, שהביטוי $\cap (F)$ אינו מוגדר כש- F קבוצה ריקה. זאת, כיוון שלפי ההגדרה הפורמלית לעיל, $\cap (\varnothing)$ היה צריך להיות אוסף כל העצמים, ואוסף זה אינו קבוצה. לעומת זאת אם $B \in F$, אז :

$$\cap (F) = \{x \in B \mid \forall z \in F \,.\, x \in z\}$$

ולכן $\cap (F)$ הינו קבוצה, כש- F אינה ריקה (לפי עיקרון הקומפרהנסיה המוגבל). לעומת זאת, כדי להבטיח ש- $\cup (F)$ הינו קבוצה אנו זקוקים לעקרון מיוחד:

VI. עיקרון האיחוד (נוסח מלא):

אם F היא קבוצה אז כך גם האיחוד שלה[10].

בניסוח המתייחס לשפה ננסח את העיקרון כך:
אם A ביטוי המייצג קבוצה, אז כך גם הביטוי $\{x \mid \exists z \in A.\, x \in z\}$, דהיינו: $\cup (A)$.

דוגמאות:

(1) נניח A ו- B קבוצות, ותהי $F = \{A, B\}$. אז:

$\cup (F) = A \cup B$ ו- $\cap (F) = A \cap B$ (לבדוק!) .

פעולות האיחוד והחיתוך הכלליות, שהגדרנו כאן, הן לכן הכללה של הפעולות, שהגדרנו קודם.

(2) $$\cup (\{ \{n\} \mid n \in \mathbf{N}\}) = \mathbf{N}$$

הוכחה:

לכיוון האחד, נניח $x \in \mathbf{N}$. אז $x \in \{x\}$, ו- $\{x\} \in \{\{n\} \mid n \in \mathbf{N}\}$. לכן :

$$\exists z \,.\, z \in \{\{n\} \mid n \in \mathbf{N}\} \wedge x \in z \qquad (z = \{x\})$$

---

[10] קורא קשוב שם לב, אני מקווה, שהושמט כאן התנאי ש- F היא קבוצה של קבוצות. למעשה, הן ההגדרות והן העיקרון ישימים גם כש- F היא קבוצה כלשהי, לאו דווקא "משפחה". באופן מעשי משתמשים בהם, אבל, רק במקרה המשפחתי.

כלומר:

$$x \in \cup(\{\{n\} \mid n \in \mathbf{N}\})$$

לכיוון ההפוך, נניח $x \in \cup(\{\{n\} \mid n \in \mathbf{N}\})$ אז יש z ב- $\{\{n\} \mid n \in \mathbf{N}\}$ כך ש- $x \in z$. z זה הוא מהצורה $\{n\}$ עבור איזה $n \in \mathbf{N}$. מזה נובע, שקיים n ב- $\mathbf{N}$ כך ש- $x \in \{n\}$. אבל $x \in \{n\}$ פירושו ש- $x = n$. ממילא, אם $n \in \mathbf{N}$, אז גם $x \in \mathbf{N}$.

(3) $$\bigcup\left(\left\{\{x \in \mathbf{R} \mid \frac{1}{n} \le x \le 1 - \frac{1}{n}\} \mid n \in \{k \in \mathbf{N} \mid k \ge 2\}\right\}\right) = \{x \in \mathbf{R} \mid 0 < x < 1\}$$

$$\bigcap\left(\left\{\{x \in \mathbf{R} \mid \frac{1}{n} \le x \le 1 - \frac{1}{n}\} \mid n \in \{k \in \mathbf{N} \mid k \ge 2\}\right\}\right) = \{\tfrac{1}{2}\}$$

הבנת דוגמה זו הינה תרגיל טוב בהבנת הנקרא בשפה, אותה אנו לומדים. הבה נסביר אותה בפרוטרוט. ובכן מדובר פה ב- $\cup(F)$ וב- $\cap(F)$, כאשר:

$$F = \left\{\{x \in \mathbf{R} \mid \frac{1}{n} \le x \le 1 - \frac{1}{n}\} \mid \mathrm{n} \in \{\mathrm{k} \in \mathbf{N} \mid \mathrm{k} \ge 2\}\right\}$$

עתה לביטוי F הצורה, שמרשה עיקרון ההחלפה, כלומר צורתו היא $\{t \mid n \in A\}$, כאשר:

$$t = \{\mathrm{x} \in \mathbf{R} \mid \frac{1}{n} \le x \le 1 - \frac{1}{n}\}$$

$$A = \{k \in \mathbf{N} \mid k \ge 2\}$$

A עצמה היא קבוצה לפי עיקרון הקומפרהנסיה המוגבל. t הינו ביטוי, שבו n מופיע כמשתנה חופשי (הוא לא חופשי ב- F כולו, כמובן). לכל $n \in \mathbf{N}$ ספציפי t מגדיר קבוצה (שוב, לפי עיקרון הקומפרהנסיה המוגבל). קבוצה זו מסמנים, כזכור, בצורה: $[\frac{1}{n}, 1 - \frac{1}{n}]$. מכאן שהביטוי F אכן מיצג קבוצה לפי עקרון ההחלפה, ו- $\cup(F)$ מוגדר אפוא. הקבוצה שהוא מיצג היא קבוצת כל המספרים הממשיים השיכים לקטע כלשהו מהצורה: $[\frac{1}{n}, 1 - \frac{1}{n}]$, כש- $n \in \{k \in \mathbf{N} \mid k \ge 2\}$ (כלומר: כש- $n \ge 2$). כיוון שלכל מספר x בין 0 ל- 1 (לא כולל 0 ו- 1), אפשר למצוא $\mathrm{n} \ge 2$ כך ש- $\frac{1}{n} \le x \le 1 - \frac{1}{n}$, ורק למספרים אלו תכונה זו, $\cup(F)$ הינו הקטע הפתוח $(0,1)$ (כלומר: $\{x \in \mathbf{R} \mid 0 < x < 1\}$). את העובדה ש- $\cap(F) = \{\frac{1}{2}\}$ נשאיר כתרגיל לקוראים.

הערה:

קבוצה כמו F בדוגמה האחרונה מתוארת בדרך-כלל בספרות כך :

$$F = \{[\frac{1}{n}, 1 - \frac{1}{n}] \mid n \in \mathbf{N} \wedge n \geq 2\}$$

יש כאן שימוש בסימון המקוצר המקובל עבור קטעים על הישר הממשי, יחד עם שילוב של הסימונים עבור עקרונות ההחלפה והקומפרהנסיה. השילוב מרשה לרשום: $\{t \mid x \in A \wedge \varphi\}$ במקום: $\{t \mid x \in \{x \in A \mid \varphi\}\}$.

סימונים מקובלים נוספים :

1. כאשר F סופית, $F = \{A_1, ..., A_n\}$, אז במקום $\cup(F)$ מקובל לכתוב $A_1 \cup A_2 \cup \ldots \cup A_n$. למשל: $\cup(\{A, B, C, D\}) = A \cup B \cup C \cup D$. בדומה, במקום $\cap(F)$ כותבים $A_1 \cap A_2 \cap \ldots \cap A_n$. קל לראות שסימון זה מתאים לפירוש הקודם, שנתנו לביטוי $A_1 \cup A_2 \cup \ldots \cup A_n$.
2. כמו שכבר ראינו בדוגמאות (2) ו- (3), ברוב המקרים בהם עוסקים ב"משפחה" של קבוצות, ובמיוחד כשעושים חיתוך או איחוד של משפחה כזו, המשפחה מתוארת לפי עיקרון ההחלפה. במלים אחרות: היא מיוצגת בצורה $\{A(\mathrm{i}) \mid i \in I\}$. במקרה זה מקובל לכתוב $\underset{i \in I}{\cup} A(i)$ במקום $\cup(\{A(i) \mid i \in I\})$. לחיתוך יש סימון דומה.

   למשל:

   בדוגמא מספר 2 כותבים: $\bigcup_{n \in N} \{n\}$ (זה שווה, כזכור, ל- **N**).

   בדוגמא מספר 3 כותבים: $\bigcup_{n \in \{k \in \mathbf{N} | k \geq 2\}} \left\{x \in \mathbf{R} \,|\, \frac{1}{n} \leq x \leq 1 - \frac{1}{n}\right\}$

   או אפילו (בעזרת (β)): $\bigcup_{n \geq 2} \left\{x \in \mathbf{R} \,|\, \frac{1}{n} \leq x \leq 1 - \frac{1}{n}\right\}$

   ובעוד יותר קיצור: $\bigcup_{n \geq 2} \left[\frac{1}{n}, 1 - \frac{1}{n}\right]$

המקרים הנפוצים ביותר הם כש- $I = \{1, 2, ..., n\}$ או כש- $I = \{n \in \mathbf{N} \mid n \geq k\}$ לאיזה k טבעי. במקרים אלו כותבים לעתים $\bigcup_{i=1}^{n} A(i)$ ו- $\bigcup_{i=k}^{\infty} A(i)$ (בהתאמה). בסימון דומה משתמשים עבור $\cap$.

את הדוגמא השלישית, למשל, נכתוב בדרך כלל כך:

$$\bigcup_{n=2}^{\infty}\left[\frac{1}{n},1-\frac{1}{n}\right]=(0,1)\ ,\quad \bigcap_{n=2}^{\infty}\left[\frac{1}{n},1-\frac{1}{n}\right]=\left\{\frac{1}{2}\right\}$$

כאשר משתמשים בסימונים הנ"ל יש לזכור את שתי העובדות הבסיסיות הבאות:

$$\boxed{\begin{array}{lcl} x\in\bigcup_{i\in I}A_i & \Leftrightarrow & \exists\, i\in \mathrm{I}\,.\,x\in A_i \\ x\in\bigcap_{i\in I}A_i & \Leftrightarrow & \forall\, i\in \mathrm{I}\,.\,x\in A_i \end{array}}$$

(הלכנו כאן בעקבות הנוהג המקובל בכל הטקסטים וכתבנו A_i במקום $A(i)$ בכתיבה זו קוראים ל- "i" שב- A_i "אינדקס" ול- I "קבוצת אינדקסים". אין למונחים אלו שום חשיבות, למעשה).

%% הערה:

אם נתרגם את $x\in\bigcup_{i\in I}A_i$ (למשל) לפי ההגדרות, באופן שיטתי, נקבל משהו מסובך יותר:

$$x\in\bigcup_{i\in I}A_i \;\Leftrightarrow\; \exists Y(x\in Y \wedge \exists i\in I.Y=A_i)$$

זה שקול למה שכתוב למעלה לפי כללים לוגיים, הקשורים בשוויון, אותם לא ניסחנו במפורש בקורס זה. עם זאת, די בחשיבה אינטואיטיבית כדי להבין שקילות זו. למעשה, אפשר להבין את נכונות העקרונות שבמסגרת למעלה ישירות מהגדרת "איחוד של משפחה של קבוצות" ו "חיתוך של משפחת קבוצות", כמו שניסחנו אותם בעברית.

%%

תכונות של הפעולות המוכללות:

הפעולות המוכללות של איחוד וחיתוך מקיימות הכללות טבעיות של כמה מהכללים בטבלה מספר ב.1. הכללות אלו מסוכמות בטבלה מספר ב.2. המספרים בה תואמים את אלה שבטבלה מספר ב.1. הכללים מתוארים בצורה המקובלת בספרות (שהיא השימושית יותר). בתוך הסוגרים המרובעים נמצאת אבל הצורה הבסיסית יותר של הכלל. אנו נוכיח, לדוגמה, את כלל 9a:

$$A\cap(\bigcup_{i\in I}B_i)=\bigcup_{i\in I}(A\cap B_i)$$

$(\subseteq)$:

נניח $x \in A \cap (\bigcup_{i \in I} B_i)$. אז $x \in A$ ו- $x \in \bigcup_{i \in I} B_i$. מכאן שקיים $i \in I$ כך ש- $x \in B_i$. עבור אותו i , $x \in A \cap B_i$, ולכן $x \in \bigcup_{i \in I} (A \cap B_i)$.

$(\supseteq)$:

נניח $x \in \bigcup_{i \in I} (A \cap B_i)$. אז קיים $i \in I$ כך ש- $x \in A \cap B_i$. מכאן ש- $x \in A$ ו- $x \in B_i$. מהעובדה האחרונה נובע, ש- $x \in \bigcup_{i \in I} B_i$. מזה ומ- $x \in A$ נובע ש- $x \in A \cap (\bigcup_{i \in I} B_i)$.

טבלה ב.2: תכונות יסודיות של האופרציות המוכללות

1	$\forall j \in I.\ \bigcap_{i \in I} A_i \subseteq A_j$	$[\ \forall X \in F.\ \cap(F) \subseteq X\ ]$
2	$\forall j \in I.\ A_j \subseteq \bigcup_{i \in I} A_i$	$[\ \forall X \in F.\ X \subseteq \cup(F)\ ]$
5a	$C \subseteq \bigcap_{i \in I} B_i \Leftrightarrow \forall i \in I.\ C \subseteq B_i$	$[\ C \subseteq \cap(F) \Leftrightarrow \forall X \in F.\ C \subseteq X\ ]$
5b	$\bigcup_{i \in I} B_i \subseteq C \Leftrightarrow \forall i \in I.\ B_i \subseteq C$	$[\cup(F) \subseteq C \Leftrightarrow \forall X \in F.\ X \subseteq C\ ]$
8a	$(\bigcap_{i \in I_1} A_i) \cap (\bigcap_{i \in I_2} A_i) = \bigcap_{i \in I_1 \cup I_2} A_i$	$[(\cap(F_1)) \cap (\cap(F_2)) = \cap(F_1 \cup F_2)\ ]$
8b	$(\bigcup_{i \in I_1} A_i) \cup (\bigcup_{i \in I_2} A_i) = \bigcup_{i \in I_1 \cup I_2} A_i$	$[\ (\cup(F_1)) \cup (\cup(F_2)) = \cup(F_1 \cup F_2)\ ]$
9a	$A \cap (\bigcup_{i \in I} B_i) = \bigcup_{i \in I} (A \cap B_i)$	$[\ A \cap (\cup F) = \cup(\{A \cap X \mid X \in F\})\ ]$
9b	$A \cup (\bigcap_{i \in I} B_i) = \bigcap_{i \in I} (A \cup B_i)$	$[\ A \cup (\cap F) = \cap(\{A \cup X \mid X \in F\})\ ]$
13a*	$\overline{\bigcap_{i \in I} A_i} = \bigcup_{i \in I} \overline{A_i}$	$[\overline{\cap(F)} = \cup(\{\overline{X} \mid X \in F\})\ ]$
13b*	$\overline{\bigcup_{i \in I} A_i} = \bigcap_{i \in I} \overline{A_i}$	$[\overline{\cup(F)} = \cap(\{\overline{X} \mid X \in F\})\ ]$

(ד) קבוצות זרות, איחוד זר

הגדרה:

קבוצות A ו- B נקראות זרות זו לזו (Disjoint) אם אין להן שום איבר משותף, דהיינו אם $A \cap B = \varnothing$.

סימון:

כאשר A ו- B זרות מקובל (אך לא הכרחי) לסמן את האיחוד שלהן בצורה הבאה: $A \uplus B$. קוראים לזה אז ***"איחוד זר"***. יש להדגיש כי אין הבדל בין ערך הביטוי $A \uplus B$ (כשהוא מוגדר) ובין ערך $A \cup B$. השימוש ב- $\uplus$ בא רק לציין שאנו יודעים (או מניחים לצורך החישוב) ש- A ו- B זרות זו לזו. למעשה, $\cup$ ו- $\uplus$ שונים זה מזה רק בתחום ההגדרה שלהם: הביטוי $A \cup B$ מוגדר עבור כל שתי קבוצות A ו- B, בעוד ש- $A \uplus B$ מוגדר רק כאשר A ו- B קבוצות זרות.

את המושג והסימון ניתן (ומקובל) להכליל גם לאיחוד של יותר משתי קבוצות. כללית, הביטוי $\biguplus_{i \in I} A_i$ מוגדר אם (ורק אם):

$$\forall i \in I \ \ \forall j \in I \ . \ A_i \cap A_j = \varnothing$$

(דהיינו – אם איברי הקבוצה $\{A_i | \, i \in I\}$ זרים זה לזה בזוגות). כאשר $\biguplus_{i \in I} A_i$ מוגדר, אין ערכו שונה מזה של $\bigcup_{i \in I} A_i$.

דוגמאות:

$\biguplus_{x \in Z} [x, x+1) = \mathbf{R}$, בעוד $\biguplus_{x \in Z} [x, x+1]$ הוא ביטוי בלתי מוגדר, כיוון ש(לדוגמה): $[2,3] \cap [3,4] \neq \varnothing$.

(ה) זוגות סדורים, מכפלה קרטזית

הגדרה:

הזוג הסדור של a ו- b הוא הקבוצה $\{\{a\},\{a,b\}\}$, והוא מסומן ב $<a,b>$ (בטקסטים אחדים כותבים (a,b), אך איננו רוצים ליצור כאן בלבול עם הסימן המקובל לקטעים פתוחים על הישר הממשי).

התכונה העיקרית של זוגות סדורים ניתנת על ידי הטענה הבאה:

טענה:

$$<a,b> = <c,d> \Leftrightarrow a = c \wedge b = d$$

הוכחה: תרגיל.

מהותו של הזוג סדור $<a,b>$ הינה, שהוא יוצר עצם חדש משני עצמים, שבו הסדר של שני העצמים הללו הינו ***חשוב***. זהו ההבדל בינו ובין הזוג הלא סדור $\{a,b\}$ (שגם הוא עצם הנוצר על-ידי קומבינציה של שני עצמים). לגבי זוגות לא סדורים $\{a,b\} = \{b,a\}$, לעומת זאת, $<a,b> \neq <b,a>$ (אלא אם $a = b$, כמובן).

הגדרה:

המכפלה הקרטזית של שתי קבוצות A ו- B (סימון $A \times B$) היא קבוצת כל הזוגות הסדורים $<a,b>$ כך ש- $a \in A$ ו- $b \in B$.

$$A \times B = \{ z \mid \exists a \in A \, \exists b \in B . z = <a,b> \}$$

בעיה:
האם נוכל להראות ש- $A \times B$ אכן קבוצה, כאשר ידוע ש- A ו- B קבוצות?

התשובה חיובית:

ברור שאם $a \in A$ ו- $b \in B$ אז $a \in A \cup B$ ו- $b \in A \cup B$.

לכן $\{a\} \subseteq A \cup B$ ו- $\{a,b\} \subseteq A \cup B$. מכאן ש- $\{a\} \in P(A \cup B)$ ו- $\{a,b\} \in P(A \cup B)$.

לכן $\{\{a\},\{a,b\}\} \subseteq P(A \cup B)$ וממילא $<a,b> \in P(P(A \cup B))$. מכאן ש:

$$A \times B = \{ z \in P(P(A \cup B)) \mid \exists a \in A \, \exists b \in B . z = <a,b> \}$$

ולכן $A \times B$ קבוצה לפי עיקרון הקומפרהנסיה המוגבל (יחד עם עיקרון האיחוד ועיקרון קבוצת החזקה).

אזהרה: בדרך-כלל $A \times B \neq B \times A$ ו- $A \times (B \times C) \neq (A \times B) \times C$.

סימונים נוספים:

1. כאשר $A = B$ מקובל לכתוב A^2 במקום $A \times A$. כך לדוגמה: $<-3, \pi> \in \mathbf{R}^2$.

2. במקום ביטוי מהצורה: $\{z \mid \exists x \exists y.\ z = <x,y> \wedge \varphi\}$
מקובל לכתוב את הדבר הבא: $\{<x, y> \mid \varphi\}$.
בצורת כתיבה זו קושר האופרטור $\{\ \ldots \mid \ldots\ \}$ בבת אחת גם את x וגם את y !
כלל (β) מקבל כאן את הצורה:

$$<t, s> \in \{<x, y> \mid \varphi\} \Leftrightarrow \varphi(t/x, s/y)$$

(ובלבד ש- t חופשי להצבה במקום x ב- φ ו- s חופשי שם להצבה במקום y).
קבוצות חלקיות ל- $A \times B$ המוגדרות לפי עיקרון הקומפרהנסיה המוגבל מיוצגות בדרך כלל על ידי ביטויים מהצורה:

$$\{<x, y> \in A \times B \mid \varphi\}$$

זהו כמובן קיצור של:

$$\{z \in A \times B \mid \exists x \exists y.\ z = <x, y> \wedge x \in A \wedge y \in B \wedge \varphi\}$$

לכן:

$$<t, s> \in \{<x, y> \in A \times B \mid \varphi\}$$

אם ורק אם: $t \in A \wedge s \in B \wedge \varphi(t/x, s/y)$.

דוגמה:

קבוצת הנקודות במישור, הנמצאות בפנים עיגול היחידה (העיגול שמרכזו בראשית הצירים ורדיוסו 1) מיוצגת בגיאומטריה אנליטית כך:

$$\{<x, y> \in \mathbf{R}^2 \mid x^2 + y^2 < 1\}$$

מושג חשוב נוסף הקשור בזוגות סדורים הוא מושג ה*היטל* :

הגדרה:

הפעולות π_1 ו- π_2 על זוגות סדורים מוגדרות על ידי הנוסחאות:

$$\pi_2(<a, b>) = b \qquad\qquad \pi_1(<a, b>) = a$$

פעולות אלו מספקות את ה***היטלים*** של הזוג $<a, b>$. כדאי לציין, שאם z הינו זוג סדור, אז $z = <\pi_1(z), \pi_2(z)>$.

הכללות:

1. ***n*-יות** – אלו מוגדרות, באופן כללי, בצורה הבאה:

$$< a_1, \ldots, a_n > = < a_1, < a_2 < \ldots < a_{n-1}, a_n > \ldots >>>$$

למשל הרביעייה $<a,b,c,d>$ מוגדרת כך:

$$< a,b,c,d > = < a, < b, < c,d >>>$$

התכונה החשובה של n-יות היא, שוב, ש:

$$<a_1, \ldots, a_n> = <b_1, \ldots, b_n> \Leftrightarrow a_1 = b_1 \wedge a_2 = b_2 \wedge \ldots \wedge a_n = b_n$$

(תרגיל: להוכיח זאת באינדוקציה על n, תוך שימוש במקרה $n = 2$).

2. אם $A_1, A_2, \ldots, A_n$ הינן קבוצות, אז ***מכפלתן הקרטזית***: $A_1 \times A_2 \times \ldots \times A_n$ היא קבוצת כל ה- ***n*-יות** $<a_1,\ldots, a_n>$, כך ש-

$$a_1 \in A_1 \wedge a_2 \in A_2 \wedge \ldots \wedge a_n \in A_n$$

תרגיל: $A_1 \times A_2 \times \ldots \times A_n = A_1 \times (A_2 \times (\ldots \times A_n)) \ldots)$

גם כאן כותבים A^n כאשר $A_1 = A_2 = \ldots = A_n = A$.

3. את הסימון $\{<x, y> \mid \varphi\}$ אפשר להכליל ל- ***n*-יות** כך:

$$\{<x_1, \ldots, x_n> \mid \varphi\}$$

או בשימוש בעיקרון הקומפרהנסיה המוגבל:

$$\{<x_1, \ldots, x_n> \in A_1 \times \ldots \times A_n \mid \varphi\}$$

נשאיר לקוראים את ניסוח הגרסה המתאימה של כלל (β) .

4. פונקציות ההיטל π_i^n מוגדרות בצורה הבאה:

$$\pi_i^n(<a_1, \ldots, a_n>) = a_i$$

למשל: $\pi_2^5(<-3, 2, \varnothing, \{\varnothing\}, e>) = 2$

טבלה מספר ב.3 מסכמת את הסימונים השונים עבור קבוצות, שלמדנו עד כה (כן נכללים בה אי-אלו סימונים נוספים, אותם נלמד בהמשך).

טבלה ב.3: סימונים עבור קבוצות

I. <u>שמות מיוחדים (וקבועים)</u>: $\mathbf{N}$, $\mathbf{N}^+$, $\mathbf{Z}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$, $\varnothing$

II. <u>קבוצות סופיות מפורטות</u>: $\{t_1, t_2, \ldots, t_n\}$

III. <u>עיקרון הקומפרהנסיה המוגבל</u>: $\{x \in S \mid \varphi\}$

IV. <u>עיקרון ההחלפה</u>: $\{t \mid x \in S\}$

V. <u>קבוצת החזקה</u>: $P(S)$ או 2^S

VI. <u>פעולות בינאריות</u>: $A \cup B$ איחוד, $A \cap B$ חיתוך, $A - B$ הפרש, $A \Delta B$ הפרש סימטרי, $A \times B$ מכפלה קרטזית

VII. <u>משלים יחסי</u>: $\overline{A}_E$ או A_E^c ולפעמים פשוט $\overline{A}$ או A^c

VIII. <u>הכללת הפעולות הבינאריות</u>:

$\bigcup (A)$ $\quad$ $\bigcap (A)$

$\bigcap\limits_{\alpha \in I} A_\alpha$ $\quad$ $\bigcup\limits_{\alpha \in I} A_\alpha$

סימונים נוספים, שיילמדו בהמשך:

$A \to B$ או B^A – קבוצת הפונקציות מ- A ל- B

$A/\equiv$ – קבוצת המנה של A ביחס ל- $\equiv$

ב.4. פונקציות

I. מושג הפונקציה

בפרק הזה נדון באחד המושגים הבסיסיים ביותר של המתמטיקה ושל מדעי המחשב: מושג הפונקציה. זהו, ללא ספק, אחד הפרקים החשובים ביותר בספר זה.

הגדרה

(1) ***פונקציה*** מעל קבוצה A היא התאמה, המתאימה לכל איבר x ב- A עצם אחד ויחיד. עצם זה נקרא ***ערך הפונקציה*** ב- x.

(2) ***פונקציה חלקית מעל קבוצה*** A היא פונקציה מעל תת-קבוצה כלשהי של A.

הערות

1. בעצם, אין אלו "הגדרות" ממש, כי מושג ה"התאמה", המופיע בהן, אינו בהיר הרבה יותר (אם בכלל) ממושג הפונקציה עצמו. יש לראותן יותר בגדר הסבר. מאוחר יותר ניתן הגדרה פורמלית במונחי תורת הקבוצות בלבד.

2. ברור מה"הגדרה", שכל פונקציה מעל A היא גם פונקציה חלקית מעל A. לפעמים מדגישים את ייחודן של הפונקציות מעל A על-ידי זה, שמכנים אותן "פונקציות טוטאליות" (מלאות) מעל A. אצלנו לא יהיה אבל הבדל בין המושגים של "פונקציה" ושל "פונקציה מלאה".

3. עוד ברור מה"הגדרה", כי מושג הפונקציה (מלאה או חלקית) כולל שני מרכיבים עיקריים:

   (א) הקבוצה עליה הפונקציה מוגדרת. קבוצה זו נקראת ***התחום*** של הפונקציה.
   (ב) ההתאמה עצמה, שבדרך-כלל ניתנת על-ידי ***כלל התאמה*** כלשהו.

סימונים

1. את תחום ההגדרה של פונקציה (מלאה או חלקית) f מסמנים ב- $\mathrm{Dom}(f)$.

2. את ערך הפונקציה f עבור איבר t של תחום ההגדרה מסמנים בדרך-כלל בצורה $f(t)$ (באי-אלו שפות מחשב, כמו LISP ו- SCHEME, כותבים דווקא (f, t)). בביטויים מהצורה $f(t)$ מקובל לקרוא ל- t ה"ארגומנט" של הביטוי, ול- $f(t)$ כולו "ערך הפונקציה f עבור הארגומנט t".

אינפורמציה חשובה נוספת על פונקציות (מלאות/חלקיות) היא מאין הן לוקחות את הערכים שלהן.

הגדרה

(1) תהי f פונקציה. קבוצה B נקראת ***טווח*** של f אם מתקיים:

$$\forall x \in \mathrm{Dom}(f).\, f(x) \in B$$

(כלומר, B כוללת את כל הערכים ש-f "מקבלת").

(2) פונקציה חלקית מעל A, ש- B היא טווח שלה, נקראת ***פונקציה (חלקית) מ- A ל- B***.

סימונים

1. הסימון המקובל לכך, ש- "f היא פונקציה מ- A ל- B" הוא $f: A \to B$.
2. בעקבות מספר שפות מחשב מתקדמות (כמו SML), נסמן אנו את קבוצת הפונקציות מ- A ל- B ב- $A \to B$. עם סימון זה אין הבדל, למעשה, בין $f: A \to B$ ובין $f \in A \to B$.[11]
סימון מקובל אחר בספרות עבור $A \to B$ (שגם אנו נשתמש בו לעתים) הוא B^A.
3. אנו נסמן את קבוצת הפונקציות החלקיות מ- A ל- B ב- $A \xrightarrow{p} B$. סימון זה הוא עבור ספר זה בלבד, ואין לו בסיס בטקסטים אחרים.

הערות

(1) $A \to B$ ו- $A \xrightarrow{p} B$ הן אכן קבוצות כאשר A ו- B קבוצות. דבר זה יתברר בהמשך. (כמו כן נובע גם מהאקסיומות של תורת הקבוצות, שלכל פונקציה יש טווח. זהו, למעשה, ניסוח אלטרנטיבי לעיקרון ההחלפה, אך לא ניכנס לפרטים).

---

[11] באותן שפות מחשב משתמשים גם ב- ":" במקום ב- "$\in$". כך כותבים, למשל, 3:Int במקום $3 \in \mathbf{Z}$. לפיכך, משמעות "$f: A \to B$" בשפות אלו היא *בדיוק* זו של תורת הקבוצות (אם כי היסטורית, סימון זה קדם בהרבה לסימונים ולמושגים של תורת הקבוצות).

(2) כדאי לשים לב, שלא הגדרנו את מושג הטווח של פונקציה f, אלא "טווח" סתם. בעניין זה, *אין* ספר זה תואם את ההגדרות הרשמיות במרבית הטקסטים, אבל הוא תואם מאוד את האופן בו משתמשים *באמת* במושגים. כך, למשל, מעטים המתמטיקאים, אם בכלל, שירצו באמת ובתמים להבדיל בין "שתי" הפונקציות הבאות:

$f : \mathbf{Z} \to \mathbf{Z}$ המוגדרת על-פי הכלל $f(z) = z^2$.

$g : \mathbf{Z} \to \mathbf{N}$ המוגדרת על-פי הכלל $g(z) = z^2$.

למרות זאת, נדבר גם אנו לפעמים על ה"טווח" של פונקציה מסוימת. זה יקרה כשחשוב יהיה לציין תחום מסוים, ממנו היא לוקחת את ערכיה.

בין הטווחים הרבים של פונקציה יש אחד, שיש לו חשיבות מיוחדת. זהו ה*מינימלי* ביניהם:

הגדרה

ה*תמונה* של פונקציה f היא הקבוצה הבאה:

$$\mathrm{Im}(f) = \{f(x) \mid x \in \mathrm{Dom}(f)\}$$

דהיינו:

$$\mathrm{Im}(f) = \{y \mid \exists x \in \mathrm{Dom}(f).\ y = f(x)\}$$

הערות

(1) ברור שאם $f : A \to B$, אז $\mathrm{Im}(f) = \{y \in B \mid \exists x \in A.\ y = f(x)\}$.

(2) קבוצה B היא טווח של פונקציה f אם"ם $\mathrm{Im}(f) \subseteq B$. בפרט $\mathrm{Im}(f)$ היא טווח של f.

דוגמאות

(1) נגדיר $f : \mathbf{R} \to \mathbf{R}$ על-ידי $f(x) = x^2$. ההצהרה "$f : \mathbf{R} \to \mathbf{R}$" אומרת כאן, שהתחום של הפונקציה הוא $\mathbf{R}$, ושהפונקציה מקבלת ערכים ממשיים בלבד. "$f(x) = x^2$" נותן את כלל ההתאמה, והוא אכן קומפטיבילי עם ההצהרה הקודמת: ניתן להפעיל כלל זה על כל מספר ממשי והתוצאה תהיה אכן מספר ממשי (יש לבדוק דברים כאלה!). לאותו f נכון גם ש- $f : \mathbf{R} \to \{x \in \mathbf{R} \mid x \geq -1\}$, כיוון שכמו $\mathbf{R}$, גם $\{x \in \mathbf{R} \mid x \geq -1\}$ הוא טווח אפשרי של הפונקציה f. התמונה של f, לעומת זאת, היא $\{x \in \mathbf{R} \mid x \geq 0\}$. ברור לכן, שמתקיים גם $f : \mathbf{R} \to \{x \in \mathbf{R} \mid x \geq 0\}$.

(2) נגדיר $g:\mathbf{R}\overset{p}{\to}\mathbf{R}$ על-ידי הכלל $g(x)=\frac{1}{x}$. זוהי אכן רק פונקציה חלקית מ- $\mathbf{R}$ ל- $\mathbf{R}$, כי $g(0)$ אינו מוגדר. התחום של g הינו $\{x\in\mathbf{R}\mid x\neq 0\}$. לכן $g:\{x\in\mathbf{R}\mid x\neq 0\}\to\mathbf{R}$ (כאן כבר מדובר בפונקציה מלאה, אם כי עדיין נכון גם ש- $g:\{x\in\mathbf{R}\mid x\neq 0\}\overset{p}{\to}\mathbf{R}$). גם התמונה של g היא $\{x\in\mathbf{R}\mid x\neq 0\}$.

הערות

(1) שתי הדוגמאות היו של פונקציות ממשיות, דהיינו: פונקציות מקבוצות חלקיות של $\mathbf{R}$ אל $\mathbf{R}$. היסטורית, מושג הפונקציה התפתח מפונקציות כאלו, והן אלו העומדות במרכז החשבון הדיפרנציאלי והאינטגרלי. נדגיש אבל כבר כאן, שכל ***קבוצה יכולה להיות תחום של פונקציה, וכל קבוצה יכולה לשמש כטווח.*** עוד מעט נראה אכן דוגמאות רבות של פונקציות, שאינן מ- $\mathbf{R}$ ל- $\mathbf{R}$.

(2) בבעיות של "חקירת פונקציות" בחדו"א נותנים למעשה כלל התאמה, המגדיר פונקציה חלקית מ- $\mathbf{R}$ ל- $\mathbf{R}$. חקירת הפונקציה החלקית נפתחת במציאת תחום ההגדרה של פונקציה חלקית זו.

(3) תמיד $A\to B\subseteq A\overset{p}{\to}B$, ולמעשה $A\to B\subset A\overset{p}{\to}B$ (אלא אם כן $A=\varnothing$, אך במקרה זה כדאי לדון רק אחרי שניתן את ההגדרה הרשמית של פונקציות).

(4) אם $f:A\overset{p}{\to}B$ אז $f:\mathrm{Dom}(f)\to B$ (ומתקיים, כמובן, $\mathrm{Dom}(f)\subseteq A$).

שוויון פונקציות

העיקרון לגבי שוויון פונקציות דומה לזה של שוויון קבוצות: ממש כמו שזהות קבוצה אינה תלויה באופן בו מגדירים אותה אלא רק בזהות איבריה, כך גם זהות פונקציה אינה תלויה באופן בו אנו מתארים את כלל ההתאמה, אלא רק בזהות תחומה והערכים שהיא מקבלת. מכאן מקבלים אנו את ***עיקרון האקסטנציונליות עבור פונקציות***:

$$\boxed{f=g \iff \mathrm{Dom}(f)=\mathrm{Dom}(g)\wedge \forall x\in\mathrm{Dom}(f).\, f(x)=g(x)}$$

דוגמה טריוויאלית: אם התחום של f ו- g הוא $\mathbf{R}$ ו-

$$g(x)=x^2+2x+1 \qquad f(x)=(x+1)^2$$

אז $f=g$.

II. הגדרת פונקציות וסימונן

הדרכים לסימון פונקציות ולהגדרתן דומות מאוד לאלו הנהוגות לגבי קבוצות, אותן ראינו בפרק ב.2. הבה נסקור אותן.

שיטת סימון מס' 1: שמות קבועים

לפונקציות חשובות רבות יש שמות *קבועים* מיוחדים. המספר כאן גדול לאין ערוך מזה של הקבוצות בעלי שמות מיוחדים, ולא ננסה כלל לתת רשימה מלאה. הנה דוגמאות: $\sin$, tg, $\sqrt{\ }$, $|\,|$, $!$, t. כיוון שסימונים אלו הוכנסו לשימוש לפני פיתוח התיאוריה הכללית, לעתים קרובות אין משתמשים בהם באופן הסטנדרטי. כך כותבים אנו $\sqrt{9}$ ולא $\sqrt{\ }(9)$ $(\sqrt{\ }: \{x \in \mathbf{R} \mid x \geq 0\} \to \mathbf{R})$, $7!$ ולא $!(7)$ $(! : \mathbf{N} \to \mathbf{N})$, $|-3|$ ולא $|\,|(-3)$ $(|\,| : \mathbf{C} \to \mathbf{R})$, $\begin{bmatrix} 1 & 0 \\ 3 & 5 \end{bmatrix}^t$ ולא $t\left(\begin{bmatrix} 1 & 0 \\ 3 & 5 \end{bmatrix}\right)$ (הפונקציה t המתאימה לכל מטריצה את המטריצה המוחלפת שלה, והיא פונקציה מקבוצת כל המטריצות הממשיות (נאמר) אל קבוצת כל המטריצות הממשיות). כדאי לציין, שדבר זה מתוקן באי-אלו שפות מחשב, בהן כותבים, למשל, SQRT(9) במקום $\sqrt{9}$ או fact(9) במקום $7!$.

שיטת סימון מס' 2: שמות זמניים

הדרך הסטנדרטית ביותר בטקסטים להתייחס אל פונקציות, שאין להן שם קבוע, היא על-ידי מתן שמות ***זמניים***. בדרך זו נקטנו גם אנו בשתי הדוגמאות הראשונות, שהבאנו למעלה ($f(x) = x^2$, $g(x) = 1/x$). השמות הפופולריים ביותר הם f, g, h, F, G, H (אולי בתוספת אינדקסים: f_1, g_2). הניסוח הסטנדרטי הוא: "תהי $f: A \to B$ מוגדרת על-ידי $f(x) = \ldots$ ". נזכיר, שבצורת הגדרה זו המשתנה x הינו קשור על-ידי כמת אוניברסלי סמוי $\forall$, ויש להשתמש בהגדרה כאילו כתוב: "$\forall x \in A.\ f(x) = \ldots$" (ההקבלה בקבוצות: הגדרת קבוצה A על-ידי: "$\forall x.\ x \in A \Leftrightarrow \varphi(x)$").

שיטת סימון מס' 3: טבלאות

לגבי קבוצות ***סופיות***, ראינו את הדרך הפשוטה של ***פירוט*** איבריהן, כמו $\{1,2,3,4\}$. דרכי פירוט דומות קיימות לגבי פונקציות, אם תחום הפונקציה סופי. שימושית וקומפקטית במיוחד היא דרך התיאור בעזרת ***טבלה***. לדוגמה, פונקציה, שהתחום שלה

הוא קבוצת הסטודנטים ב"מתמטיקה בדידה" והטווח – המספרים הטבעיים בין 0 ל- 100 [12], מתוארת בעזרת טבלה כך:

סטודנט	ציון
סטיבן הנדרי	85
ג'ימי ווייט	90
סטיב דייויס	100
דייגו מרדונה	20

(זה היה בסמסטר קיץ, ומספר הסטודנטים היה קטן במיוחד).

שיטת סימון מס' 4: סימון למדא

הדרך של מתן שמות זמניים אינה נחשבת (ובצדק!) מספקת ויעילה עבור קבוצות, ולכן הומצאה שיטת האבסטרקציה: $\{x|\varphi\}$. לגבי פונקציות המצב דומה. שיטת השמות הזמניים אינה מספקת ולעתים קרובות אינה יעילה. למעשה, היא עוד הרבה פחות יעילה לגבי פונקציות מאשר לגבי קבוצות. אי לכך המציאו הלוגיקאים שיטת אבסטרקציה גם לצורך הגדרת פונקציות. השיטה ידועה בשם "סימון למדא", כיוון שמשתמשים בה באות היוונית λ (למדא). סימון זה טרם נקלט, למרבה הצער, אצל המתמטיקאים, אך משתמשים בו הרבה במדעי המחשב התיאורטיים, ובעיקר בשפות תכנות פונקציונליות. הרעיון: אם t ביטוי, אז $\lambda x.t$ מתאר את הפונקציה (אולי חלקית), המתאימה לכל עצם a את הערך של t עבור $x = a$. (בצורה אחרת: $\lambda x.t$ היא הפונקציה, שעבור x מסוים כקלט מחזירה כפלט את תוצאת חישוב t עבור x זה). לדוגמה: $\lambda x.x^2$ היא הפונקציה המתאימה לכל מספר x את ריבועו.

$\lambda x.t$ נותן בעצם רק את *הכלל*. כדי שסימון למדא יתאר פונקציה באופן מלא, מכניסים גם את התחום לסימון וכותבים $\lambda x \in A.t$ (או $\lambda x{:}A.t$) כדי לציין ש- A היא התחום (זה מקביל מאוד לסימון $\{x \in A \,|\varphi\}$ עבור קבוצות). כשרוצים לציין גם טווח, כותבים $\lambda x.t{:}\,A \to B$ או $\lambda x \in A.t \in B$, אך אנו נסתפק, בדרך-כלל, בסימון $\lambda x \in A.t$. מציאת טווח לפונקציה תיעשה על-ידי ניתוח הביטוי t. כשיהיה ברור באיזה תחום מדובר, נשמיט אפילו אותו ונכתוב רק $\lambda x.t$.

---

[12] כמובן, היינו צריכים לומר "אחד הטווחים", אבל במקרה הנוכחי אין זה נוח, וזה גם מחטיא את המטרה. השימוש בהא הידוע ("הטווח") נוח יותר. במקרה זה הוא בא לציין, שמראש ידוע שהציונים לא יחרגו מהטווח הנ"ל, אך כל איבר בו אפשרי (אם כי לא בהכרח מתקבל).

♦ דוגמאות ראשונות

הפונקציה בדוגמה הראשונה למעלה היא $\lambda x \in \mathbf{R}.x^2$. הפונקציה החלקית בדוגמה השנייה היא $\lambda x \in \mathbf{R}.\frac{1}{x}$. אם ברצוננו לתארה כפונקציה, עלינו לדייק ולכתוב: $\lambda x \in (\mathbf{R}\text{-}\{0\}).\frac{1}{x}$. (אם לא יצוין בפירוש אחרת, ביטויי-למדא ישמשו אצלנו לתיאור פונקציות בלבד).

האופרטור λ הוא אופרטור קשירה. המשתנה שהוא קושר נכתב מיד אחריו (כמו במקרה של $\forall$ ו- $\exists$). טווח הקשירה נקבע כמו במקרה של ι ו- μ (כלומר: הקטע הארוך ביותר אחרי הנקודה, שהוא תקין תחבירית, כשהוא עומד בפני עצמו). כיוון ש- λ הוא אופרטור קשירה, כלל (α) תופש, כמובן. כך אין כל הבדל בין $\lambda x \in \mathbf{R}.x^2$ ובין $\lambda y \in \mathbf{R}.y^2$. שני הביטויים זהים במשמעותם (ומתארים אותה פונקציה בדיוק!)

כשם שהעובדה היסודית ביותר הקשורה בסימון $\{x|\varphi\}$ היא כלל (β):

$$t \in \{x|\varphi\} \Leftrightarrow \varphi(t/x)$$

כך העובדה היסודית בדבר סימון למדא, זו *שמגדירה* אותו בעצם, היא כלל (β), עבורו:

$$\boxed{(\beta) \qquad (\lambda x.t)(s) = t(s/x)}$$

בתנאי, כמובן, ש- s הוא ביטוי החופשי להצבה ב- t במקום x. עבור $\lambda x \in A.\,t$ יש כלל דומה (ותנאי נוסף: ש- s הוא ביטוי המייצג איבר של A).

דוגמאות:

$$(\lambda x \in \mathbf{R}.x^2)(3) = 3^2 = 9$$

$$(\lambda x \in \mathbf{R}.x^2)(a+b) = (a+b)^2$$

$$(\lambda x.\int_0^x xydy)(z+1) = \int_0^{z+1}(z+1)ydy$$

אבל

$$(\lambda x.\int_0^x xydy)(y+1) \neq \int_0^{y+1}(y+1)ydy$$

למעשה:

$$(\lambda x.\int_0^x xydy)(y+1) = (\lambda x.\int_0^x xzdz)(y+1) = \int_0^{y+1}(y+1)z\,dz$$

הסברים:
כלל (β) משמש לצורך פישוט של ביטויים. צורת הפעלתו היא כדלקמן: בהינתן ביטוי מהצורה $(\lambda x.t)(s)$ (דהיינו: אפליקציה של ביטוי-λ לארגומנט s), מוחקים את "$\lambda x.$" מהביטוי $\lambda x.t$, ובמה שנשאר מציבים את s במקום כל מופע חופשי של x.

בחישוב האחרון השתמשנו באמצע בכלל (α) כדי להפוך את הביטוי $y + 1$ שלא היה חופשי להצבה, לביטוי, שהוא כן חופשי להצבה.

כמו כל ביטוי או נוסחה אחרים, גם ביטוי-למדא יכול להכיל פרמטרים. אם נסמן את $\{x \in \mathbf{R} \mid x > 0\}$ ב- $\mathbf{R}^+$, אז הביטוי $\lambda x \in \mathbf{R}^+.x^a$ הוא ביטוי עבור פונקציה, שערכו המדויק תלוי בערך הפרמטר a. כאשר $a = 2$, זו תהיה הפונקציה $\lambda x \in \mathbf{R}^+.x^2$, וכש- $a = 1/2$ – הפונקציה $\lambda x \in \mathbf{R}^+.\sqrt{x}$. את כלל (β) ניתן, כמובן, להפעיל גם כשיש פרמטרים. לדוגמה:

$$(\lambda x \in \mathbf{R}^+.x^a)(3) = 3^a$$

(כיוון שזו ***זהות***, הכוונה בעצם ל- $\forall a \in \mathbf{R}[(\lambda x \in \mathbf{R}^+.x^a)(3) = 3^a]$).

נצעד עתה צעד נוסף קדימה, ונשים לב שהביטוי $\lambda a \in \mathbf{R}.\ \lambda x \in \mathbf{R}^+.x^a$ מתאר פונקציה, שהתחום שלה הוא $\mathbf{R}$, והיא מתאימה לכל $a \in \mathbf{R}$ ***פונקציה*** מ- $\mathbf{R}^+$ ל- $\mathbf{R}$ (או ל- $\mathbf{R}^+$):

$$\lambda a \in \mathbf{R}.\ \lambda x \in \mathbf{R}^+.x^a : \ \mathbf{R} \to (\mathbf{R}^+ \to \mathbf{R})$$

כך, למשל: $((\lambda a \in \mathbf{R}.\ \lambda x \in \mathbf{R}^+.x^a)(2))(3) = (\lambda x \in \mathbf{R}^+.x^2)(3) = 3^2 = 9$

דוגמה נוספת:

$$\lambda n \in \mathbf{N}.\ \lambda x \in [0,1].\ x^n(1 - x^n) : \mathbf{N} \to ([0,1] \to [0,1])$$

הביטוי כאן משמאל מגדיר פונקציה, שמתאימה לכל מספר טבעי פונקציה מסוימת מהקטע $[0,1]$ אל $[0,1]$. כך, למשל:

$$\underbrace{((\lambda n \in \mathbf{N}.\ \lambda x \in [0,1].\ x^n(1 - x^n))(3))}(0.2) = (\lambda x \in [0,1].\ x^3(1 - x^3))(0.2) =$$

$$= 0.2^3(1 - 0.2^3)$$

(הערה: ב- "$\underbrace{\qquad}$" סימנו את אותו חלק בביטוי המורכב, עליו הופעל כלל (β)).

לסיום הצגה ראשונה זו של סימון λ, נביא אנלוגיה נוספת בין הסימון $\{x|\varphi\}$ ובין הסימון $\lambda x.t$. לגבי $\{x|\varphi\}$ היה לנו, כזכור, כלל η, שקבע כי $\{x|x \in S\} = S$, בתנאי ש- x אינו חופשי ב- S. לגבי λ יש לנו כלל דומה. הוא נובע מההבחנה הבאה: אם f היא פונקציה, אז הפונקציה, המחזירה לכל x בתחום של f את $f(x)$, זהה לפונקציה f (כי זו עושה, כמובן, אותו דבר בדיוק). פורמלית:

$$\forall f(\lambda x.\, f(x) = f)$$

בנוסחה זו אין התייחסות לתחום של $\lambda x.\, f(x)$, אבל ברור, שתחום ההגדרה של הפונקציה, שביטוי זה מתאר, זהה לתחום ההגדרה של f.

מהנוסחה האחרונה נובע, בעזרת הכלל היסודי של $\forall$, שאם t ביטוי עבור פונקציות, שאינו מכיל את x כפרמטר, אזי:

$$\boxed{(\eta) \qquad \lambda x.\, t(x) = t}$$

(וכמובן $\lambda x \in A.\, t(x) = t$ בתנאי הנוסף, ש- t מגדיר פונקציה, שהתחום שלה הוא A).

דוגמאות:

(א) $\lambda x.\, \sin x = \sin$

(ב) אם g משתנה עבור פונקציות, אז $\lambda x.\, g(x) = g$.

(ג) נקח $t = \lambda y \in \mathbf{R}.y^2$, אז נקבל (אם נרשום גם את התחום של $\lambda x.\, t(x)$):

$$\lambda x \in \mathbf{R}.\, (\lambda y \in \mathbf{R}.\, y^2)(x) \overset{\eta}{=} \lambda y \in \mathbf{R}.\, y^2$$

את האינסטנציה הזו של כלל (η) יכולנו לקבל, במקרה זה, ישירות מכללי (α) ו- (β) כך:

$$\lambda x \in \mathbf{R}.\, \underbrace{(\lambda y \in \mathbf{R}.\, y^2)(x)} \overset{\beta}{=} \lambda x \in \mathbf{R}.\, x^2 \overset{\alpha}{=} \lambda y \in \mathbf{R}.\, y^2$$

הכללים α, β ו- η (ביחד עם עיקרון האקסטנציונליות, עליו ניתן להראות שהוא שקול לכלל η) הם הכללים היסודיים של ביטויי-λ. היסטורית, למעשה, השמות (והניסוח המפורש של הכללים) נעשו קודם כל עבור ביטויים אלה דווקא. טבלה מס' ב.4 מסכמת את הכללים הללו הן עבור קבוצות והן עבור פונקציות. כמו כן הוכנסו לטבלה שלושה כללים פשוטים נוספים, המשמשים לפישוטים, אותם הכרנו בפרק הקודם.

טבלה ב.4: כללי היסוד לפישוטים

(α) $\{y\|\varphi\} = \{x\|\varphi(x/y)\}$ בתנאי ש- x משתנה חדש, שאינו מופיע ב- φ.	(α) $\lambda y.\ t = \lambda x.\ t(x/y)$ בתנאי ש- x משתנה חדש, שאינו מופיע ב- t.
(β) $s \in \{x\|\varphi\} \Leftrightarrow \varphi(s/x)$ בתנאי ש- s חופשי להצבה במקום x ב- φ.	(β) $(\lambda x.\ t)(s) = t(s/x)$ בתנאי ש- s חופשי להצבה במקום x ב- t.
(η) $\{x \mid x \in S\} = S$ בתנאי ש- x אינו מופיע חופשי ב- S.	(η) $\lambda x.\ t(x) = t$ בתנאי ש- x אינו מופיע חופשי ב- t.
(Ext) $\forall A \forall B[A = B \Leftrightarrow \forall x.\ x \in A \Leftrightarrow x \in B]$	(Ext) $\forall f \forall g\ [f = g \Leftrightarrow \forall x.\ f(x) = g(x)]$ וביתר דיוק: $\forall f \forall g\ [f = g \Leftrightarrow \mathrm{Dom}(f) = \mathrm{Dom}(g) \wedge$ $\wedge\ \forall x \in \mathrm{Dom}(f).\ f(x) = g(x)]$
$\pi_1(<x, y>) = x$ $\pi_2(<x, y>) = y$ $z = <\pi_1(z),\ \pi_2(z)>$ (הנוסחה האחרונה נכונה בתנאי ש- z הוא זוג סדור).	

שיטת סימון מס' 5: סימון למדא עבור פונקציות של כמה משתנים

במתמטיקה נוהגים לעתים קרובות להתעסק בפונקציות של שני משתנים או יותר. כך, למשל, אפשר למצוא בטקסטים בחדו"א הגדרה של פונקציה g, הנראית כך:

$$g(x, y) = xy^2$$

רשמית, רואים בדרך-כלל בפונקציות כאלה לא משהו מסוג חדש, אלא פונקציות *במובן הרגיל*, שהארגומנטים שלהן הם *זוגות* של מספרים. פונקציה g כמו זו שבדוגמה האחרונה היא אפוא פונקציה מ- $\mathbf{R} \times \mathbf{R}$ אל $\mathbf{R}$. למעשה, היה צריך, לאור זאת, לכתוב בעצם:

$$g(<x, y>) = xy^2$$

הסימון המקובל בטקסטים רחוק אבל, לרוע המזל, מלהיות עקבי, וכמעט תמיד כותבים $g(x, y)$, היכן שהיה צריך לכתוב $g(<x, y>)$.
את הסימון $\{\_\ |\ldots\}$ עבור קבוצות מסגלים, כזכור, לצורה מיוחדת, נוחה, כשרוצים להגדיר קבוצה של זוגות (כותבים: $\{<x,y>|\ \varphi(x,y)\}$). בדומה, גם את סימון-למדא מרחיבים לצורך טיפול נוח בפונקציות, שהתחום שלהן הוא מכפלה קרטזית (או קבוצה כלשהי אחרת של זוגות). כאשר רוצים לתאר את *הכלל* בלבד, כותבים: $\lambda x,y.\ t$

(לדוגמה: $(\lambda x,y.xy^2$). כאשר רוצים לציין גם את התחום, ותחום זה הוא המכפלה הקרטזית $A \times B$, כותבים:

$$\lambda x \in A\ ,\ y \in B.\ t$$

לדוגמה: הביטוי $\lambda x \in \mathbf{R}\ ,\ y \in \mathbf{R}.\ xy^2$ מייצג את הפונקציה g למעלה מ- $\mathbf{R}^2$ אל $\mathbf{R}$.

את הסימון המיוחד עבור פונקציות של שני משתנים ניתן להכליל ללא קושי לפונקציות של מספר כלשהו של משתנים. כללית, $\lambda x_1 \in A_1, x_2 \in A_2, \ldots, x_n \in A_n.\ t$ הוא ביטוי המתאר פונקציה, שהתחום שלה הוא $A_1 \times A_2 \times \ldots \times A_n$ ומתאימה לכל $<x_1, x_2, \ldots, x_n>$ ב- $A_1 \times \ldots \times A_n$ את הערך המתאים של הביטוי t.

נקודות שחשוב לזכור כאן:

(i) בביטוי כמו $\lambda x \in A\ ,\ y \in B.\ t$, האופרטור λ קושר בבת-אחת בתוך t גם את x וגם את y.

(ii) חשוב לשים לב להבדל בין השימוש בפסיק לבין השימוש בנקודה: פסיקים מפרידים בין המשתנים השונים ש- λ קושר, בעוד הנקודה מציינת את תחילתו של טווח הקשירה.

(iii) חשוב מאוד להבדיל בין $\lambda x \in A\ ,\ y \in B.\ t$ ובין $\lambda x \in A.\ \lambda y \in B.\ t$. כך, למשל, $\lambda x \in \mathbf{R}, y \in \mathbf{R}.\ xy^2$ מציין פונקציה מ- $\mathbf{R} \times \mathbf{R}$ אל $\mathbf{R}$, בעוד $\lambda x \in \mathbf{R}.\ \lambda y \in \mathbf{R}.\ xy^2$ מציין פונקציה מ- $\mathbf{R}$ אל $\mathbf{R} \to \mathbf{R}$! (כשמכניסים, לדוגמה, 2 כקלט לביטוי האחרון, התוצאה היא $\lambda y \in \mathbf{R}.\ 2y^2$, וזו היא פונקציה מ- $\mathbf{R}$ ל- $\mathbf{R}$).

(iv) כלל (β) לפונקציות של שני משתנים נראה כך:

$$(\lambda x,y.\ t)\ (<u,v>) = t(u/x,\ v/y)$$

בתנאי ש- u חופשי להצבה במקום x ב- t, ו- v חופשי להצבה במקום y ב- t (במקום $\lambda x,y.\ t$ יכול כאן, כמובן, להיות רשום משהו מהצורה המלאה יותר: $\lambda x \in A, y \in B.\ t$).

דוגמאות:

$$(\lambda x \in \mathbf{R}\ ,\ y \in \mathbf{R}.\ xy^2)\ (<3,2>) = 3 \cdot 2^2$$
$$(\lambda x \in \mathbf{R}\ ,\ y \in \mathbf{N}.\ x^y)\ (<x+1,\ 2y>) = (x+1)^{2y}$$

מעשית, נכתוב גם אנו לעתים קרובות $(\lambda x,y.\ t)(u,v)$ במקום $(\lambda x,y.\ t)(<u,v>)$. למשל: $(\lambda x \in \mathbf{R}\ ,\ y \in \mathbf{R}.\ xy^2)(3,2)$ או $(\lambda x \in \mathbf{R}\ ,\ y \in \mathbf{N}.\ x^y)(x+1,\ 2y)$.

את כלל (β) ניתן להכליל בצורה דומה לפונקציות של n משתנים ($n \in \mathbf{N}^+$).

(v) שתי פונקציות חשובות במיוחד בהקשר הנוכחי הן פונקציות ההטל, π_1 ו- π_2[13], אותן הכרנו כבר בסוף הפרק הקודם:

$$\pi_1 = \lambda x,y.\ x \qquad \pi_2 = \lambda x,y.\ y$$

ולכן:

$$\boxed{\pi_1(<x,y>) = x \qquad \pi_2(<x,y>) = y}$$

אם נדייק, מה שתיארנו כאן הוא שוב רק *הכלל*. כדי שנוכל לדבר על *פונקציות*, עלינו לציין מה התחום. לאמיתו של דבר אין, בעצם, פונקציות "π_1" ו- "π_2", אלא לכל שתי קבוצות A ו- B יש לנו:

$$\pi_1^{A,B} = \lambda x \in A\ , y \in B.\ x : A \times B \to A$$
$$\pi_2^{A,B} = \lambda x \in A\ , y \in B.\ y : A \times B \to B$$

שיטת סימון מס' 6: סדרות

בחשבון דיפרנציאלי ואינטגרלי עיקר העניין, בדרך-כלל, הוא בפונקציות מ- $\mathbf{R}$ ל- $\mathbf{R}$. התכונות החשובות ביותר בו של פונקציות כאלו הן רציפות וגזירות. במתמטיקה בדידה, לעומת זאת, יש עניין גדול הרבה יותר בפונקציות, שהתחום שלהן הוא $\mathbf{N}$ או $\mathbf{N}^+$. לפונקציות כאלו קוראים *סדרות*[14]. לפעמים קוראים בשם זה גם לפונקציות, שהתחום שלהן הוא מהצורה $\{1,\ldots, n\}$ (או אולי $\{0, \ldots, (n-1)\}$). שם אחר לפונקציות כאלו הוא "רשימות"). המספר n נקרא אז אורך הסדרה או הרשימה.

עבור סדרות מקובלים מאוד, לדאבוננו, סימונים, שהינם שונים מאלה המקובלים לגבי שאר הפונקציות. מקובל, למשל, לכתוב "f_n" במקום $f(n)$ (ולקרוא לזה "האיבר ה- n-י של הסדרה" במקום: "ערכה של f עבור הארגומנט n"). יתר על כן, במקום להשתמש בסימון-למדא ולכתוב $\lambda n \in \mathbf{N}^+.\ t(n)$ (או רק $\lambda n.\ t(n)$, אם ברור שהתחום הוא $\mathbf{N}^+$), כותבים $\{t_n\}_{n=1}^{\infty}$ (או $(t_n)_{n=1}^{\infty}$). לדוגמה: במקום לכתוב $\lambda n \in \mathbf{N}^+.\frac{1}{n}$, כותבים בטקסטים בחדו"א $\{\frac{1}{n}\}_{n=1}^{\infty}$. אין למעשה שום הבדל במובן של שני הסימונים. אי לכך ברור,

---

[13] כדאי להעיר שב- LISP (או SCHEME) כותבים, משום מה, CAR במקום π_1 ו- CDR במקום π_2.

[14] שני סוגים של סדרות, שיש להם חשיבות רבה גם בחשבון אינפיניטסימלי, הן סדרות עם טווח $\mathbf{R}$ (ונקראות "סדרות של מספרים ממשיים"), וסדרות עם טווח $\mathbf{R} \xrightarrow{p} \mathbf{R}$ (ונקראות "סדרות של פונקציות ממשיות").

שבביטוי מהצורה $\{t(n)\}_{n=1}^{\infty}$ המשתנה n הינו קשור. כלל (α) תופס אפוא כאן: אין הבדל בין הסדרה $\{\frac{1}{n}\}_{n=1}^{\infty}$ והסדרה $\{\frac{1}{k}\}_{k=1}^{\infty}$.

לעובדה, שעבור סדרות משתמשים בצורת סימון שונה, יש מספר השלכות לא נעימות. ראשית, למרות שאין הבדל בין $\lambda n \in \mathbf{N}^+.\, t(n)$ ובין $\{t(n)\}_{n=1}^{\infty}$, השימוש בכלל (β) אינו מוכר עבור הסימון האחרון, ולכן צורות הניסוח של טענות וחישובים בטקסטים בחדו"א מסובכות לעתים קרובות ללא כל הצדקה. יתר על כן: קיימת נטייה גדולה שם להוריד את הסוגריים המסולסלים ה"מעצבנים", ולכתוב: $a_n \to a$ במקום $\lambda n \in N^+.\, a_n \to a$ (או, לפחות, $\{a_n\}_{n=1}^{\infty} \to a$ או $\lim\limits_{n\to\infty} a_n = a$). צורת כתיבה זו פשוט מזמינה בלבול: כך בנוסחה $\frac{1}{n} < \varepsilon$ הביטוי "$\frac{1}{n}$" מייצג *מספר* (שערכו תלוי בערכו של n), בעוד ב"נוסחה" $\frac{1}{n} \to 0$ הביטוי "$\frac{1}{n}$" לא מייצג בעצם מספר, כי אם את ***ה**סדרה* $\lambda n \in \mathbf{N}^+.\, \frac{1}{n}$. אין אבל שום דבר בצורת הביטויים, שמרמז על ההבדל המהותי, או על כך, ש- n הינו קשור ב- "$\frac{1}{n} \to 0$" וחופשי ב- "$\frac{1}{n} < \varepsilon$". יש לנהוג אפוא זהירות יתרה בסימונים המקובלים (בעיקר כשיש פרמטרים נוספים), ובכל מקרה של חשש לבלבול – להשתמש בסימונים שלמדנו (ובמיוחד בסימון-למדא)! (למען ההגינות נציין, שכאשר מדובר בסדרות של *פונקציות*, הרי יש לסימון המקובל יתרון פסיכולוגי מסוים. אם $f: \mathbf{N} \to (\mathbf{R} \to \mathbf{R})$, הרי נוח יותר לעתים לדבר על $f_7(\sqrt{2})$, למשל, מאשר על $((f(7))(\sqrt{2})$.)

III. הגדרת פונקציות כסוג מיוחד של קבוצות

בסעיף זה נראה, כיצד ניתן להגדיר פונקציות במונחים היסודיים של תורת הקבוצות. מכאן ואילך יהיו ההגדרות, שנכניס בסעיף זה, ההגדרות הרשמיות של המושגים הקשורים בפונקציות (הגם שאינן חיוניות לצורך הבנת מרבית מה שבא בהמשך). מושג המפתח הוא המושג הבא:

הגדרה:
הגרף של פונקציה f, שתחומה A, הוא הקבוצה $G(f) = \{<x, f(x)> \mid x \in A\}$.

לגרף $G(f)$ של פונקציה f יש התכונות הבאות:

(א) אם $z \in G(f)$ אז קיימים a ו- b כך ש- $z = <a,b>$.

(ב) אם $<a, b_1> \in G(f)$ ו- $<a, b_2> \in G(f)$ אז $b_1 = b_2$.

בתורת הקבוצות (ובמתמטיקה המודרנית בכלל) נהוג *לזהות* פונקציה עם הגרף שלה (שהוא קבוצה, כמובן). בהתאם, מגדירים פונקציה כקבוצה, שיש לה התכונות שפירטנו לעיל.

הגדרה:

1. פונקציה f היא קבוצה של זוגות סדורים, כך שאם $<a,b_1> \in f$ ו- $<a, b_2> \in f$, אז $b_1 = b_2$.
2. אם f פונקציה, אז הקבוצה $\{x| \exists y. <x,y> \in f\}$ נקראת ה**תחום** של f ומסומנת ב- $\mathrm{Dom}(f)$. כאשר $A = \mathrm{Dom}(f)$, אומרים ש- f היא פונקציה **מעל** A. אם $\mathrm{Dom}(f) \subseteq A$, אומרים ש-$f$ היא פונקציה חלקית מעל A.

תרגיל

הראו ש- $\mathrm{Dom}(f) \in P(\cup(\cup(f)))$, ולכן $\mathrm{Dom}(f)$ אכן קבוצה.

הגדרה:

נניח ש-f פונקציה. אם $f \subseteq A \times B$, אומרים ש-f היא פונקציה חלקית מ- A ל- B, וש- B היא טווח של f. אם, בנוסף, $A = \mathrm{Dom}(f)$, אומרים ש-f פונקציה מ- A ל- B, ומסמנים $f : A \to B$ או $f \in A \to B$. (הגדרה מדויקת של $A \to B$ וכן של $A \overset{p}{\to} B$ כקבוצות ניתנת בטבלה ב.5 למטה).

תרגיל

הראו ש- $(A \to B) \in P(P(A \times B))$.

נשים לב, שאם $f : A \to B$, אז:

$$\forall x \in A\ \exists! y \in B.\ <x,y> \in f$$

כאשר $\exists! y.\psi$ פירושו, שיש y יחיד שיש לו התכונה ψ, דהיינו:

$$\exists! y.\ \psi =_{Df} \exists y.\ \psi \wedge \forall y_1\ \forall y_2 .\psi(y_1/y) \wedge \psi(y_2/y) \Rightarrow y_1 = y_2$$

או, בניסוח שקול לוגית אחר:

$$\exists! y.\ \psi =_{Df} \exists y(\psi \wedge \forall z.\ \psi(z/y) \to z = y))$$

סימון:

אם f פונקציה ו- $x \in \mathrm{Dom}(f)$, אז נסמן ב- $f(x)$ את ה- y היחיד, כך ש- $<x,y> \in f$:

$$f(x) =_{Df} \iota y.\ <x,y> \in f$$

נפנה עתה לסוגייה של שוויון פונקציות. בתחילת פרק זה ניסחנו את עיקרון האקסטנציונליות עבור פונקציות. באותו שלב מילא עיקרון זה, בעצם, תפקיד של הגדרה: הוא הגדיר מתי שתי פונקציות נחשבות שוות. עתה, כאשר פונקציות הינן קבוצות, הרי שוויון ביניהן כמוהו כשוויון בין כל שתי קבוצות אחרות, דהיינו $f = g$ אם"ם $\forall x.\, x \in f \Leftrightarrow x \in g$. בשלב זה הופך, אפוא, העיקרון הנ"ל *לטענה*, שיש להוכיחה:

טענה (עיקרון האקסטנציונליות עבור פונקציות):
אם f ו- g פונקציות, אז:

$$f = g \Leftrightarrow [\mathrm{Dom}(f) = \mathrm{Dom}(g) \wedge \forall x \in \mathrm{Dom}(f).\, f(x) = g(x)]$$

הוכחה:
($\Rightarrow$) נניח $f = g$. אז לפי עקרונות הלוגיקה, כל מה שנכון לגבי f נכון לגבי g, ולהיפך, בפרט $\mathrm{Dom}(f) = \mathrm{Dom}(g)$, ולכל x $f(x)$ מוגדר אם"ם $g(x)$ מוגדר והם שווים (בתנאי שאחד מהם אכן מוגדר).
($\Leftarrow$) נניח ש- $\mathrm{Dom}(f) = \mathrm{Dom}(g)$, ושלכל $x \in \mathrm{Dom}(f)$ מתקיים ש- $f(x) = g(x)$. נוכיח ש- $f = g$, כלומר: ש- $f \subseteq g$ ו- $g \subseteq f$. נראה, לדוגמה, ש- $f \subseteq g$ (ההוכחה ש- $g \subseteq f$ היא דומה). נניח אפוא, ש- $z \in f$. כיוון ש- f פונקציה, קיימים x ו- y כך ש- $z = <x,y>$. מהגדרת $\mathrm{Dom}(f)$ נובע לכן, ש- $x \in \mathrm{Dom}(f)$ ו- $y = f(x)$. כיוון ש- $\mathrm{Dom}(f) = \mathrm{Dom}(g)$, נובע מזה ש- $x \in \mathrm{Dom}(g)$, ולכן (לפי הנתון השני) $y = f(x) = g(x)$. אבל העובדה ש- $x \in \mathrm{Dom}(g)$ ו- $y = g(x)$ פירושה, מהגדרותינו, ש- $<x,y> \in g$. במלים אחרות: $z \in g$.

לבסוף, סימון λ גם הוא אינו, עקרונית, אלא סימון מקוצר. למעשה:

$$\lambda x.\, t = \{z \mid \exists x.\, z = <x, t>\}$$
$$\lambda x \in A.\, t = \{<x, t> \mid x \in A\}$$

הערות:

א. t יכיל כאן, בדרך-כלל, מופעים חופשיים של x.

ב. $\lambda x.\, t$ הוא פונקציה אח"ח הביטוי באגף ימין של הגדרתו הרשמית הינו ביטוי כשר, המתאר קבוצה. עם זאת, בתפקידו כמתאר *כלל*, יש ל- $\lambda x.\, t$ מובן גם אם אינו קבוצה.

ג. $\lambda x \in A.\, t$ מייצג תמיד פונקציה ***חלקית*** מעל A. הוא מייצג ***פונקציה*** מעל A אם t מוגדר עבור כל $x \in A$.

ד. אפשר להראות ללא קושי, שעם ההגדרות הרשמיות למעלה, כללי α, β ו- η עבור λ נובעים מהכללים המקבילים עבור קבוצות.

טבלה ב.5 מסכמת את ההגדרות הרשמיות הבסיסיות הקשורות בפונקציות. היא כוללת גם את ההגדרות הרשמיות של הרכבת פונקציות ושל פונקציה הפוכה (מושגים אותם נלמד בהמשך).

טבלה ב.5: פונקציות – הגדרות בסיסיות

(1)	קבוצה f נקראת ***פונקציה*** אם מתקיים: (א) $\forall z \in f \exists a\, \exists b.\ z = \langle a,b\rangle$ (ב) $\forall a\, \forall b_1\, \forall b_2.\ \langle a, b_1\rangle \in f \wedge \langle a, b_2\rangle \in f \Rightarrow b_1 = b_2$
(2)	**תהי** f **פונקציה, אז:** (i) $\underline{\mathrm{Dom}(f)} = \{x \mid \exists y.\ \langle x,y\rangle \in f\}$ (ii) f היא פונקציה חלקית מעל A אם $\mathrm{Dom}(f) \subseteq A$ (iii) f היא פונקציה מעל A אם $\mathrm{Dom}(f) = A$ (iv) f היא פונקציה חלקית מ- A ל- B אם $f \subseteq A \times B$ (v) f היא פונקציה מ- A ל- B אם $f \subseteq A \times B \wedge \mathrm{Dom}(f) = A$
(3)	**אם** f **פונקציה ו-** $x \in \mathrm{Dom}(f)$**, אז:** $\underline{f(x)} =_{Df} \iota y.\ \langle x,y\rangle \in f$
(4)	**אם** $f: A \to B$ **ו-** $g: B \to C$**, אז:** $\underline{g \circ f} = \{\langle a,c\rangle \in A \times C \mid \exists b \in B.\ \langle a,b\rangle \in f \wedge \langle b,c\rangle \in g\}$
(5)	**אם** $f: A \to B$ **פונקציית שקילות, אז:** $\underline{f^{-1}} = \{\langle b,a\rangle \in B \times A \mid \langle a,b\rangle \in f\}$
(6)	**טענה: אם** f **ו-** g **פונקציות, אז:** $f = g \Leftrightarrow [\mathrm{Dom}(f) = \mathrm{Dom}(g) \wedge \forall x \in \mathrm{Dom}(f).\ f(x) = g(x)]$
(7)	(i) $\lambda x.\ t = \{z \mid \exists x.\ z = \langle x, t\rangle\}$ (ii) $\lambda x \in A.\ t = \{\langle x, t\rangle \mid x \in A\}$
(8)	(i) $A \to B = \{f \in P(A \times B) \mid (f \text{ פונקציה}) \wedge \mathrm{Dom}(f) = A\}$ (ii) $A \xrightarrow{p} B = \{f \in P(A \times B) \mid (f \text{ פונקציה}) \wedge \mathrm{Dom}(f) \subseteq A\}$

IV. דוגמאות של פונקציות חשובות

א. תהי A קבוצה. פונקצית ***הזהות*** על A היא הפונקציה:

$$i_A = \lambda x \in A.\, x$$

במלים אחרות: $i_A(x) = x$ לכל $x \in A$.

ב. תהיינה A ו- B קבוצות. פונקציה ***קבועה*** מ- A ל- B היא פונקציה מהצורה $\lambda x \in A.\, c$, כאשר c איבר של B. לדוגמה: $\lambda x \in R.2$ היא פונקציה המחזירה לכל מספר ממשי את הערך 2 ("$y = 2$" כותבים לפעמים בחדו"א).

חשוב להבדיל בין c, שהוא איבר של B, ובין פונקציה הקבועה עליו, שאינה איבר של B, אלא איבר בקבוצה מהצורה $A \to B$. כך, למשל, אין זה נכון שהנגזרת של $\lambda x \in R.2$ היא אפס! מה שנכון הוא:

$$(\lambda x \in \mathbf{R}.\, 2)' = \lambda x \in \mathbf{R}.\, 0$$

ג. תהי A קבוצה. ***הפונקציה האופיינית*** של A מסומנת בדרך-כלל על-ידי χ_A ומוגדרת בצורה הבאה:

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

הגדרה מסוג זה נקראת הגדרה לפי מקרים. אלטרנטיבית נוכל להשתמש בסגנון ההגדרה הבא:

$$\chi_A = \lambda x.\ \text{if } x \in A \text{ then } 1 \text{ else } 0$$

כך או כך, ברור ש:

$$\forall x (x \in A \Leftrightarrow \chi_A(x) = 1)$$

מכאן, שהפונקציה האופיינית של A כשמה כן היא: היא מאפיינת את A לחלוטין. הכרה מושלמת של χ_A פירושה הכרה מושלמת של A, ולהפך. ניתן היה לכן לקחת דווקא את מושג הפונקציה כמושג יסודי, ולראות בקבוצות סוג מיוחד של פונקציות: אלו עם טווח $\{0,1\}$ (בתורת הפונקציות הְחְשִׁיבוֹת (computable) זו אכן הפרוצדורה המקובלת).

שאלה אחת בקשר ל- χ_A נשארה פתוחה למעלה: מהו, בעצם, תחום ההגדרה שלה? ממבט ראשון זהו "כל העולם". ברם, ראינו, שאוסף כל העצמים אינו מהווה

קבוצה ולכן אינו יכול להיות תחום הגדרה של פונקציה. χ_A אינה, לכן, פונקציה במובן שהגדרנו למעלה.[15] ניתן אבל להופכה לכזו אם מגבילים אותה לאיזו קבוצה מקיפה E, כך ש- $A \subseteq E$. אם נגדיר, במקרה כזה,

$$\chi_A^{(E)} = \lambda x \in E.\ \text{if } x \in A \text{ then } 1 \text{ else } 0$$

אז:

$$\chi_A^{(E)} \in E \to \{0,1\}$$

הערה:
בדרך-כלל משמיטים את ההתייחסות ל- E, אם היא ברורה מהקונטקסט, או אם זהותה המדויקת אינה חשובה (השווי לדיון ב- $\overline{A}$ בפרק ב.3).

ד. ניתן ללכת צעד נוסף, מעבר למה שנעשה בדוגמה הקודמת, ולהגדיר את הפונקציה $\lambda A \in P(E).\chi_A^{(E)}$. זוהי פונקציה מ- $P(E)$ אל $E \to \{0,1\}$. במלים אחרות: זוהי פונקציה, שהארגומנטים שלה (ה"קלטים", בלשון מדעי המחשב) הם קבוצות, וערכיה (ה"פלטים") הם עצמם פונקציות. נדגיש שוב, שפונקציות כמו $\lambda A \in P(E).\chi_A^{(E)}$ הן עצמים לגיטימיים לחלוטין ואף מועילים. לרוע מזלנו, לימודינו בבית-הספר התיכון הרגילו אותנו להבין במושג "פונקציה" רק התאמה מקבוצה אחת של מספרים אל קבוצה אחרת של מספרים. הכרחי לכן להתגבר על הרגל זה ולעכל את הרעיון, שאפשריות וקיימות פונקציות *מכל* קבוצה (אפילו קבוצה של קבוצות או קבוצה של פונקציות!) אל *כל* קבוצה אחרת! אגב, אין כאן רק "מוזרות מתמטית". בשפות מחשב פונקציונליות, כמו LISP, SCHEME או SML, פונקציות הינן אכן "אזרחיות מדרגה ראשונה", דהיינו: הן יכולות לשמש הן בתפקיד של קלט והן בתפקיד של פלט, ממש כמו מספרים!

הבה נשתמש בדוגמה הנוכחית למטרה נוספת. נשים לב, שהפונקציה, שהוגדרה בה (אם נכתוב אותה באופן מלא, ללא השימוש בשם "$\chi_A^{(E)}$"), היא:

$$(*)\quad \lambda A \in P(E).\ \lambda x \in E.\ \text{if } x \in A \text{ then } 1 \text{ else } 0$$

הבה ננתח ביטוי זה כדוגמה, כיצד נעשה ניתוח כזה באופן כללי. ובכן, הביטוי (*) מתחיל ב- "$\lambda A \in P(E)$". זה מלמד, ש- (*) מייצג פונקציה, שהתחום שלה הוא $P(E)$. ההמשך (מה שבא אחרי הנקודה הראשונה) מלמד, מה פונקציה זו מתאימה לכל איבר A בתחום זה, ומניתוחו נוכל למצוא טווח שלה. ובכן, מה ש- (*) מתאים

[15] χ_A היא פונקציה במובן רחב יותר מזה שהגדרנו, מובן שבו הדגש הוא על הכלל, *החוקיות*. סימון למדא נועד, בעצם, קודם כל, לתאר פונקציות במובן רחב זה!

ל- $A \in P(E)$ הוא משהו, שמתחיל ב- "$\lambda x \in E$". מכאן למדים אנו, ש- (*) מתאים לכל A ב- $P(E)$ פונקציה מסוימת, שהתחום שלה E. זהותה של פונקציה זו תתברר מתוך ניתוח ההמשך: if $x \in A$ then 1 else 0, אנו רואים מיד, שערך ביטוי זה יכול להיות 0 או 1 (תלוי בערכו של x). מכאן אנו למדים, שטווח ודאי של הפונקציה, ש- (*) מתאים לאיזה $A \in P(E)$, הוא $\{0,1\}$. סיכום כל זאת מראה, ש-(*) הוא פונקציה, שתחומה $P(E)$, ומתאימה לכל $A \in P(E)$ פונקציה, שתחומה E, ו- $\{0,1\}$ הוא טווח שלה. (*) מגדיר לכן איבר של $P(E) \to (E \to \{0,1\})$. (*) כולל, כמובן, גם אינפורמציה מדויקת על ***פעולתו*** של איבר זה (כפונקציה מעל $P(E)$).

דוגמה:

ניקח את E בתור $\mathbf{N}$, ונסמן את קבוצת המספרים הטבעיים הזוגיים ב- $\mathbf{N}_{\text{even}}$. אז:

$$\Big(\underbrace{(\lambda A \in P(\mathbf{N}).\ \lambda x \in \mathbf{N}.\ \text{if}\ \ x \in A\ \ \text{then 1 else 0})(\mathbf{N}_{\text{even}})}\Big)(7)$$

$$\stackrel{\beta}{=} (\lambda x \in \mathbf{N}.\ \text{If}\ \ x \in \mathbf{N}_{\text{even}}\ \ \text{then 1 else 0})\ (7)$$

$$\stackrel{\beta}{=} \text{If}\ \ 7 \in \mathbf{N}_{\text{even}}\ \ \text{then 1 else 0}$$

$$= 0$$

כמו כן:

$$(\lambda A \in P(\mathbf{N}).\ \lambda x \in \mathbf{N}.\ \text{If}\ \ x \in A\ \ \text{then 1 else 0})\ (\{n \in \mathbf{N} \mid n \le 20\})$$

$$= \lambda x \in \mathbf{N}.\ \text{If}\ \ x \in \{n \in \mathbf{N} \mid n \le 20\}\ \text{then 1 else 0}$$

$$= \lambda x \in \mathbf{N}.\ \text{If}\ \ x \le 20\ \ \text{then 1 else 0}$$

זוהי, כמובן, פונקציה מ- $\mathbf{N}$ אל $\{0,1\}$ (הלא היא $\chi_{\{n \in \mathbf{N} \mid n \le 20\}}$).

ה. אופרטור הגזירה D (המסומן בדרך-כלל על-ידי תג אחרי שם הפונקציה, אותה גוזרים) הינו דוגמה מצוינת לפונקציה, המקבלת פונקציות (חלקיות) מ- $\mathbf{R}$ ל- $\mathbf{R}$ כקלטים, ומחזירה פונקציות כאלה כפלטים:

$$D: (\mathbf{R} \stackrel{p}{\to} \mathbf{R}) \to (\mathbf{R} \stackrel{p}{\to} \mathbf{R})$$

כך, לדוגמה:

$$D(\lambda x.\ x^2) = \lambda x.\ 2x$$

(השמטנו כאן, כמקובל, את ההתייחסות ל- $\mathbf{R}$, ולא כתבנו $\lambda x \in \mathbf{R}.\ 2x$, כי התחום כאן ברור, ובמקרה כזה מוטב לקצר).

השימוש בסימון $D(f)$ לציון הנגזרת של הפונקציה f מקובל באמת רק בשיטות מתמטיות מתקדמות (בהן חיוני להתייחס אל D כאל עצם, עליו מבצעים פעולות שונות). בדרך-כלל, כידוע, כותבים $f\,'$ במקום $D(f)$. נוהג זה אינו שונה, במהותו, מהנוהג לכתוב $n!$ במקום $!(n)$, וברוב המקרים אינו יוצר בעיות מיוחדות. מה שיוצרת גם יוצרת בעיות היא העובדה, שזהויות כמו $(\lambda x.\ x^2)' = \lambda x.\ 2x$ כותבים בטקסטים בחדו"א בצורה המאוד לא-מוצלחת הבאה: $(x^2)' = 2x$. ביטויים כמו "x^2" ו- "$2x$" הינם באופן טבעי ביטויים עבור **_מספרים_**, וערכם תלוי בערך המשתנה החופשי x, המופיע בהם. בנוסחאות כמו $(x^2)' = 2x$, לעומת זאת, משתמשים ב- "x^2" וב- "$2x$" כביטויים עבור **_פונקציות_**. למעלה מזאת: המשתנה x קשור בהם (אי אפשר, למשל, להציב במקומו ערכים קונקרטיים: "$(7^2)' = 2 \cdot 7$" הוא, כפי שכבר ראינו, צירוף **_חסר משמעות_**. כמו כן, כלל α ישים: ל- $(z^2)' = 2z$ אותה משמעות כמו ל- $(x^2)' = 2x$). למרות זאת, אין רואים שום אופרטור קשירה, שקושר את x! השימוש הכפול בביטוי כמו "x^2", פעם כמספר, פעם כפונקציה, ואי-השימוש באופרטור קשירה לציון משתנים קשורים, הם (שניהם ביחד וכל אחד לחוד) ערובה ודאית לשגיאות ובלבולים – ובמיוחד כשמשתנים נוספים מעורבים בעניין (כמו ב"נוסחה" $(x^a)' = ax^{a-1}$). יתר על כן: זה גורם לכך, שאפילו ניסוחן של עובדות פשוטות מאוד הופך עניין מסורבל וקשה. נניח, לדוגמה, שאנו רוצים לבטא בנוסחה את העובדה, ש- z הוא מספר, שהנגזרת של "x^2" בו שווה ל- $z + 1$. איך נעשה זאת? בטקסטים הרגילים, הדרך היחידה הפתוחה בפנינו היא להשתמש בשם זמני, תוך היעזרות במלים בעברית, למשל: "z הוא מספר, כך שאם $f(x) = x^2$, אז $f\,'(z) = z + 1$". עם סימון נכון, המאפשר להבדיל בין הביטוי "x^2" עבור מספרים, ו- "$\lambda x.\ x^2$" עבור פונקציות, נכתוב פשוט, בנוסחה אחת:

$$(\lambda x.\ x^2)'(z) = z + 1$$

(או, אפילו, אם יתחשק לנו, $(\lambda z.\ z^2)'(z) = z + 1$: אין הרי קשר בין שני המופעים הראשונים של z כאן, שהם קשורים, ובין השניים האחרונים, שהם חופשיים!).

ועוד: כאשר משתמשים אנו בסימון נכון, ברור לנו, שהנוסחה $(\lambda x.\ x^2)' = \lambda x.\ 2x$ שקולה ל- $(\lambda x.\ x^2)' = \lambda z.\ 2z$. לעומת זאת, הנוסחה $(x^2)' = 2z$ תיחשב כמעט בוודאות כשגויה. למה? דומה, שמפסיקים להתייחס בה ל- "$2z$" כאל ביטוי המייצג פונקציה, או (מה שגרוע יותר) – לא מתייחסים ל- "$2x$" ול- "$2z$" כביטויים המייצגים אותה פונקציה (למרות שברור, שכאשר מגדירים $f(x) = 2x$ ו- $g(z) = 2z$, אזי f ו- g הן אותה פונקציה – ואין מורה, שלא ידגיש זאת!).

לסיום עניין הנגזרת, נחזור לבעיה של הנוסחה $(x^a)' = ax^{a-1}$, בה דשנו כבר בפרק א.5. שם, כזכור, פירשנו אותה כמבטאת זהות בין שני מספרים (שערכם תלוי בערך של a ושל x). כעת נוכל לתת לה פירוש אחר, הקרוב יותר לאופן, שבו משתמשים בה בדרך-כלל. הפעם נפרשה כמבטאת שוויון בין שתי *פונקציות* (שזהותן תלויה בערך של a). באופן מלא ונכון צריכה הנוסחה להיכתב כך:

$$\forall a[(\lambda x.\ x^a)' = \lambda x.\ ax^{a-1}]$$

גם בפירוש זה ברור מיד, למה לא נוכל להציב במקום a ביטוי, המכיל את x כמשתנה חופשי: גם כאן ביטוי כזה אינו חופשי להצבה במקום a בתוך הסוגריים המרובעים, ולכן כלל (14) אינו ישים עבורו. הפעם, אבל, נקשר x על-ידי אופרטורי למדא, לא על-ידי "$\forall x$".

ו. האינטגרל המסוים הינו פונקציה. כדי להבין זאת, הבה נהרהר רגע מה אנו צריכים לדעת כדי שנוכל להתחיל במלאכת חישוב אינטגרל מסוים. ובכן, עלינו לדעת, מהי הפונקציה לה אנו עושים אינטגרל, ואת הגבול התחתון והעליון שלו ("מאין ועד היכן עושים את האינטגרל"). עלינו לקבל אפוא פונקציה ושני מספרים כקלט, ולקבל בחזרה מספר כפלט. ברור, לכן, שאינטגרל מסוים הוא פונקציה "של שלושה משתנים" (שהיא חלקית, כיוון שלא לכל פונקציה אפשר לעשות אינטגרל מכל מספר לכל מספר). ליתר דיוק:

$$\int : ((\mathbf{R} \overset{p}{\rightarrow} \mathbf{R}) \times \mathbf{R} \times \mathbf{R}) \overset{p}{\rightarrow} \mathbf{R}$$

לדוגמה, בסימון האחיד לפונקציות, הביטוי $\int_1^2 x^2dx$ ייכתב כך: $\int(\lambda x.\ x^2, 1, 2)$.

נשים לב, שהמשפט היסודי של החשבון הדיפרנציאלי והאינטגרלי אומר ש:

$$\forall f \in \mathbf{R} \overset{p}{\rightarrow} \mathbf{R}\ \forall a \in \mathbf{R}\ \forall y \in \mathbf{R}\ [(a,y) \text{ -רציפה ב } f \Rightarrow (D\,(\lambda x.\int(f, a, x)))\,(y) = f(y)]$$

ז. גם אינטגרל לא מסוים הינו פונקציה. הפעם עלינו להכניס כקלט פונקציה (חלקית) מ- $\mathbf{R}$ ל- $\mathbf{R}$, ומה שנקבל כפלט יהיה *קבוצה* שלמה של פונקציות. לכן:

$$\int : (\mathbf{R} \overset{p}{\rightarrow} \mathbf{R}) \rightarrow P(\mathbf{R} \overset{p}{\rightarrow} \mathbf{R})$$

(אנו משתמשים, כנהוג, באותו הסימן עבור האינטגרל המסוים והלא מסוים. זהו נוהג, שיש לו גם חסרונות וגם מעלות). למשל, כשאנו כותבים:

$$\int x^2 dx = \frac{x^3}{3} + c$$

כוונתנו, בעצם, היא ש:

$$\int (\lambda x.\ x^2) = \{\lambda x.\ \frac{x^3}{3} + c \mid c \in \mathbf{R}\}$$

V. מושגים בסיסיים הקשורים בפונקציות

א. צמצום של פונקציה

נניח $f: A \to B$ ו- $X \subseteq A$ אז f/X, ***הצמצום של f ל- X***, היא הפונקציה $\lambda x \in X.\ f(x)$. f/X היא אפוא פונקציה הזהה ל- f (כמעט), אך עם תחום הגדרה מצומצם יותר.

נשים לב:

1. $f/X: X \to B$
2. במונחי תורת הקבוצות, $f/X \subseteq f$. למעשה: $f/X = \{\langle x, f(x)\rangle \mid x \in X\}$
3. $f/\mathrm{dom}(f) = f$
4. אם $X \subseteq A$ אז $(\lambda x \in A.\ t) / X = \lambda x \in X.\ t$

נעיר עוד, שניתן להכליל כל זאת לפונקציות חלקיות, בלי לשנות דבר.

ב. מקור ותמונה

נניח $f: A \to B$.

1. כאשר $f(x) = y$ (עבור איזה $x \in A$ ו- $y \in B$), אז קוראים ל- y ***התמונה*** של x לפי y, בעוד x מכונה ***מקור*** של y לפי f (נשים לב, שלכל x ב- A יש תמונה יחידה ב- B, בעוד מספר המקורות של $y \in B$ אינו מוגבל, ויכול גם להיות 0).

2. נגדיר $\hat{f}: P(A) \to P(B)$ על-ידי:

$$\hat{f} = \lambda X \in P(A).\ \{f(x) \mid x \in X\}\ (= \lambda X \in P(A).\ \{y \in B \mid \exists x \in X.\ f(x) = y\})$$

$\hat{f}$ היא פונקציה, המתאימה לכל קבוצה חלקית של A את קבוצת התמונות של איבריה לפי f. אם $X \subseteq A$, אז ל- $\hat{f}(X)$ קוראים ***התמונה*** של X לפי f. מקובל מאוד לסמן את התמונה של X לפי f ב- $f(X)$ במקום $\hat{f}(X)$. זהו נוהג גרוע, כיוון שהוא עלול לגרום לדו-משמעויות ולבלבולים. ברם, כיוון שהוא ***מאוד*** מקובל,

נאמץ אותו בדרך-כלל גם אנו (סימון עדיף, מקובל למדי, הוא $f[X]$, ומדי פעם נשתמש גם בו).

נדגיש שוב: אם $x \in A$, אז $f(x)$ התמונה של x לפי f, היא איבר ב- B ($f(x) \in B$). לעומת זאת, אם $X \subseteq A$, אז $f(X)$ היא איבר ב- $P(B)$ (כלומר $f(X) \subseteq B$). כאשר $x \in A \cap P(A)$ (כלומר, כאשר x גם שייך ל- A וגם חלקי ל- A בעת ובעונה אחת), אז הסימון $f(x)$ הוא דו-משמעי, וזוהי צרה. (הסימון $f[x]$ אינו יוצר בעיה זו!).

דוגמאות:

(א) $(\lambda x \in R.x^2)((-\infty, 3]) = [9,\infty)$

(ב) אם $f = \lambda x \in P(\mathbf{N}).\ 1$ אז:

$f(\varnothing) = 1$ כאשר מסתכלים על $\varnothing$ כעל איבר של $P(\mathbf{N})$.

$f(\varnothing) = \varnothing$ כאשר המובן של "$f(\varnothing)$" הוא "$\hat{f}(\varnothing)$", ומסתכלים על $\varnothing$ כעל קבוצה חלקית של $P(\mathbf{N})$.

במצב כזה נעדיף לכתוב: $f(\varnothing) = 1$ ו- $f[\varnothing] = \varnothing$.

(ג) אם $f: A \to B$ אז $f(A) = \mathrm{Im}(f)$.

3. נניח שוב, ש- $f: A \to B$. נגדיר $\breve{f}: P(B) \to P(A)$ על-ידי:

$$\breve{f} = \lambda X \in P(B).\ \{x \in A \mid f(x) \in X\}$$

$\breve{f}$ היא פונקציה, המתאימה לכל קבוצה חלקית של B את קבוצת מקורותיה לפי f. ל- $\breve{f}(X)$ קוראים *המקור* של X לפי f, ומסמנים אותה בדרך-כלל ב- $f^{-1}(X)$. סימון עדיף, אם כי מקובל פחות, הוא $f^{-1}[X]$ (אנו נשתמש בשניהם).

נדגיש: כאשר $x \in B$, אזי ייתכנו לו מספר מקורות לפי f, וייתכן גם, שאין לו שם שום מקור. לעומת זאת, כש- $X \subseteq B$, אז יש ל- X מקור יחיד לפי f. המונח "מקור של x לפי f" הופך להיות דו-משמעי כש- $x \in B \cap P(B)$. בהמשך נראה, שגם *הסימון* $f^{-1}(x)$ (אך לא $f^{-1}[x]$!) עלול להיות דו-משמעי במקרים מסוימים.

דוגמאות:

(1) אם $f = \lambda x \in \mathbf{R}.\ x^2$, אז

$$f^{-1}([0,1)) = (-1,1)$$

$$f^{-1}([9, \infty)) = (-\infty, -3] \cup [3, \infty)$$

$$f^{-1}([-\frac{1}{2}, 1)) = (-1,1)$$

הבה נפרט את הדוגמה האחרונה. כיוון שכאן $f: \mathbf{R} \to \mathbf{R}$ ו- $[-\frac{1}{2}, 1) \subseteq \mathbf{R}$,

אז לפי הגדרה:

$$\begin{aligned} f^{-1}([-\tfrac{1}{2}, 1)) &= \{x \in \mathbf{R} \mid f(x) \in [-\tfrac{1}{2}, 1)\} \\ &= \{x \in \mathbf{R} \mid x^2 \in [-\tfrac{1}{2}, 1)\} \\ &= \{x \in \mathbf{R} \mid -\tfrac{1}{2} \le x^2 < 1\} \\ &= \{x \in \mathbf{R} \mid x^2 < 1\} \\ &= \{x \in \mathbf{R} \mid -1 < x < 1\} \\ &= (-1, 1) \end{aligned}$$

(2) נניח $g = \lambda x \in \mathbf{N}.\, 1$. אז $g : \mathbf{N} \to \mathbf{N}$ ומתקיים:

$$g^{-1}(\{1,2\}) = \mathbf{N}$$
$$g^{-1}(\{2,3\}) = \varnothing$$

תרגיל

נניח $f: A \to B$ הוכיחי:

(i) $\forall X \in P(A).\, X \subseteq f^{-1}(f(X))$

(ii) $\forall X \in P(B).\, f(f^{-1}(X)) \subseteq X$

מצאי דוגמאות בהן ההכלה ב- (i) וב- (ii) היא הכלה ממש.

ג. הרכבה של פונקציות

אם $f: B \to C$ ו- $g: A \to B$, אז ***ההרכבה*** שלהן, $f \circ g$, הינה הפונקציה $\lambda x \in A.\, f(g(x))$.

נשים לב:

1. $f \circ g : A \to C$
2. $\forall x \in A. (f \circ g)(x) = f(g(x))$
3. שתי ההבחנות הקודמות מהוות ביחד הגדרה שקולה ל- $f \circ g$.
4. הרכבה $f \circ g$ של שתי פונקציות f ו- g מוגדרת אם"ם התמונה של g חלקית לתחום של f.

5. ייתכן ש- $f \circ g$ תהיה מוגדרת, בעוד $g \circ f$ לא.
6. לפי כלל (α), יכולנו בהגדרת $f \circ g$ להשתמש במשתנה אחר במקום x לדוגמה: $f \circ g = \lambda y. f(g(y))$. יש אבל להיזהר מלהשתמש במשתנה, המופיע חופשי בביטויים המייצגים את הפונקציות f ו- g.

דוגמאות:

(1)
$$(\lambda x. x^2) \circ (\lambda x. x+1) = \lambda x. (\lambda x. x^2) \underbrace{((\lambda x. x+1)(x))}$$
$$= \lambda x. \underbrace{(\lambda x. x^2)(x+1)}$$
$$= \lambda x. (x+1)^2$$

הסברים:

א. שוב כתבנו לשם הנוחות רק $\lambda x. x^2$ במקום $\lambda x \in \mathbf{R}. x^2$ וכו'.

ב. כדי להבין את החישוב, סימנו בכל שלב ב- $\underbrace{\qquad}$ את תת הביטוי עליו מופעל כלל (β).

ג. לצורך הנוחות יכולנו להפעיל גם את כלל (α) בכל מקום בו x נקשר. היה אכן יכול להיות ברור יותר, אולי, אם היינו מתחילים כך:
$$(\lambda x. x^2) \circ (\lambda x. x+1) = (\lambda y. y^2) \circ (\lambda z. z+1) = \ldots$$

בדוגמה הבאה נדגים זאת.

(2)
$$(\lambda x. x+1) \circ (\lambda x. x^2) \overset{\alpha}{=} \lambda x. (\lambda y. y+1) \underbrace{((\lambda z. z^2)(x))}$$
$$\overset{\beta}{=} \lambda x. \underbrace{(\lambda y. y+1)(x^2)}$$
$$\overset{\beta}{=} \lambda x. x^2 + 1$$

תכונות של פעולת ההרכבה:

חוק הקיבוץ:

אם $h : A \to B$, $g : B \to C$, $f : C \to D$, אז
$$f \circ (g \circ h) = (f \circ g) \circ h$$

הוכחה

תחום ההגדרה של שני האגפים הוא A (ו- D הוא טווח).

כמו כן, לכל $x \in A$ מתקיים:

$$(f \circ (g \circ h))(x) = f((g \circ h)(x)) = f(g(h(x)))$$

$$((f \circ g) \circ h)(x) = (f \circ g)(h(x)) = f(g(h(x)))$$

לכן $(f \circ (g \circ h))(x) = ((f \circ g) \circ h)(x)$ לכל $x \in A$. מכאן שהפונקציות שוות.

כרגיל, חוק הקיבוץ מאפשר לנו לכתוב פשוט $f \circ g \circ h$ בלי לשים סוגריים, כי מקום הסוגריים אינו משנה. כאשר $f : A \to A$, הרי ניתן להרכיב את f עם עצמו מספר כלשהו של פעמים: $f \circ f \circ \ldots \circ f$. ביטוי זה מקצרים ל- f^n (כאשר n הוא מספר הפעמים ש-f מופיע כאן). למשל: $f^4 = f \circ f \circ f \circ f$.

מה בדבר החוק הקומוטטיבי (חוק החילוף)? ובכן, כאמור למעלה, בדרך כלל $g \circ f$ אינו מוגדר כלל כאשר $f \circ g$ מוגדר. שניהם מוגדרים רק כש- $f : A \to B$ ו- $g : B \to A$. אפילו אז ברוב המקרים $f \circ g \neq g \circ f$, כמו שמראות כבר שתי הדוגמאות הפשוטות, שהבאנו למעלה.

שני מקרים, שחשוב (וקל) לזכור, בהם כן מתקיים החוק הקומוטטיבי, הם:

(i) אם $f : A \to A$, אז $f^m \circ f^n = f^n \circ f^m$. הסיבה: שני האגפים שווים פשוט ל- f^{m+n}. למשל:

$$f^3 \circ f^4 = (f \circ f \circ f) \circ (f \circ f \circ f \circ f) = f \circ f \circ f \circ f \circ f \circ f \circ f$$

(ההוכחה של הנוסחה הכללית היא דומה. נעיר, שגם הנוסחה $(f^n)^m = f^{n \cdot m}$ מתקיימת כאן, כש- $n,m \in \mathbf{N}^+$. ההוכחה שוב קלה).

(ii) אם $f : A \to A$, אז $f \circ i_A = i_A \circ f = f$

ואכן לכל $x \in A$:

$$(f \circ i_A)(x) = f(i_A(x)) = f(x)$$

$$(i_A \circ f)(x) = i_A(f(x)) = f(x)$$

נהוג להגדיר $f^0 = i_A$ כאשר $f : A \to A$. עם הגדרה זו הנוסחאות $f^m \circ f^n = f^{m+n}$ ו- $(f^n)^m = f^{n \cdot m}$ נכונות לכל n ו- m ב- $\mathbf{N}$.

את העובדות על i_A, שראינו כאן, אפשר להכליל באופן הבא:

טענה:

אם $f: A \to B$, אז $f \circ i_A = f$ ו- $i_B \circ f = f$.

את ההוכחה נשאיר כתרגיל לקורא.

%%

תת-סדרות

דוגמה חשובה לשימוש בהרכבת פונקציות היא מה שנקרא ***"תת-סדרות"***. אם a סדרה של מספרים ממשיים (כלומר $a : \mathbf{N}^+ \to \mathbf{R}$), ו- n הינה ***סדרה*** (***לא*** מספר!) עולה של מספרים טבעיים (כלומר $n : \mathbf{N}^+ \to \mathbf{N}$, ומתקיים ש- $\forall i \forall j.\ i < j \Rightarrow n_i < n_j$ (דהיינו: $\forall i \forall j.\ i < j \Rightarrow n(i) < n(j)$), אז הרכבתן $a \circ n$ נקראת תת-סדרה של a.

לדוגמה: אם $a = \lambda n \in \mathbf{N}^+ . \frac{1}{n}$ $\left(= \left\{ \frac{1}{n} \right\}_{n=1}^{\infty} \right)$ ו- $n = \lambda k \in \mathbf{N}^+ . k^2$ $\left(= \left\{ k^2 \right\}_{k=1}^{\infty} \right)$,

אז

$$
\begin{aligned}
a \circ n &= \lambda k \in \mathbf{N}^+ . a(n(k)) && (= \lambda k \in \mathbf{N}^+ . a_{n_k}) \\
&= \lambda k \in \mathbf{N}^+ . a(k^2) && (= \lambda k \in \mathbf{N}^+ . a_{k^2}) \\
&= \lambda k \in \mathbf{N}^+ . \frac{1}{k^2} \\
&\overset{\alpha}{=} \lambda n \in \mathbf{N}^+ . \frac{1}{n^2}
\end{aligned}
$$

המסקנה היא, ש- $\lambda n. \frac{1}{n^2}$ היא תת-סדרה של $\lambda n. \frac{1}{n}$ (או, במלים יותר סטנדרטיות: $\left\{ \frac{1}{n^2} \right\}_{n=1}^{\infty}$ היא תת-סדרה של $\left\{ \frac{1}{n} \right\}_{n=1}^{\infty}$).

כמו שנרמז בסוגריים בחישובים למעלה, במקום לכתוב $a(n(k))$ או $(a \circ n)(k)$, כותבים a_{n_k}. חשוב מאוד להבין, שהשוני בסימון אינו שוני בתוכן. גם כשכותבים "a_{n_k}", a ו- n הם שמות של סדרות (או משתנים עבור סדרות), בעוד k – משתנה עבור מספרים טבעיים. $\left\{ a_{n_k} \right\}_{k=1}^{\infty}$ אינו אלא $\lambda k.\ a(n(k))$!

הערה:

כדאי לשים לב, שבחישוב למעלה אין שום קשר בין המשתנה n בשורה האחרונה $(\lambda n \in \mathrm{N} . \frac{1}{n^2})$ לבין ה- n שבביטוי $a \circ n$!

%%

ד. חד-חד-ערכיות, על, שקילות, פונקציה הפוכה

הגדרה:

1. פונקציה f נקראת ***חד-חד-ערכית*** (***ח.ח.ע.***) אם

$$\forall x \in \mathrm{Dom}(f)\ \forall y \in \mathrm{Dom}(f).\ (f(x) = f(y) \Rightarrow x = y)$$

או, בניסוח שקול לוגית:

$$\forall x \in \mathrm{Dom}(f)\ \forall y \in \mathrm{Dom}(f)\ (x \neq y \Rightarrow f(x) \neq f(y))$$

2. פונקציה $f: A \to B$ נקראת **על**[16] B אם B הינה התמונה של A לפי f, דהיינו: כל ערך של B מתקבל. פורמלית, f היא על B אם:

$$\forall y \in B\ \exists x \in A.\ y = f(x)$$

3. פונקציה $f: A \to B$, שהיא גם ח.ח.ע. וגם על B, נקראת ***פונקציית שקילות*** בין A ל- B (או מ- A על B). f היא לכן פונקציית שקילות בין A ל- B אם:

$$\forall y \in B\ \exists! x \in A.\ y = f(x)$$

דוגמאות:

1. נתחיל בפונקציות מ- **R** ל- **R**.

(i) $\lambda x.\, e^x$ היא ח.ח.ע., אך לא על **R** (כי $e^x > 0$ לכל x).

(ii) $\lambda x.\, x^2$ היא לא ח.ח.ע. ולא על **R** ($(\lambda x.\, x^2)(-1) = (\lambda x.\, x^2)(1) = 1$ ו- $\forall x.\, x^2 > 0$).

(iii) $f = \lambda x.\, x^2(x-1)$ היא על **R**, אך היא אינה ח.ח.ע. ($f(0) = f(1) = 0$, ולכן f אינה ח.ח.ע.. שהיא על **R** נובע מהעובדה, שהיא רציפה וש-

$$\lim_{x \to -\infty} f(x) = -\infty \quad , \lim_{x \to \infty} f(x) = \infty$$

זה מאפשר להפעיל את משפט ערך הביניים של החשבון הדיפרנציאלי).

(iv) $\lambda x.\, x + 1$ היא גם ח.ח.ע. וגם על **R**.

---

[16] לא לבלבל עם "מעל" A ! (באנגלית "on" ו- "onto").

2. לכל A הפונקציה i_A היא פונקציית שקילות מ- A על A.

3. הפונקציה: $\lambda x \in \{\varnothing,\{\varnothing\}\}.\ \text{if } x = \varnothing \text{ then } 1 \text{ else } 2$ היא פונקציית שקילות מ- $\{\varnothing,\{\varnothing\}\}$ על $\{1,2\}$. היא איננה על $\{1,2,3\}$ (אם זה נלקח בתור הטווח), אך עדיין ח.ח.ע. בטווח זה (ובכל טווח אחר).

4. $\lambda A \in P(E).\chi_A^{(E)}$ (ראה סעיף (IV) למעלה) היא פונקציית שקילות בין $P(E)$ ובין $E \to \{0,1\}$. (***תרגיל:*** להוכיח זאת!).

5. אם f היא פונקציה ח.ח.ע., אז f היא פונקציית שקילות בין התחום של f והתמונה של f.

המשפט הבא הוא שימושי, עת רוצים להוכיח, שפונקציה מסוימת היא פונקציית שקילות:

משפט:

$f : A \to B$ היא פונקציית שקילות מ- A על B אם"ם קיימת פונקציה $g : B \to A$, כך ש- $g \circ f = i_A$ ו- $f \circ g = i_B$. יתר על כן: אם פונקציה כזו קיימת, אז היא יחידה.

הוכחה:

$(\Rightarrow)$ נניח שפונקציה g כנ"ל קיימת. נראה ש-f פונקציית שקילות מ- A על B.

ח.ח.ע.:

נניח ש- $f(x) = f(y)$ ל- y,x כלשהם ב- A.

$$\Rightarrow g(f(x)) = g(f(y))$$
$$\Rightarrow (g \circ f)(x) = (g \circ f)(y)$$
$$\Rightarrow i_A(x) = i_A(y)$$
$$\Rightarrow x = y$$

על B:

יהי $y \in B$. נקח $x = g(y)$. אז $x \in A$ ומתקיים:

$$f(x) = f(g(y)) = (f \circ g)(y) = i_B(y) = y$$

לכן $\exists x \in A.\ y = f(x)$.

בטרם נוכיח את הכיוון ההפוך נראה, שאם g כנ"ל קיימת, אז היא יחידה. נניח אפוא, ש- $h : B \to A$ מקיימת אותם תנאים (כלומר: $h \circ f = i_A$, $f \circ h = i_B$). נראה ש- $h = g$. יהי אפוא $y \in B$ כלשהו. אזי:

$$y = i_B(y) = (f \circ h)(y) = (f \circ g)(y)$$

ולכן

$$f(h(y)) = f(g(y))$$

הראינו אבל מקודם, שאם g כנ"ל קיימת, אז f הינה ח.ח.ע.. לכן, מהשוויון האחרון נובע, ש- $h(y) = g(y)$. כיוון שזה נכון לכל $y \in B$, הרי $h = g$.

($\Leftarrow$) נניח $f : A \to B$ ח.ח.ע. ועל B. נגדיר:

$$g = \lambda y \in B.\ \iota x \in A.\ f(x) = y$$

(כלומר: g מתאימה לכל y ב- B את האיבר היחיד ב- A כך ש- $f(x) = y$. איבר כזה קיים, כי f הינה על B, והוא יחיד, כי f ח.ח.ע.). אז:

$$f \circ g = \lambda y \in B.\ f(g(y)) = \lambda y \in B.\ f(\iota x \in A.\ f(x) = y) = \lambda y \in B.\ y = i_B$$

$$g \circ f = \lambda x \in A. g(f(x)) \overset{(\alpha)}{=} \lambda z \in A. \underbrace{g(f(z))}$$

$$\overset{(\beta)}{=} \lambda z \in A. \iota x \in A. f(x) = f(z)$$

$$= \lambda z \in A. z$$

$$= i_A$$

(הסתמכנו כאן על העובדה הטריביאלית, שאם f היא ח.ח.ע., אז $(\iota x \in A.\ f(x) = f(z)) = z$, כלומר: ה- x היחידי, שערך f בו שווה אז לערך של f ב- z, הוא z עצמו).

מ.ש.ל.

הגדרה:

אם $f : A \to B$ היא פונקציית שקילות, אז לפונקציה g, שקיומה הובטח במשפט הקודם, קוראים הפונקציה *ההפוכה* של f, ומסמנים אותה ב- f^{-1}. (במקום להגיד, ש- f פונקציית שקילות, אומרים לכן לעתים קרובות, ש- f *הפיכה*).

נשים לב:

(i) $f^{-1}: B \to A$ (או, בצורה כללית יותר: אם f ח.ח.ע., אז $f^{-1}: \mathrm{Im}(f) \to B$).

(ii) $f^{-1} \circ f = i_A$, $f \circ f^{-1} = i_B$.

(iii) f^{-1} מתאפיינת על-ידי התכונה:

$$\boxed{\forall x \in A.\ \forall y \in B.\ y = f(x) \Leftrightarrow x = f^{-1}(y)}$$

(iv) לפי ההגדרות הרשמיות ו- (iii):

$$f^{-1} = \{<y,x> \mid <x,y> \in f\}$$

(v) אם $f: A \to A$ היא פונקציית שקילות, אז את (ii) אפשר לרשום בצורה:

$$f \circ f^{-1} = f^{-1} \circ f = f^0$$

אזהרות:

1. הביטוי f^{-1} מייצג פונקציה אך ורק כש- f פונקציה ח.ח.ע.. f^{-1} הינה אז פונקציה מהתמונה של f על התחום של f.

2. במקום אחר בפרק זה השתמשנו בביטוי f^{-1} במובן אחר. הבה נבהיר אפוא: אם $f: A \to B$, אז עבור $x \in B$, הביטוי $f^{-1}(x)$ הוא בעל מובן, רק אם f הינה ח.ח.ע. ועל B. בהתקיים תנאי זה, $f^{-1}(x) \in A$ ומסמן את הערך של הפונקציה ההפוכה ל-f עבור הארגומנט x. כאשר $X \subseteq B$, לעומת זאת, $f^{-1}(X) \in P(A)$ ומוגדר אפילו אם f איננה ח.ח.ע. ועל (ולמעשה אפילו כש-f רק חלקית!). $f^{-1}(X)$ מייצג אז את המקור של X לפי f. לבסוף, אם f ח.ח.ע. ועל B, ו- $x \in B \cap P(B)$, אז הביטוי $f^{-1}(x)$ הוא דו-משמעי!

דוגמאות:

1. $(\lambda x \in \mathbf{R}.\ x + 1)^{-1} = \lambda x \in \mathbf{R}.\ x - 1$

שאלה 1

כיצד מוכיחים זאת?

תשובה

מרכיבים את $\lambda x.\ x + 1$ ואת $\lambda x.\ x - 1$ בשתי הצורות האפשריות, ורואים, שבשתיהן יוצא $i_{\mathbf{R}}$ (כלומר: $\lambda x \in \mathbf{R}.\ x$).

שאלה 2

כיצד מוצאים זאת?

תשובה

מציבים $y = x + 1$ ומנסים "לבודד" את x, דהיינו: לבטא את x בעזרת y. כאן זה קל: $x = y - 1$. המסקנה היא, ש- $\lambda y.\, y - 1$ היא הפונקציה ההפוכה, או, בעזרת כלל (α): $\lambda x.\, x - 1$.

---

2. $i_A^{-1} = i_A$

3. $(\lambda x \in \{\varnothing,\{\varnothing\}\}.\ \text{if } x = \varnothing \text{ then } 1 \text{ else } 2)^{-1} = \lambda x \in \{1,2\}.\ \text{if } x = 1 \text{ then } \varnothing \text{ else } \{\varnothing\}$

4. $$\begin{aligned}(\lambda A \in P(E).\chi_A^{(E)})^{-1} &= \lambda f \in E \to \{0,1\}.f^{-1}(\{1\}) \\ &= \lambda f \in E \to \{0,1\}.\{x \in E \mid f(x) = 1\}\end{aligned}$$

עובדות פשוטות על הפונקציה ההפוכה

טענה:

(1) אם $f : A \to B$ פונקציית שקילות, אז $f^{-1} : B \to A$ גם היא פונקציית שקילות, ו- $(f^{-1})^{-1} = f$.

(2) אם $f : A \to B$ ו- $g : B \to C$ הן פונקציות שקילות, אז כך גם $g \circ f$, ומתקיים:

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}$$

הוכחה

(1) מהשוויונות $f \circ f^{-1} = i_B$ ו- $f^{-1} \circ f = i_A$ נובע, שיש פונקציה $g : A \to B$, כך ש- $f^{-1} \circ g = i_A$, $g \circ f^{-1} = i_B$ $(g = f)$. מהמשפט הקודם נובע לכן, ש- f^{-1} היא פונקציית שקילות, וש- f היא הפונקציה ההפוכה לה, דהיינו $(f^{-1})^{-1} = f$.

(2) ברור ש- $f^{-1} \circ g^{-1} : C \to A$. כמו כן:

$$(g \circ f)(f^{-1} \circ g^{-1}) = (g \circ (f \circ f^{-1})) \circ g^{-1} = (g \circ i_B) \circ g^{-1} = g \circ g^{-1} = i_C$$

$$(f^{-1} \circ g^{-1})(g \circ f) = (f^{-1} \circ (g^{-1} \circ g)) \circ f = (f^{-1} \circ i_B) \circ f = f^{-1} \circ f = i_A$$

לכן, לפי המשפט הקודם, $g \circ f$ הפיכה, ו- $f^{-1} \circ g^{-1}$ היא הפונקציה ההפוכה.

מסקנה:

אם $f:A\to A$ היא פונקציית שקילות, אז $(f^{n})^{-1}=(f^{-1})^{n}$.

הוכחה:

$$(f^n)^{-1}=(f\circ f\circ\ldots\circ f)^{-1}\overset{(2)}{=}f^{-1}\circ f^{-1}\circ\ldots\circ f^{-1}=(f^{-1})^n$$

סימון: f^{-n} יסמן את $(f^{-1})^{n}$ ואת $(f^{n})^{-1}$ (השווים לפי המסקנה האחרונה).

בשלב זה מתחיל אולי להתחוור לקוראים פשר הסימון f^{-1} לפחות כאשר $f:A\to A$ ו-f הינה הפיכה, מוביל סימון זה לנכונות כללי החזקות הבאים עבור כל n ו- m ***שלמים*** (כלומר $n,m\in\mathbf{Z}$):

$$\boxed{\begin{aligned}f^n\circ f^m&=f^{n+m}\\ (f^n)^m&=f^{n\cdot m}\end{aligned}}$$

תרגיל

הוכח זהויות אלו.

אזהרה:

הזהות $(f\circ g)^n=f^n\circ g^n$ ***אינה*** נכונה בדרך-כלל, אפילו אם גם f וגם g הן פונקציות מ- A ל- A (היא נכונה אבל כאשר $f\circ g=g\circ f$).

לסיום סעיף זה נביא מספר הכללות חשובות של עובדות, שראינו קודם. את ההוכחות נשאיר לקוראים.

טענה:

1. אם $f:A\to B$ ו- $g:B\to C$ הן ח.ח.ע., אז $g\circ f$ היא ח.ח.ע..
2. אם $f:A\to B$ היא על B ו- $g:B\to C$ היא על C, אז $g\circ f$ היא על C.
3. אם $f:A\to B$, $g:B\to A$ ו- $g\circ f=i_A$, אז f ח.ח.ע.,ו- g היא על A.

טבלה ב.6 מסכמת את ההגדרות של המושגים החשובים ביותר הקשורים בפונקציות.

טבלה ב.6: פונקציות – מושגים חשובים

(1)		פונקציית ***הזהות*** מעל קבוצה A היא $i_A = \lambda x \in A.\, x$.
(2)		אם $E \subseteq A$, אז הפונקציה ***האופיינית*** של A יחסית ל- E היא: $\chi_A^{(E)} = \lambda x \in E.$ if $x \in A$ then 1 else 0
(3)		אם $f: A \to B$ ו- $X \subseteq A$, אז ***הצמצום*** של f ל- X הוא: $f/X = \lambda x \in X.\, f(x)$ $\quad (f/X : X \to B)$
(4)		אם $f: A \to B$ ו- $g: B \to C$, אז ***ההרכבה*** $g \circ f$ היא: $g \circ f = \lambda x \in A.\, g(f(x))$ $\quad (g \circ f: A \to C)$
(5)	(i)	אם $f: A \to B$ ו- $X \subseteq A$, אז ***התמונה*** של X לפי f היא: $f[X] = \{f(x) \mid x \in A\}$ $\quad (f[X] \subseteq B)$
	(ii)	אם $f: A \to B$ ו- $X \subseteq B$, אז ***המקור*** של X לפי f הוא: $f^{-1}[X] = \{x \in A \mid f(x) \in X\}$ $\quad (f^{-1}[X] \subseteq A)$
(6)	(i)	פונקציה f מעל A נקראת ***חד-חד-ערכית*** (חח"ע) אם: $\forall x \in A\ \forall y \in A.\, f(x) = f(y) \Rightarrow x = y$
	(ii)	פונקציה $f: A \to B$ נקראת ***על*** B אם $B = f[A]$, כלומר: $\forall y \in B\ \exists x \in A.\, y = f(x)$
	(iii)	$f: A \to B$ נקראת ***פונקציית שקילות*** (בין A ל- B) אם היא חח"ע ועל B.
(7)		אם $f: A \to B$ היא פונקציית שקילות, אז ***הפונקציה ההפוכה ל-*** f היא: $f^{-1} = \lambda x \in B.\, \iota y \in A.\, f(y) = x$ $\quad (f^{-1} : B \to A)$

ה. פונקציית Curry ושימושיה

אחת הדרכים המקובלות בשפות מחשב מתקדמות לטיפול בפונקציות של שני משתנים (או יותר), היא לעשות להן רדוקציה לפונקציה של משתנה אחד. דבר זה נעשה על-ידי כך שמספקים את הקלטים באופן סדרתי, זה אחר זה, ולא את כולם בבת-אחת. כאשר מגיע הקלט הראשון, המחשב יוצר בעזרתו פונקציה ***חד***-מקומית. פונקציה חד-מקומית זו מופעלת בהגיע הקלט השני, וכך מתקבל הפלט הסופי. מה שהמחשב עושה, אפוא, הוא לקחת פונקציה דו-מקומית $f: A \times B \to C$ ולעשות לה טרנספורמציה לפונקציה ב- $A \to (B \to C)$, הנקראת f^{curry}. הרעיון עליו מבוססת טרנספורמציה זו הוא פשוט ביותר. לפי כלל (η), אם $f \in A \times B \to C$, אזי

$$f = \lambda x \in A, y \in B.\, f(x,y)$$

מה שהטרנספורמציה עושה, בעיקרו של דבר, הוא להחליף את הפסיק בנקודה:

$$f^{\text{Curry}} = \lambda x \in A.\ \lambda y \in B.\, f(x,y)$$

לדוגמה, אם f היא פונקציית החיבור +, אז:

$$+^{\text{curry}} = \lambda x \in \mathbf{R}.\ \lambda y \in \mathbf{R}.\ x + y$$

לכן

$$+^{\text{curry}}(3) \overset{\beta}{=} \lambda y \in \mathbf{R}.\ 3 + y$$

אנו רואים, אם כך, שבעוד הפונקציה + אינה יכולה לעשות דבר, לפני שיעמדו לרשותה שני הארגומנטים, הפונקציה $+^{\text{curry}}$ עושה משהו מועיל עם קלט אחד בלבד: עם הקלט 3, למשל, היא יוצרת את הפונקציה $\lambda y \in \mathbf{R}.\ 3 + y$. עתה:

$$(+^{\text{curry}}(3))(4) = (\lambda y \in \mathbf{R}.\ 3 + y)\,(4) \overset{\beta}{=} 3 + 4 = 7$$

אם נסכם: בצורה של $+^{\text{curry}}$ פונקציית החיבור מקבלת תחילה את הקלט הראשון (3 בדוגמה) ומחזירה פונקציה ב- $\mathbf{R} \to \mathbf{R}$ (פונקציית החיבור עם 3, בדוגמה שלנו). עם קבלת הקלט השני (4 בדוגמה שלנו) מופעלת פונקציה זו מ- $\mathbf{R} \to \mathbf{R}$ על קלט זה, וכך מתקבלת התוצאה הסופית.

הטרנספורמציה מ- f ל- f^{Curry} היא בעצמה פונקציה, ששמה המלא הוא פונקציית Curry. אנו נסמן אותה, מטעמי חיסכון, ב- Cu. נפרט:

$$Cu : (A \times B \to C) \to (A \to (B \to C))$$

ו- $Cu(f) = f^{\text{Curry}}$ עבור $f \in A \times B \to C$.

או, ביתר פירוט:[17]

$$Cu = \lambda f \in A \times B \to C.\ \lambda x \in A.\ \lambda y \in B.\ f(x,y)$$

הפונקציה Cu היא פונקציית שקילות בין $A \times B \to C$ ו- $A \to (B \to C)$. נוכיח זאת על-ידי שנראה, שיש לה פונקציה הפוכה. פונקציה הפוכה זו ידועה בשם UnCurry, ואנו נסמנה בקיצור ב- U. הגדרת U הינה:

$$U = \lambda g \in A \to (B \to C).\ \lambda x \in A, y \in B.\ (g(x))(y)$$

קל לוודא, שאכן $U : (A \to (B \to C)) \to (A \times B \to C)$. נראה, שהיא הפוכה ל- Cu. דבר זה לא כרוך ביותר מאשר חישוב פשוט בעזרת הכללים α, β ו- η:

(i) אם $h \in A \to (B \to C)$, אז

$$\begin{aligned}
(Cu \circ U)(h) &= Cu(\underbrace{U(h)}) \\
&\overset{\beta}{=} \underbrace{Cu(\lambda x \in A, y \in B.(h(x))(y))} \\
&\overset{\beta}{=} \lambda x \in A.\lambda y \in B.\underbrace{(\lambda x \in A, y \in B.(h(x))(y))(x,y)} \\
&\overset{\beta}{=} \lambda x \in A.\underbrace{\lambda y \in B.(h(x))(y)} \\
&\overset{\eta}{=} \underbrace{\lambda x \in A.h(x)} \\
&\overset{\eta}{=} h
\end{aligned}$$

מכאן ש- $Cu \circ U = i_{A \to (B \to C)}$.

הערה:

השורה השלישית בחישוב זה עלולה להיות מבלבלת. יכולנו, לכן, לשנות לפניה את הגדרת U בעזרת כלל (α) ל:

$$U = \lambda g \in A \to (B \to C).\ \lambda a \in A, b \in B.\ (g(a))(b)$$

---

[17] למעשה, היינו צריכים לכתוב $Cu^{A,B,C}$, כיוון שלכל שלוש קבוצות A,B,C יש את פונקציית Curry שלהן (ממש כשם שלכל קבוצה A יש את פונקציית הזהות שלה, ולכן הסימון i_A). לצורך נוחות הקריאה, אנו משמיטים את האזכור המפורש של A, B ו- C.

ואז היינו מקבלים בשורה השלישית:

$$\lambda x \in A.\lambda y \in B.\underbrace{(\lambda a \in A, b \in B.(h(a))(b))(x,y)}$$

ההמשך היה ללא שינוי.

(ii) אם $h \in A \times B \to C$, אז

$$
\begin{aligned}
(U \circ Cu)(h) &= U(\underbrace{Cu(h)}) \\
&\overset{\alpha,\beta}{=} U(\lambda a \in A.\, \lambda b \in B.\, h(a,b)) \\
&\overset{\beta}{=} \lambda x \in A, y \in B.\, (\underbrace{(\lambda a \in A.\, \lambda b \in B.\, h(a,b))(x)})\,(y) \\
&\overset{\beta}{=} \lambda x \in A, y \in B.\, \underbrace{(\lambda b \in B.\, h(x,b))\,(y)} \\
&\overset{\beta}{=} \lambda x \in A, y \in B.\, h(x,y) \\
&\overset{\eta}{=} h
\end{aligned}
$$

מכאן ש- $U \circ Cu = i_{A \times B \to C}$.

מ.ש.ל.

לפונקציית Curry יש חשיבות רבה במתמטיקה (הגם שהיא לא ידועה שם בשם זה). כדוגמה נביא את נושא ***הנגזרת***. המושג הבסיסי בנושא נגזרות, ממנו נגזרים כל השאר, הוא המושג של ***נגזרת של פונקציה בנקודה***. זהו ***מספר***, שאינטואיטיבית שווה לשיפוע המשיק של גרף הפונקציה באותה נקודה. עתה, כדי שנוכל לחפש נגזרת של פונקציה בנקודה, עלינו לדעת שני דברים:

(א) באיזו פונקציה מדובר.

(ב) באיזו נקודה מדובר.

נסמן את הפעולה של מציאת נגזרת של פונקציה בנקודה על-ידי D^*. D^* היא פונקציה המצפה לשני קלטים: הראשון – פונקציה חלקית מ- $\mathbf{R}$ ל- $\mathbf{R}$, השני – מספר ממשי. הפלט הוא מספר ממשי. מכאן:

$$D^* : ((\mathbf{R} \overset{p}{\to} \mathbf{R}) \times \mathbf{R}) \overset{p}{\to} \mathbf{R}$$

כאשר D^* מוגדר, כידוע, על-ידי:

$$D^*(f, x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

דהיינו:

$$D^* = \lambda f \in \mathbf{R} \overset{p}{\to} \mathbf{R},\ x \in R.\ \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

בשלב הבא בלימודי חדו"א עוברים לדבר על ***הפונקציה הנגזרת***, f' (או $D(f)$) של פונקציה f. אם f פונקציה חלקית מ- **R** ל- **R**, אז נגזרתה, $D(f)$, גם היא פונקציה חלקית מ- **R** ל- **R**, ומתקיים:

$$D(f) = \lambda x \in \mathbf{R}.\ D^*(f, x)$$

מכאן ש- D היא פונקציה ב- $(\mathbf{R} \overset{p}{\to} \mathbf{R}) \to (\mathbf{R} \overset{p}{\to} \mathbf{R})$, ומתקיים:

$$D = \lambda f \in \mathbf{R} \overset{p}{\to} \mathbf{R}.\ \lambda x \in \mathbf{R}.\ D^*(f, x)$$

אנו רואים אפוא, ש- $D = Cu(D^*)$: הפעולה של גזירת פונקציות אינה אלא התוצאה של הפעלת פונקציית Curry על הפעולה של מציאת נגזרת של פונקציה בנקודה.

דוגמה זו מבליטה באופן מיוחד את יעילותו האפשרית של הרעיון לטפל בקלטים אחד אחד, במקום בבת-אחת. מציאה ישירה של נגזרת של פונקציה בנקודה הינה בעיה לא פשוטה של חישוב גבולות. למשל:

$$D^*(\lambda x.\ x^2, 3) = \lim_{h\to 0} \frac{(3+h)^2 - 9}{h} = \lim_{h\to 0} \frac{6h + h^2}{h} = 6$$

לעומת זאת, בחישובים מעשיים אנו תחילה מטפלים ב- $(\lambda x.\ x^2)$ ומחשבים את נגזרתה, שהיא פונקציה, ***ואז*** מפעילים פונקציה נגזרת זו על 3:

$$D(\lambda x.\ x^2) = \lambda x.\ 2x$$

ולכן

$$D^*(\lambda x.\ x^2, 3) = (D(\lambda x.\ x^2))(3) = (\lambda x.\ 2x)(3) = 2 \cdot 3 = 6$$

%%

שימוש נוסף: נגזרות חלקיות

אחד הכלים החשובים בחקר "פונקציות של שני משתנים" הוא מושג הנגזרת החלקית $\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$. הרעיון ב- $\frac{\partial F}{\partial x}$, למשל, הוא שגוזרים את F כאילו היתה פונקציה של x בלבד, כאשר y משמש כפרמטר. ב- $\frac{\partial F}{\partial y}$ מתחלפים, כמובן, התפקידים. לדוגמה:

$$\text{(i)} \quad \frac{\partial x^y}{\partial x} = yx^{y-1} \qquad\qquad \text{(ii)} \quad \frac{\partial x^y}{\partial y} = x^y \ln x$$

מה שכתבנו למעלה הוא צורה סטנדרטית של הצגת נגזרות חלקיות וזהויות הקשורות בהן. נבדוק עתה מה באמת מתרחש כאן.

ממבט ראשון התשובה פשוטה. בחישוב $\frac{\partial x^y}{\partial y}$, למשל, אנו מסתכלים בפונקציה $\lambda x.\, x^y$ (שבה x הוא משתנה, ו- y – פרמטר) וגוזרים אותה. לכן:

$$\frac{\partial x^y}{\partial y} = D(\lambda x.\, x^y)$$

תשובה פשוטה זו היא מטעה! ראשית, אם נתבונן ב"זהות" המתקבלת:

$$D(\lambda y.\, x^y) = x^y \cdot \ln x$$

הרי ברור, שמשהו כאן לא בסדר. $D(\lambda y.\, x^y)$ הוא ביטוי המתאר ***פונקציה*** (שבו x מופיע כפרמטר). $x^y \cdot \ln x$ ***אינו*** ביטוי עבור פונקציה. זהו ביטוי עבור ***מספר***.
מה שנכון הוא ש-

$$D(\lambda y.\, x^y) = \lambda y.\, x^y \ln x$$

נוסחה זו היא אכן מדויקת – אך לא לכך הכוונה בנגזרות חלקיות! נגזרת חלקית של "פונקציה של שני משתנים" אמורה להיות ***בעצמה*** "פונקציה של שני משתנים", לא רק אחד. השימוש בנגזרות חלקיות שניות, כמו $\frac{\partial^2 F}{\partial x \partial y}$, מלמד על כך. מה שהנוסחה (ii) בעמוד הקודם רוצה באמת לבטא, הוא ש:

$$\frac{\partial}{\partial y}(\lambda x, y.\, x^y) = \lambda x, y.\, x^y \ln x$$

זה מחזיר אותנו לשאלה המקורית: מה זה בעצם $\frac{\partial}{\partial y}$?

התשובה היא, שראשית, את מה שקיבלנו על-ידי הפעלת D, $\lambda y.\, x^y \ln x$, אנו רוצים להפוך לפונקציה של שני משתנים. לכך נחוצה אבסטרקציה על x. אנו נרצה לעבור לכן ל- $\lambda x.\, \lambda y.\, x^y \ln x$ על מנת לקשור את x. זה אבל לא מספיק. מה שמתקבל הוא פונקציה חלקית ב- $\mathbf{R} \xrightarrow{p} (\mathbf{R} \xrightarrow{p} \mathbf{R})$, ולא ב- $\mathbf{R} \times \mathbf{R} \xrightarrow{p} \mathbf{R}$. לכן יש עוד להפעיל את U, הפונקציה ההפוכה ל- Curry. ואכן:

$$U(\lambda x.\, D(\lambda y.\, x^y)) = \lambda x,y.\, x^y \ln x$$

כללית,

$$\frac{\partial F}{\partial y} = U(\lambda x.\, D(\lambda y.\, F(x,y))) = U(\lambda x.\, D((Cu(F))(x)))$$

ומכאן

$$\frac{\partial}{\partial y} = \lambda F \in \mathbf{R} \times \mathbf{R} \xrightarrow{p} \mathbf{R}.\, U(\lambda x.\, D(\lambda y.\, F(x, y)))$$

לגבי $\frac{\partial}{\partial x}$, הנוסחה מסובכת טיפה יותר:

$$\frac{\partial}{\partial x} = \lambda F \in \mathbf{R} \times \mathbf{R} \xrightarrow{p} \mathbf{R}.\, U^*(\lambda y.\, D(\lambda x.\, F(x, y)))$$

כאשר

$$U^* = \lambda G \in \mathbf{R} \xrightarrow{p} (\mathbf{R} \xrightarrow{p} \mathbf{R}).\, \lambda x \in R, y \in R.\, G(y, x)$$

כדאי לשים לב, שכיוון שהמשתנים x ו- y בנוסחאות אלו הם קשורים, ניתן להחליפם במשתנים אחרים. אין לכן, למעשה, כל קשר בין הפונקציה $\frac{\partial}{\partial y}$ (למשל) ובין המשתנה y! ב- $\frac{\partial}{\partial y}$ הכוונה בעצם הינה לגזירה "לפי המקום השני".

%%

ב.5. יחסים

I. הגדרת המושג "יחס"

מושג ה"יחס" (relation) הוא אחד המושגים הבסיסיים של המתמטיקה והידע האנושי בכלל.

דוגמאות:

(א) יחס הסדר בין מספרים ממשיים.

(ב) יחס ההתחלקות בין מספרים טבעיים ("a מתחלק ב- b").

(ג) יחס החפיפה בין משולשים.

(ד) היחס "בין" בין נקודות ("הנקודה A נמצאת בין הנקודות B ו- C").

(ה) יחס קרבת המשפחה בין בני האדם.

(ו) יחס ההרשמה בין סטודנטים וקורסים.

כרגיל, בתורת הקבוצות, אנו משתדלים להגדיר כל מושג בעזרת המושגים הבסיסיים של קבוצה ושייכות. בנוגע ליחס, מה שחשוב הוא מי מתייחס למי. כשמדובר ביחסים ***בינאריים*** (כלומר: דו-מקומיים), ניתן לתאר כל עובדה מסוג זה על-ידי ***זוג סדור***, בו הרכיב הראשון מתייחס לרכיב השני ביחס בו מדובר. זה מוביל להגדרה הבאה:

הגדרה

(1) יחס (בינארי) R הוא קבוצה של זוגות סדורים, דהיינו: $\forall z \in R\ \exists a \exists b.\ z = \langle a,b\rangle$.

(2) קבוצה R היא ***יחס מקבוצה A לקבוצה B***, אם $R \subseteq A \times B$, כלומר:

$$\forall z \in R\ \exists a\ \exists b.\ z = \langle a,b\rangle \wedge a \in A \wedge b \in B$$

(3) יחס מ- A ל- A נקרא גם יחס ***על*** A.

תרגיל

כל יחס R הוא יחס על - $\cup(\cup(R))$.

דוגמאות:

1. $< \; = \{\langle a,b\rangle \in \mathbf{R}\times\mathbf{R} \mid \exists z \in \mathbf{R}^+.\ b = a + z^2\}$

2. $\backslash \; = \{\langle a,b\rangle \in \mathbf{N}\times\mathbf{N} \mid \exists k \in \mathbf{N}.\ b = k \cdot a\}$

3. אם T היא קבוצת המשולשים במרחב, אזי:
$\cong \; = \{\langle a,b\rangle \in T\times T \mid$ b חופף ל- $a\}$

4. אם S קבוצת הסטודנטים ו- Co קבוצת הקורסים, אז יחס ההרשמה בין סטודנטים ובין קורסים הוא:
$\mathrm{Reg} = \{\langle x,y\rangle \in S\times Co \mid$ y רשום ל- $x\}$

הערות:

1. מקובל מאוד לכתוב aRb במקום $\langle a,b\rangle \in R$. כך כותבים $a<b$ במקום $\langle a,b\rangle \in <$, $\Delta ABC \cong \Delta DEF$ במקום $\langle \Delta ABC, \Delta DEF\rangle \in \cong$ וכו'.

2. רוב היחסים המעניינים במתמטיקה הם בינאריים, אך יש יוצאים מן הכלל. כך, למשל, היחס "בין" בין נקודות (דוגמה (ד) למעלה) הוא יחס **טרנארי** (תלת-מקומי). את ההגדרות שנתנו ניתן להכליל בקלות ליחסים כלליים יותר:

יחס n-מקומי הוא קבוצה של n-יות (עצמים מהצורה $\langle a_1, \ldots, a_n\rangle$).
יחס n-מקומי על $A_1, \ldots, A_n$ הוא קבוצה חלקית של $A_1 \times A_2 \times \ldots \times A_n$
(בפרט יחס חד-מקומי על קבוצה A הוא פשוט קבוצה חלקית של A).
יחס n-מקומי על A הוא קבוצה חלקית של A^n.

בקורס זה נעסוק בעיקר ביחסים בינאריים.[18]

3. %% הערנו בפרק הקודם, שמושג הפונקציה הכללי ביותר חורג לעתים מהמושג הרשמי. דבר זה נכון ביתר שאת לגבי יחסים באופן כללי. אנו מדברים על יחס השוויון $=$, על יחס ההכלה $\subseteq$ ועל יחס השייכות $\in$, למרות שאי-אפשר לראות בהם קבוצות של זוגות, והבנתם קודמת להבנת מושג "הזוג הסדור" (הרי יהיה זה מטופש לטעון, שהנוסחה "$1 \in \mathrm{N}$" אינה אלא קיצור של "$\langle 1,\mathrm{N}\rangle \in \in$", שהיא בעצמה רק קיצור של "$\langle\langle 1,\mathrm{N}\rangle,\in\rangle \in \in$", שהיא בעצמה...). לא ניכנס כאן לדיון

---

[18] תחום יישומי חשוב במדעי המחשב, שבו יש חשיבות רבה ליחסים n-מקומיים כלשהם, הוא "בסיסי נתונים טבלאיים". בבסיסי נתונים כאלה מייצגים יחסים בשיטת הטבלה, שתוארה עבור פונקציות בפרק הקודם.

מעמיק בסוגייה זו. נציין רק, שברוב המכריע של המקרים, המושגים השונים שנגדיר וההגדרות עצמן טובים גם ליחסים כמו $\in$, $=$ ו- $\subseteq$, כלומר ליחסים במובן רחב יותר מזה של ההגדרה הרשמית.

%%

4. אם נבדוק את ההגדרות הרשמיות של פונקציות בפרק הקודם (בטבלה מס' ב.5), ניווכח, שפונקציות הן סוג מיוחד של יחסים. ליתר דיוק:

(i) ***פונקציה*** היא יחס f המקיים את תנאי החד-ערכיות:

$$\forall a \forall b_1 \forall b_2 .\ \langle a,b_1 \rangle \in f \wedge \langle a,b_2 \rangle \in f \Rightarrow b_1 = b_2$$

(זהו תנאי (1)(ב) בטבלה, בעוד תנאי (1)(א) שם אומר, ש- f יחס).

(ii) ***פונקציה חלקית מ- A ל- B*** היא יחס מ- A ל- B, המקיים את תנאי החד-ערכיות הנ"ל.

(iii) ***פונקציה מ- A ל- B*** היא פונקציה חלקית f מ- A ל- B, המקיימת:

$$\forall a \in A\ \exists b \in B .\ \langle a,b \rangle \in f$$

כזכור, כאשר יחס f הוא פונקציה, אנו נוהגים לכתוב $y = f(x)$ במקום $\langle x,y \rangle \in f$ או xfy

II. פעולות על יחסים

א. היפוך יחס:

אם R יחס, אז היחס ההפוך, R^{-1}, מוגדר על-ידי:

$$R^{-1} = \{\langle b,a \rangle \mid \langle a,b \rangle \in R\}$$

ברור ש- R^{-1} אכן יחס כש- R יחס. יתר על כן, אם R הינו יחס מ- A ל- B, אז R^{-1} הינו יחס מ- B ל- A. R^{-1} מתאפיין על-ידי העובדה הבאה:

$$bR^{-1}a \Leftrightarrow aRb$$

דוגמאות:

(1) $<^{-1} = >$

(2) היחס ההפוך ליחס "הורה של" הוא היחס "הילד של", כלומר: x הוא הורה של y אם"ם y הוא ילד של x.

ב. הרכבת יחסים

אם S יחס מ- A ל- B, ו- R יחס מ- B ל- C, או הרכבתם, $R \circ S$, היא היחס הבא מ- A ל- C:

$$R \circ S = \{<a,c> \in A \times C \mid \exists b \in B.\ <a,b> \in S \wedge <b,c> \in R\}$$

במלים אחרות:

$$\forall a \in A\ \forall c \in C.\ a\ R \circ S\ c \Leftrightarrow \exists b \in B.\ aSb \wedge bRc$$

דוגמאות:

(1) ההרכבה של היחס "הורות" עם עצמו נותנת את יחס ה"סבאות".

(2) ההרכבה של היחס "הורה של" עם היחס "אח/אחות של" נותנת את היחס "דוד/דודה של".

שימו לב:
השוואה של ההגדרות, שניתנו פה למושגים של הרכבת יחסים ושל היחס ההפוך, להגדרות של אותם מושגים עבור פונקציות (בטבלה ב.5), מראה שההגדרות ***זהות***. מה שהוגדר כאן אינו אלא הכללה של ההגדרות עבור פונקציות. עם זאת, יש לציין את העובדות הבאות:

א. אם f ו- g הן פונקציות, אז הרכבתן $f \circ g$ אינה סתם יחס, ***אלא היא פונקציה בעצמה***. כאן יש, לכן, התאמה מושלמת בין ראיית $f \circ g$ כהרכבה של פונקציות (כמו בפרק הקודם), ובין ראיית $f \circ g$ כהרכבה של יחסים.

ב. המצב בעניין f^{-1}, לעומת זאת, עדין יותר. היחס ההפוך R^{-1} מוגדר עבור ***כל*** יחס R. מכאן, שאם f פונקציה (ולכן יחס), אז היחס f^{-1} מוגדר אף הוא. ברם, ***יחס זה אינו תמיד פונקציה בעצמו***: תנאי החד-ערכיות יכול שלא להתקיים עבור f^{-1}. רק אם הפונקציה f הינה חד-חד-ערכית, יהיה היחס f^{-1} פונקציה גם הוא. הוא יהווה פונקציה, שהתחום שלה שווה לתמונה של f, והתמונה שלה שווה לתחום של f. היוצא מזה הוא, שאם f פונקציה, אז בעוד ***הנוסחאות*** $<x,y> \in f^{-1}$ (או $xf^{-1}y$) הן בעלות משמעות עבור ***כל*** ערך של x ו- y, הרי ***הביטוי*** $f^{-1}(x)$ הוא בעל משמעות רק בתנאי ש- f הינה פונקציה ח.ח.ע., ו- x שייך לתמונה של f.

התרגיל הבא מראה, שלפעולות המוכללות של הרכבה והפיכה יש תכונות דומות לאלה, שהיו להן במקרה המיוחד של פונקציות:

תרגיל

א. $(R \circ S) \circ T = R \circ (S \circ T)$

ב. $(R^{-1})^{-1} = R$

ג. $(R \circ S)^{-1} = S^{-1} \circ R^{-1}$

ד. אם R יחס מ- A ל- B, אז $i_B \circ R = R \circ i_A = R$.

הנושא של פעולות על יחסים (לא רק בינאריים) הינו בעל אפליקציות רבות (למשל – במערכות בסיסי נתונים), אבל חורג ממה שנזדקק לו בהמשך קורס זה.[19]

III. תכונות של יחסים

בסעיף זה נציג תכונות חשובות, שעשויות להיות ליחסים.

1. (א) יחס R על קבוצה A נקרא ***רפלקסיבי*** (חוזר) על A אם $\forall x \in A.\, xRx$.

   (ב) יחס R נקרא ***אי-רפלקסיבי*** אם $\forall x.\, \neg(xRx)$. היחס נקרא אי-רפלקסיבי על A אם $\forall x \in A.\, \neg(xRx)$.

דוגמאות:

(א) יחסים רפלקסיביים: $=$, $\leq$ על $\mathbf{R}$, $\backslash$ על $\mathbf{N}$, יחס הדמיון על קבוצות המשולשים במישור.

(ב) יחסים אי-רפלקסיביים: $\neq$, $<$ על $\mathbf{R}$, יחס ההורות.

(ג) דוגמה ליחס שאינו רפלקסיבי, אך גם אינו אי-רפלקסיבי:

$$S = \{<x,y> \in \mathbf{R}^2 \mid y = x^2\}$$

יחס זה אינו רפלקסיבי (כי $<2,2> \notin S$), אך גם אינו אי-רפלקסיבי (כי $<1,1> \in S$).

[19] חומר נוסף (כולל פעולות, שלא הגדרנו כאן) ניתן למצוא בחלק המוקדש לתורת הקבוצות בקורס "מתמטיקה בדידה" של האוניברסיטה הפתוחה.

הערה:
נשים לב, שרפלקסיביות יחס אינה משהו אבסולוטי, אלא תלויה בקבוצה, עליה רואים את היחס כמוגדר בקונטקסט מסוים. היחס $\{<n,n> \mid n \in \mathbf{N}\}$ הוא רפלקסיבי בתור יחס על $\mathbf{N}$, אך אינו רפלקסיבי בתור יחס על $\mathbf{R}$. ברם, כמעט תמיד, כשנתייחס ליחס R, נציין גם קבוצות A ו- B, שיש באותו קונטקסט לראות את R כיחס מהאחת לשנייה. בדרך-כלל נציין לכן פשוט, ש- R הוא יחס רפלקסיבי או לא-רפלקסיבי (להבדיל מאי-רפלקסיבי!), ויהיה ברור אז מהקונטקסט, יחסית לאיזו קבוצה מדובר.

%%
כדי להימנע מסיבוכים מסוג זה, מכניסים טקסטים רבים את הקבוצות A ו- B לתוך ***ההגדרה*** של יחס. לפי טקסטים אלו, אם R_1 הוא יחס מ- A_1 ל- B_1, ו- R_2 הוא יחס מ- A_2 ל- B_2, הרי די ש- $A_1 \neq A_2$, או ש- $B_1 \neq B_2$, כדי שהיחסים R_1 ו- R_2 ייחשבו כשונים – אפילו אם R_1 ו- R_2 שווים כקבוצות (של זוגות). לגישה זו מחיר משלה. אם ניישם אותה לפונקציות (דבר מקובל מאוד גם הוא), הרי מתברר, שאין פונקציה אקספוננציאלית $\lambda x \in \mathbf{R}.\ e^x$ אחת, אלא הרבה. אנו נאלצים להבדיל, למשל, בין הפונקציה $\mathrm{Exp}_1 = \lambda x.\ e^x : \mathbf{R} \to \mathbf{R}$ ובין $\mathrm{Exp}_2 = \lambda x.\ e^x : \mathbf{R} \to \mathbf{R}^+$ (כש- $\mathbf{R}^+ =_{df} \{x \in \mathbf{R} \mid x > 0\}$). איש אינו עושה אבל הפרדה כזו, כיוון שהיא היתה הופכת את כתיבתם וקריאתם של טקסטים לבלתי נסבלת! המחיר, שאנו משלמים כאן, הוא זניח לעומת המחיר של הגישה האלטרנטיבית. **%%**

2. יחס R נקרא ***טרנסיטיבי*** אם:

$$\forall x\ \forall y\ \forall z.\ xRy \wedge yRz \Rightarrow xRz$$

דוגמאות:
$=$, $\leq$, $>$, חפיפה, דמיון, התחלקות וצאצאות הם יחסים טרנסיטיביים (לדוגמה: אם $a \setminus b$, ו- $b \setminus c$, אז $a \setminus c$. אם a צאצא של b, ו- b צאצא של c, אז a צאצא של c).
לעומת זאת, יחס ההורות, יחס האהבה, $\neq$ והיחס S למעלה, אינם טרנסיטיביים.

אזהרה:
כשאנו אומרים, שיחס האהבה, למשל, אינו טרנסיטיבי, ***אין*** כוונתנו ***שבכל*** מקרה בו a אוהב את b ו- b אוהב את c, a בהכרח אינו אוהב את c, אלא רק ש- a, b ו- c כאלו ***קיימים***.

3. (א) יחס R נקרא ***סימטרי*** אם $\forall x\ \forall y.\ xRy \Rightarrow yRx$.
(ב) יחס R נקרא ***אנטי סימטרי במובן החזק*** אם $\forall x\ \forall y.\ xRy \Rightarrow \neg(yRx)$.

(ג) יחס R נקרא ***אנטי סימטרי*** (או אנטי סימטרי במובן החלש) אם

$$\forall x\ \forall y.\ xRy \wedge yRx \Rightarrow x = y$$

דוגמאות:

$=$, $\neq$, חפיפה, יחס הדמיון, יחס הנישואין, היחס $\{<x,y> \in \mathbf{R}^2 \mid x^2 + y^2 = 1\}$ הם יחסים סימטריים. $<$ ויחס ההורות הם אנטי סימטריים במובן החזק. $\leq$, $=$, $\subseteq$, יחס ההתחלקות והיחס S למעלה, הם אנטי סימטריים (כך, למשל, אם $a \leq b$ ו- $b \leq a$, אז $a = b$).

4. יחס R, שהינו רפלקסיבי על A, אנטי סימטרי וטרנסיטיבי, נקרא ***יחס סדר חלקי*** על A אם, בנוסף, מתקיים ש- $\forall x \in A\ \forall y \in A.\ xRy \vee yRx$, אז אומרים ש- R סדר ***מלא*** (או ***טוטאלי***, או ***לינארי***) על A.

דוגמאות:

$=$, $\leq$ על $\mathbf{R}$, $\subseteq$ על $P(E)$ (E קבוצה כלשהי) ויחס ההתחלקות על $\mathbf{N}$ הם יחסי סדר חלקיים. מתוכם $\leq$ הוא יחס סדר מלא על $\mathbf{R}$. כמו כן, היחס:

$$\{<x,y> \in \mathbf{R}^2 \mid x \leq y \wedge \sin \pi x = \sin \pi y\}$$

הינו יחס סדר חלקי על $\mathbf{R}$ (הוכיחי!), שהינו סדר מלא על $\mathbf{N}$, אך לא על $\mathbf{R}$ כולו.

להגדרות הללו יש גרסאות חזקות יותר:

יחס ***סדר חלקי חזק*** (strict) על A הוא יחס טרנסיטיבי ואי-רפלקסיבי על A. יחס כזה נקרא ***מלא*** (או ליניארי, או טוטאלי) על A אם

$$\forall x \in A\ \forall y \in A.\ xRy \vee x = y \vee yRx$$

דוגמאות:

$<$ הוא יחס סדר חזק ומלא על $\mathbf{R}$, בעוד $\subset$ הינו יחס סדר חלקי חזק על $P(E)$, שאינו מלא אם ב- E שני איברים לפחות (אם $a \in E$, $b \in E$, ו- $a \neq b$, אז $\{a\} \neq \{b\}$, $\{a\} \not\subset \{b\}$, ו- $\{b\} \not\subset \{a\}$).

תרגיל

(1) אם R יחס סדר חלקי על A, אז היחס $R^* = \{<x,y> \in A^2 \mid xRy \wedge x \neq y\}$ הינו יחס סדר חלקי חזק על A. אם, בנוסף, R הוא מלא על A, אז גם R^* מלא על A.

(2) אם R יחס סדר חלקי חזק על A, אז היחס $R' = \{<x,y> \in A^2 \mid xRy \vee x = y\}$ הוא יחס סדר חלקי על A. אם, בנוסף, R הינו מלא על A, אז כך גם R'.

5. יחס על קבוצה A, שהוא רפלקסיבי על A, סימטרי וטרנסיטיבי, נקרא יחס *שקילות* (אקויולנציה) על A.

דוגמאות: $=$, דמיון משולשים, חפיפת משולשים.

דוגמאות נוספות:

(א) $\equiv(2) = \{<a,b> \in \mathbf{N} \times \mathbf{N} \mid 2 \setminus |a-b|\}$.

למשל: $<11,5> \in \equiv(2)$ ו- $<5,11> \in \equiv(2)$, כיוון ש- $|11-5|$ ו- $|5-11|$ מתחלקים ב- 2. לעומת זאת, $<4,7> \notin \equiv(2)$, כיוון ש- $|7-4|$ אינו מתחלק ב- 2.

מקובל לסמן $a \equiv b (\mathrm{mod}\ 2)$ במקום $a \equiv (2)\ b$ או $<a,b> \in \equiv(2)$.

הערה: באופן דומה מגדירים את היחס $\equiv(n)$ לכל מספר טבעי n.

(ב) $Q^* = \{<<a,b>, <c,d>> \in (\mathbf{N}^+ \times \mathbf{N}^+)^2 \mid ad = bc\}$.

תרגיל

להוכיח ש- $\equiv(2)$ (ו- $\equiv(n)$ באופן כללי) הוא יחס שקילות על $\mathbf{N}$, בעוד Q^* הוא יחס שקילות על $\mathbf{N}^+ \times \mathbf{N}^+$.

התכונות החשובות השונות, שהגדרנו, מסוכמות בטבלה ב.7 בעמוד הבא. אנו מניחים שם, ש- R יחס על A, ולכן התוספת "על A" הושמטה מהמונחים השונים.

IV. יחסי שקילות ופירוקים, קבוצות מנה

הגדרה

תהי A קבוצה. ***פירוק***[20] (partition) של A היא קבוצה F של קבוצות חלקיות לא ריקות של A (כלומר $F \subseteq P(A) - \{\varnothing\}$), המקיימת:

(i) $\cup(F) = A$

(ii) $\forall X \in F\, \forall Y \in F.\, X \neq Y \Rightarrow X \cap Y = \varnothing$

---

[20] יש המשתמשים במונח "חלוקה" במקום "פירוק".

טבלה ב.7: תכונות חשובות של יחסים

יהי R יחס על A $(R \subseteq A^2)$. ההגדרות בטבלה הן יחסית ל- A.		
.1	(א)	R נקרא **רפלקסיבי**, אם: $\forall a \in A.\ aRa$
	(ב)	R נקרא **אי-רפלקסיבי**, אם: $\forall a \in A.\ \neg(aRa)$
.2	(א)	R נקרא **סימטרי**, אם: $\forall a \in A\ \forall\ b \in A.\ aRb \Rightarrow bRa$
	(ב)	R נקרא **אנטי סימטרי**, אם: $\forall a \in A\ \forall\ b \in A.\ (aRb \wedge bRa) \Rightarrow a = b$
	(ג)	R נקרא **אנטי סימטרי חזק**, אם: $\forall a \in A\ \forall\ b \in A.\ aRb \Rightarrow\ \neg(bRa)$
.3		R נקרא **טרנסיטיבי**, אם: $\forall a \in A\ \forall b \in A\ \forall c \in A.\ (aRb \wedge bRc) \Rightarrow aRc$
.4	(א)	R נקרא ***יחס סדר חלקי***, אם R רפלקסיבי, אנטי-סימטרי וטרנסיטיבי.
	(ב)	R נקרא ***יחס סדר מלא***, אם R יחס סדר חלקי, ובנוסף מתקיים: $\forall a \in A\ \forall b \in A.\ aRb \vee bRa$
	(ג)	R נקרא ***יחס סדר חלקי חזק***, אם R אי-רפלקסיבי וטרנסיטיבי.
	(ד)	R נקרא ***יחס סדר מלא חזק***, אם R יחס סדר חלקי חזק, ובנוסף: $\forall a \in A\ \forall b \in A.\ aRb \vee a = b \vee bRa$
.5		R נקרא *יחס שקילות* (או יחס *אקוויוולנציה*), אם R הינו רפלקסיבי, סימטרי וטרנסיטיבי.

הערות:

1. תנאי (ii) שקול ל:

$$\forall X \in F\, \forall Y \in F.\, X \cap Y \neq \varnothing \Rightarrow X = Y$$

זוהי, בדרך-כלל, צורה נוחה יותר להוכחה.

2. משמעות התנאים היא, ש- F הינו פירוק של A אם"ם כל איבר ב- A שייך לאיבר אחד בדיוק של F (שייך – בגלל (i), לאחד בלבד – בגלל (ii)). בטרמינולוגיה אחרת: F הוא פירוק של A אם $A = \uplus(F)$ ו- $X \neq \varnothing$ לכל $X \in F$.

דוגמה

$\{\{1,3,5\},\{6\},\{2,4,7,8\}\}$ היא פירוק של $\{1,2,3,4,5,6,7,8\}$.

סימונים:

1. את קבוצת הפירוקים של קבוצה A נסמן ב- $\mathrm{par}(A)$.
2. את קבוצת יחסי השקילות על קבוצה A נסמן ב- $\mathrm{Eq}(A)$.

תרגילים

1. $\mathrm{par}(A) \in P(P(P(A)))$
2. $\mathrm{Eq}(A) \in P(P(A \times A))$

נראה עתה, שלמושג הפירוק קשר הדוק למושג של יחס שקילות:

הגדרה:

1. יהי R יחס שקילות על קבוצה A, ונניח $x \in A$. ***מחלקת השקילות של*** x לפי R היא הקבוצה:

$$[x]_R = \{y \in A \mid xRy\}$$

2. יהי R יחס שקילות על A. ***קבוצת המנה של*** A ***לפי*** R היא הקבוצה:

$$A/R = \{[x]_R \mid x \in A\}$$

במלים אחרות: A/R היא קבוצת מחלקות השקילות לפי R של איברי A.

טענה

יהי R יחס שקילות על A. אז לכל x ו- y ב- A מתקיים:

(0) $x \in [x]_R$

(1) $y \in [x]_R \iff xRy$

(2) $y \in [x]_R \iff yRx$

(3) $y \in [x]_R \iff x \in [y]_R$

(4) $[y]_R \subseteq [x]_R \iff xRy$

(5) $[x]_R = [y]_R \iff xRy$

(6) אם $Z \in A/R$ ו- $x \in Z$, אז $Z = [x]_R$.

הוכחה

(0) מיידי מהגדרת $[x]_R$ והעובדה, ש- R רפלקסיבי.

(1) זוהי בעצם הגדרת $[x]_R$.

(2) מיידי מ- (1) ומסימטריות R.

(3) מיידי מ- (2) ומהגדרת $[y]_R$: $x \in [y]_R \overset{(1)}{\iff} yRx \overset{(2)}{\iff} y \in [x]_R$.

(4) $(\Rightarrow)$. נניח $[y]_R \subseteq [x]_R$. כיוון ש- $y \in [y]_R$ (לפי (0)), נובע מזה, ש- $y \in [x]_R$, ולכן, לפי (1), xRy.

$(\Leftarrow)$. נניח xRy. יהי $z \in [y]_R$, אז yRz (מ- (1)). מזה ומ- xRy נובע xRz, בגלל ש- R טרנסיטיבי. לכן $z \in [x]_R$ לפי (1). סה"כ $[y]_R \subseteq [x]_R$.

(5) $(\Rightarrow)$. אם $[x]_R = [y]_R$, אז $[y]_R \subseteq [x]_R$, ולכן xRy לפי (4).

$(\Leftarrow)$. נניח xRy. אז גם yRx (כי R סימטרי). לכן מ- (4) נובע, ש- $[y]_R \subseteq [x]_R$ וגם ש- $[x]_R \subseteq [y]_R$. סה"כ $[x]_R = [y]_R$.

(6) נניח $Z \in A/R$. אז $Z = [y]_R$ לאיזה $y \in A$. לכן, אם $x \in Z$, אז $x \in [y]_R$. מכאן, לפי (2), xRy. לכן, לפי (5), $[x]_R = [y]_R$. כלומר: $[x]_R = Z$.

מסקנות:

1. אם R יחס שקילות ו- $x \in A$, אז $[x]_R = \iota Z \in A/R.\, x \in Z$ (מ- (0) ו- (6)).
2. $\lambda x \in A.\, [x]_R$ היא פונקציה מ- A **על** A/R.

משפט

אם R יחס שקילות על A, אז A/R היא פירוק של A.

הוכחה:

כיוון ש- $[x]_R \subseteq A$ לכל $x \in A$, A/R היא קבוצה של קבוצות חלקיות של A. אף אחת מהן אינה ריקה, כי $x \in [x]_R$ לכל $x \in A$ (סעיף (0) בטענה האחרונה). מהעובדה, ש- $\forall x \in A.\ x \in [x]_R$ נובע גם, שכל איבר x ב- A שייך לאחת הקבוצות ב- A/R (הלא היא $[x]_R$). לכן $A \subseteq \cup(A/R)$. מצד שני, נובע מהעובדה ש- $X \subseteq A$ לכל $X \in A/R$, ש- $\cup(A/R) \subseteq A$ (טבלה ב.2, כלל $(5)_b$). סה"כ $\cup(A/R) = A$. נראה, לבסוף, שמדובר באיחוד זר. נניח, אפוא, ש- $X_1 \in A/R$, $X_2 \in A/R$ ו- $X_1 \cap X_2 \neq \varnothing$ יהי $x \in X_1 \cap X_2$ מהסעיף השישי בטענה האחרונה: $X_1 = [x]_R$ ו- $X_2 = [x]_R$. לכן $X_1 = X_2$.

דוגמאות

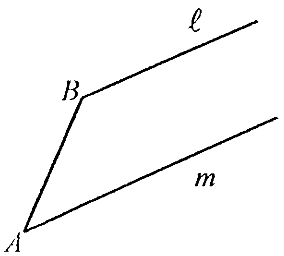
1. ב- $\mathbf{N} / \equiv (2)$ יש שתי מחלקות שקילות:

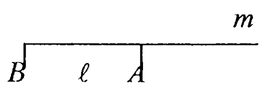
קבוצת המספרים הזוגיים: $N_{\text{even}} = \{2n \mid n \in \mathbf{N}\}$

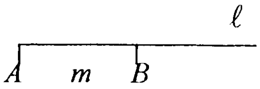
קבוצת המספרים האי-זוגיים: $N_{\text{odd}} = \{2n+1 \mid n \in \mathbf{N}\}$

מתקיים ש- $\mathbf{N}_{\text{even}} = [0]_{\equiv(2)} = [2]_{\equiv(2)} = [4]_{\equiv(2)} = \dots$, $\mathbf{N}_{\text{odd}} = [1]_{\equiv(2)} = [3]_{\equiv(2)} = \dots$

2. נגדיר על קבוצת הקרניים במישור את היחס D הבא: $\ell D m$ אם: הקרן ℓ מוכלת בקרן m, או הקרן m מוכלת בקרן ℓ, או ℓ ו- m מקבילות ונמצאות באותו צד של הישר, המחבר את נקודת ההתחלה שלהן:







קל לברר, שזהו יחס שקילות על קבוצת הקרניים במישור. למחלקת השקילות $[\ell]_D$ של הקרן ℓ נוהגים לקרוא במקרה זה ***הכיוון*** של ℓ. היחס היסודי בין עצמים ומחלקות השקילות שלהן (סעיף (5) בטענה למעלה):

$$[\ell]_D = [m]_D \Leftrightarrow \ell D m$$

מְתרגם כאן לעובדה, שקרן ℓ וקרן m נמצאות ביחס D אם ורק אם יש להן ***אותו כיוון***. מחלקת השקילות של קרן ℓ היא קבוצת כל הקרניים, שיש להן אותו כיוון כמו ℓ (גם אינטואיטיבית וגם לפי ההגדרה הרשמית כאן).

דוגמה 2 מדגימה תהליך נפוץ למדי של יצירת ישויות (או עצמים) חדשים על-ידי מעבר מעצמים בקבוצה מסוימת למחלקות השקילות שלהן, לפי יחס שקילות מסוים. העניין לא מסתיים אבל בדרך-כלל בכך. על העצמים החדשים מגדירים בדרך-כלל פעולות ויחסים (כדי שיהיה להם שימוש) משלהם. הדבר נעשה כמעט תמיד על-ידי שימוש ב***מייצגים***. לצורך עניין זה, כל איבר של מחלקת שקילות נחשב ***מייצג*** של אותה מחלקת שקילות (לפי סעיף (6) של הטענה למעלה, x הוא מייצג של Z אם"ם $Z = [x]_R$). ניקח, לדוגמה, את המושג של ***זווית בין כיוונים***. היא נמדדת, כידוע, כך: לוקחים קרן כלשהי בכיוון האחד וקרן כלשהי בכיוון השני. הזווית בין הקרניים היא הזווית בין הכיוונים. כדי שהגדרה זו תהיה בעלת משמעות, הכרחי אבל, שהתוצאה הסופית ***לא תהיה תלויה בקרניים, שבחרנו כמייצגות של הכיוונים*** (תנאי זה נקרא "אי-תלות במייצגים", ועוד נחזור אליו רבות בהקשרים אחרים). בדוגמה של כיוונים קל להוכיח, שאכן כך הדבר.

3. **%%** נחזור ליחס Q^*, שהבאנו כדוגמה ליחס שקילות על $(\mathbf{N}^+)^2$:

$$<a,b>\ Q^*\ <c,d> \ \Leftrightarrow\ ad = bc$$

מהי, למשל, מחלקת השקילות של $<3,5>$? ובכן, קל לברר, ש-

$$[<3,5>]_{Q^*} = \{<3,5>, <6,10>, <9,15>, \ldots\} = \{<3k, 5k> \mid k \in \mathbf{N}^+\}$$

עתה, במקום לכתוב $[<3,5>]_{Q^*}$, אנו כותבים, עוד מאז ימי בית-הספר היסודי, $\frac{3}{5}$.
כללית, אנו מסמנים:

$$\frac{a}{b} =_{Df} [< a, b >]_{Q^*}$$

למחלקות שקילות אלו אנו קוראים "מספרים רציונליים חיוביים". נשים לב, שאכן השוויון

$$\frac{a}{b} = \frac{c}{d}$$

פירושו:

$$[<a,b>]_{Q^*} = [<c,d>]_{Q^*}$$

וזה קורה אם"ם $<a,b>\ Q^*\ <c,d>$ (לפי האיפיון המרכזי בטענה למעלה).
כלומר אם"ם: $ad = bc$.

את קבוצת המספרים הרציונליים החיוביים אנו מסמנים ב- Q^+. Q^+ היא, אם כך, קבוצת מחלקות השקילות של היחס Q^*, דהיינו $(\mathbf{N}^+ \times \mathbf{N}^+) / Q^*$.

הנקודה המרכזית מכל זה היא, שקבוצת המספרים הרציונליים החיוביים נבנית כקבוצת מנה של יחס שקילות מסוים על $\mathbf{N}^+ \times \mathbf{N}^+$, ומספר רציונלי הוא, בעצם, קבוצה אינסופית של זוגות. כל זוג כזה יכול לשמש כמייצג של המספר הרציונלי. כך, למשל, הזוג $\langle 6,10\rangle$ יכול לשמש כמייצג של המספר הרציונלי $\frac{3}{5}$ (כי $\frac{6}{10} = \frac{3}{5}$. המדובר פה, נשים לב, בשוויון של ***קבוצות***!).

כאמור למעלה, לעתים קרובות, כשמדובר בקבוצת מנה, מעוניינים אנו להגדיר עליה פעולות ויחסים, הדומים לאלה של קבוצת המקור. כך, לדוגמה, מעוניינים אנו להגדיר חיבור של מספרים רציונליים, הדומה בתכונותיו לחיבור של מספרים טבעיים ומכליל אותו. הדרך הרגילה לעשות דברים כאלו היא על-ידי שימוש במייצגים. עם זאת, יש תמיד להיזהר ולהראות, שמה שהגדרנו אינו תלוי במייצגים, בהם אנו משתמשים. כדי להבין למה הכוונה, נדמה בנפשנו מישהו המגדיר "פעולה" חדשה $\oplus$ על Q^+ על-ידי:

$$\frac{a}{b} \oplus \frac{c}{d} =_{Df} \frac{a+c}{b+d}$$

הגדרה כזו יש להבין כך: כדי לבצע $\oplus$ על-ידי שני איברים X ו- Y ב- Q^+, עלינו למצוא מייצג $\langle a,b\rangle$ עבור X (דהיינו $X = \frac{a}{b}$) ומייצג $\langle c,d\rangle$ עבור Y (ולכן $Y = \frac{c}{d}$), ולחשב ב- $\mathbf{N}^+$ את $a + c$ ואת $b + d$. תוצאת החיבור $\oplus$ של X ו- Y היא אז $\frac{a+c}{b+d}$ – מחלקת השקילות של $\langle a + c,\ b + d\rangle$. הצרה בהגדרה זו הינה, שבחירות שונות של המייצגים של X ו- Y נותנות תשובות שונות לשאלה, כמה שווה $X \oplus Y$. לדוגמה, למרות ש- $\frac{1}{2} = \frac{2}{4}$, הרי $\frac{1}{3} \oplus \frac{2}{4} \neq \frac{1}{3} \oplus \frac{1}{2}$! במקרים כגון אלו נהוג לומר, ש- $\oplus$ ***אינה מוגדרת היטב***.

לעומת זאת, אם נגדיר חיבור על Q^+ באופן הבא:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd}$$

אזי נקבל פעולה מוגדרת היטב, כיוון שאם $\langle a_1, b_1\rangle$ מייצג אחר של $\frac{a}{b}$ (כלומר:

אם $<a,b>Q^*<a_1, b_1>$) ו-$<c_1, d_1>$ מייצג אחד של $\frac{c}{d}$ (כלומר: $<c,d>Q^*<c_1, d_1>$),

אזי

$$<a_1d_1 + b_1c_1, b_1d_1> Q^* <ad + bc, bd>$$

וממילא:

$$\frac{ad+bc}{bd} = \frac{a_1d_1+b_1c_1}{b_1d_1}$$

במלים אחרות: החלפת המייצג $<a,b>$ של $\frac{a}{b}$ ב- $<a_1, b_1>$ והמייצג $<c,d>$ של $\frac{c}{d}$ ב- $<c_1, d_1>$, לא תשנה את תוצאת הפעולה.

%%

*נספח:
עוד על הקשר בין יחסי שקילות ופירוקים

ראינו בפרק האחרון, שאם R יחס שקילות, אז A/R הינה פירוק של A. מכאן, שכל יחס שקילות על A מגדיר באופן טבעי פירוק של A. נראה עתה, ראשית, שגם ההיפך נכון:

הגדרה

יהי F פירוק של A. נגדיר:

$$A/F = \{<x,y> \in A^2 \mid \exists Z \in F.\ x \in Z \wedge y \in Z\}$$

(כלומר $x(A/F)y$ אם"ם x ו- y נמצאים באותו איבר של F).

טענה

אם F פירוק של A, אז A/F יחס שקילות על A.

הוכחה

(א) רפלקסיביות: מהגדרת A/F, לכל $x \in A$ מתקיים:

$$\begin{aligned} <x,x> \in A/F &\Leftrightarrow \exists Z \in F.\ x \in Z \wedge x \in Z \\ &\Leftrightarrow \exists Z \in F.\ x \in Z \end{aligned}$$

וזה נכון, כיוון ש- F פירוק.

(ב) **סימטריות:** לכל $x,y \in A$:

$$<x,y> \in A/F \Leftrightarrow \exists Z \in F.\ x \in Z \wedge y \in Z$$
$$\Leftrightarrow \exists Z \in F.\ y \in Z \wedge x \in Z$$
$$\Leftrightarrow <y,x> \in A/F$$

(ג) **טרנסיטיביות:** נניח $<x,y> \in A/F$ ו- $<y,z> \in A/F$. אז קיים $Z_1 \in F$ כך ש- $x \in Z_1 \wedge y \in Z_1$ וקיים $Z_2 \in F$ כך ש- $y \in Z_2 \wedge z \in Z_2$. מכאן ש- $y \in Z_1 \cap Z_2$. זה גורר (כיוון ש- F פירוק), ש- $Z_1 = Z_2$. מכאן ש- $x \in Z_1 \wedge z \in Z_1$, ולכן קיים $Z \in F$ $(Z = Z_1)$, כך ש- $x \in Z \wedge z \in Z$. במלים אחרות: $<x,z> \in A/F$.

סה"כ קיבלנו, שלכל יחס שקילות על A מתאים פירוק של A, ולכל פירוק של A מתאים יחס שקילות על A. מתעוררת עתה שאלה טבעית: נניח, שאנו מתחילים ביחס שקילות על A. ליחס זה מתאים פירוק של A. לפירוק זה, מצדו, מתאים יחס שקילות על A. האם זה יחס השקילות בו התחלנו? שאלה דומה מתעוררת, כמובן, אם מתחילים בפירוק F. התשובה לשתי השאלות הינה חיובית:

משפט

1. $A/(A/R) = R$ (כאשר R יחס שקילות על A).
2. $A/(A/F) = F$ (כאשר F פירוק של A).

הוכחה

1. נניח ש- R הוא יחס שקילות על A. אז A/R פירוק של A. אם נציב $F = A/R$ בהגדרת A/F, נקבל:

$$A/(A/R) = \{<x,y> \in A^2 \mid \exists Z \in A/R.\ x \in Z \wedge y \in Z\}$$

$$= \{<x,y> \in A^2 \mid \exists Z \in A/R.\ Z = [x]_R \wedge Z = [y]_R\}$$

(לפי סעיף (6) בטענה המרכזית בעמוד 149 על A/R ו- $[x]_R$)

$$= \{<x,y> \in A^2 \mid [x]_R = [y]_R\}$$

(לפי עקרונות לוגיים ברורים הקשורים בשוויון)

$$= \{<x,y> \in A^2 \mid <x,y> \in R\}$$

(לפי סעיף (5) של אותה טענה)

$$= \{<x,y> \mid <x,y> \in R\}$$

(כי $R \subseteq A^2$)

$$\overset{(\text{ח})}{=} R$$

2. נוכיח תחילה את *הלמה* הבאה: אם F_1 ו- F_2 פירוקים של A ו- $A/F_1 = A/F_2$, אז $F_1 = F_2$.

הוכחת הלמה:

נוכיח, למשל, ש- $F_1 \subseteq F_2$ (ההוכחה ש- $F_2 \subseteq F_1$ היא דומה). יהי אפוא $U \in F_1$. כיוון ש- U לא ריק, יש איבר x ב- U. כיוון ש- F_2 פירוק, יש $W \in F_2$ כך ש- $x \in W$. נראה ש- $U = W$, ומזה ינבע ש- $U \in F_2$. בשביל זה מספיק שנוכיח, ש- $U \subseteq W$ ו- $W \subseteq U$. נוכיח, למשל, ש- $U \subseteq W$ (ההוכחה ש- $W \subseteq U$ דומה). יהי אפוא $y \in U$, נראה ש- $y \in W$. כיוון ש- $x \in U$, $y \in U$ ו- $U \in F_1$, נובע, מהגדרת A/F_1, ש- $<x,y> \in A/F_1$. כיוון ש- $A/F_1 = A/F_2$, נקבל ש- $<x,y> \in A/F_2$. מהגדרת A/F_2, פירוש הדבר, שקיים $Z \in F_2$ כך ש- $x \in Z$ ו- $y \in Z$. אבל x גם שייך ל- W, ולכן $x \in W \cap Z$. כיוון ש- F_2 פירוק ו- $W,Z \in F_2$, נובע מזה, ש- $Z = W$. כיוון ש- $y \in Z$, פירוש הדבר, ש- $y \in W$. הראינו, שאם $y \in U$, אז $y \in W$ דהיינו: $U \subseteq W$. באותו אופן מוכיחים ש- $W \subseteq U$. סה"כ $W = U$.

סיום הוכחת 2:

אם F פירוק של A, אז מהמשפט הקודם, A/F יחס שקילות על A. לכן, מחלק 1 של משפט זה נובע ש:

$$A/(A/(A/F)) = A/F$$

משוויון זה ומהלמה (כש- $F_1 = A/(A/F)$ ו- $F_2 = F$) נובע

$$A/(A/F) = F$$

מ.ש.ל.

את המשפט האחרון נוכל להציג מנקודת ראות של פונקציות. נזכיר, ש- $\mathrm{Eq}(A)$ היא קבוצת יחסי השקילות על A ו- $\mathrm{par}(A)$ היא קבוצת הפירוקים של A.

נוכל עתה להתבונן בפונקציות הבאות:

$$H = \lambda R \in \mathrm{Eq}(A).\ A/R \qquad (H : \mathrm{Eq}(A) \to \mathrm{par}(A))$$
$$G = \lambda F \in \mathrm{par}(A).\ A/F \qquad (H : \mathrm{par}(A) \to \mathrm{Eq}(A))$$

המשפט האחרון אומר, בעצם, ש- $H \circ G = i_{\mathrm{par}(A)}$ ו- $G \circ H = i_{\mathrm{Eq}(A)}$.

מסקנה:

$\lambda R \in \mathrm{Eq}(A).\ A/R$ היא פונקציית שקילות בין $\mathrm{Eq}(A)$ ו- $\mathrm{par}(A)$, בעוד $\lambda F \in \mathrm{par}(A).\ A/F$ היא הפונקציה ההפוכה שלה.

ג. קומבינטוריקה כללית

ג.1 עוצמות ושוויון עוצמות

קומבינטוריקה היא ענף במתמטיקה, העוסק בעיקר בשאלה: "כמה איברים יש בקבוצה נתונה?". התואר "כללית", שנלווה למלה "קומבינטוריקה" בכותרת חלק זה, בא להדגיש, שבשלב ראשון נעסוק בעקרונות קומבינטוריים, החלים על קבוצות *כלשהן*, כולל קבוצות אינסופיות.

הצהרה אחרונה זו מעוררת, מן הסתם, תמיהה עוד בטרם נתחיל. מה פירוש "מספר איברים בקבוצה אינסופית"? מה עוד ניתן לפרט כאן פרט לכך, ש"מספר" האיברים בקבוצה כזו הוא אינסופי? אנו נראה בפרק זה, שיש אכן מקום לפירוט, ושיש אכן גדלים שונים גם בתחום האינסוף. דבר זה יוביל אותנו להרחבת מאגר המספרים, בהם אנו משתמשים למדידת קבוצות סופיות (הלא הוא N), למאגר רחב יותר, שיאפשר גם מדידת קבוצות אינסופיות. אנו נעשה זאת, כמובן, באופן, שיכליל בצורה טבעית את מושג המספר הטבעי המונה ("סופר")[1].

כדי שהרחבה מהסוג שהבטחנו לפני רגע תיעשה כראוי, היה רצוי מאוד, שתהיה לנו הגדרה ראויה לשמה של מושג המספר הטבעי. בשלב זה, לרוע המזל, אין בידינו הגדרה כזו. כדי להיווכח בכך, נסו להסביר לעצמכם, או (להבדיל) לילד קטן, מה זה "חמש", "שלושים" או "מספר". ספק בלבי אם תצליחו. אין פירוש הדבר, אבל, שאין לכולנו *הבנה* טובה של משמעות מושגים אלו. הדרך שנלך בה, עת נכליל את מושג המספר, תסתמך על ניתוח הבנה זו.

כדי לקבל פרספקטיבה טובה יותר על מה שאנו עומדים לעשות, הבה נחזור לדוגמה של מושג ה"כיוון של קרן", אותו הגדרנו בפרק הקודם. נניח שמישהו מבין הקוראים היה נשאל: "מה זה כיוון של קרן?", *לפני* שקרא את הפרק הקודם. האם היתה ההגדרה הרשמית, אותה הבאנו שם, הדבר הראשון, שהיה עולה על דעתו? קשה להאמין. למעשה, ספק אם היה מצליח להגיע להגדרה כלשהי מתקבלת על הדעת. ייתכן, שהיה אומר, שכיוון של קרן זה המספר, שמודד את הזווית בינה ובין קרן ייחוס מסוימת. זוהי אינה הגדרה מוצלחת במיוחד. ראשית, מושג הכיוון ברור לנו, ואנו משתמשים בו, עוד

---

[1] הכוונה למספרים הטבעיים, כשהם בצורה של "אחת, שתיים, שלוש, ארבע...". בצורה זו משמשים הם למניית מספר האיברים בקבוצות סופיות. למספרים הטבעיים בצורתם "ראשון, שני, שלישי, רביעי..." קוראים, לעומת זאת, מספרים סודרים, כי הם משמשים לצורך סידור איברים בתוך קבוצה.

בטרם נדע למדוד אותו (אנו הרי מודדים *משהו*, לא?). שנית, אותו כיוון עצמו ניתן למדוד בעזרת מספרים שונים: המספר תלוי ביחידת המידה ובקרן הייחוס. כיוון שכך, ברור שהכיוון אינו יכול להיות דבר, *זהה* עם המספר. לעומת זאת, אין כל בעיה להגדיר במדויק, מתי יש לשתי קרניים *אותו כיוון*. קצת חשיבה תוביל אותנו בדיוק ליחס D של הפרק הקודם: לשתי קרניים יש אותו כיוון, אם אחת מהן מוכלת בשנייה, או אם הן מקבילות ונמצאות באותו צד של הישר, המחבר את נקודות ההתחלה שלהן. יחס זה הינו, כזכור, יחס שקילות.

המצב, שתיארנו כרגע, הינו מצב שחוזר הרבה. אנו מתחילים *ביחס* מוגדר היטב בין עצמים מסוימים. כשיחס זה הינו יחס *שקילות*, אנו חשים, שיש משהו המשותף לכל שני עצמים, המתייחסים לפיו זה לזה. בצעד הבא אנו מבצעים תהליך של *הפשטה* (אבסטרקציה). בלי להגדיר במדויק את אותו "דבר משותף", אנו מכניסים לשימוש מונח חדש עבורו. במקום לדבר על קיום היחס בין שני עצמים, אנו מתחילים לדבר אז על *השוויון* עבורם של אותו דבר משותף. כך במקום להגיד, ששתי קרניים נמצאות ביחס D זו לגבי זו, אנו אומרים, שיש להן אותו כיוון. אנו עושים זאת, בטרם הגדרנו מה זה בכלל "כיוון". הבנת המושג "שוות כיוון" קודמת אפוא להבנת מושג הכיוון!

אם נסמן ב- $\text{direction}(a)$ את הכיוון של הקרן a, הרי הרעיון המרכזי סביב מושג הכיוון הוא, ש- aDb אם"ם $\text{direction}(a) = \text{direction}(b)$. תהליך דומה מקובל עבור כל יחס R, שהוא יחס שקילות: אנו עוברים בתהליך הפשטה מטענה מהצורה aRb לטענה מהצורה $d_R(a) = d_R(b)$, כש- d_R הוא מושג חדש, שהתקבל על-ידי *הפשטה*. ל- d_R אין בדרך-כלל הגדרה. כל שאנו יודעים עליו הוא, ש:

$$d_R(a) = d_R(b) \quad \Leftrightarrow \quad aRb$$

למה תהליך זה מקובל? בראש ובראשונה, משום שפסיכולוגית, נוח לנו יותר לעבוד עם שוויונות מאשר עם יחסים. כך ברור לנו, למשל, שאם $d_R(a) = d_R(b)$ ו- $d_R(b) = d_R(c)$, אז $d_R(a) = d_R(c)$. פחות רגילים אנו לרעיון, ש"היחס R הוא טרנסיטיבי" (שזה אותו הדבר בדיוק!). סיבה חשובה נוספת היא, שאחרי שאנו קונים הבנה אינטואיטיבית של העצמים החדשים, אותם יוצר תהליך ההפשטה, אנו מתחילים לעשות עמם מה שאנו עושים עם כל סוג של עצמים: אנו מגדירים יחסים מועילים ופעולות מועילות על עצמים אלה. גילוין של עובדות חשובות (גם כאלה, שאינן מתייחסות ישירות לעצמים החדשים) יכול להיות אז קל מאוד. בדוגמה של כיוונים, למשל, מוביל מושג הכיוון למושג הוקטור (במרחב) ואחר-כך לתחשיב הוקטורים (שהינו, בדרך-כלל, מכשיר הרבה יותר נוח לעבודה מאשר הגיאומטריה האוקלידית הרגילה).

ומה אם מישהו לוחץ ושואל שאלה כמו: "אז מה זה בכל זאת 'כיוון'?". לאחד כזה ניתן לתת שתי תשובות. האחת, המעשית, היא: "מה זה בכלל חשוב, מה הם כיוונים? העיקר הוא, שאנו יכולים לעבוד איתם!". אנלוגיה טובה נוכל להביא כאן משחמט: אפשר להסתבך קשות בניסיון לענות על השאלה: מהו "פרש" בשחמט? זה בוודאי לא הכלי המגולף, שאנו מזיזים על הלוח (אפשר הרי לשחק שחמט בעל-פה, ללא שימוש בכלים פיזיים כלל!). בסופו של דבר, מה שקובע את מהותו של הכלי בשחמט אלה הכללים, הקשורים בו ובצורת הזזתו, כלומר: הפעולות שעושים אתו.

הצורה השנייה לענות על שאלה מהסוג "מה זה כיוון?" (או כל מונח אחר, שמקורו בתהליך אבסטרקציה), היא לספק בכל זאת הגדרה כלשהי, אפילו מסובכת, כך שיובטח הקשר הבסיסי:

$$aRb \quad \Leftrightarrow \quad d_R(a) = d_R(b)$$

זוהי דרכם של המתמטיקאים, כיוון שאלה מרגישים חובה לא להשאיר דברים תלויים באוויר באופן מיסטי. יש להם אפילו דרך סטנדרטית איך לעשות זאת, אותה ראינו בפעולה בפרק הקודם. היא מתבססת על כך, שאם R יחס שקילות, אז:

$$aRb \quad \Leftrightarrow \quad [a]_R = [b]_R$$

השוואה של נוסחה זו עם הקשר הבסיסי, אותו אנו מחפשים, מראה, שנוכל להגדיר פשוט:

$$d_R(a) =_{Df} [a]_R$$

ההפשטה המבוקשת יכולה אפוא להינתן על-ידי מחלקות השקילות, אותן היחס משרה. בדוגמה של "כיוון", למשל, אנו יכולים להגדיר את הכיוון של קרן a בתור מחלקת השקילות של a ביחס ל- D, דהיינו: הכיוון של קרן a הוא קבוצת כל הקרניים, שיש להן "אותו כיוון" כמו a. כך בדיוק הגדרנו אכן את הכיוון של a בפרק הקודם! (נעיר שוב, שאנו מבינים היטב מתי לשתי קרניים יש "אותו כיוון" גם בלי להגדיר מה זה "כיוון", ושמושג ה"כיוון" היה היסטורית מובן ושימושי, עוד בטרם ניתנה לו ההגדרה הרשמית הזו).

נחזור עתה אל מושג המספר. גם ביחס אליו, מה שאנו מבינים תחילה הוא מה פירוש הדבר, שלשתי קבוצות יש *"אותו מספר"* של איברים. מושג המספר עצמו מובן רק אחר-כך, ומתקבל בתהליך הפשטה.[2]

[2] טעות נפוצה היא לחשוב, שכדי להבין מה זה "אותו מספר" צריך תחילה להבין את פשר המלים "אותו" ו"מספר". אין זה נכון. רובנו, למשל, מבינים את פירוש הביטוי "ישב על המדוכה" בלי שיהיה לנו מושג מה זה "מדוכה"...

ובכן, מתי נאמר, ששתי קבוצות סופיות הן שוות בגודלן? אולי תאמרו: אם נקבל אותה תוצאה, כאשר נספור כמה איברים יש בכל אחת. זה נכון, אמנם, אבל מבוסס כבר על *אמצעי*, שפותח לצורך מתן תשובה לשאלה זו עצמה (לא תמיד האמצעי היעיל ביותר, אגב!). לאמתו של דבר, אפשר לדעת על שתי קבוצות, שהן שוות בגודלן, בלי לדעת כלל לספור (וילדים קטנים אכן עושים זאת לפעמים). נניח, לדוגמה, שאנו נכנסים לאולם ריקודים ומגלים בו זוגות רוקדים. נניח עוד, שכולם מעורבים, וההתלהבות בשיאה: איש אינו יושב בצד. בלי שום ספירה ברור לנו מיד, שמספר הנשים בחדר (לפני שנכנסנו אנו) היה שווה למספר הגברים בו. בדומה, אין כל צורך בספירה כדי לדעת, שבכל לוח שחמט יש מספר שווה של משבצות לבנות ושחורות!

ברור משתי הדוגמאות, ששתי קבוצות הן "שוות-גודל" אם ניתן לבצע התאמה של זוגות בין אחת לשנייה. זוהי ההבנה היסודית. רק מאוחר יותר בהיסטוריה האנושית (או בהתפתחותו של ילד) נעשות *אבסטרקציות*. אז ניתן, למשל, השם "חמש" למשותף לכל הקבוצות, שהן שוות-גודל עם קבוצת האצבעות ביד ימין של אדם נורמלי, והשם "עשר" – למשותף לאלה, שהן שוות-גודל לקבוצת האצבעות בשתי הידיים. מושג המספר עצמו נוצר בשעתו על-ידי אבסטרקציה נוספת. שום הגדרות לא ניתנו אז!

הגיע הזמן, אבל, *שאנו* ניתן הגדרה כלשהיא.

הגדרה:

שתי קבוצות A ו- B נקראות *אקויפוטנטיות* (או שוות-עוצמה) אם קיימת פונקציית שקילות ביניהן.

סימון: את העובדה ש- A ו- B אקויפוטנטיות נסמן ב- $A \sim B$.

טענה:

יחס האקויפוטנטיות בין קבוצות הוא יחס שקילות.

הוכחה:

(א) רפלקטיביות: $A \sim A$ לכל A, כי i_A פונקציית שקילות מ- A על A.

(ב) סימטריות: נניח $A \sim B$. תהי f פונקציית שקילות מ- A על B. אז f^{-1} היא פונקציית שקילות מ- B על A. לכן $B \sim A$.

(ג) טרנסיטיביות: נניח $A \sim B$ ו- $B \sim C$. תהי f פונקציית שקילות מ- A על B, ותהי g פונקציית שקילות מ- B על C. אז $g \circ f$ היא פונקציית שקילות מ- A על C. לכן $A \sim C$.

%% הערה:

יחס האקויפוטנטיות בין קבוצות אינו, למעשה, יחס במובן הרשמי, כפי שהגדרנוהו בפרק הקודם: אי אפשר לתאר אותו בתור *קבוצה* של זוגות בלא להסתבך בפרדוקסים. כמו $\in$, $=$ ו- $\subseteq$, הוא יחס במובן *מוכלל*, מעל "אוסף כל הקבוצות". בהתאם, משמעות הטענה האחרונה היא, שיחס האקויפוטנטיות הוא "יחס שקילות" במובן מוכלל (דהיינו: רפלקסיבי, טרנסיטיבי ואנטי-סימטרי).

%%

דוגמאות פשוטות של קבוצות אקויפוטנטיות הן קבוצת הימים בשבוע וקבוצת הגמדים של שלגייה, כמו גם שתי קבוצות המשחקות זו נגד זו במשחק כדורגל (בתנאי ששום שחקן לא הורחק על-ידי השופט). דוגמאות חשובות יותר מהבחינה המתמטית (יחד עם פונקציות השקילות המתאימות) נכללות בטבלה ג.1[3].

טבלה ג.1: דוגמאות לקבוצות שוות-עוצמה

	קבוצות	*פונקציית שקילות*
(1)	$\mathbf{N} \sim \mathbf{N}^+$	$\lambda n \in \mathbf{N}.\ n+1$
(2)	$\mathbf{N}_{\text{even}} \sim \mathbf{N}_{\text{odd}}$	$\lambda n \in \mathbf{N}_{\text{even}}.\ n+1$
(3)	$\mathbf{N} \sim \mathbf{N}_{\text{even}}$	$\lambda n \in \mathbf{N}.\ 2n$
(4)	$\mathbf{N} \times \mathbf{N} \sim \mathbf{N}$	$\lambda m \in \mathbf{N},\ k \in \mathbf{N}.\ \frac{(m+k)(m+k+1)}{2} + m$
(5)	$(a<b)\quad [a,b] \sim [0,1]$	$\lambda x \in [a,b].\ \frac{x-a}{b-a}$
(6)	$(a<b)\quad (a,b) \sim (0,1)$	$\lambda x \in (a,b).\ \frac{x-a}{b-a}$
(7)	$\mathbf{R} \sim (-1,1)$	$\lambda x \in \mathbf{R}.\ \frac{x}{1+\lvert x\rvert}$
(8)	$\mathbf{R}^+ = (0,\infty) \sim (0,1)$	$\lambda x \in (0,\infty).\ \frac{x}{1+x}$
(9)	$A \times B \sim B \times A$	$\lambda a \in A, b \in B.\ \langle b, a\rangle$
(10)	$A \times B \to C \sim A \to (B \to C)$	$Cu = \lambda f \in A \times B \to C.\ \lambda x \in A.\ \lambda y \in B.\ f(x,y)$
(11)	$P(E) \sim E \to \{0,1\}$	$\lambda A \in P(E).\ \chi_A^{(E)}$

[3] כזכור, $\mathbf{N}^+$ היא קבוצת המספרים הטבעיים בלי אפס, $\mathbf{N}_{\text{even}}$ – קבוצת המספרים הזוגיים, ו- $\mathbf{N}_{\text{odd}}$ – קבוצת המספרים האי-זוגיים.

בשתי הדוגמאות האחרונות בטבלה ג.1 נתקלנו כבר בפרק ב.4 (ראה סעיף ה שם אודות פונקציית Curry ודוגמה (4) בעמוד 130). את מרבית הדוגמאות האחרות נשאיר כתרגילים לקורא. אנו נסתפק כאן בהוכחת דוגמה מס' (7).

תהי אפוא $f = \lambda x \in R.\dfrac{x}{1+|x|}$. ברור ש- f מוגדרת אכן לכל $x \in \mathbf{R}$. כמו כן, לכל $x \in \mathbf{R}$ מתקיים: $|f(x)| = \dfrac{|x|}{1+|x|}$, ולכן $0 \le |f(x)| < 1$, וממילא $f(x) \in (-1,1)$. מכאן, שאכן $f: \mathbf{R} \to (-1,1)$. כדי להראות ש- f פונקציית שקילות, די שנראה ש- $g = \lambda x \in (-1,1).\dfrac{x}{1-|x|}$ היא פונקציה הפוכה ל- f. ברור ש- g אכן מוגדרת לכל $x \in (-1,1)$, וש- $\mathbf{R}$ היא טווח שלה. חישוב פשוט מראה, ש- $g(f(x)) = x$ לכל $x \in \mathbf{R}$ ו- $f(g(x)) = x$ לכל $x \in (-1,1)$. לכן $f \circ g = i_{(-1,1)}$ ו- $g \circ f = i_{\mathrm{R}}$. מכאן, שאכן $g = f^{-1}$.

הערות:

1. עבור מרבית שאר הדוגמאות קל למצוא פונקציה הפוכה מתאימה, ולכן קל להוכיחן באופן דומה לזה של הוכחת דוגמה (7). דוגמה מס' (4) היא יוצאת מן הכלל כאן. פה יהיה קל יותר להוכיח, שהפונקציה בה מדובר היא ח.ח.ע. ועל $\mathbf{N}$. דוגמה זו מהווה תרגיל, שהינו יחסית קשה יותר, ועוד נחזור אליה מאוחר יותר (בפרק ג.4).

2. הפונקציה בדוגמה (2) היא הצמצום של הפונקציה בדוגמה (1) ל- $\mathbf{N}_{\text{even}}$. בדומה, הפונקציה בדוגמה (6) היא הצמצום של הפונקציה בדוגמה (5) לקטע (a, b), והפונקציה בדוגמה (8) היא הצמצום של הפונקציה בדוגמה (7) לקטע $(0, \infty)$. כל זאת מדגים את העובדה, שהשימוש בצמצומים של פונקציות שקילות הוא נפוץ לצורך בניית פונקציות שקילות אחרות. שתי העובדות הפשוטות הבאות אודות צמצומים כאלה עשויות אז להועיל. את הוכחתן נשאיר לקורא.

טענה:

1. אם f היא פונקציה ח.ח.ע. מ- A אל B ו- $X \subseteq A$, אז f/X היא פונקציית שקילות בין X ל- $f(X)$ (התמונה של X לפי f), ולכן $X \sim f(X)$. בפרט $A \sim f(A)$ במקרה זה.

2. תהי $f: A \to B$ פונקציית שקילות. נניח ש- $X \subseteq A$ ו- $Y \subseteq B$ הן שתי קבוצות, כך ש- $f(X) \subseteq Y$ ו- $f^{-1}(Y) \subseteq X$, אז $X \sim Y$, ו- f/X היא פונקציית שקילות ביניהן.[4]

דוגמה

בדוגמה מס' (7) ראינו ש- $g = \lambda x. \frac{x}{1-|x|}$ היא הפונקציה ההפוכה ל- f (כאשר $f = \lambda x. \frac{x}{1+|x|}$). כדי להסיק את דוגמה (8) מדוגמה (7) די לכן להראות (על-פי החלק השני של הטענה האחרונה), שאם $x \in (0, \infty)$ אז $f(x) \in (0,1)$, ושאם $x \in (0,1)$ אז $g(x) \in (0, \infty)$. זה קל.

דוגמאות (3), (7) ו- (8) עלולות להיראות פרדוקסליות במבט ראשון. דוגמאות אלה מציגות מקרים, בהם קבוצה היא שוות-עוצמה עם קבוצה, החלקית לה ממש. מהגדרותינו יוצא אכן, ש"יש אותו מספר של מספרים טבעיים זוגיים ושל מספרים טבעיים בכלל", וכן ש"יש אותו מספר של מספרים ממשיים בין 0 ל- 1 ומספרים ממשיים חיוביים בכלל". זה סותר את האינטואיציה הבסיסית שלנו, שאם $A \subset B$, אז ב- A יש פחות איברים מאשר ב- B. ברם, אינטואיציה זו נקנתה מתוך התנסות בקבוצות *סופיות* בלבד, ואילו בדוגמאות הללו מדובר בקבוצות *אינסופיות*. כדאי, שהקוראים יסכינו מרגע זה ואילך עם העובדה, שאינטואיציות הנכונות לקבוצות סופיות אינן תמיד נכונות לקבוצות אינסופיות, ועוד נראה דוגמאות רבות לכך.

אחרי שהגדרנו, מתי שתי קבוצות הן "שוות-עוצמה", נעבור בשלב הבא לבצע את תהליך ההפשטה, אותו תיארנו בראשית פרק זה. כתוצאה מכך יתווסף מושג חדש לעולם הקבוצות:

"הגדרה":

לדבר המשותף לקבוצה A ולכל הקבוצות, שהן אקויפוטנטיות עמה, קוראים *העוצמה* של A (או "מספר האיברים ב- A", או ה"קרדינליות של A"), ומסמנים אותה ב- $|A|$ (או ב- $\overline{\overline{A}}$).

התכונה העיקרית של מושג העוצמה הינה:

$$\boxed{|A| = |B| \iff A \sim B}$$

[4] את הביטוי $f^{-1}(Y)$ אפשר להבין כאן בשתי צורות: כמקור של Y לפי X, או התמונה של Y לפי f^{-1}. קל לראות, ששני הפירושים הם זהים במקרה זה (הוכח!).

ברור, *שלא* הגדרנו כאן בדיוק, מה זה $|A|$. הגדרנו רק, מה פירוש הדבר ש- $|A| = |B|$. בהמשך נגדיר עוד יחסים אחרים על עוצמות (כמו $|A| < |B|$) ופעולות עליהן (כמו $|A| + |B|$). אנו נשתמש בכל אלה בהצלחה, כשאנו מסתמכים אך ורק על מושג העוצמה, שמספק תהליך ההפשטה, ומבלי שניתן למושג זה הגדרה מדויקת. בכך משחזרים אנו, למעשה, את דרכו של קנטור, אבי תורת הקבוצות. נרגיע אבל את קוראינו, שעובדה זו מדאיגה אותם (ובצדק!), שבמסגרת תורת הקבוצות האקסיומטית ניתנת גם ניתנת הגדרה מדויקת למושג "העוצמה של קבוצה" במונחים של תורת הקבוצות בלבד. הטענה בתוך המסגרת למעלה הופכת שם להיות *משפט* (עם הוכחה). ברם, בטרם אפשר יהיה להציג הגדרה זו, יש ללמוד תחילה חומר, שהוא הרבה מעבר לטווח של קורס זה.

כדאי לציין בהקשר זה, שבטקסטים לא מעטים, העוסקים בתורת הקבוצות הנאיבית, מפעילים את אותה הפרוצדורה, שהפעלנו במקרה של הגדרת "כיוון". במלים אחרות: מגדירים את $|A|$, העוצמה של A, ***בתור מחלקת השקילות של A לפי יחס האקויפוטנציה***. הבעיה עם "הגדרה" זו הינה, שיחס האקויפוטנציה הינו, כזכור, יחס רק במובן *מוכלל*: אי אפשר לתאר אותו בתור *קבוצה* של זוגות! יתר על כן, גם "מחלקות השקילות" שלו אינן קבוצות (לו היו קבוצות, היינו מסתבכים בסתירות לוגיות, הדומות לזו של פרדוקס ראסל). כיוון שכך, אין הן עצמים ואין הן יכולות, לכן, להיות איברים בשום קבוצה – גם לא N. (היסטורית, ההגדרה המדויקת הראשונה, שניסה מישהו לתת למושג "עוצמה", היתה בדיוק זו – והיא אכן הובילה לסתירות!).

ג.2 עוצמות סופיות ואינסופיות

לא פעם במהלך הקורס עד כאן השתמשנו במושגים של "קבוצה סופית" ו"קבוצה אינסופית", מתוך הנחה, שאלו מושגים ברורים. כאשר מפתחים את תורת הקבוצות בצורה מסודרת, מגדירים, כמובן, גם מושגים אלה. הדבר יכול להיעשות בשני מסלולי התקדמות אפשריים. במסלול האחד מגדירים תחילה מהי קבוצה סופית (על סמך אפיון, המופיע כ"מסקנה" בסוף פרק זה). לאחר מכן מוגדרים המספרים הטבעיים בתור העוצמות של קבוצות סופיות, ותכונותיהם הבסיסיות מוכחות על סמך זאת. במסלול השני מגדירים תחילה את המספרים הטבעיים (בצורה אחרת, כמובן), מוכיחים את תכונותיהם הבסיסיות על סמך הגדרה זו, ו*אז* מגדירים קבוצות סופיות בעזרתם. הדבר נעשה כך, שהמספרים הטבעיים ישמשו כעוצמות של קבוצות סופיות, וקבוצה תהיה סופית, אם ורק אם עוצמתה הינה מספר טבעי. בקורס זה נלך במסלול השני. יהיה רק הבדל משמעותי אחד: אנו ***לא*** נגדיר את המספרים הטבעיים, אלא נניח, שאנו מכירים אותם ואת תכונותיהם הבסיסיות (אותן למדנו בבית-הספר העממי). כיוון שכך, יש להתייחס להגדרה של קבוצות סופיות, שניתן מיד, *כהגדרת עבודה*. אין לראותה בגדר הגדרה במובן המדוקדק של המלה.

הגדרה:

1. **קבוצה A נקראת *סופית* אם קיים מספר טבעי n כך ש- A אקויפוטנטית עם $\mathbf{N}_n = \{0, 1, \ldots, n-1\}$. n זה נחשב אז כעוצמת הקבוצה A (או "מספר איבריה"), דהיינו $|A| = n$ במקרה זה.**
2. **קבוצה תיקרא אינסופית אם אינה סופית.**

הערות:

1. מההגדרה ברור, שהמספרים הטבעיים הינם העוצמות של הקבוצות הסופיות.
2. הגדרתנו כוללת את המקרה $n = 0$. במקרה זה הקבוצה $\mathbf{N}_n = \{0, 1, \ldots, n-1\}$ היא פשוט $\varnothing$. הקבוצה הריקה נחשבת, לכן, קבוצה סופית, ומספר איבריה הוא המספר הטבעי 0.[5]
3. החלק השני של הגדרת קבוצה סופית, זה המתייחס למספר איבריה, מכיל הנחה סמויה: שלכל קבוצה A ייתכן מספר טבעי n אחד לכל היותר, כך ש- $A \sim \mathbf{N}_n$

---

[5] זו הסיבה, שבמסגרת תורת הקבוצות טבעי לקבל את אפס בתור מספר טבעי.

ההנחה מתבטאת בצירוף "n זה", שמניח יחידותו. אנו מניחים, אם כן, שלא ייתכן, למשל, שקבוצה A תהיה אקויפוטנטית גם עם $\{0, \ldots, 1{,}000{,}000\}$ וגם עם $\{0, \ldots, 999{,}999\}$. על מה מבוססת אמונה זו? לרובנו – על ניסיוננו עם קבוצות קטנות. למעשה, אפשר (וצריך!) להוכיח זאת:

משפט:

אם $A \sim \mathbf{N}_n$ ו- $A \sim \mathbf{N}_k$ $(n, k \in \mathbf{N})$, אז $n = k$.

את ההוכחה נביא בנספח לפרק זה.

4. מההגדרה ברור, שאם A סופית ו- $B \sim A$, אז B סופית (וכמובן $|B| = |A|$).

טבלה ג.2 בעמוד הבא כוללת רשימה של תכונות ידועות היטב של קבוצות סופיות. אנו נקבל תכונות אלו ללא הוכחה (הוכחות מלאות לאותן תכונות, שאינן מיידיות מההגדרות, ניתנות במסגרת קורס מלא בתורת הקבוצות האקסיומטית. חלק מהן ניתנות במסגרת הנספח לפרק זה).

מרשימה זו ומהגדרת קבוצה אינסופית, נובעת מיידית רשימת התכונות הבאה של קבוצות אינסופיות:

1. אם A אינסופית ו- $B \sim A$, אז B אינסופית.
2. אם A אינסופית ו- $A \subseteq B$, אז B אינסופית.
3. אם A אקויפוטנטית עם קבוצה חלקית ממש שלה, אז A אינסופית.
4. אם A אינסופית ו- f פונקציה ח.ח.ע. מ- A ל- B, אז B אינסופית.
5. אם A אינסופית ו- f פונקציה מ- B על A, אז B אינסופית.

האם יש בכלל קבוצות אינסופיות? אינטואיטיבית, ברור שכן. למעשה, זה גם נובע ממה שראינו עד עכשיו! כך ראינו, ש- $\mathbf{N}$ אקויפוטנטית עם קבוצה חלקית ממש שלה ($\mathbf{N}_{\text{even}}$). לכן $\mathbf{N}$ אינסופית, מתכונה 3 ברשימה האחרונה. מאותה סיבה גם $\mathbf{R}^+$ אינסופית: היא אקויפוטנטית עם $(0,1)$, שהיא קבוצה חלקית ממש שלה (דוגמה (8) בטבלה ג.1). כיוון ש- $\mathbf{N}_{\text{even}} \sim \mathbf{N}$ ו- $(0,1) \sim \mathbf{R}^+$, גם $\mathbf{N}_{\text{even}}$ ו- $(0,1)$ אינסופיות. כיוון ש- $\mathbf{R}^+ \subseteq \mathbf{R}$, גם $\mathbf{R}$ אינסופית (תכונה 2 ברשימה האחרונה).

השאלה המתבקשת עתה הינה: האם כל הקבוצות האינסופיות הינן אקויפוטנטיות, או שמא קיימת יותר מעוצמה אינסופית אחת? אנו נראה עתה, שהאפשרות השנייה היא הנכונה. זה יפתח את הפתח לתיאוריה עשירה ומרתקת של עוצמות אינסופיות. הפרקים הבאים יוקדשו בעיקרם לתיאוריה זו.

טבלה ג.2: תכונות בסיסיות של קבוצות סופיות

(1)	(i) לכל $n \in \mathbf{N}$, $\mathbf{N}_n = \{0, 1, \ldots, n-1\}$ הינה סופית (בפרט $\varnothing$ הינה סופית). (ii) אם $A \sim \mathbf{N}_n$ ו- $A \sim \mathbf{N}_k$ $(n, k \in \mathbf{N})$, אז $n = k$.
(2)	אם A סופית ו- $A \sim B$, אז B סופית.
(3)	(i) אם A סופית ו- $B \subseteq A$, אז B סופית ו- $\|B\| \le \|A\|$. (ii) אם A סופית ו- $B \subset A$, אז B סופית ו- $\|B\| < \|A\|$.
(4)	(i) קבוצה סופית אינה אקויפוטנטית עם שום קבוצה חלקית-ממש שלה. (ii) אם $\|B\| = \|A\|$, A סופית, ו- $f: A \to B$ ח.ח.ע. **או** על B, אז f פונקציית שקילות מ- A על B.
(5)	אם A סופית ו- f פונקציה ח.ח.ע. מ- B ל- A, אז B סופית ו- $\|B\| \le \|A\|$.
(6)	אם A סופית ו- f פונקציה מ- A על B, אז B סופית ו- $\|B\| \le \|A\|$.
(7)	אם A סופית ו- R יחס שקילות על A, אז A/R סופית ו- $\|A/R\| \le \|A\|$.
(8)	אם A ו- B סופיות, אז גם $A \cup B$, $A \times B$ ו- $A \to B$ הן סופיות, ומתקיים: (i) $A \cap B = \varnothing \;\Rightarrow\; \|A \cup B\| = \|A\| + \|B\|$ (ii) $\|A \times B\| = \|A\| \cdot \|B\|$ (iii) $\|A \to B\| = \|B\|^{\|A\|}$ (ובסימון אחר: $\|B^A\| = \|B\|^{\|A\|}$)

משפט:

$|\mathbf{N}^+| \neq |(0,1]|$. כלומר: $\mathbf{N}^+$ וקבוצת המספרים הממשיים בין 0 ל- 1 (כולל 1) אינן אקויפוטנטיות.

הוכחה:

אנו נסתמך על העובדה הבאה בדבר מספרים ממשיים בין 0 ל- 1: כל מספר כזה אפשר לייצג בצורה יחידה על-ידי שבר עשרוני אינסופי מהצורה:

$$0.\, d_1 d_2 d_3 d_4 \ldots$$

כאשר d_1, d_2, d_3, וכו' הן ספרות בין 0 ו- 9, וכך שאין מקום, שממנו ואילך כל הספרות הן אפסים (1/8, למשל, תיוצג על-ידי ...0.12499999 ולא על-ידי ...0.125000. 1 עצמו הוא ...0.9999). אנו נוכיח, שאין בכלל פונקציה מ- N^+ *על* [0,1) (וממילא אין פונקציית שקילות מ- N^+ על [0,1)). נוכיח זאת בדרך השלילה. נניח אפוא, ש- f היא פונקציה מ- N^+ על [0,1). נסמן ב- $(f(i))_j$ את הספרה ה- j של $f(i)$. נגדיר עתה מספר b בצורה הבאה:

$$b = 0.\, b_1 b_2 b_3 \ldots$$

כאשר

$$b_i = \begin{cases} 1 & (f(i))_i \neq 1 \\ 2 & (f(i))_i = 1 \end{cases}$$

עתה $b \in [0,1)$, אבל $b \neq f(i)$ לכל i, כיוון ש- b ו- $f(i)$ שונים בספרה ה- i-ית שלהם. זוהי סתירה לכך ש- f היא על [0,1), כיוון ש- b אינו בתמונה של f (נשים לב, שבייצוג $0.\, b_1 b_2 b_3 \ldots$ אין אפסים כלל, ולכן הוא ייצוג כשר למהדרין של מספר ממשי בקטע [0,1)).

הערה:

שיטת ההוכחה של המשפט האחרון נקראת "שיטת האלכסון". הסיבה היא, שאינטואיטיבית, מה שעושים בה, הוא הדבר הבא: מסדרים את $f(1)$, $f(2)$,... במטריצה האינסופית הבאה:

$$\begin{array}{llllll}
f(1) = & 0.f(1)_1 & f(1)_2 & f(1)_3 & f(1)_4 & \cdots \\
f(2) = & 0.f(2)_1 & f(2)_2 & f(2)_3 & f(2)_4 & \cdots \\
f(3) = & 0.f(3)_1 & f(3)_2 & f(3)_3 & f(3)_4 & \cdots \\
f(4) = & 0.f(4)_1 & f(4)_2 & f(4)_3 & f(4)_4 & \cdots \\
\vdots & & \ddots & & &
\end{array}$$

המספר b נוצר (אינטואיטיבית) על-ידי שמתקדמים לאורך האלכסון של מטריצה אינסופית זו, ומחליפים את הספרות, שמוצאים שם, בספרות אחרות (ואין משתמשים ב- 0). כך מבטיחים, ש- b יהיה שונה בספרה אחת לפחות מכל אחד מהמספרים, המיוצגים במטריצה.

סימונים: את העוצמה של N^+ מסמנים ב- $\aleph_0$.

את העוצמה של [0,1) מסמנים ב- c או ב- $\aleph$.

המשפט האחרון הראה, לכן, ש- $\aleph_0 \neq$ א.

הערה:

כיון ש- $\lambda n \in \mathbf{N}.\ n+1$ היא פונקציית שקילות בין $\mathbf{N}$ ל- $\mathbf{N}^+$, גם $|\mathbf{N}| = \aleph_0$.

כיוון ש- $\mathbf{N} \sim \mathbf{N}_{even}$ ו- $\mathbf{N}_{even} \sim \mathbf{N}_{odd}$, גם $|\mathbf{N}_{even}| = |\mathbf{N}_{odd}| = \aleph_0$.

הגדרה:

לקבוצות שעוצמתן $\aleph_0$ קוראים קבוצות *בנות-מניה* (*ב"מ*, בקיצור). במלים אחרות: קבוצה A נקראת בת-מניה אם"ם היא אקויפוטנטית עם $\mathbf{N}$ (או עם $\mathbf{N}^+$).

המשפט העיקרי של פרק זה מספק אפיונים של קבוצות אינסופיות. לצורך הוכחתו נזדקק ללמה הפשוטה הבאה, שיש לה חשיבות בפני עצמה:

למה:

אם $A \cap B = \varnothing$, $A' \cap B' = \varnothing$, $A \sim A'$ ו- $B \sim B'$, אז $A \cup B \sim A' \cup B'$.

הוכחה:

נניח ש- f היא פונקציית שקילות מ- A על A', וש- g היא פונקציית שקילות מ- B על B'. נגדיר $h : (A \cup B) \to (A' \cup B')$ על-ידי:

$$h(x) = \begin{cases} f(x) & x \in A \\ \\ g(x) & x \in B \end{cases}$$

(בסימון λ: $h = \lambda x \in A \cup B.\ \text{if } x \in A \text{ then } f(x) \text{ else } g(x)$). כיוון ש- $A \cap B = \varnothing$, h הינה מוגדרת היטב.[6] נראה ש- h היא פונקציית שקילות.

h ח.ח.ע.: נניח $h(x_1) = h(x_2)$. לא ייתכן ש- $x_1 \in A$ ו- $x_2 \in B$, כי אז $h(x_1) \in A'$, בעוד $h(x_2) \in B'$. נקבל אז לכן סתירה לכך, ש- $A' \cap B' = \varnothing$. בדומה, לא ייתכן גם ש- $x_1 \in B$ ו- $x_2 \in A$. אם, לעומת זאת, $x_1 \in A$ וגם $x_2 \in A$, אז השוויון $h(x_1) = h(x_2)$ פירושו $f(x_1) = f(x_2)$. כיוון ש- f ח.ח.ע., נקבל ש- $x_1 = x_2$. גם כאשר $x_1 \in B$ ו- $x_2 \in B$, נקבל ש- $x_1 = x_2$, הפעם בגלל חד-חד-ערכיות g. סה"כ $x_1 = x_2$ בכל המקרים האפשריים.

h על: נניח $y \in A' \cup B'$. אז $y \in A'$ או $y \in B'$. אם $y \in A'$, אז יש $x \in A$, כך ש- $f(x) = y$ (כיוון ש- f על A'). לכן $h(x) = y$. אם $y \in B'$, אז יש $x \in B$ כך ש- $g(x) = y$ (כי g על B'), ולכן $h(x) = y$. בכל מקרה יש $x \in A \cup B$ כך ש- $h(x) = y$.

[6] לו היה איבר y ב- $A \cap B$, אזי לא היה ברור, אם $h(y) = f(y)$, או שמא $h(y) = g(y)$!

הערה:

אלטרנטיבית, לא קשה להוכיח ש- $\lambda x \in A' \cup B'.\ \text{if}\ \ y \in A'\ \ \text{then}\ \ f^{-1}(y)\ \text{else}\ \ g^{-1}(y)$
הינה פונקציה הפוכה ל- h.

משפט:

התכונות הבאות של קבוצה A הינן שקולות:

(א) A מכילה קבוצה חלקית בת-מניה.

(ב) A אקויפוטנטית עם איזו תת-קבוצה חלקית-ממש שלה.

(ג) A הינה אינסופית.

הוכחה:

(א) $\Leftarrow$ (ב):

נראה תחילה, שאם B בת-מניה, אז ל- B קבוצה חלקית-ממש, שהיא אקויפוטנטית עמה. זה ודאי נכון ל- $\mathbf{N}$, כי $\mathbf{N}^+ \subset \mathbf{N} \sim \mathbf{N}^+$. עתה, אם $\mathbf{N} \sim B$, אז יש פונקציית שקילות f מ- $\mathbf{N}$ על B. תהי $B^+ = f(\mathbf{N}^+)$. אז $\mathbf{N}^+ \sim B^+$, וברור ש- $B^+ \subset B$ (כי $f(0) \notin B^+$). כיוון ש- $B \sim \mathbf{N} \sim \mathbf{N}^+ \sim B^+$, אנו מקבלים ש- $B \sim B^+$, ולכן B^+ קבוצה כמבוקש.

נניח עתה, ש- A מכילה קבוצה ב"מ B. כפי שראינו זה עתה, יש $B^+ \subset B$ כך ש- $B^+ \sim B$. עתה $(A - B) \cap B = \varnothing$, וממילא $(A - B) \cap B^+ = \varnothing$ (כי $B^+ \subset B$). כמו כן $(A - B) \sim (A - B)$ ו- $B \sim B^+$. מכל זה ומהלמה לפני המשפט נקבל:

$$A = (A - B) \cup B \ \sim\ (A - B) \cup B^+$$

כמו כן, $(A - B) \cup B^+ \subset A$, כי אם $x \in B - B^+$, אז $x \notin A - B$ ו- $x \notin B^+$, ולכן $x \in A$ ו- $x \notin (A - B) \cup B^+$. סה"כ $(A - B) \cup B^+$ קבוצה חלקית-ממש ל- A, שהינה שוות-עוצמה עם A.

(ב) $\Leftarrow$ (ג):

זה מיידי מהעובדה, שציינו למעלה, שקבוצה סופית אינה יכולה להיות אקויפוטנטית עם קבוצה חלקית שלה.

(ג) $\Leftarrow$ (א):

הוכחה מדויקת מסתמכת כאן על אקסיומה מיוחדת, שלא נלמדת בצורה מסודרת בקורס זה, ונקראת "אקסיומת הבחירה". לכן לא נביא הוכחה זו במלואה. ***הרעיון*** שמאחוריה הוא כזה: כיוון ש- A אינסופית, $A \neq \varnothing$ לכן יש בה איזה איבר. יהי a_0 איבר כזה. כיוון ש- A אינסופית, $A \neq \{a_0\}$ לכן יש איבר a_1 ב- $A - \{a_0\}$. עתה $A \neq \{a_0, a_1\}$ (כי A אינסופית), ולכן יש איבר $a_2 \in A - \{a_0, a_1\}$. בצורה זו נמשיך ונקבל סדרה

$a_0, a_1, a_2, a_3, \ldots$ של איברים מ- A, שכולם שונים זה מזה. הקבוצה B, המורכבת מאיברי סדרה זו, היא תת-קבוצה ב"מ של A.

%%

שאלה: מדוע מה שתיארנו למעלה, אינו הוכחה שלמה?

התשובה נעוצה במלים "בצורה זו נמשיך" המופיעות שם. עלינו להמשיך כאן בתהליך אינסופי, שבו בכל שלב אנו בוחרים באופן אקראי איבר חדש של A. לבסוף אנו מסתכלים במה שהתקבל ב*כל* התהליך, כאילו ***התהליך הושלם*** (מעשית, הוא לא יושלם לעולם, כמובן !). נשים לב, שאם נקח איבר x מסוים של A וננסה לברר, האם הוא נמצא ב- B, הרי לא תהיה לנו כל דרך לעשות זאת. אין ביכולתנו לדעת מראש, אם x ייבחר באופן אקראי באיזה שלב עתידי בבניית B! אקסיומת הבחירה מבטיחה, אינטואיטיבית, שקבוצות המתקבלות בתהליך אינסופי כזה של בחירות אקראיות, אכן קיימות – למרות שאין לנו כל דרך לתאר אותן או להכיר אותן. בכך היא שונה מאוד מכל העקרונות, שראינו בחלק הקודם. העקרונות ההם סיפקו תיאור מדויק של הקבוצות, שאת קיומן הם הבטיחו, וקריטריון מדויק לשייכות בהן!

%%

מסקנה:

קבוצה הינה סופית אם"ם היא אינה אקויפוטנטית עם שום קבוצה חלקית-ממש שלה.

הערה:

מסקנה זו יכולה לשמש כבסיס להגדרה אלטרנטיבית של מושג הקבוצה הסופית, שאינה מסתמכת על המספרים הטבעיים: מגדירים קבוצה כסופית אם אינה אקויפוטנטית עם שום קבוצה חלקית-ממש שלה.

*נספח:

הוכחה שאם $A \sim \mathbf{N}_n$ ו- $A \sim \mathbf{N}_k$, אז $n = k$

טענת עזר:

אם f היא פונקציה ח.ח.ע. מ- $\mathbf{N}_n$ ל- $\mathbf{N}_n$, אז f היא על $\mathbf{N}_n$.

הוכחה:

באינדוקציה על n. עבור $n = 0$ ו- $n = 1$ זה טריביאלי. נניח שהטענה נכונה עבור n ותהי f פונקציה ח.ח.ע. מ- $\mathbf{N}_{n+1}$ אל $\mathbf{N}_{n+1}$.

אפשרות א: $n+1 \notin f(\mathbf{N}_n)$. אז $f/\mathbf{N}_n$ היא פונקציה ח.ח.ע. מ- $\mathbf{N}_n$ אל $\mathbf{N}_n$. לפי הנחת האינדוקציה, $f/\mathbf{N}_n$ היא לכן על $\mathbf{N}_n$. כיוון ש- f ח.ח.ע., נובע מזה ש- $f(n+1) \notin \mathbf{N}_n$. מכאן ש- $f(n+1) = n+1$, לכן $n+1 \in f(\mathbf{N}_{n+1})$. מזה ומהעובדה ש- $f/\mathbf{N}_n$ היא על $\mathbf{N}_n$ נקבל, ש- $\mathbf{N}_n \cup \{n+1\} = \mathbf{N}_{n+1}$ מוכלת בתמונה של f. במלים אחרות: f היא על $\mathbf{N}_{n+1}$.

אפשרות ב: $n+1 \in f(\mathbf{N}_n)$. נניח אפוא, ש- $n+1 = f(j_0)$ כש- $j_0 \in \mathbf{N}_n$.
נגדיר $g : \mathbf{N}_{n+1} \to \mathbf{N}_{n+1}$ על-ידי:

$$g(i) = \begin{cases} n+1 & i = n+1 \\ f(n+1) & i = j_0 \\ f(i) & \text{אחרת} \end{cases}$$

(g מתקבל מ- f על-ידי "החלפת" הערכים שמקבלים j_0 ו- $n+1$).
קל לראות, ש- g גם היא ח.ח.ע., ושהתמונה של g שווה לתמונה של f. על g חלה, אבל, אפשרות א, ולכן התמונה של g היא כל $\mathbf{N}_{n+1}$ (לפי מה שכבר הוכחנו). לכן גם התמונה של f היא כל $\mathbf{N}_{n+1}$. במלים אחרות: f היא על $\mathbf{N}_{n+1}$.

הערה:

טענת העזר שקולה, בעצם, לחצי מהטענה הכלולה בתכונה (4) (ii) של טבלה ג.2.

מסקנה 1: $\mathbf{N}_n$ אינה אקויפוטנטית עם שום קבוצה חלקית-ממש שלה.

מסקנה 2: אם $\mathbf{N}_n \sim \mathbf{N}_k$, אז $n = k$.

הוכחה:

אם $k < n$, אז $\mathbf{N}_k$ קבוצה חלקית-ממש של $\mathbf{N}_n$, ונקבל סתירה עם מסקנה 1. בדומה, לא ייתכן, ש- $n < k$. לכן $k = n$.

הוכחת המשפט:

אם $A \sim \mathbf{N}_n$ ו- $A \sim \mathbf{N}_k$, אז $\mathbf{N}_n \sim \mathbf{N}_k$. לכן $n = k$ לפי מסקנה 2.

הערה:

(4) (i) של טבלה ג.2 נובעת בקלות ממסקנה 1 ומהגדרת קבוצה סופית (הוכיחו!).

ג.3 סדר על עוצמות

לאחר שהגדרנו מתי שתי עוצמות הן שוות, השלב הטבעי הבא הוא לנסות לסדר אותן, באופן שנוכל לקבוע, מי גדולה יותר ומי קטנה יותר (ממש כשם שהדבר נעשה לגבי קבוצות סופיות). הקו המנחה יהיה שוב ניתוח המושג של "קבוצה גדולה יותר" בתחום המוכר של קבוצות סופיות. הבה נחזור אפוא לאולם הריקודים שלנו. הפעם נניח אבל, שבהיכנסנו לאולם *לא* כל הנוכחים מרקדים, אלא חלקם יושבים על הספסל בהמתנה. נניח עוד, שכל חובשי הספסל הם גברים. מה יש בחדר יותר: גברים או נשים? התשובה המיידית היא: גברים, ושוב אין כל צורך בספירה בשביל להגיע למסקנה זו. די לנו בכך, שיש פונקציית שקילות בין קבוצות הנשים בחדר ובין קבוצה חלקית-ממש לקבוצת הגברים (או פונקציה ח.ח.ע. מקבוצת הנשים אל קבוצת הגברים, שאיננה על).

ניסיון לא זהיר להכליל עקרון זה לקבוצות כלשהן יביא, מן הסתם, ל"הגדרה" הבאה: "קבוצה A היא קטנה ממש מקבוצה B אם קיימת פונקציה ח.ח.ע. מ- A על קבוצה חלקית-ממש של B". ברם, לפי זה נקבל, למשל, ש- $\mathbf{N}$ הינה קטנה-ממש מעצמה (כי $\mathbf{N}$ אקויפוטנטית, כזכור, ל- $\mathbf{N}_{\text{even}}$, החלקית לה ממש), בעוד יחס סדר חזק אמור להיות אי-רפלקסיבי! הלקח הינו, שהכללה של מושגים מהתחום הסופי אל התחום האינסופי צריכה להיעשות בזהירות מופלגת. לא כל עקרון, שהוא "מובן מאליו" בתחום הסופי, נשאר נכון לגבי קבוצות אינסופיות.

עם זאת, מה שעדיין נראה אינטואיטיבית כנכון, ללא ספק, הוא העיקרון הבא, אותו נהפוך להגדרה:

הגדרה:

יהיו a ו- b עוצמות, כך ש- $a=|A|$ ו- $b=|B|$. נאמר ש- $a \le b$ ("a קטנה או שווה ל- b"), אם קיימת פונקציה ח.ח.ע. מ- A ל- B.

בהגדרה זו יש בעייתיות מסוימת, האופיינית כמעט לכל ההגדרות, הקשורות בעוצמות. הבה נסבירה אפוא בפרוטרוט, ונראה איך מתגברים עליה. מה שמתרחש כאן הוא, שכדי להגדיר מושג הקשור בעוצמות (במקרה הזה, $a \le b$), אנו משתמשים בקבוצות ***מייצגות*** עבור עוצמות אלה (כאן: קבוצות A ו- B כך ש- $|A|=a$ ו- $|B|=b$) ומגדירים את המושג באמצעותן. הבעיה היא, שלעוצמה אחת יש מייצגים רבים, וחייבים אנו להיות בטוחים, שההגדרה אינה תלויה במייצגים שבחרנו. שוו, למשל, בנפשכם מצב, בו יש קבוצות A_1, A_2, B_1 ו- B_2, כך ש- $|A_1|=a$, $|A_2|=a$, $|B_1|=b$, $|B_2|=b$, יש פונקציה ח.ח.ע. מ- A_1 ל- B_1, אך ***אין*** פונקציה ח.ח.ע. מ- A_2 ל- B_2. במצב כזה, שמעון, שהיה

מסתמך על A_1 ו- B_1, היה מגיע למסקנה ש- $a \le b$. לעומתו לוי, שהיה מסתמך על A_2 ו- B_2, היה מגיע למסקנה ההפוכה – ושניהם היו צודקים לפי ה"הגדרה" למעלה! במצב עניינים כזה היינו אומרים, שהיחס $\le$ ***אינו מוגדר היטב***, כיוון שה"הגדרה" שניתנה תלויה במייצגים. המשימה הראשונה לכן, בכל פעם שנותנים הגדרה כזו (עקרונית: עוד *לפני* שנותנים הגדרה כזו!), היא להראות, שמצב כזה אינו נוצר, ושהמושג שהוגדר – אכן הוגדר היטב.

לגבי $\le$, אי-תלות כזו מובטחת בטענה הבאה:

טענה:

היחס $\le$ על עוצמות מוגדר היטב: אם $|A_1| = |A_2|$, $|B_1| = |B_2|$ וקיימת פונקציה ח.ח.ע. f מ- A_1 אל B_1, אז קיימת גם פונקציה ח.ח.ע. מ- A_2 אל B_2.

הוכחה:

כיוון ש- $|A_1| = |A_2|$, אז קיימת פונקציה ח.ח.ע. F מ- A_1 על A_2. בדומה, קיימת פונקציה ח.ח.ע. G מ- B_1 על B_2. F ו- G הן כמובן הפיכות, והפונקציות ההפוכות, F^{-1} ו- G^{-1}, אף הן חד-חד-ערכיות. עתה, $G \circ f \circ F^{-1}$ היא פונקציה מ- A_2 ל- B_2, והיא ח.ח.ע., כיוון שהיא הרכבה של פונקציות חד-חד-ערכיות.

הערות:

1. רעיון ההוכחה מודגם על-ידי הדיאגרמה הבאה, שבה חצים עבים רציפים מייצגים פונקציות, שקיומן מובטח ישירות על-ידי הנתונים:

$$\begin{tikzcd}
A_1 \arrow[r, shift left, "F"'] \arrow[d, "f"'] & A_2 \arrow[l, shift left, "F^{-1}"'] \arrow[d, dotted, "?"] \\
B_1 \arrow[r, shift right, "G"'] & B_2 \arrow[l, shift right, "G^{-1}"']
\end{tikzcd}$$

ברור מהדיאגרמה, שאפשר "ללכת" מ- A_2 ל- B_2 במסלול:

$$\begin{tikzcd}
A_1 \arrow[d, "f"'] & A_2 \arrow[l, "F^{-1}"'] \\
B_1 \arrow[r, "G"'] & B_2
\end{tikzcd}$$

לכן טבעי לשער, ש- $G \circ f \circ F^{-1}$ היא הפונקציה המבוקשת.

דיאגרמות מסוג זה יכולות להיות לעזר רב במציאת הוכחות ובהבנתן.

2. ניסוח אחר של הטענה האחרונה הוא: אם $|A_1| = |A_2|$, $|B_1| = |B_2|$, ו- $|A_1| \le |B_1|$, אז $|A_2| \le |B_2|$.
3. לגבי מספרים טבעיים, היחס $\le$, כמו שהוגדר כאן, והיחס $\le$, המוכר לנו, הם זהים.

נביא עתה אפיונים נוספים ליחס $\le$ על עוצמות:

משפט:

תהיינה a ו- b עוצמות. נניח $a = |A|$, $b = |B|$. אז:

(א) $a \le b$ אם"ם A אקויפוטנטית עם קבוצה חלקית כלשהי של B.

(ב) $a \le b$ אם"ם $A = \varnothing$ או יש פונקציה מ- B על A.

הוכחה:

(א) נובע מיידית מההגדרות והעובדה, שאם f פונקציה ח.ח.ע. מ- A ל- B, אז $A \sim f(A) \subseteq B$.

(ב) ($(\Leftarrow)$): נניח $a \le b$ ו- $A \ne \varnothing$ אז יש פונקציה ח.ח.ע. מ- A ל- B, ויש איבר $a_0 \in A$. נגדיר:

$$g = \lambda x \in B.\ \text{if}\ \ x \in f(A)\ \ \text{then}\ \ (\iota z \in A.\ f(z) = x)\ \ \text{else}\ \ a_0$$

(השימוש באופרטור ι (מפרק א.5) בהגדרת g מוצדק כאן, כיוון ש- f ח.ח.ע.).

מיידי, כשלכל $y \in A$ מתקיים ש- $y = g(f(y))$, ולכן $y \in g(B)$. לכן g על A.

($\Rightarrow$): אם $A = \varnothing$, אז $\varnothing$ היא פונקציה ח.ח.ע. מ- A ל- B ($\varnothing$ היא פונקציה מ- $\varnothing$ אל כל קבוצה, והיא ח.ח.ע. באופן טריביאלי!).

אם קיימת פונקציה g מ- B על A, אז לכל $x \in A$ הקבוצה $g^{-1}(\{x\})$ אינה ריקה. נבחר אפוא לכל x איבר b_x ב- $g^{-1}(\{x\})$. אז $g(b_x) = x$. נגדיר עתה $f: A \to B$ על-ידי שנתאים לכל $x \in A$ את b_x (רשמית: $f = \{\langle x, b_x\rangle \mid x \in A\}$). f זו הינה ח.ח.ע., כיוון ש:

$$\begin{aligned} f(x) = f(y) &\Rightarrow b_x = b_y \\ &\Rightarrow g(b_x) = g(b_y) \\ &\Rightarrow x = y \end{aligned}$$

%%

הערה:

בהוכחת כיוון ה"אם" ($\Rightarrow$) בחלק ב' של המשפט האחרון, השתמשנו באופן מפורש ב*אקסיומת הבחירה*: את b_x הרי לא הגדרנו במפורש (בעזרת סימון-למדא, למשל),

ולא נתנו שום כלל איך "לבחור" אותו. עתה, אקסיומת הבחירה מבטיחה, שאם $\{B_x | x \in A\}$ היא קבוצה של קבוצות לא ריקות, אז קיימת פונקציה f מ- A אל $\bigcup_{x\in A} B_x$ כך ש- $\forall x \in A. f(x) \in B_x$. במקרה הנוכחי $B_x = g^{-1}(\{x\})$, $\bigcup_{x\in A} B_x = B$ והאקסיומה היא שסיפקה, למעשה, את ה- f בה השתמשנו.

%%

טענה:

(1) תהיינה a ו- b עוצמות. אז $a \le b$ אם"ם קיימות קבוצות A ו- B, כך ש- $|A| = a$, $|B| = b$, ו- $A \subseteq B$.

(2) נניח a ו- b עוצמות ו- $|B| = b$. אז $a \le b$ אם"ם יש $A \subseteq B$ כך ש- $|A| = a$.

הוכחה: תרגיל.

נעבור עתה לבדוק את השאלה אם היחס $\le$ על עוצמות הינו אכן (כמו שהסימון מרמז) יחס סדר חלקי.

טענה:

היחס $\le$ על עוצמות הינו רפלקסיבי וטרנסיטיבי.

הוכחה:

רפלקסיביות: תהי a עוצמה, ונניח $|A| = a$. כיוון ש- $A \subseteq A$ ו- $A \sim A$, הרי $a \le a$ לפי המשפט הקודם.

טרנסיטיביות: נניח a, b, c עוצמות כך ש- $a \le b$ ו- $b \le c$. נניח $|A| = a$, $|B| = b$ ו- $|C| = c$. אז יש פונקציה ח.ח.ע. f מ- A ל- B ופונקציה ח.ח.ע. g מ- B ל- C. עתה, $g \circ f$ הינה פונקציה ח.ח.ע. מ- A ל- C. לכן $a \le c$.

מה בקשר לתכונה השלישית, אנטי-סימטריות? ובכן, בניגוד לרפלקסיביות ולטרנסיטיביות, ההוכחה שלה אינה טריביאלית כלל ועיקר! זהו התוכן, למעשה, של המשפט הבא, שהינו אחד המשפטים המרכזיים של תורת הקבוצות:

משפט קנטור-ברנשטיין-(שרודר):

היחס $\le$ על עוצמות הוא אנטי-סימטרי: אם a ו- b עוצמות כך ש- $a \le b$ ו- $b \le a$, אז $a = b$.

ניסוחים שקולים:

(א) אם A ו- B קבוצות, כך ש- $|A| \le |B|$ ו- $|B| \le |A|$, אז $|A| = |B|$.

(ב) אם A ו- B קבוצות, כך ש- A אקויפוטנטית עם קבוצה חלקית של B, ו- B אקויפוטנטית עם קבוצה חלקית של A, אז A ו- B אקויפוטנטיות.

(ג) אם A ו- B קבוצות, וקיימות פונקציות ח.ח.ע. $f: A \to B$ ו- $g: B \to A$, אז קיימת פונקציה ח.ח.ע. h מ- A על B.

העובדה, שארבעת הניסוחים אכן שקולים, מיידית מההגדרות של שוויון עוצמות, של היחס $\le$ בין עוצמות ושל אקויפוטנטיות בין קבוצות. את הוכחת המשפט עצמו אפשר למצוא בנספח לפרק זה.

מסקנה:

היחס $\le$ על עוצמות הינו יחס סדר חלקי.

הוכחה:

מיידי מהמשפט האחרון ומהטענה שקדמה לו.

%%

הערה:

בעיה, שמתעוררת באופן טבעי בנקודה זו, היא: האם היחס $\le$ בין עוצמות הוא יחס סדר ***מלא*** (דהיינו: האם לכל שתי עוצמות a ו- b מתקיים, ש- $a \le b$ או $b \le a$)? התשובה הינה חיובית, אך ההוכחה (הנעזרת באקסיומת הבחירה) חורגת ממה שאפשר להביא בקורס זה. בהמשך לא נסתמך לכן על עובדה זו.

%%

כדי להדגים את הכוח של משפט קנטור-ברנשטיין, נוכיח את המשפט הבא (נזכיר לקוראים ש- $|(0,1)| = \aleph$):

משפט:

אם A היא קבוצה חלקית של $\mathbf{R}$, המכילה קטע פתוח (דהיינו: $(a,b) \subseteq A \subseteq \mathbf{R}$ עבור מספרים ממשיים a,b כלשהם כך ש- $a < b$), אז $|A| = \aleph$.

הוכחה:

נשתמש בשתי למות, שאת הוכחתן ראינו בפרק ג.1 (ראה טבלה ג.1):

למה 1:

$(a,b) \sim (0,1)$ לכל $a,b \in \mathbf{R}$ כך ש- $a<b$.

למה 2:

$\mathbf{R} \sim (-1,1)$.

עתה, מלמה 1 נובע, ש- $(0,1) \sim (-1,1)$, ולכן, מלמה 2, $\mathbf{R} \sim (0,1)$. לכן, שוב מלמה 1, $(a,b) \sim \mathbf{R}$ עבור כל קטע (a,b) כך ש- $a<b$. לכן $|\mathbf{R}| = |(a,b)|$ עבור כל קטע (a,b) כזה. עתה, אם $(a,b) \subseteq A \subseteq \mathbf{R}$, אז $|(a,b)| \le |A| \le |\mathbf{R}|$. אבל כיוון ש- $|(a,b)| = |\mathbf{R}|$, אנו מקבלים ש- $|\mathbf{R}| \le |A|$ ו- $|A| \le |\mathbf{R}|$. לכן, לפי משפט קנטור-ברנשטיין, $|A| = |\mathbf{R}|$. נותר עוד לכן להראות ש- $|\mathbf{R}| = א$. אבל כיוון ש- $(0,1) \subseteq [0,1) \subseteq \mathbf{R}$, נובע ממה שהראינו, שגם $|\mathbf{R}| = |[0,1)|$, כלומר: $|\mathbf{R}| = א$.

מסקנה:

ל- $\mathbf{R}$ ולכל קטע עליו (סופי או אינסופי, פתוח, סגור, או חצי פתוח חצי סגור) יש אותה עוצמה: א.

הערות:

(1) ממה שהראינו נובע בפרט, ש- $|[0,1)| = |(0,1)|$. לכן יש פונקציה ח.ח.ע. מ- $(0,1)$ על $[0,1)$. אין זה כה פשוט אבל להציג פונקציה כזו ישירות, ללא שימוש במשפט קנטור-ברנשטיין!

(2) אם נבדוק את הוכחת המשפט האחרון יתברר לנו, שהשתמשנו בעיקרון השימושי הבא:

אם $A \subseteq B \subseteq C$ ו- $|A| = |C|$, אז $|A| = |B| = |C|$

לעיקרון זה אפשר לקרוא ***"עיקרון הסנדויץ"*** עבור עוצמות, והוא מסקנה מיידית ממשפט קנטור-ברנשטיין.

כמו בכל מקרה בו מוגדר יחס סדר חלקי $\le$, ניתן לגזור ממנו יחס סדר חזק:

הגדרה:

יהיו a ו- b עוצמות. $a<b$ אם $a \le b$, אך $a \ne b$.

ברור ש- $|A| < |B|$, אם קיימת פונקציה ח.ח.ע. מ- A אל B, אך לא מ- A על B (ולפי קנטור-ברנשטיין, גם לא פונקציה ח.ח.ע. מ- B אל A). כמו כן ברור, ש- $<$ הוא יחס סדר *חזק* על עוצמות.

דוגמאות:

(1) **לגבי מספרים טבעיים, היחס $<$, כמו שהוגדר כאן, והיחס $<$ המוכר לנו מקודם, הם זהים.**

(2) **אם n טבעי, אז $n < \aleph_0$.**

הוכחה:

כיוון ש- $\{0,1, \ldots, n-1\} \subseteq \mathbf{N}$, $n = |\{0, \ldots, n-1\}| \leq |\mathbf{N}| = \aleph_0$. לא יתכן אבל ש- $n = \aleph_0$ (כלומר ש- $\{0,1, \ldots, n-1\} \sim \mathbf{N}$), כי $\mathbf{N}$ אינסופית (הוכחנו!), בעוד $\{0, \ldots, n-1\}$ הינה סופית.

(3) **$\aleph_0 <$ א.**

הוכחה:

$\mathbf{N} \subseteq \mathbf{R}$. לכן: $\aleph_0 = |\mathbf{N}| \leq |\mathbf{R}| =$ א. כיוון שהראינו בפרק הקודם ש- א $\neq \aleph_0$, נובע מזה, ש- $\aleph_0 <$ א.

דוגמה (3) אינה מקרית. היא מקרה פרטי של המשפט הבא:

משפט:

(1) **אם a עוצמה אינסופית, אז $\aleph_0 \leq a$.**

(2) **אם a עוצמה אינסופית ו- $a \neq \aleph_0$, אז $\aleph_0 < a$.**

הוכחה:

(1) בפרק הקודם ראינו, שכל קבוצה אינסופית מכילה קבוצה בת-מניה. מכאן, שאם a אינסופית ו- $a = |A|$, אז יש קבוצה $B \subseteq A$ כך ש- $|B| = \aleph_0$. לכן $a = |A| \geq |B| = \aleph_0$. את הכיוון ההפוך (ש- $\aleph_0 \leq a \Leftarrow a$ אינסופית) נשאיר לקורא.

(2) מיידי מ- (1).

מסקנה:

אם a עוצמה כך ש- $a < \aleph_0$, אז a סופית (כלומר: $a \in \mathbf{N}$).

עד כה הכרנו שתי עוצמות אינסופיות בלבד: $\aleph_0$ ו- $\aleph$. המשפטים, שהוכחנו עד כה, עלולים ליצור את הרושם, שאלו הן כל העוצמות האינסופיות. המשפט הבא הוא משפט מפתח המראה, *שאין הדבר כך* !

משפט קנטור:

$|P(A)| > |A|$ לכל קבוצה A.

הוכחה:

$\lambda x \in A.\{x\}$ היא פונקציה ח.ח.ע. מ- A אל $P(A)$. לכן $|A| \le |P(A)|$. נותר להראות, ש- $|P(A)| \ne |A|$. בשביל זה מספיק להראות, שלא תיתכן פונקציה מ- A על $P(A)$. תהי אפוא F פונקציה מ- A אל $P(A)$, ונראה ש- F לא על $P(A)$. לצורך זאת נתבונן בקבוצה הבאה:

$$\mathbf{C}_F = \{x \in A \mid x \notin F(x)\}$$

ברור ש- $C_F \in P(A)$. עתה, לו היתה F על $P(A)$, כי אז היה $a \in A$, כך ש- $F(a) = C_F$. אבל אז היינו מקבלים ש:

$$\begin{aligned} a \in F(a) &\Leftrightarrow a \in C_F \\ &\Leftrightarrow a \in \{x \in A \mid x \notin F(x)\} \\ &\Leftrightarrow a \notin F(a) \end{aligned}$$

קיבלנו ש- $a \in F(a) \Leftrightarrow a \notin F(a)$, וזו סתירה לוגית. מכאן ש- F אינה יכולה להיות על $P(A)$.

הסבר נוסף להוכחה:

הוכחת משפט קנטור נותנת דרך קונסטרוקטיבית איך, בהינתן $F: A \to P(A)$, אפשר לבנות איבר ב- $P(A)$, שאינו בתמונה של F. נדגים זאת עם $A = \{1,2\}$. במקרה זה $P(A) = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$, ויש 16 פונקציות ב- $A \to P(A)$. ניקח שתיים כדוגמה:

$$\begin{aligned} F_1 &= \lambda x \in A.\ \text{If}\ x = 1\ \text{then}\ \{1\}\ \text{else}\ \{1,2\} \\ F_2 &= \lambda x \in A.\ \text{If}\ x = 1\ \text{then}\ \{2\}\ \text{else}\ \{1,2\} \end{aligned}$$

עתה, $C_{F_1} = \varnothing$ (כי $1 \in F_1(1) = \{1\}$ ו- $2 \in F_1(2) = \{1,2\}$), בעוד $C_{F_2} = \{1\}$ (כי $1 \notin F_2(1) = \{2\}$, בעוד $2 \in F_2(2) = \{1,2\}$). ואכן: $\varnothing$ אינו בתמונה של F_1 (שהיא $\{\{1\},\{1,2\}\}$), ו- $\{1\}$ אינו בתמונה של F_2 (שהיא $\{\{2\},\{1,2\}\}$).

מסקנות ממשפט קנטור:

(1) אין עוצמה מקסימלית.

(2) $\aleph_0 = |\mathbf{N}| < |P(\mathbf{N})| < |P(P(\mathbf{N}))| < |P(P(P(\mathbf{N})))| < \ldots$

מהמסקנה השנייה ברור, שיש אינסוף עוצמות אינסופיות. אלה, אגב, אינן כולן. אם נגדיר, למשל, $P^n(\mathbf{N}) = \overbrace{P(P(\ldots P}^{n \text{ פעמים}}(\mathbf{N}))\ldots)$ (P מופעל n פעמים על $\mathbf{N}$) ו- $\mathbf{N}_\infty = \bigcup_{n=0}^{\infty} P^n(\mathbf{N})$, אז לכל n $\mathbf{N}_\infty \supseteq P^{n+1}(\mathbf{N})$, ולכן $|\mathbf{N}_\infty| \geq |P^{n+1}(\mathbf{N})| > |P^n(\mathbf{N})|$, לפי משפט קנטור. $|\mathbf{N}_\infty|$ היא לכן עוצמה הגדולה מכל אלה בסדרה למעלה, ולפי מסקנה (1) גם היא אינה מקסימלית!

*נספח:
הוכחת משפט קנטור-ברנשטיין

משפט עזר:

נניח כי $f: A \to D$ היא פונקציה ח.ח.ע., וש- $D \subseteq A$. אז $A \sim D$.

הוכחת משפט העזר:

נגדיר:

$$A^* = \{x \in D \mid \exists\, n \in \mathbf{N}\ \exists a \in A - D.\ x = f^n(a)\}$$
$$F = \lambda x \in A.\ \text{If}\ \ x \in A^*\ \ \text{then}\ \ f(x)\ \ \text{else}\ x$$

במלים: A^* מכילה את כל הנקודות, הנמצאות על "מסלול" מהצורה:

$$x = f^0(x), f(x), f^2(x), f^3(x), \ldots$$

המתחיל בנקודה הנמצאת ב- $A - D$ ($f^n(x)$ מוגדר לכל $n \geq 0$, כי $D \subseteq A$ ו- $f: A \to D$, ולכן $f: A \to A$). בפרט $A - D \subseteq A^*$ (המקרה $n = 0$ בהגדרת A^*). אינטואיטיבית, מה ש- F עושה הוא להזיז "ימינה" במסלול כל נקודה הנמצאת על מסלול כזה, ולהשאיר במקומן את הנקודות האחרות.

נראה עתה ש- F היא פונקציית שקילות מ- A על D.

(א) $F: A \to D$

ברור ש- A היא התחום של F. נראה ש- $F(x) \in D$ לכל $x \in A$. ואכן, אם $x \in A^*$, אז $F(x) = f(x) \in D$ (כי $f: A \to D$). אם $x \notin A^*$, אז $x \in D$ (כי $A - D \subseteq A^*$), ולכן $F(x) = x \in D$ גם במקרה זה.

(ב) F היא על D

יהי $y \in D$. נראה, שקיים $x \in A$ כך ש- $y = f(x)$.

(i) אם $y \in A^*$, אז מהגדרת A^* והעובדה ש- $y \notin A - D$ (כי $y \in D$), נובע, שיש $a \in A - D$ ו- $n \geq 1$ כך ש- $y = f^n(a)$. עתה, $f^{n-1}(a) \in A^*$, שוב מהגדרת A^*. לכן, אם נקח $x = f^{n-1}(a)$ נקבל:

$$F(x) = f(x) = f(f^{n-1}(a)) = f^n(a) = y$$

(ii) אם $y \notin A^*$ אז $F(y) = y$, ולכן נוכל במקרה זה לקחת $x = y$.

(ג) F ח.ח.ע.

נניח $F(x_1) = F(x_2)$, כאשר $x_1, x_2 \in A$. נראה ש- $x_1 = x_2$.

(i) נניח $x_1 \in A^*$, $x_2 \notin A^*$. אז מהנתון $F(x_1) = F(x_2)$ והגדרתה של F נובע אז, ש- $f(x_1) = x_2$. במקרה זה קיימים אבל $a \in A - D$ ו- $n \in \mathbf{N}$, כך ש- $x_1 = f^n(a)$. מכאן ש- $x_2 = f(x_1) = f^{\,n+1}(a)$, ולכן גם $x_2 \in A^*$. זוהי סתירה לנתון. מקרה זה הינו לכן בלתי אפשרי.

(ii) נניח $x_1 \notin A^*$, $x_2 \in A^*$. נקבל כאן שוב סתירה, בדיוק כמו במקרה הקודם.

(iii) נניח $x_1 \in A^*$ וגם $x_2 \in A^*$. אז מ- $F(x_1) = F(x_2)$ נובע כאן ש- $f(x_1) = f(x_2)$, ולכן $x_1 = x_2$ (כי f ח.ח.ע.).

(iv) נניח $x_1 \notin A^*$ וגם $x_2 \notin A^*$. אז הנתון $F(x_1) = F(x_2)$ פירושו $x_1 = x_2$, כי $F(x_1) = x_1$ ו- $F(x_2) = x_2$ במקרה זה.

הוכחת משפט קנטור-ברנשטיין:

נניח ש- $f: A \to B$ ו- $g: B \to A$ הן פונקציות ח.ח.ע. תהי $D = g[B]$. אז $D \subseteq A$ ו- $g \circ f$ היא פונקציה ח.ח.ע. מ- A ל- D. לפי משפט העזר, $A \sim D$. כיוון ש- g פונקציית שקילות מ- B על D, גם $B \sim D$. משתי השקילויות נובע, כי $A \sim B$ (ולכן יש פונקציית שקילות מ- A על B).

ג.4 פעולות על עוצמות

אחרי שהכללנו בפרק הקודם את יחס הסדר של N לעוצמות כלשהן, נכליל בפרק זה כמה מהפעולות הבסיסיות על מספרים טבעיים: חיבור, כפל וחזקה. המפתח להכללות יהיה תכונה מס' (8) בטבלת התכונות הבסיסיות של קבוצות סופיות (טבלה ג.2).

נפתח בחיבור:

הגדרה:

יהיו a ו- b עוצמות. אז $a + b = |A \cup B|$, כאשר A ו- B הן קבוצות, כך ש- $|A| = a$, $|B| = b$ ו- $A \cap B = \varnothing$.

כמו שקורה בכל ההגדרות הקשורות בעוצמות, יש קודם כל להצדיק את ההגדרה ולהראות, שחיבור עוצמות אכן מוגדר היטב. בשביל זה יש, ראשית חוכמה, להראות, שהתוצאה הסופית, $a + b$, אינה תלויה במייצגים A ו- B (השווה עם הדיון אחרי הגדרת הסדר על עוצמות בפרק הקודם, ועם הטענה, שבאה אחרי הדיון הנ"ל). זאת עשינו כבר, למעשה, בפרק קודם, במסגרת הלמה, אותה הוכחנו מיד אחרי ההגדרה של קבוצות בנות-מניה (בפרק ג.2). ברם, בניגוד למה שקורה עם הטענה המקבילה במקרה של $\leq$ (ולמה שיקרה בהגדרות של כפל וחזקה), הלמה הנ"ל ***אינה*** מספיקה בשביל להראות, שחיבור עוצמות מוגדר היטב! מה שהיא מראה הוא, ***שאם*** $a + b$ מוגדר, אז הוא מוגדר היטב (כלומר, באופן יחיד). עדיין לא ברור אבל, שכל שתי עוצמות ניתן אכן לחבר! בשביל זה יש להראות את הדבר הבא:

טענה 1:

אם a ו- b עוצמות, אז ***יש*** קבוצות A ו- B כך ש- $|A| = a$, $|B| = b$ ו- $A \cap B = \varnothing$.

הוכחה:

כיוון ש- a ו- b עוצמות, אז קיימות קבוצות A' ו- B', כך ש- $|A'| = a$, ו- $|B'| = b$. נגדיר $A = A' \times \{0\}$, $B = B' \times \{1\}$. ברור ש- $A \cap B = \varnothing$. כמו כן, $\lambda x \in A'. \langle x,0\rangle$ הינה פונקציית שקילות מ- A' על A. לכן $a = |A'| = |A|$. באופן דומה נראה גם, ש- $|B| = |B'| = b$.

מסקנה מהלמה (בפרק ג.2) ומהטענה: חיבור כל שתי עוצמות מוגדר היטב.

דוגמאות:

(1) אם $n, m \in \mathbf{N}$, אז $n+m$ הוא בדיוק תוצאת החיבור הרגיל.

(2) $\aleph_0 + \aleph_0 = \aleph_0$.

הוכחה:

$\mathbf{N} = \mathbf{N}_{\text{odd}} \cup \mathbf{N}_{\text{even}}$ ו- $\mathbf{N}_{\text{odd}} \cap \mathbf{N}_{\text{even}} = \varnothing$. לכן

$$\aleph_0 = |\mathbf{N}| = |\mathbf{N}_{\text{odd}}| + |\mathbf{N}_{\text{even}}| = \aleph_0 + \aleph_0$$

(כזכור, $|\mathbf{N}_{\text{even}}| = |\mathbf{N}_{\text{odd}}| = \aleph_0$, כיוון ש- $\mathbf{N} \sim \mathbf{N}_{\text{odd}} \sim \mathbf{N}_{\text{even}}$).

מסקנה: $|\mathbf{Z}| = \aleph_0$.

הוכחה:

$\mathbf{Z} = \mathbf{N} \cup \{-n | \, n \in \mathbf{N}^+\}$ ו- $\mathbf{N} \cap \{-n | \, n \in \mathbf{N}^+\} = \varnothing$. לכן:

$$|\mathbf{Z}| = |\mathbf{N}| + |\{-n | \, n \in \mathbf{N}^+\}|$$

ברור אבל, ש- $|\{-n | \, n \in \mathbf{N}^+\}| = \aleph_0$ (למה?). לכן $|\mathbf{Z}| = \aleph_0 + \aleph_0 = \aleph_0$.

(3) $\aleph + \aleph = \aleph$.

הוכחה:

$(0,2) = (0,1) \cup [1,2)$ ו- $(0,1) \cap [1,2) = \varnothing$. לכן:

$$\aleph = |(0,2)| = |(0,1)| + |[1,2)| = \aleph + \aleph.$$

תרגיל:

$\aleph = \aleph_0 + \aleph$. (רמז: כדאי להשתמש במשפט על עוצמת קבוצות חלקיות של $\mathbf{R}$ המכילות קטע פתוח).

נעבור עתה לכפל עוצמות.

הגדרה:

יהיו a ו- b עוצמות. אז $a \cdot b = |A \times B|$, כאשר A ו- B קבוצות, כך ש- $a = |A|$, $b = |B|$.

טענה 2:

כפל עוצמות מוגדר היטב.

הוכחה:

במקרה זה יש להוכיח רק אי תלות במייצגים (כי לכל עוצמה a יש קבוצה A כך ש- $a = |A|$). כלומר: עלינו להראות, שאם $A \sim A'$ ו- $B \sim B'$, אז $A \times B \sim A' \times B'$.

תהי אפוא f פונקציית שקילות מ- A על A', ו- g פונקציית שקילות מ- B על B'.

אז $\lambda x \in A, y \in B.\ \langle f(x), g(y)\rangle$ היא פונקציית שקילות מ- $A \times B$ על $A' \times B'$ ($\lambda x \in A', y \in B'.\ \langle f^{-1}(x), g^{-1}(y)\rangle$ היא הפונקציה ההפוכה – בדוק!).

דוגמאות:

(1) בגלל (8) מטבלה ג.2, $m \cdot n$ הוא בדיוק הכפל הרגיל כאשר $m,n \in \mathbf{N}$.

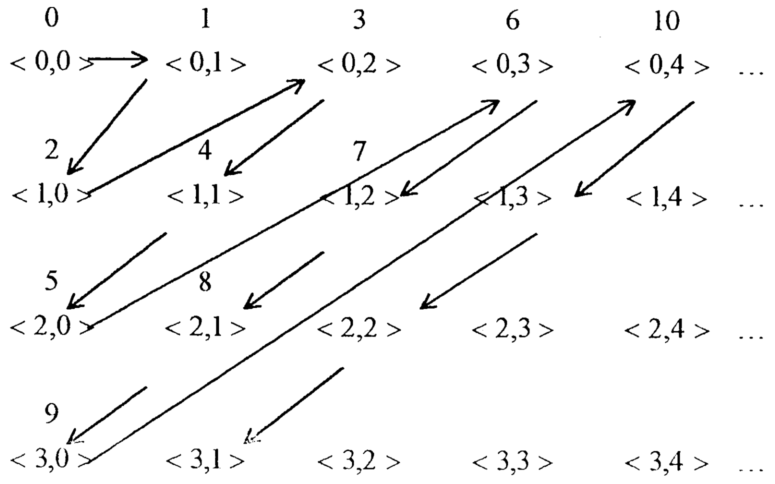
(2) $\aleph_0 \cdot \aleph_0 = \aleph_0$.

הוכחה:

לפי הגדרה, $\aleph_0 \cdot \aleph_0 = |\mathbf{N} \times \mathbf{N}|$. עתה, $\lambda n \in \mathbf{N}, k \in \mathbf{N}.\ 2^n \cdot 3^k$ היא פונקציה ח.ח.ע. מ- $\mathbf{N} \times \mathbf{N}$ אל $\mathbf{N}$, בעוד $\lambda n \in \mathbf{N}.\ \langle n,0\rangle$ היא פונקציה ח.ח.ע. בכיוון ההפוך. לפי משפט קנטור-ברנשטיין נקבל לכן ש- $\aleph_0 = |\mathbf{N}| = |\mathbf{N} \times \mathbf{N}|$.

הערה:

הוכחה זו היא קצרה, אך לא מספקת פונקציית שקילות קונקרטית בין $\mathbf{N} \times \mathbf{N}$ ו- $\mathbf{N}$. פונקציה כזו ניתנה בטבלה ג.1. היא מתקבלת על-ידי התבוננות בדיאגרמה הבאה:



הדיאגרמה מגדירה פונקציה $F : \mathbf{N} \to \mathbf{N} \times \mathbf{N}$, כך ש- $F(0) = \langle 0,0\rangle$, $F(1) = \langle 0,1\rangle$, $F(2) = \langle 1,0\rangle$, $F(3) = \langle 0,2\rangle$, ..., $F(7) = \langle 1,2\rangle$, ... ברור ש-

F זו אכן ח.ח.ע. ועל. לגבי הפונקציה ההפוכה, $F^{-1}: \mathbf{N} \times \mathbf{N} \to \mathbf{N}$, מתקיים: $F^{-1}(<0,0>) = 0$, $F^{-1}(<0,1>) = 1$, $F^{-1}(<1,0>) = 2$, ..., $F^{-1}(<1,2>) = 7$, ... לא קשה להיווכח, שאכן:

$$F^{-1} = \lambda m \in \mathbf{N}, k \in \mathbf{N}. \frac{(m+k)(m+k+1)}{2} + m$$

(3) $\aleph = \aleph_0 \cdot \aleph$. ואכן: $\lambda x \in [0,1), z \in \mathbf{Z} \cdot x + z$ היא פונקציית שקילות מ- $[0,1) \times \mathbf{Z}$ על $\mathbf{R}$. לכן $\aleph = |\mathbf{R}| = |[0,1) \times \mathbf{Z}| = \aleph_0 \cdot \aleph$.

נעבור לבסוף לחזקה של עוצמות.

הגדרה:

יהיו a ו- b עוצמות. אז $a^b = |B \to A|$, כאשר A ו- B קבוצות, כך ש- $|A| = a$ ו- $|B| = b$.

הערות:

(1) כשמדובר בחזקות, הסימון A^B (במקום $B \to A$) נוח אולי יותר. בסימון זה – $|A^B| = |A|^{|B|}$.

(2) שוב, ל- $n,m \in \mathbf{N}$ m^n היא החזקה הרגילה, לפי (8) של טבלה ג.2.

טענה 3:

חזקה של עוצמות מוגדרת היטב.

הוכחה:

עלינו להוכיח, שאם $A \sim A'$ ו- $B \sim B'$, אז $A \to B \sim A' \to B'$.

מהנתונים קיימת פונקציה F ח.ח.ע. מ- A על A', ופונקציה G ח.ח.ע. מ- B על B'. תהי $H = \lambda f \in A \to B.\ G \circ f \circ F^{-1}$. ברור ש- $H : (A \to B) \to (A' \to B')$. נראה ש- H פונקציית שקילות. בשביל זה די להראות ש- $H' = \lambda g \in A' \to B'.\ G^{-1} \circ g \circ F$ הינה הפוכה ל- H. ואכן, $H': (A' \to B') \to (A \to B)$, ואם $g \in A' \to B'$, אז:

$$\begin{aligned}(H \circ H')(g) &= H(H'(g)) \\ &= H(G^{-1} \circ g \circ F) \\ &= G \circ (G^{-1} \circ g \circ F) \circ F^{-1} \\ &= (G \circ G^{-1}) \circ g \circ (F \circ F^{-1}) \\ &= i_{B'} \circ g \circ i_{A'} \\ &= g\end{aligned}$$

לכן $H \circ H' = i_{A' \to B'}$. באופן דומה מראים ש- $H' \circ H = i_{A \to B}$.

הערה:
כדי להבין כיצד מגיעים להגדרה זו של H, כדאי להתבונן שוב בדיאגרמה בפרק הקודם, המופיעה אחרי הוכחת העובדה, שהיחס $\le$ בין עוצמות מוגדר היטב (וכמו כן לעיין בהוכחה של העובדה הנ"ל).

דוגמה:
אם $|E| = a$, אז $|P(E)| = 2^a$, או ביתר קיצור: $|P(E)| = 2^{|E|}$.

הוכחה:
ראינו בעבר, ש- $P(E) \sim E \to \{0,1\}$ (ראה טבלה ג.1 בפרק ג.1). לכן:

$$|P(E)| = |E \to \{0,1\}| = |\{0,1\}|^{|E|} = 2^a$$

עיקר העניין והתועלת בפעולות על המספרים הטבעיים היא העובדה, שיש חוקים פשוטים המקשרים ביניהם. טבלה ג.3 בעמוד הבא מרכזת את העובדות החשובות ביותר הקשורות לפעולות החיבור, הכפל והחזקה של מספרים טבעיים, ומסתבר שהן נשארות נכונות גם לגבי עוצמות כלשהן!

משפט:

כל התכונות המופיעות בטבלה ג.3 נכונות לעוצמות a,b,c,d כלשהן.

הוכחת המשפט:
הבה נתחיל ב- (1) כדוגמה. ההוכחה מתחילה בתרגום הטענות על עוצמות לטענות בדבר קבוצות, שאינן מזכירות כלל את מושג העוצמה. כך, למשל, משמעות $(1)_{(i)}$ היא, שאם ניקח קבוצות A ו- B כך ש- $|A| = a$, $|B| = b$ ו- $A \cap B = \varnothing$, אז ל- $A \cup B$ ול- $B \cup A$ תהיה אותה עוצמה, דהיינו $A \cup B \sim B \cup A$. באופן דומה, $(1)_{(ii)}$ פירושו, שאם $|A| = a$ ו- $|B| = b$, אז $A \times B \sim B \times A$. כיוון שהטענות כאן הן עבור כל עוצמה a וכל עוצמה b, ולכל קבוצה יש עוצמה כלשהי, הרי $(1)_{(i)}$ פירושו, שאם $A \cap B = \varnothing$, אז $A \cup B \sim B \cup A$, ו- $(1)_{(ii)}$ פירושו, שלכל A, B $A \times B \sim B \times A$. עתה $(1)_{(i)}$ נכון פשוט משום *שלכל* A ו- B (לא רק כאלו שחיתוכן ריק) מתקיים אפילו ש- $A \cup B = B \cup A$ (וממילא $A \cup B \sim B \cup A$). $(1)_{(ii)}$ נכון כיוון שהראינו בפרק ג.1, ש- $\lambda x \in A, y \in B. \langle y,x \rangle$ היא פונקציית שקילות מתאימה (טבלה ג.1).

טבלה ג.3: התכונות היסודיות של הפעולות על עוצמות

(1)	(i)	$a+b=b+a$	(ii)	$a\cdot b=b\cdot a$			
(2)	(i)	$(a+b)+c=a+(b+c)$	(ii)	$(a\cdot b)\cdot c=a\cdot(b\cdot c)$			
(3)	$a\cdot(b+c)=a\cdot b+a\cdot c$						
(4)	(i)	$a+0=a$	(ii)	$a\cdot 0=0$	(iii)	$a\cdot 1=a$	
	(iv)	$a^0=1$	(v)	$a^1=a$	(vi)	$0^a=0 \ (a\neq 0)$	
	(vii)	$1^a=1$					
(5)	$a^c\cdot b^c=(a\cdot b)^c$						
(6)	$a^b\cdot a^c=a^{b+c}$						
(7)	$(a^b)^c=a^{b\cdot c}$						
(8)	(i)	$a\le b\wedge c\le d \Rightarrow a+c\le b+d$					
	(ii)	$a\le b\wedge c\le d \Rightarrow a\cdot c\le b\cdot d$					
	(iii)	$a\le b\wedge c\le d \Rightarrow a^c\le b^d$		(פרט למקרה $a=c=b=0$, $d\neq 0$)			
(9)	$2^a>a$						

טבלה ג.4: תכונות מקבילות של קבוצות

(1)	(i) $A \cup B = B \cup A$ (ii) $A \times B \sim B \times A$
(2)	(i) $(A \cup B) \cup C = A \cup (B \cup C)$ (ii) $(A \times B) \times C \sim A \times (B \times C)$
(3)	$A \times (B \cup C) = (A \times B) \cup (A \times C)$ (יחד עם: $B \cap C = \varnothing \Rightarrow (A \times B) \cap (A \times C) = \varnothing$)
(4)	(i) $A \cup \varnothing = A$ (ii) $A \times \varnothing = \varnothing$ (iii) $A \times \{\varnothing\} \sim A$ (iv) $\varnothing \to A = \{\varnothing\}$ (v) $\{\varnothing\} \to A \sim A$ (vi) $A \to \varnothing = \varnothing \; (A \neq \varnothing)$ (vii) $A \to \{\varnothing\} = \{\lambda x \in A.\, \varnothing\}$
(5)	$(C \to A) \times (C \to B) \sim C \to (A \times B)$
(6)	$(B \to A) \times (C \to A) \sim B \cup C \to A$ if $B \cap C = \varnothing$
(7)	$C \to (B \to A) \sim B \times C \to A$ וזה שקול ל: $C \to (B \to A) \sim C \times B \to A$
(8)	(i) $A \prec B \wedge C \prec D \Rightarrow A \cup C \prec B \cup D$ if $B \cap D = \varnothing,\ A \cap C = \varnothing$ (ii) $A \prec B \wedge C \prec D \Rightarrow A \times C \prec B \times D$ (iii) $A \prec B \wedge C \prec D \Rightarrow C \to A \prec D \to B$ (פרט למקרה $A = C = B = \varnothing$, $D \neq \varnothing$)
(9)	$A \prec P(A) \wedge \neg(A \sim P(A))$

באופן דומה ניתן לתרגם כל טענה על עוצמות בטבלה ג.3 לטענה על קבוצות, שאינה מזכירה כלל את מושג העוצמה. תרגום זה ניתן (לכל הטענות בטבלה ג.3) בטבלה ג.4. התרגום הוא מיידי מההגדרות, ולכן ניתן להבין את הקשר בין טבלה ג.3 וטבלה ג.4 בקלות, בתנאי שנשים לב לנקודות הבאות:

- במקומות בהם הטענה של עוצמות נובעת באופן מיידי מטענה חזקה יותר בדבר שוויון קבוצות, מובאת בטבלה ג.4 הזהות על קבוצות. $(1)_{(i)}$ הוא דוגמה לכך.
- בתור קבוצה שעוצמה 1 בחרנו כאן (באופן אקראי) ב- $\{\varnothing\}$.
- בסעיף (7) הסתמכנו (במעבר "וזה שקול ל-...") על $(1)_{(ii)}$ (ועל טענה 3 מפרק זה).
- בסעיף (9) הסתמכנו כבר על כך, ש- $|P(A)| = 2^{|A|}$.
- הסימון $A \prec B$ פירושו, שיש פונקציה ח.ח.ע. מ- A ל- B. למעשה, משמעותו היא בדיוק כמו של $|A| \le |B|$, אך הוא אינו מזכיר "עוצמות".

כדאי לציין, שאת סעיפי (8) בטבלה ג.4 ניתן להחליף בטענות נוחות יותר להוכחה על סמך משפט שהוכחנו, והוא: $a \le b$ אם"ם קיימות קבוצות B ו- A כך ש- $b = |B|$, $a = |A|$ ו- $A \subseteq B$. מזה נובע בקלות, ש:

$(8)_{(i)}$ בטבלה ג.3 נובע מ: $A \subseteq B \wedge C \subseteq D \wedge B \cap D = \varnothing \;\Rightarrow\; A \cup C \subseteq B \cup D$

$(8)_{(ii)}$ בטבלה ג.3 נובע מ: $A \subseteq B \wedge C \subseteq D \;\Rightarrow\; A \times C \subseteq B \times D$

$(8)_{(iii)}$ בטבלה ג.3 נובע מ: $A \subseteq B \wedge C \subseteq D \;\Rightarrow\; C \to A \prec D \to B$

(פרט למקרה: $B = C = \varnothing$, $D \neq \varnothing$).

בצורה זו $(8)_{(i)}$ ו- $(8)_{(ii)}$ הופכים לטריביאליים.

נעבור עתה לדון בהוכחות של הטענות בטבלה ג.4. ראשית נעיר, שכל השוויונות שם הם טריביאליים, ועל רובם (אם לא כולם) עמדנו בעבר. לכן לא נוכיחם כאן. היוצא מן הכלל היחיד הוא אולי $4_{(iv)}$, שגם הוא טריביאלי, אבל אין זה אולי טריביאלי להבין, שזה טריביאלי. נסביר אפוא: הקבוצה הריקה $\varnothing$ היא קבוצה של זוגות (ריקה, כמובן), ולכן הינה יחס בין כל שתי קבוצות. כיחס היא מקיימת באופן ריק את תנאי החד-ערכיות, ולכן היא *פונקציה חלקית* מכל קבוצה לכל קבוצה. לבסוף, היא *פונקציה* מ- $\varnothing$ לכל קבוצה A, כיוון שהיא מתאימה משהו מ- A לכל איבר ב- $\varnothing$ (גם זה נכון באופן ריק, כמובן). לכן $\varnothing \in \varnothing \to A$. ברור, שאין שום פונקציה אחרת ב- $\varnothing \to A$, ולכן $\varnothing \to A = \{\varnothing\}$.

אשר ל- (9), אין הוא אלא משפט קנטור.

לגבי הסעיפים ב- (7)-(2) העוסקים באקויפוטנציות, יש צורך בשביל הוכחתם לתת פונקציות שקילות מתאימות. טבלה ג.5 בעמוד הבא מציגה פונקציית שקילות כזו בכל מקרה ומקרה, יחד עם הפונקציה ההפוכה לה. העובדה, ששתי הפונקציות הן אכן הפוכות זו לזו (ובין הקבוצות המתאימות) היא כמעט עניין של חישוב טהור, דומה לזה שעשינו בפרק ב.4 לגבי פונקציית Curry (Cu) וההפוכה שלה (U). החישובים למקרים של (5) ו- (6) מובאים כנספח לפרק זה. מומלץ לעשות אותם תחילה כתרגילים. (אשר ל- (7), הפונקציות Cu ו- U הן בדיוק מה שנדרש להוכחתו, ולכן (7) כבר הוכח למעשה בפרק ב.4!).

לבסוף, לגבי (8) יש (אם מוכיחים אותו ישירות) לתת פונקציות ח.ח.ע. מתאימות בכל אחד משלושת הסעיפים, בהנחה, שנתונות פונקציה ח.ח.ע. F מ- A ל- B ופונקציה ח.ח.ע. G מ- C ל- D (ועם ההנחות הנוספות ב- (i) וב- (iii)). טבלה ג.5 כוללת בסופה גם פונקציות כאלו.

הערות חשובות לטבלה ג.5:

(א) בטבלה נעשה שימוש בפונקציות ההטל π_1 ו- π_2 על זוגות. כדאי לחזור על תכונותיהן היסודיות לפני ההוכחות. במיוחד חשוב ש- $\pi_1(<x,y>) = x$, $\pi_2(<x,y>) = y$, וש- $z = <\pi_1(z), \pi_2(z)>$, כאשר z הינו זוג. כמו כן, נעשה בטבלה שימוש בפונקציות מצומצמות (F/X), וכדאי לחזור גם על תכונות פונקציות כאלה.

(ב) כמו שנרמז בתוך הטבלה, הפונקציה ב- $(2)_{(ii)}$ היא הגדרה רשמית של הפונקציה f מ- $(A \times B) \times C$ ל- $A \times (B \times C)$, שאינטואיטיבית היינו מגדירים כך:

$$f(<<x,y>, z>) = <x, <y,z>> \quad (\text{עבור } x \in A,\ y \in B,\ z \in C)$$

צורת הגדרה מהסוג שתיארנו הרגע ידועה כהגדרה בעזרת pattern matching. היא אפשרית כאן, בגלל שלכל איבר w ב- $(A \times B) \times C$ קיימים $x \in A$, $y \in B$ ו- $z \in C$ *יחידים* כך ש- $w = <<x,y>, z>$.

טבלה ג.5: פונקציות מתאימות לטבלה ג.4

סעיף		פונקציה מתאימה	פונקציה הפוכה
(1)	(ii)	$\lambda x \in A,\ y \in B.\ \langle y,x\rangle$	$\lambda y \in B,\ x \in A.\ \langle x,y\rangle$
(2)	(ii)	$\lambda z \in (A \times B) \times C.$ $\langle \pi_1(\pi_1(z)), \langle \pi_2(\pi_1(z)), \pi_2(z)\rangle\rangle$ $[\, f(\langle\langle a,b\rangle, c\rangle) = \langle a, \langle b,c\rangle\rangle \,]$	$\lambda z \in A \times (B \times C).$ $\langle\langle \pi_1(z), \pi_1(\pi_2(z))\rangle, \pi_2(\pi_2(z))\rangle$ $[\, g(\langle a,\langle b,c\rangle\rangle) = \langle\langle a,b\rangle, c\rangle \,]$
(4)	(iii)	$\lambda z \in A \times \{\varnothing\}.\ \pi_1(z) \quad [f(\langle x,\varnothing\rangle) = x\,]$	$\lambda x \in A.\ \langle x, \varnothing\rangle$
	(v)	$\lambda f \in \{\varnothing\} \to A.\ f(\varnothing)$	$\lambda x \in A.\ \lambda y \in \{\varnothing\}.\ x$
(5)		$\lambda f \in C \to A,\ g \in C \to B.\ \lambda x \in C.$ $\langle f(x), g(x)\rangle$	$\lambda g \in C \to A \times B.$ $\langle \lambda y \in C.\ \pi_1(g(y)), \lambda y \in C.\ \pi_2(g(y))\rangle$
(6)		$\lambda f \in B \to A,\ g \in C \to A.\ \lambda x \in B \cup C.$ If $x \in B$ then $f(x)$ else $g(x)$	$\lambda h \in B \cup C \to A.\ \langle h/B, h/C\rangle$
(7)		$Cu\colon (C \times B \to A) \to (C \to (B \to A)) =$ $\lambda f \in C \times B \to A.\ \lambda x \in C.\ \lambda y \in B.\ f(x,y)$	$U\colon (C \to (B \to A) \to (C \times B \to A)) =$ $\lambda g \in C \to (B \to A).\ \lambda x \in C, y \in B.\ (g(x))(y)$
(8)		נניח ש- $G: C \to D$, $F: A \to B$ הן פונקציות ח.ח.ע.:	
	(i)	$\lambda x \in A \cup C.$ If $x \in A$ then $F(x)$ else $G(x)$	
	(ii)	$\lambda x \in A, y \in C.\ \langle F(x), G(y)\rangle$	
	(iii)	$\lambda h \in C \to A.\ \lambda z \in D. \begin{cases} F(h(G^{-1}(z))) & z \in G(C) \\ b_0 & z \notin G(C) \end{cases}$ (כאשר b_0 איבר כלשהו של B)	

(ג) כאמור, את (8) ניתן להחליף בדברים קלים יותר להוכחה (ומה שיש בטבלה הוא יותר בגדר תרגיל). למעשה, את $(8)_{(i)}$ ו- $(8)_{(ii)}$ כבר הוכחנו, ול- $(8)_{(iii)}$ נתנו ניסוח פשוט יותר. לגבי ניסוח זה די להראות, שאם $A \subseteq B$ ו- $C \subseteq D$, אז:

$$H = \lambda h \in C \to A.\ \lambda z \in D.\ \text{If } z \in C \text{ then } h(z) \text{ else } b_0$$

היא פונקציה ח.ח.ע. מ- $C \to A$ ל- $D \to B$, כאשר b_0 איבר כלשהו של B. ואכן, נניח ש- $h_2, h_2 \in C \to A$ ונניח: $H(h_1) = H(h_2)$ אז:

$$\lambda z \in D.\ \text{If } z \in C \text{ then } h_1(z) \text{ else } b_0 = \lambda z \in D.\ \text{If } z \in C \text{ then } h_2(z) \text{ else } b_0$$

ולכן:

$$\begin{aligned} &\forall z \in D.\ \ z \in C \Rightarrow h_1(z) = h_2(z) \\ &\Rightarrow \forall z \in C.\ h_1(z) = h_2(z) \quad (\text{כי } C \subseteq D) \\ &\Rightarrow h_1 = h_2 \end{aligned}$$

הערות:

1. H היא אכן פונקציה מ- $C \to A$ ל- $D \to B$, בגלל ש- $A \subseteq B$ ו- $C \subseteq D$.
2. אם B ריקה, אז לא נוכל למצוא b_0 כנדרש, ו- H למעלה אינה מוגדרת היטב. במקרה זה אבל $A = \varnothing$. אם בנוסף $C \neq \varnothing$, אז $C \to A = \varnothing$, ולכן $\varnothing$ הינה פונקציה ח.ח.ע. מ- $C \to A$ לכל קבוצה, כולל $D \to B$. אם $C = \varnothing$ ו- $D = \varnothing$, אז $A \to B = C \to D = \{\varnothing\}$. לבסוף, אם $A = B = C = \varnothing$ ו- $D \neq \varnothing$, אז $A \to C = \{\varnothing\}$, בעוד $D \to B = \varnothing$, ואין אז לכן פונקציה מ- $A \to C$ אל $D \to B$.

גישה שונה במעט להוכחת $a \le b \wedge c \le d \Rightarrow a^c \le b^d$ היא לפרק זאת לשני חלקים הבאים, שהוכחת כל אחד קלה יותר. את הפרטים נשאיר לקורא:

(א) $a \le b \Rightarrow a^c \le b^c$

(ב) $c \le d \Rightarrow b^c \le b^d$ (פרט למקרה $b = c = 0$ ו- $d \neq 0$).

דוגמאות לשימושים בחוקי הפעולות

(א) אם $n \subset \mathrm{N}$, אז

$$a \cdot n = \overbrace{a + a + \ldots + a}^{n \text{ פעמים}}$$

$$a^n = a \cdot a \cdot a \cdot \ldots \cdot a$$

הוכחה:

$$a \cdot n = a(1 + 1 + \ldots + 1) = a \cdot 1 + a \cdot 1 + \ldots + a \cdot 1 = a + a + \ldots + a$$

$$a^n = a^{1+1+\ldots+1} = a^1 \cdot a^1 \cdot \ldots \cdot a^1 = a \cdot a \cdot a \cdot \ldots \cdot a$$

%%

הערה:

בפרט נכון הדבר ל- $a \in \mathbf{N}$, ואנו רואים אפוא, שלמספרים הטבעיים פעולות הכפל והחזקה שהגדרנו, אכן זהות לאלה הידועות. זוהי הוכחה לכך, ללא הסתמכות על תכונה מס' (8) בטבלה ג.2 (טבלת התכונות של קבוצות סופיות), ולכן קיבלנו כאן ***הוכחה*** לנוסחאות שם! (תכונה מס' (8) הנ"ל סיפקה אומנם ***מוטיבציה*** להגדרות שלנו, אך מאז ***לא*** הסתמכנו עליה בהוכחות).

%%

(ב) $\aleph_0 \cdot n = \aleph_0$ כאשר $n \in \mathbf{N}^+$

$\aleph \cdot n = \aleph$ כאשר $n \in \mathbf{N}^+$

הוכחה ראשונה:

$\aleph_0 \cdot n = \aleph_0 + \aleph_0 + \ldots + \aleph_0 = \aleph_0$ (כי $\aleph_0 + \aleph_0 = \aleph_0$).

הוכחה שנייה:

$\aleph_0 \cdot 1 \le \aleph_0 \cdot n \le \aleph_0 \cdot \aleph_0 = \aleph_0$ (כי $1 \le n \le \aleph_0$). לכן $\aleph_0 \cdot n = \aleph_0$.

ההוכחות לגבי $\aleph$ דומות (על סמך הנוסחאות $\aleph + \aleph = \aleph$ ו- $\aleph \cdot \aleph_0 = \aleph$, אותן כבר הוכחנו).

(ג) $\aleph_0{}^n = \aleph_0$ כש- $n \in \mathbf{N}^+$.

הוכחה:

אינדוקציה על n [7].

(ד) $\aleph_0 + n = \aleph_0$ כש- $n \in \mathbf{N}$.

הוכחה:

תרגיל.

(ה) אם $a \le \aleph$, אז $\aleph + a = \aleph$.

הוכחה:

$0 \le a \le \aleph \;\Rightarrow\; \aleph = \aleph + 0 \le \aleph + a \le \aleph + \aleph = \aleph$

מסקנה: $\aleph + \aleph_0 = \aleph$, ו- $\aleph + n = \aleph$ ל- $n \in \mathbf{N}$.

---

[7] למעשה, גם בשביל הוכחה 1 של (ב) יש צורך באינדוקציה כדי להראות ש- $\aleph_0 + \aleph_0 + \ldots + \aleph_0 = \aleph_0$ (העובדה שמדובר במספר סופי של מחוברים היא קריטית לצורך זה!). גם הוכחת (א) היא למעשה באינדוקציה.

(ו) $\mathbf{N} \to P(\mathbf{N}) \sim P(\mathbf{N})$

הוכחה:

$$\left|\mathbf{N} \to P(\mathbf{N})\right| = \left|P(\mathbf{N})\right|^{|\mathbf{N}|} = \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0} = \left|P(\mathbf{N})\right|$$

(ז) $P(\mathbf{N}) \times P(\mathbf{N}) \sim P(\mathbf{N})$

הוכחה:

$$\left|P(\mathbf{N}) \times P(\mathbf{N})\right| = \left|P(\mathbf{N})\right| \cdot \left|P(\mathbf{N})\right| = 2^{\aleph_0} \cdot 2^{\aleph_0} = 2^{\aleph_0 + \aleph_0} = 2^{\aleph_0} = \left|P(\mathbf{N})\right|$$

(ח) אם $n \in \mathbf{N}$ ו- $n \geq 2$, אז $\aleph_0^{\aleph_0} = n^{\aleph_0} = 2^{\aleph_0}$.

הוכחה:

$2 \leq n \leq \aleph_0 \leq 2^{\aleph_0}$. לכן: $2^{\aleph_0} \leq n^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq \left(2^{\aleph_0}\right)^{\aleph_0}$.

אבל $\left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0}$. לכן כל האי-שוויונות בשרשרת האחרונה הם למעשה שוויונות.

מסקנה: $(\mathbf{N} \to \mathbf{N}) \sim (\mathbf{N} \to \{0,1\})$.

הערה:

דוגמאות (ו), (ז) והמסקנה ב- (ח) מדגימות היטב את הכוח הרב שבשימוש בחוקי עוצמות. להוכיח עובדות אלו ישירות (על-ידי בניית פונקציות שקילות) אינו דבר של מה בכך – ואילו כאן ניתנת הוכחה של שורה אחת בדיוק... (למעשה, עצם ***גילוי*** העובדות הללו התאפשר בעיקר בגלל השימוש בעוצמות!).

נשתמש עתה בדוגמה (ח) כדי להוכיח את המשפט הבא:

משפט:

$\aleph = 2^{\aleph_0}$ (ולכן $|\mathbf{R}| = 2^{\aleph_0}$).

הוכחה:

נראה ש- $2^{\aleph_0} \leq \aleph \leq 10^{\aleph_0}$. כיוון ש- $10^{\aleph_0} = 2^{\aleph_0}$ לפי דוגמה (ח), נקבל ש- $\aleph = 2^{\aleph_0}$.

עתה $|(0,1)| = \aleph$, ולכל מספר בקטע הזה מתאים פיתוח עשרוני אינסופי יחיד. פיתוח זה אינו בעצם אלא פונקציה מ-$\mathbf{N}^+$ ל- $\{0,1,\ldots,9\}$: אם $x \in (0,1)$, אז $x = 0.x_1x_2x_3\ldots$, ו- $f_x = \lambda i \in \mathbf{N}^+ . x_i$ היא פונקציה כזו. יתר על כן, אם $x \neq y$, אז $f_x \neq f_y$ (כי $x \neq y$ אם"ם הם שונים באיזו ספרה בפיתוח). על כן, $\lambda x \in (0,1) . f_x$ היא פונקציה ח.ח.ע. מ-$(0,1)$ אל $N^+ \to \{0,1,\ldots,9\}$ ומכאן $\aleph = |(0,1)| \leq \left|N^+ \to \{0,\ldots,9\}\right| = 10^{\aleph_0}$.

מצד שני, נוכל להגדיר פונקציה $G : (N^+ \to \{1,2\}) \to (0,1)$ על-ידי:

$$G(f) = 0.f(1)f(2)f(3)\ldots$$

הפונקציה נותנת פיתוחים עשרוניים אינסופיים (ללא אפסים כלל) וברור שהיא ח.ח.ע..
לכן:

$$2^{\aleph_0} = \left|N^+ \to \{1,2\}\right| \le \left|(0,1)\right| = \aleph$$

קיבלנו אכן, ש- $\aleph \le 10^{\aleph_0} = 2^{\aleph_0}$ ו- $2^{\aleph_0} \le \aleph$. לכן $\aleph = 2^{\aleph_0}$.

מסקנה 1: $\mathbf{R} \sim P(\mathbf{N})$.

מסקנה 2: $\aleph = \aleph \cdot \aleph$.

הוכחה:

$\aleph = 2^{\aleph_0} = 2^{\aleph_0 + \aleph_0} = 2^{\aleph_0} \cdot 2^{\aleph_0} = \aleph \cdot \aleph$.

מסקנה 3: $\mathbf{R} \times \mathbf{R} \sim \mathbf{R}$.

הוכחה:

$|\mathbf{R}| = \aleph = \aleph \cdot \aleph = |\mathbf{R} \times \mathbf{R}|$.

תרגילים:

(א) $\mathbf{R}^n \sim \mathbf{R}$ לכל $n \in \mathbf{N}^+$.

(ב) $\mathbf{N} \to \mathbf{R} \sim \mathbf{R}$.

על השערת הרצף

העוצמה האינסופית הקטנה ביותר היא $\aleph_0$. מבין כל העוצמות האינסופיות האחרות, אותן הכרנו עד עתה, הקטנה ביותר היא $\aleph = 2^{\aleph_0}$. שאלה המתבקשת לכן היא: האם באמת אין שום עוצמה ביניהן? ההשערה שהתשובה היא חיובית – כלומר: שאין עוצמה a בין $\aleph_0$ ו- $2^{\aleph_0}$ (או: כל קבוצה חלקית אינסופית של $\mathbf{R}$ היא ב"מ, או שוות-עוצמה עם $\mathbf{R}$) – ידועה בשם ***השערת הרצף*** (CH). קנטור עשה מאמצים נואשים להוכיחה, אך ללא הצלחה. היום אנו יודעים למה: מעבודותיהם של גדל (1940) וכוהן (1963) ***ידוע*** לנו, שמערכת האקסיומות המקובלת של תורת הקבוצות אינה מספיקה בשביל להוכיח או להפריך השערה זו. לעומת זאת, ***לא ידוע*** לנו אם היא נכונה או לא, וספק אם אי-פעם נדע (יש כאלה שסוברים, שזוהי שאלה חסרת משמעות. זה מכניס אותנו אבל רחוק לתחום הפילוסופיה של המתמטיקה, ואין באפשרותנו לדון בסוגיות כאלה במסגרת קורס זה).

על חיסור וחילוק

לסיום פרק זה נעשה דיון קצר בשאלה הבאה: הרחבנו את פעולות החיבור, הכפל והחזקה לעוצמות כלשהן. מדוע, כך תמה אולי הקורא, לא ניסינו להכליל גם את החיסור והחילוק? התשובה היא, שהדבר לא ניתן להיעשות (לעוצמות אינסופיות) בצורה פרודוקטיבית. ניקח לדוגמה את החיסור. במספרים הטבעיים זוהי פעולה הפוכה לחיבור במובן הבא: אם $a \geq b$, אז:

$$x = a - b \Leftrightarrow b + x = a$$

כלומר: אם $a \geq b$, אז $a - b$ הוא ה- x ה*יחיד* כך ש- $b + x = a$. ברם, כשעוברים לעוצמות אינסופיות, שוב אין בהכרח למשוואה $b + x = a$ פתרון יחיד. למשל: למשוואה $\aleph_0 + x = \aleph_0$ יש הפתרונות הבאים: $0,1,2,\ldots$ ואפילו $\aleph_0$ עצמו. כל אחד מהם זכאי באותה מידה לתואר "$\aleph_0 - \aleph_0$". בהמשך נראה (ולא קשה להוכיח זאת כתרגיל כבר עכשיו), שלכל a אינסופי יש למשוואה $a + x = a$ לפחות את הפתרונות $x = 0,1,2,\ldots$ ו- $x = \aleph_0$. לכן ברור, שאין משמעות לדבר על "$a - a$".

ניתן להציג את הבעיה עם חיסור גם מנקודת ראות אחרת. הגדרה "טבעית" לחיסור $a - b$ יכלה להיראות כך: $a - b = |A - B|$, כאשר A ו- B קבוצות כך ש- $B \subseteq A$, $a = |A|$ ו- $b = |B|$. ברם, פעולת "חיסור" כזו *אינה* מוגדרת היטב, כיוון שיש בהגדרתה תלות במייצגים. אם נרצה, למשל, לחשב לפיה את $\aleph_0 - \aleph_0$, הרי אם, מצד אחד, ניקח $A = B = \mathbf{N}$, נקבל $0 = |\varnothing| = |\mathbf{N} - \mathbf{N}| = "\aleph_0 - \aleph_0"$. מצד שני, אם ניקח $A = \mathbf{N}$, $B = \mathbf{N}_{\text{even}}$, אז נקבל:

$$"\aleph_0 - \aleph_0" = |\mathbf{N} - \mathbf{N}_{\text{even}}| = |\mathbf{N}_{\text{odd}}| = \aleph_0$$

כיוון ש- $\aleph_0 \neq 0$ (ובגדול), פעולת החיסור אינה מוגדרת היטב (שימו לב, שלגבי קבוצות *סופיות* ההגדרה הנ"ל היא תקפה!).

באופן דומה אפשר להסביר, למה אין כל טעם לדבר על חילוק של עוצמות אינסופיות. כיוון ש- $\aleph_0 \cdot n = \aleph_0$ ו- $\aleph_0 \cdot \aleph_0 = \aleph_0$, הרי "$\aleph_0 / \aleph_0$" היה צריך להיות שווה ל- 1, ל- 2, ל- 3, ... ואפילו ל- $\aleph_0$!

נספח: הוכחת (5) ו- (6) בטבלה ג.5

הוכחת (5):

נסמן:

$$F = \lambda f \in C \to A, g \in C \to B.\ \lambda x \in C.\ \langle f(x), g(x)\rangle$$

$$G = \lambda h \in C \to (A \times B).\ \langle \lambda y \in C.\ \pi_1(h(y)), \lambda y \in C.\ \pi_2(h(y))\rangle$$

ברור ש- $F : ((C \to A) \times (C \to B)) \to (C \to A \times B)$
ו- $G : (C \to (A \times B)) \to ((C \to A) \times (C \to B))$.

(i) נניח $<f,g> \in (C \to A) \times (C \to B)$, אז:

$$(G \circ F)(< f, g >) \quad = G(F(< f,g >))$$

$$= G(\lambda x \in C. < f(x), g(x) >)$$

$$= < \lambda y \in C.\ \pi_1 \underbrace{((\lambda x \in C. < f(x), g(x) >)(y))}, \lambda y \in C.\ \pi_2 \underbrace{((\lambda x \in C. < f(x), g(x) >)(y))} >$$

$$= < \lambda y \in C.\ \underbrace{\pi_1(< f(y), g(y) >)}, \lambda y \in C.\ \underbrace{\pi_2(< f(y), g(y) >)} >$$

$$= < \lambda y \in C. f(y), \lambda y \in C. g(y) >$$

$$\overset{\eta}{=} < f, g >$$

לכן $G \circ F = i_{(C \to A) \times (C \to B)}$.

(ii) נניח: $h : C \to (A \times B)$, אז:

$$(F \circ G)(h) \quad = F(G(h))$$

$$= F(< \lambda y \in C.\ \pi_1(h(y)), \lambda y \in C.\ \pi_2(h(y)) >$$

$$= \lambda x \in C.\ < \underbrace{(\lambda y \in C. \pi_1(h(y)))(x)}, \underbrace{(\lambda y \in C. \pi_2(h(y)))(x)} >$$

$$= \lambda x \in C. < \pi_1(h(x)), \pi_2(h(x)) >$$

$$= \lambda x \in C.\ h(x) \qquad \text{(כי } h(x) \text{ הינו זוג)}$$

$$\overset{\eta}{=} h$$

לכן $F \circ G = i_{C \to A \times B}$.

הוכחת (6)

נסמן:

$F = \lambda f \in B \to A, g \in C \to A.\ \lambda x \in B \cup C.$ If $x \in B$ then $f(x)$ else $g(x)$

$G = \lambda h \in B \cup C \to A. < h/B, h/C >$

אנו מניחים: $B \cap C = \varnothing$.

(i) נניח $h \in B \cup C \to A$, אז:

$$\begin{aligned}
(F \circ G)(h) &= F(G(h)) \\
&= F(h/B,\ h/C) \\
&= \lambda x \in B \cup C.\ \text{If } x \in B \text{ then } (h/B)(x) \text{ else } (h/C)(x) \\
&= \lambda x \in B \cup C.\ \text{If } x \in B \text{ then } h(x) \text{ else } h(x) \quad (\text{כי } x \notin B \Rightarrow x \in C) \\
&= \lambda x \in B \cup C.\ h(x) \\
&\overset{\eta}{=} h
\end{aligned}$$

לכן $F \circ G = i_{B \cup C \to A}$.

(הערה: בשביל כיוון זה אין צורך להניח ש- $B \cap C = \varnothing$!).

(ii) נניח $<f,g> \in (B \to A) \times (C \to A)$, אז:

$$\begin{aligned}
(G \circ F)(<f,g>) &= G(F(<f,g>)) \\
&= G(\lambda x \in B \cup C.\ \text{If } x \in B \text{ then } f(x) \text{ else } g(x)) \\
&= <(\lambda x \in B \cup C.\ \text{If } x \in B \text{ then } f(x) \text{ else } g(x)) / B, \\
&\qquad (\lambda x \in B \cup C.\ \text{If } x \in B \text{ then } f(x) \text{ else } g(x)) / C> \\
&= <(\lambda x \in B.\ \text{If } x \in B \text{ then } f(x) \text{ else } g(x)), \\
&\qquad \lambda x \in C.\ \text{If } x \in B \text{ then } f(x) \text{ else } g(x)> \\
&= <\lambda x \in B.\ f(x),\ \lambda x \in C.\ g(x)>
\end{aligned}$$

(הרכיב השני – כיוון שאם $x \in C$, אז $x \notin B$, ולכן ה- "else" תופס. זהו המקום היחיד בחישוב בו משתמשים בנתון $B \cap C = \varnothing$!).

$$\overset{\eta}{=} <f,g>$$

לכן $G \circ F = i_{(B \to A) \times (C \to A)}$.

ג.5 קבוצות בנות מניה ותכונותיהן

בפרק זה נדון בעיקר בתכונותיהן של הקבוצות האינסופיות הפשוטות ביותר: הקבוצות בנות המניה. נזכיר, שקבוצה היא בת-מניה (ב"מ), אם עוצמתה היא $\aleph_0$. נזכיר כמו כן, שבפרק ג.3 ראינו, שאם a עוצמה אינסופית, אז $\aleph_0 \leq a$, וש- $a < \aleph_0$ אם"ם a היא סופית (כלומר: $a \in \mathbf{N}$). התכונות של קבוצות בנות-מניה, שנוכיח בפרק זה, מסוכמות בטבלה ג.6 בעמוד 203.

משפט 1:

נניח A ב"מ. אם $B \subseteq A$, או אם יש פונקציה ח.ח.ע. מ- B אל A, או אם יש פונקציה מ- A על B, אז B סופית או ב"מ.

הוכחה:

כל התנאים פירושם הוא, ש- $|B| \leq |A|$ (פרק ג.3). כיוון ש- $|A| = \aleph_0$, $|B| \leq \aleph_0$. מכאן, שאו ש- $|B| = \aleph_0$, או ש- $|B| < \aleph_0$. במלים אחרות: או ש- B ב"מ, או ש- B סופית.

%%

הערה:

את העובדה, שאם יש פונקציה מ- A על B, אז $|B| \leq |A|$, הוכחנו בשעתו על סמך אקסיומת הבחירה. ברם: אם A ב"מ, אז אין צורך באקסיומה זו. ניתן אז לבנות פונקציה ח.ח.ע. מ- B ל- A בצורה מפורשת. ואכן, אם F היא פונקציית שקילות מ- $\mathbf{N}$ על A, ו- f פונקציה מ- A על B, אז $\lambda x \in B.\ F(\mu n \in \mathbf{N}.\ f(F(n)) = x)$ היא פונקציה ח.ח.ע. מ- B ל- A.

%%

משפט 2:

אם A אינסופית ו- B סופית או ב"מ, אז $|A \cup B| = |A|$.

הוכחה:

ברור ש- A ו- $B - A$ הן קבוצות זרות זו לזו, שאיחודן שווה ל- $A \cup B$. לכן:

$$|A \cup B| = |A| + |B - A| \qquad \text{(I)}$$

כיוון ש- $B-A \subseteq B$, $B-A$ היא סופית או ב"מ. לכן:

$$\aleph_0 + |B-A| = \aleph_0 \qquad \text{(II)}$$

כמו כן, מהעובדה ש- A אינסופית נובע שקיימת קבוצה C חלקית ל- A, שהינה בת מניה. כיוון ש- $A-C$ ו- C הן זרות זו לזו ואיחודן הוא A, $|A| = |A-C| + |C|$. לכן:

$$|A| = |A-C| + \aleph_0 \qquad \text{(III)}$$

מ- (I)-(III) נקבל:

$$\begin{aligned}
|A \cup B| &\overset{\text{(I)}}{=} |A| + |B-A| \\
&\overset{\text{(III)}}{=} \left(|A-C| + \aleph_0\right) + |B-A| \\
&= |A-C| + \left(\aleph_0 + |B-A|\right) \\
&\overset{\text{(II)}}{=} |A\text{-}C| + \aleph_0 \\
&\overset{\text{(III)}}{=} |A|
\end{aligned}$$

מסקנה 1:

אם a עוצמה אינסופית, אז $a + \aleph_0 = a$ ו- $a + n = a$ לכל $n \in \mathbf{N}$. במלים אחרות:

$$a \geq \aleph_0 \wedge b \leq \aleph_0 \Rightarrow a + b = a$$

מסקנה 2:

אם A אינסופית, $B \subseteq A$ סופית או ב"מ, ו- $A-B$ אינסופית, אז $|A| = |A-B|$.

הוכחה:

תמיד $|A| = |A-B| + |B|$ כאשר $B \subseteq A$. אבל לפי משפט 2, אם $A-B$ אינסופית, ו- B סופית או ב"מ, אז: $|A-B| + |B| = |A-B|$. לכן $|A| = |A-B|$ במקרה זה.

תרגיל:

אם A אינסופית ו- B סופית, אז $|A| = |A-B|$.

טבלה ג.6: עובדות על קבוצות סופיות או בנות מניה

(1)	(i) $a \le \aleph_0 \Rightarrow a \in \mathbf{N} \vee a = \aleph_0$ (ii) $a < \aleph_0 \Rightarrow a \in \mathbf{N}$
(2)	נניח A ב"מ. (i) אם $B \subseteq A$, אז B סופית או ב"מ. (ii) אם יש פונקציה ח.ח.ע. מ- B ל- A, אז B סופית או ב"מ. (iii) אם יש פונקציה מ- A על B, אז B סופית או ב"מ.
(3)	אם A ב"מ, $B \neq \emptyset$ סופית או ב"מ, אז $A \times B$ ב"מ.
(4)	איחוד סופי או ב"מ של קבוצות סופיות או ב"מ הוא סופי או ב"מ.
(5)	אם A אינסופית ו- B סופית או ב"מ, אז $A \cup B \sim A$ $(a \ge \aleph_0 \Rightarrow a + \aleph_0 = a + n = a)$
(6)	אם A אינסופית, B סופית או ב"מ ו- $A - B$ אינסופית, אז $A - B \sim A$. מסקנות: (i) אם $\lvert A\rvert \ge \aleph_0$ ו- B סופית, אז $A - B \sim A$. (ii) אם $\lvert A\rvert > \aleph_0$ ו- B ב"מ, אז $A - B \sim A$.

מסקנה 3:

אם A אינסופית ולא ב"מ, ו- $B \subseteq A$ סופית או ב"מ, אז $|A - B| = |A|$.

הוכחה:

בנתונים אלה, לא ייתכן ש- $A - B$ סופית, כי אחרת

$$|A| = |A - B| + |B| \leq \aleph_0 + \aleph_0 = \aleph_0$$

וזה סותר את הנתון, ש- A אינסופית ולא ב"מ. מכאן ש- $A - B$ אינסופית, ולכן $|A - B| = |A|$ לפי מסקנה 2.

משפט 3:

אם A היא ב"מ, ו- B קבוצה לא ריקה, שהיא סופית או ב"מ, אז $A \times B$ היא ב"מ.

הוכחה:

מיידי מכך ש- $\aleph_0 \cdot \aleph_0 = \aleph_0$ וש- $\aleph_0 \cdot n = \aleph_0$.$\forall n \in \mathbf{N}^+$.

מסקנה 4:

(א) קבוצת המספרים הרציונליים היא ב"מ.

(ב) קבוצת המספרים האירציונליים היא קבוצה שעוצמתה $\aleph$.

הוכחה:

(א) ראינו ש- $\mathbf{Z}$ ב"מ ולכן גם $\mathbf{Z}\text{-}\{0\}$ ב"מ (למה?). לפי משפט 3 נובע לכן, ש- $\mathbf{Z} \times \mathbf{Z}\text{-}\{0\}$ היא ב"מ. עתה, $\lambda x \in \mathbf{Z}, y \in \mathbf{Z}\text{-}\{0\}.\ x/y$ היא פונקציה מ- $\mathbf{Z} \times (\mathbf{Z}\text{-}\{0\})$ על $\mathbf{Q}$. לפי משפט 1 נובע מזה, ש- $\mathbf{Q}$ סופית או בת-מניה. ברור, שאינה סופית. מכאן שהיא ב"מ.

(ב) המדובר בקבוצה $\mathbf{R}\text{-}\mathbf{Q}$, וידוע ש: $|\mathbf{R}| = \aleph > \aleph_0$. ממסקנה 3 והחלק הראשון של המסקנה הנוכחית נובע לכן, ש- $|\mathbf{R}\text{-}\mathbf{Q}| = |\mathbf{R}| = \aleph$.

הערה:

מהמסקנה האחרונה נובע, שיש הרבה יותר מספרים אי-רציונליים מרציונליים!

משפט 4:

איחוד סופי או ב"מ של קבוצות סופיות או ב"מ הוא סופי או ב"מ:

$$(|I| \leq \aleph_0 \wedge \forall i \in I. |A_i| \leq \aleph_0) \Rightarrow \left|\bigcup_{i \in I} A_i\right| \leq \aleph_0$$

הוכחה:

נבחר לכל $i \in I$ פונקציה f_i מ- **N** על A_i (אפשרי כי $|A_i| \le \aleph_0$).

נגדיר: $G = \lambda i \in I, n \in \mathbf{N}. f_i(n)$. G הינה פונקציה מ- $I \times \mathbf{N}$ אל $\bigcup_{i \in I} A_i$. עתה, אם $x \in \bigcup_{i \in I} A_i$, אז קיים $\ell \in I$, כך ש- $x \in A_\ell$. כיוון ש- f_ℓ היא פונקציה מ- **N** על A_ℓ, יש $n \in \mathbf{N}$ כך ש- $x = f_\ell(n)$. לכן $x = G(\ell, n)$. מכאן ש- G היא פונקציה מ- $I \times \mathbf{N}$ על $\bigcup_{i \in I} A_i$. לכן:

$$\left|\bigcup_{i \in I} A_i\right| \le |I \times \mathbf{N}| = |I| \cdot |\mathbf{N}| \le \aleph_0 \cdot \aleph_0 = \aleph_0$$

הערות:

(1) ההוכחה הנ"ל מניחה, ש- $A_i \ne \varnothing$ לכל $i \in I$ (היכן?). השלימי אותה למקרה, שהנחה זו אינה מתקיימת!

(2) אם בנוסף להנחות של המשפט ידוע גם, ש- $\bigcup_{i \in I} A_i$ אינסופית (למשל: אם A_i אינסופית לאיזה i, או אם $\forall i \in I.\ A_i \subset A_{i+1}$, או אם ידוע, שלכל $n \in \mathbf{N}$ קיים $i \in I$ כך ש- $|A_i| \ge n$), אז $\left|\bigcup_{i \in I} A_i\right| = \aleph_0$.

נביא עתה מספר מסקנות של משפט 4, החשובות במיוחד למדעי המחשב:

משפט 5:

יהי נתון א"ב סופי (או אפילו בן-מניה), שיש בו a סימבולים ($1 \le a \le \aleph_0$). אז קבוצת כל המלים (או ה"מחרוזות") הסופיות, שאפשר להרכיב בעזרתו היא ב"מ. הדבר נכון גם לקבוצות המשפטים, המאמרים והספרים, שאפשר ליצור בעזרת א"ב כזה.

הוכחה:

כל מלה בת k אותיות אפשר לראות כפונקציה מ- $\{1, 2, \ldots, k\}$ אל L – קבוצת הסימבולים בא"ב הנתון. (לדוגמה: אם L היוה הא"ב העברי, אז את המלה "שלום" אפשר לראות כפונקציה f מ- $\{1,2,3,4\}$ אל L, שבה "ש" = $f(1)$, "ל" = $f(2)$, "ו" = $f(3)$, ו- "ם" = $f(4)$). עתה, קבוצת המלים, שאפשר, באופן עקרוני, לבנות בא"ב L היא $\bigcup_{i \in \mathbf{N}^+} \alpha_i$, כאשר α_i היא קבוצת המלים המורכבות מ- i אותיות בדיוק (יש לשים

לב, ש- α_i, למשל, כוללת פה מלים כמו "ףףףף"!). לאור ההבחנה שלנו, α_i (לכל $i \in \mathbf{N}^+$) היא שוות-עוצמה עם $L \to \{1,2,\ldots,i\}$. לכן:

$$|\alpha_i| = |L|^i = a^i \le \aleph_0{}^i = \aleph_0$$

לפי משפט 4 אנו מקבלים, שגם $\left|\bigcup_{i\in\mathbf{N}^+}\alpha_i\right| \le \aleph_0$. מצד שני, הפונקציה המתאימה לכל $k \in \mathbf{N}^+$ את $sss\ldots s$ (k פעמים הסימבול "s", כש- "s" איבר כלשהו של L) היא ח.ח.ע.,

ולכן $\left|\bigcup_{i\in\mathbf{N}^+}\alpha_i\right| \ge \aleph_0$. סך-הכל $\left|\bigcup_{i\in\mathbf{N}^+}\alpha_i\right| = \aleph_0$.

כדי לקבל את קבוצת המשפטים, שניתן ליצור מ- L, עלינו להוסיף סימבול אחד חדש לשפה: "רווח" (או "space"), שתפקידו להפריד בין המלים. משפט פוטנציאלי בשפה בה מדובר הוא פשוט מחרוזת סופית בשפה המורחבת ("זהו משפט", למשל, הינו מחרוזת בת 8 סימבולים). כיוון שעל-ידי הוספת סימבול אחד לא"ב סופי או ב"מ מקבלים א"ב אחר, שאף הוא סופי או ב"מ, גם אוסף המשפטים הוא ב"מ (לפי מה שהראינו על מלים).

לבסוף, כדי לקבל מאמרים או ספרים, יש להוסיף עוד אי-אלו סימני פיסוק. שוב נקבל, שכל ספר אינו אלא מלה (ארוכה למדי...) בא"ב מורחב, שהוא עדיין סופי או ב"מ. לכן גם מספר הספרים הוא ב"מ.

מסקנה:

בשפה המבוססת על א"ב סופי (או ב"מ) יש לכל היותר $\aleph_0$ מלים, $\aleph_0$ משפטים, $\aleph_0$ ספרים אפשריים וכו'.

הוכחה:

ההבדל בין שפות ובין הקבוצות, שבהן דן המשפט הקודם, הוא, שבשפה מסוימת לא כל מחרוזת אפשרית היא גם מלה תקנית (המחרוזת "ףףףף", לדוגמה, אינה מלה בשפה העברית). בדומה, לא כל משפט אפשרי הוא משפט תקין מבחינה תחבירית, וכנ"ל עם ספרים. עם זאת, קבוצת המלים התקניות היא בכל מקרה קבוצה חלקית לקבוצת המחרוזות האפשריות, וזו הינה ב"מ לפי המשפט הקודם. לכן גם היא עצמה סופית או ב"מ. אותו שיקול תופס גם לגבי קבוצת המשפטים בשפה ולגבי קבוצת הטקסטים התקינים בה.

דוגמה:
קבוצת התכניות החוקיות, שאפשר לכתוב בשפות כמו: SCHEME, PASCAL או C++, היא ב"מ.

מהדוגמה האחרונה נובעת מסקנה חשובה מאוד:

מסקנה:

קיימות פונקציות מ- $\mathbf{N}$ ל- $\mathbf{N}$, ששום תכנית ב- SCHEME אינה יכולה לחשב. הדבר נכון גם לכל שפת תכנות אחרת.

הוכחה:
$|\mathbf{N}\to\mathbf{N}|=\aleph_0^{\aleph_0}=2^{\aleph_0}>\aleph_0$, בעוד קבוצת התכניות ב- SCHEME (למשל) היא ב"מ, וממילא קבוצת הפונקציות, שאפשר לחשב בעזרתן, היא סופית או ב"מ (למה?). כיוון שכל הפונקציות הקבועות $\lambda n\in\mathbf{N}\cdot k$ $(k\in\mathbf{N})$ הן חשיבות ב- SCHEME, הקבוצה הנ"ל איננה סופית, ולכן היא ב"מ.

הערה:
למעשה, קבוצת הפונקציות החשיבות ב- SCHEME זהה לקבוצת הפונקציות החשיבות ב- PASCAL, או בכל שפת תכנות אחרת. ההוכחה לכך ניתנת בקורס במודלים חישוביים (או בקורס בתורת הרקורסיה).

תרגילים:

נניח A ב"מ. אז הקבוצות הבאות גם הן ב"מ:

(א) $\mathrm{List}(A)$: קבוצת כל הסדרות הסופיות (או "רשימות") של איברי A.

(ב) $P_{fin}(A)$: קבוצת כל תת-הקבוצות הסופיות של A.

%%

הערה על בעיית העצירה:
אחת הבעיות החשובות ביותר לגבי תכניות מחשב היא: האם תכנית מסוימת עוצרת ונותנת פלט עם קלט מסוים, או שמא ממשיכה היא לרוץ לנצח? היה זה מאוד משמח, לו יכולנו לכתוב תכנית אוניברסלית, שתענה מראש על בעיה זו. העובדה, שקבוצת התכניות, שניתן לכתוב בשפת מחשב נתונה, היא ב"מ, מאפשרת לנסח את הספציפיקציה של תכנית כזו באופן מדויק: אם $g_1,g_2,g_3,\ldots$ היא מניה אפקטיבית של כל התכניות עם קלט אחד מ- $\mathbf{N}$, אז מה שאנו רוצים היא תכנית H, שמקבלת שני קלטים: n ו- k, ומחזירה 1 אם g_n עוצרת עם הקלט k, 0 אם לא. עתה, בעזרת שיטת

האלכסון של קנטור (ראה הוכחת משפט קנטור) אפשר להראות, שתכנית כזו היא בלתי אפשרית. אכן, לו היתה אפשרית היינו יכולים לכתוב תכנית P, שעושה את הדבר הבא: היא מקבלת קלט טבעי n ומעבירה את n ו- n כקלטים ל- H. אם מה ש- H מחזירה הוא 0, P תחזיר 1. אם מה ש- H מחזירה הוא 1, P תרוץ לנצח.[8] ברור ש- P היא תכנית שמחכה לקלט אחד מ- $\mathbf{N}$, לכן $P=g_\ell$ לאיזה $\ell\in\mathbf{N}$. עתה, אם נכניס את ℓ עצמו כקלט ל- P, אז P תעצור אסם g_ℓ עוצרת על ℓ (כי $P=g_\ell$), אסם $H(\ell,\ell)=1$ (מתכונת H), אסם P לא עוצרת על ℓ (מהגדרת P). זוהי סתירה!

%%

[8] חסרים פה כמה פרטים טכניים, כמובן, שמושלמים בקורס ב"מודלים חישוביים".

ג.6 עקרונות קומבינטוריים כלליים

בפרק זה נתאר מספר עקרונות קומבינטוריים כלליים, שנשתמש בהם בעיקר בחלק הבא, העוסק בקומבינטוריקה של קבוצות סופיות. עם זאת, העקרונות, שנדון בהם בפרק זה, נכונים לכל הקבוצות, סופיות ואינסופיות כאחת. רשימת העקרונות מסוכמת בטבלה ג.7 בעמוד הבא. הטבלה כוללת מספר פעולות חדשות על עוצמות: C, P ו- E, אותן נגדיר בפרק זה. בחלק אחר אנו רק חוזרים על תכונות הידועות לנו כבר (אלה שב- (4)).

נעבור להוכחת התכונות והדגמתן:

(1) את הנוסחה ב- (1)(א) מוכיחים באינדוקציה על n. לגבי $n = 2$ זוהי פשוט הגדרת חיבור עוצמות. נניח נכונות ל- $n \geq 2$, נוכיח ל- $n + 1$. אז:

$$\biguplus_{i=1}^{n+1} A_n = \left(\biguplus_{i=1}^{n} A_i \right) \uplus A_{n+1}$$

לכן, לפי הגדרת חיבור עוצמות והנחת האינדוקציה:

$$\left|\biguplus_{i=1}^{n+1} A_i\right| = \left|\biguplus_{i=1}^{n} A_i\right| + |A_{n+1}| \underset{\substack{\uparrow \\ \text{הנחת אינדוקציה}}}{=} \sum_{i=1}^{n} |A_i| + |A_{n+1}| = \sum_{i=1}^{n+1} |A_i|$$

הוכחת (1)(ב) דומה. כאן המקרה $n = 2$ נובע מהזהות:

$$A_1 \cup A_2 = A_1 \uplus (A_2 - A_1)$$

ולכן:

$$|A_1 \cup A_2| = |A_1| + |A_2 - A_1| \leq |A_1| + |A_2|$$

(אי השוויון האחרון נובע מהעובדה, ש- $A_2 - A_1 \subseteq A_2$).
את שלב האינדוקציה נשאיר לקוראים.

טבלה ג.7: עקרונות קומבינטוריים כלליים

(1)	(א)	$\left\|\biguplus_{i=1}^{n} A_i\right\| = \sum_{i=1}^{n} \|A_i\|$
	(ב)	$(\forall i \in \{1,\dots,n\}.\|A_i\| \le a_i) \Rightarrow \left\|\bigcup_{i=1}^{n} A_i\right\| \le \sum_{i=1}^{n} a_i$
(2)	(א)	$\|A_1 \times A_2 \times \dots \times A_n\| = \|A_1\| \cdot \|A_2\| \dots \|A_n\|$
	(ב)	$(\forall \alpha \in I.\|A_\alpha\| = a) \Rightarrow \left\|\biguplus_{\alpha \in I} A_\alpha\right\| = a \cdot \|I\|$
	(ג)	$(\forall \alpha \in I.\|A_\alpha\| \le a) \Rightarrow \left\|\bigcup_{\alpha \in I} A_\alpha\right\| \le a \cdot \|I\|$
(3)	(א)	$\left\|\bigcup_{i=1}^{n} A_i\right\| > \sum_{i=1}^{n} a_i \Rightarrow \exists i \in \{1, \dots, n\}.\|A_i\| > a_i$
	(ב)	$\left\|\bigcup_{\alpha \in I} A_\alpha\right\| > a \cdot \|I\| \Rightarrow \exists \alpha \in I.\|A_\alpha\| > a$
(4)	(א)	$\|A \to B\| = \|B\|^{\|A\|}$
	(ב)	$\|P(A)\| = 2^{\|A\|}$
(5)	(א)	(i) $C(a, 1) = P(a, 1) = a$ (ii) $P(a, 0) = C(a, 0) = 1$
	(ב)	אם $a \ge b$ אז $C(a, b) \ge 1$. אחרת $C(a, b) = 0$ ו- $P(a, b) = 0$
	(ג)	$P(a, b) = C(a, b) \cdot E(b)$
	(ד)	$P(b, b) \ge E(b) \ge 1$
	(ה)	$C(a, b) \le P(a, b) \le a^b$
	(ו)	$C(a, b) \le 2^a$
	(ז)	אם $a_1 \le a_2$ אז:
		$P(a_1, b) \le P(a_2, b)$ $C(a_1, b) \le C(a_2, b)$ $E(a_1) \le E(a_2)$

הערה:

בתורת הקבוצות המתקדמת מוכיחים גם, ש:

$$\forall i \in \{1,\dots,n\}.\ |A_i| < a_i \Rightarrow \left|\bigcup_{i=1}^{n} A_i\right| < \sum_{i=1}^{n} a_i$$

אי-שוויון זה אי-אפשר להכליל לאיחוד אינסופי.

(2) את הוכחת (א) באינדוקציה נשאיר לקוראים.

לגבי (ג), תהי A קבוצה כך ש- $|A| = a$.

תהי $I' = \{\alpha \in I \mid A_\alpha \neq \varnothing\}$, אז $\bigcup_{\alpha\in I'} A_\alpha = \bigcup_{\alpha\in I} A_\alpha$. כמו כן, כיוון ש- $|A_\alpha| \le a$ ו- $A_\alpha \neq \varnothing$ לכל $\alpha \in I'$, הרי קיימת לכל $\alpha \in I'$ פונקציה f_α מ- A **על** A_α. נגדיר: $F = \lambda x \in A,\ \alpha \in I'.\ f_\alpha(x)$. ברור, ש- F היא פונקציה מ- $A \times I'$ אל $\bigcup_{\alpha\in I} A_\alpha$. נראה ש- F היא **על** $\bigcup_{\alpha\in I} A_\alpha$. נניח אפוא ש- $x \in \bigcup_{\alpha\in I} A_\alpha$. אז יש $\alpha_0 \in I'$ כך ש- $x \in A_{\alpha_0}$. כיוון ש- f_{α_0} היא פונקציה מ- A על A_{α_0}, יש $y \in A$ כך ש- $x = f_{\alpha_0}(y)$. לכן $x = F(\alpha_0, y)$.

עתה, כיוון שיש[9] פונקציה מ- $A \times I'$ על $\bigcup_{\alpha\in I} A_\alpha$, אז

$$\left|\bigcup_{\alpha\in I} A_\alpha\right| \le |A|\cdot|I'| = a\cdot|I'| \le a\cdot|I|$$

הוכחת חלק (ב) דומה מאוד. ההבדלים הם, שבמקרה זה $I' = I$, אם $\alpha = \beta$ $A_\alpha \cap A_\beta \neq \varnothing$, ואת f_α אפשר לבחור כך שתהיה גם ח.ח.ע.. מעובדות אלה נובע, ש- F, אותה הגדרנו בהוכחת סעיף (ג), היא לא רק על $\bigcup_{\alpha\in I} A_\alpha$, אלא גם ח.ח.ע..

ואכן, נניח $F(\alpha_1, x_1) = F(\alpha_2, x_2)$. כיוון ש- $F(\alpha_1, x_1) \in A_{\alpha_1}$ ו- $F(\alpha_2, x_2) \in A_{\alpha_2}$, הרי $A_{\alpha_1} \cap A_{\alpha_2} \neq \varnothing$, ולכן $\alpha_1 = \alpha_2$. מכאן, שהשוויון $F(\alpha_1, x_1) = F(\alpha_2, x_2)$ אינו אלא $f_{\alpha_1}(x_1) = f_{\alpha_1}(x_2)$. כיוון ש- f_{α_1} ח.ח.ע., $x_1 = x_2$. קיבלנו, שאם $F(\alpha_1, x_1) = F(\alpha_2, x_2)$, אז $<\alpha_1, x_1> = <\alpha_2, x_2>$, כלומר: F ח.ח.ע..

---

[9] אלא אם כן $A_\alpha = \varnothing$ לכל $\alpha \in I$, אבל במקרה זה הטענה הינה טריביאלית.

דוגמאות:

(א) $\mathbf{R} = \biguplus_{i\in\mathbf{Z}} [i,i+1)$. האיחוד פה הוא אכן איחוד זר, כי $[i, i+1) \cap [j, j+1) = \varnothing$ אם $i \neq j$ ו- j, i הם שלמים. כמו כן, $|[i, i+1)| = \aleph$ לכל $i \in \mathbf{Z}$. לכן לפי (2)(ב), $\aleph_0 \cdot \aleph = |\mathbf{Z}| \cdot \aleph = |\mathbf{R}| = \aleph$ (זו אינה עובדה חדשה, כמובן).

(ב) במטריצה בת 20 שורות, כשבכל שורה יש 5 מקומות, יש לפי (2)(ב) $20 \cdot 5 = 100$ מקומות (כאן $|I| = 20$, $a = 5$). לפי (2)(ג) מספר המספרים הרשומים במטריצה הוא קטן או שווה ל- 100 (כי אותו מספר יכול, כמובן, להופיע ביותר משורה אחת).

(ג) אם נציב ב- (2)(ג) $a = \aleph_0$ ונניח ש- I סופית או בת-מניה, נקבל ש- $\left|\bigcup_{\alpha\in I} A_\alpha\right| \le \aleph_0 \cdot |I| = \aleph_0$. במלים אחרות: האיחוד של מספר סופי או בן-מניה של קבוצות סופיות או בנות מניה הוא קבוצה סופית או ב"מ. עובדה זו ידועה לנו מהפרק הקודם (וההוכחה שם, אם נבדוק, היא מקרה פרטי של ההוכחה כאן).

(ד) מ-(2)(ב) נובע מיידית, שאם F פירוק של A, ו- $|X| = a$ לכל $X \in F$, אז $|A| = |F| \cdot a$. בפרט: אם R יחס שקילות על A, ו- $|X| = a$ לכל $X \in A/R$, אז $|A| = a \cdot |A/R|$. אם, לעומת זאת, $|X| \le a$ לכל $X \in A/R$, אז לפי (2)(ג) $|A| \le a \cdot |A/R|$.

(3) (א) אם נפעיל על (1)(ב) את העקרון הלוגי, ש- $(\varphi \Rightarrow \psi) \Leftrightarrow (\neg\psi \Rightarrow \neg\varphi)$, נקבל:

$$\neg\left(\left|\bigcup_{i=1}^{n} A_i\right| \le \sum_{i=1}^{n} a_i\right) \Rightarrow \exists i \in \{1,\dots,n\}.\neg\left(\left|A_i\right| \le a_i\right)$$

נסתמך עתה, באופן חד-פעמי, על העובדה הכללית, שלכל שתי עוצמות a ו- b מתקיים ש- $\neg(a \le b)$ אם"ם $a > b$[10], ונקבל בדיוק את (3)(א). (3)(ב) נובע באופן דומה מ- (2)(ג).

מ- (3)(ב), לדוגמה, נובע, שאם ננסה לחלק 71 אנשים ל- 7 קבוצות, הרי לפחות אחת מקבוצות אלה תכלול למעלה מ- 10 אנשים (כלומר: לפחות 11). בדוגמה זו $|I| = 7$ ו- $a = 10$.

[10] עובדה זו הינה טריביאלית בשלב זה, אם אחת מהעוצמות a או b הינה סופית או ב"מ – וזהו המקרה השימושי באמת. הטענה נכונה, כזכור, באופן כללי, אם כי לא הוכחנו זאת.

שני העקרונות ב- (3) ידועים *כעקרונות שובך-יונים*, כי שניהם מהווים הכללה ישירה של עקרון שובך היונים בצורתו הפשוטה ביותר, והיא: שאם ננסה לחלק $n+1$ עצמים (או יותר) ל- n קבוצות, הרי תהיה לפחות קבוצה אחת, שבה יהיו שניים (או יותר) מהעצמים. עקרון פשוט זה מתקבל מ- (3)(א) אם נציב $a_i=1$ לכל $1\le i\le n$, ומ-(3)(ב) אם נציב $a=1$ (השם "שובך יונים" נובע מהדימוי הספרותי, לפיו אם ננסה לפזר $n+1$ יונים בשובך שבו n תאים, הרי יהיה לפחות תא אחד, בו יהיו שתי יונים או יותר). עקרונות אלו מהווים, למרות פשטותם, כלי נשק חזק בפתרון בעיות לא מעטות. הבה נביא דוגמה.

תרגיל:

הוכח שבכל קבוצה A, שיש בה 501 מספרים טבעיים בין 1 ל- 1000, יש שני מספרים, שאחד מהם מתחלק בשני.

פתרון:

כל מספר טבעי k אפשר לפרק באופן יחיד למכפלה מהצורה $k=2^{\ell}\cdot m$, כש- m מספר אי-זוגי. כיוון שיש רק 500 מספרים אי-זוגיים בין 1 ל- 1000, יהיו לפי עקרון שובך היונים (הפשוט) לפחות שני מספרים n_1 ו- n_2 ב- A, שיש להם אותו רכיב אי-זוגי, דהיינו $n_1=2^{\ell_1}\cdot m$ ו- $n_2=2^{\ell_2}\cdot m$, כש- m אי-זוגי. עתה n_1 מתחלק ב- n_2 או n_2 מתחלק ב- n_1 (בהתאם, אם $\ell_1\ge\ell_2$ או $\ell_2\ge\ell_1$). (פורמלית, I היא כאן קבוצת המספרים האי-זוגיים בין 1 ל- 1000, ולכל $i\in I$, $A_i=\{n\in A\mid \exists\ell\in\mathbf{N}.n=2^{\ell}\cdot i\}$. אז $A=\bigcup_{i\in I}A_i$, ולכן $\left|\bigcup_{i\in I}A_i\right|=501$, בעוד $a=1$ ו- $|I|=500$).

נעבור עתה להגדיר את הפעולות C, P ו- E המוזכרות ב- (5) של טבלה ג.7.

הגדרה:

(א) $\mathrm{In}(A,B)$ היא קבוצת הפונקציות הח.ח.ע. מ- B ל- A.

(ב) $P_b(A)$ היא קבוצת הקבוצות החלקיות ל- A, שעוצמתן b:

$$P_b(A)=\{X\in P(A)\mid |X|=b\}$$

(ג) $\mathrm{Eq}(A,B)$ היא קבוצת פונקציות השקילות מ- B ל- A.

הערה:

ברור ש: $\mathrm{Eq}(A,B)\subseteq\mathrm{In}(A,B)\subseteq B\to A$, $P_b(A)\subseteq P(A)$ וש- $\mathrm{Eq}(A,B)=\varnothing$ אם $|A|\ne|B|$.

הגדרה:

יהיו a ו- b עוצמות.

(א) $P(a, b)$ היא $|\mathrm{In}(A, B)|$ כש- A ו- B קבוצות, כך ש- $|A| = a$ ו- $|B| = b$.

(ב) $C(a, b)$ (או $\binom{a}{b}$) היא $|P_b(A)|$, כאשר A קבוצה כך ש- $|A| = a$.

(ג) $E(a)$ היא $|\mathrm{Eq}(A, B)|$, כאשר A ו- B קבוצות, כך ש- $|A| = a$ ו- $|B| = a$.

משפט:

הפעולות P, C ו- E מוגדרות היטב.

הוכחה:

צריך להוכיח, שאם $A \sim A'$ ו- $B \sim B'$, אז:

(א) $\mathrm{In}(A, B) \sim \mathrm{In}(A', B')$

(ב) $\mathrm{Eq}(A, B) \sim \mathrm{Eq}(A', B')$

(ג) $P_b(A) \sim P_b(A')$

תהי אפוא F פונקציית שקילות מ- A על A', G פונקציית שקילות מ- B על B'.

לגבי (א): ראינו בהוכחה קודמת, שבנתונים אלו $H = \lambda f \in B \to A.F \circ f \circ G^{-1}$ היא פונקציית שקילות מ- $B \to A$ על $B' \to A'$. כיוון ש- $\mathrm{In}(A, B) \subseteq B \to A$, כל שצריך להראות הוא, שהתמונה לפי H של $\mathrm{In}(A, B)$ היא בדיוק $\mathrm{In}(A', B')$. ואכן, אם $f \in \mathrm{In}(A, B)$, אז $H(f) = F \circ f \circ G^{-1}$ היא פונקציה ח.ח.ע. (כי היא הרכבה של פונקציות ח.ח.ע.). לכן $H(f) \in \mathrm{In}(A', B')$. מצד שני, אם $g \in \mathrm{In}(A', B')$, אז קל לברר, ש- $g = H(F^{-1} \circ g \circ G)$ וש- $F^{-1} \circ g \circ G \in \mathrm{In}(A,B)$.

הוכחת (ב) כמעט זהה.

לבסוף, לגבי (ג), די אם נראה ש- $\lambda X \in P_b(A).\ F(X)$ ($F(X)$ – התמונה של X לפי F) היא פונקציית שקילות בין $P_b(A)$ ל- $P_b(A')$. זוהי, קודם כל, אכן פונקציה מ- $P_b(A)$ אל $P_b(A')$, כי לכל $X \in P_b(A)$ מתקיים ש- $F(X) \in P(A')$ ו- $|F(X)| = |X| = b$ (כי F ח.ח.ע.). שנית, קל לראות, ש- $\lambda Y \in P_b(A').\ F^{-1}(Y)$ היא פונקציה הפוכה, כיוון שכאשר F פונקציית שקילות, אז לכל $X \subseteq A$ ו- $Y \subseteq B$ מתקיים, ש- $F(F^{-1}(Y)) = Y$ ו- $F^{-1}(F(X)) = X$.

כאשר מדובר במספרים טבעיים, הרי $C(n, k)$ הוא בדיוק מה שמסומן בדרך-כלל ב- $\binom{n}{k}$ ומוגדר בתור מספר תת-הקבוצות בנות k איברים, שיש לקבוצה בת n איברים ("מספר האפשרויות לבחור k עצמים מתוך קבוצה בת n איברים"). $P(n, k)$ הוא מספר ה"חליפות" של k עצמים מתוך n ("מספר האפשרויות לבחור k עצמים שונים מתוך קבוצה בת n איברים עם חשיבות לסדר"), ו- $E(n)$ הוא מספר ה"תמורות" של קבוצה בת n איברים ("תמורה" של קבוצה אינה אלא פונקציית שקילות מהקבוצה על עצמה). למעשה, המשפט האחרון מוכיח, שכל הפעולות הנ"ל על k ו- n מוגדרות היטב, דבר שלא נעשה כלל בתיכון!

משפט:

לכל שתי עוצמות a,b מתקיים $P(a,b) = C(a,b).\, E(b)$.

הוכחה:

תהיינה A ו- B קבוצות כך ש- $|A| = a$ ו- $|B| = b$.

הרעיון האינטואיטיבי של ההוכחה הינו פשוט ביותר ומוכר היטב מקומבינטוריקה תיכונית: אסטרטגיה אפשרית לבניית פונקציה ח.ח.ע. מ- B ל- A היא לבחור תחילה את התמונה שלה (שעוצמתה חייבת להיות כשל B, דהיינו: b) ואחר-כך לבחור פונקציית שקילות מ- B על תמונה זו. את הבחירה הראשונה אפשר לעשות ב- $C(a,b)$ דרכים, ואת השנייה אפשר אחר-כך לעשות ב- $E(b)$ דרכים – ולכן הכפל.

הוכחה פורמלית:

אם $f \in \mathrm{In}(A, B)$, אז f היא פונקציית שקילות מ- B על התמונה שלה, $f(B)$. ממילא $b = |B| = |f(B)|$. לכן $f(B) \in P_b(A)$, בעוד $f \in \mathrm{Eq}(f(B), B)$. מכאן ש-

$$\mathrm{In}(A,B) = \biguplus_{X \in P_b(A)} \mathrm{Eq}(X,B)$$

האיחוד כאן הוא זר, כי ברור שפונקציות, שתמונתן שונה, הינן שונות זו מזו. (ממה שכתבנו נובע, בעצם, רק ש- $\mathrm{In}(A,B) \subseteq \biguplus_{X \in P_b(A)} \mathrm{Eq}(X,B)$. הכיוון ההפוך נובע מכך, שאם $X \subseteq A$ ו- $f \in \mathrm{Eq}(X, B)$, אז f היא פונקציה ח.ח.ע. מ- B ל- A. כלומר:

$\forall X \in P_b(A).\ \mathrm{Eq}(X, B) \subseteq \mathrm{In}(A, B)$. לכן $\bigcup_{X \in P_b(A)} \mathrm{Eq}(X, B) \subseteq \mathrm{In}(A, B)$. עתה, מהגדרת $E(b)$ נובע, שלכל $X \in P_b(A)$, $|\mathrm{Eq}(X, B)| = E(b)$. מכאן, לפי (2)(ב) של טבלה ג.7:

$$P(a, b) = |\mathrm{In}(A, B)| = E(b) \cdot |P_b(A)| = E(b) \cdot C(a, b)$$

הערה:

ההוכחה מדגימה תופעה כללית: שיקולים אינטואיטיביים מהסוג של ההוכחה האחרונה (דהיינו: פיצול ל"מקרים", כשכל "מקרה" כולל מספר שווה של "אפשרויות") מתורגמים בהוכחות פורמליות להסתמכות על (2)(ב) של טבלה ג.7.

המשפט האחרון סיפק, למעשה, את ההוכחה של (5)(ג) מטבלה ג.7. ההוכחה של שאר הסעיפים נובעת בקלות מההגדרות (אולי בעזרת (5)(ג)):

(א) נניח $|A| = a$. ברור ש- $\lambda x \in A.\{x\}$ היא פונקציית שקילות בין A ובין $P_1(A)$. לכן $a = |A| = |P_1(A)| = C(a,1)$. ברור גם, ש- $E(1) = 1$. לכן $P(a,1) = a$ על סמך (ג).

(ב) נניח $|B| = b$, $|A| = a$. אם $b \le a$, אז מהגדרת $\le$ יש פונקציה ח.ח.ע. מ- B אל A, כלומר: $\mathrm{In}(A, B) \neq \varnothing$, ולכן $P(a,b) \neq 0$. מ-(ג) נובע לכן, שגם $C(a,b) \neq 0$, דהיינו: $C(a,b) \ge 1$. לעומת זאת, אם $b \not\le a$, אז לא קיימת קבוצה חלקית של A שעוצמתה היא b, דהיינו: $P_b(A) = \varnothing$, ולכן $C(a,b) = 0$. ממילא גם $P(a,b) = 0$ (לפי (ג)).

(ד) נניח B קבוצה, כך ש- $|B| = b$. אז $E(b) = |\mathrm{Eq}(B, B)|$ ו- $P(b,b) = |\mathrm{In}(B, B)|$. כיוון ש- $\mathrm{Eq}(B, B) \subseteq \mathrm{In}(B, B)$, $P(b,b) \ge E(b)$. ש- $E(b) \ge 1$ נובע מזה, שתמיד (אפילו כש- B ריקה!) $i_B \in \mathrm{Eq}(B, B)$.

(ה) ש- $P(a,b) \le a^b$ נובע מכך ש- $\mathrm{In}(A, B) \subseteq B \to A$. ש- $C(a,b) \le P(a,b)$ נובע מ-(ג) ו-(ד).

(ו) נובע מכך ש- $P_b(A) \subseteq P(A)$.

(ז) נניח $a_1 \le a_2$. אז קיימות קבוצות A_1 ו- A_2 כך ש- $|A_1| = a_1$, $|A_2| = a_2$ ו- $A_1 \subseteq A_2$. מכך ש- $A_1 \subseteq A_2$ נובע, לפי הגדרה, ש- $P_b(A_1) \subseteq P_b(A_2)$. לכן $C(a_1,b) \le C(a_2,b)$. כמו כן $E(a_1) = |\mathrm{Eq}(A_1, A_1)|$, $E(a_2) = |\mathrm{Eq}(A_2, A_2)|$. נגדיר אפוא:

$$H = \lambda f \in \mathrm{Eq}(A_1, A_1).\ \lambda x \in A_2\ .\ \text{If}\ x \in A_1\ \text{then}\ f(x)\ \text{else}\ x$$

ברור ש- $H : \mathrm{Eq}(A_1, A_1) \to (A_2 \to A_2)$ קל לברר, שלכל $f \in \mathrm{Eq}(A_1, A_1)$, $H(f)$ היא פונקציית שקילות מ- A_2 ל- A_2. לכן, בעצם, $H : \mathrm{Eq}(A_1, A_1) \to \mathrm{Eq}(A_2, A_2)$. כמו כן, טריביאלי ש- H ח.ח.ע., כיוון שלכל f ב- $\mathrm{Eq}(A_1, A_1)$ מתקיים ש- $H(f) / A_1 = f$ (ולכן $H(f_1) = H(f_2)$ $\Leftarrow$ $H(f_1) / A_1 = H(f_2)/A_1$ $\Leftarrow$ $f_1 = f_2$). לכן $E(a_1) = |\mathrm{Eq}(A_1, A_1)| \le |\mathrm{Eq}(A_2, A_2)| = E(a_2)$.

לבסוף, ש- $P(a_1,b) \le P(a_2,b)$ נובע מ- (ג) ושני האי-שוויונות על C ו- E, שכבר הוכחנו.

תרגילים:

לחשב:

(1) $C(\aleph_0, n)$ ו- $P(\aleph_0, n)$ עבור $n \in \mathbf{N}$

(2) $C(\aleph_0, \aleph_0)$ ו- $P(\aleph_0, \aleph_0)$

*(3) $E(\aleph_0)$ ו- $E(\aleph)$

(4) $C(\aleph, \aleph)$ ו- $P(\aleph, \aleph)$

(5) $C(\aleph, \aleph_0)$ ו- $P(\aleph, \aleph_0)$

פתרון לחלק מתרגילים אלו מובא בנספח לפרק זה.

*נספח: משפט על $C(a,a)$ כש- a אינסופית

משפט:

אם $a+a=a$, אז $C(a,a)=2^a$.

הוכחה:

כיוון ש- $a+a=a$, אז יש קבוצות B, C כך ש-

$$|B|=|C|=|B\cup C|=a \quad \text{ו-} \quad B\cap C=\varnothing$$

נגדיר $F=\lambda X\in P(C).\,X\cup B$.

כיוון שלכל $X\in P(C)$ מתקיים ש- $B\subseteq X\cup B\subseteq B\cup C$, אזי לכל X כזה, $a=|B|\le|X\cup B|\le|B\cup C|=a$. מכאן, שלכל $X\in P(C)$, $F(X)\in P(B\cup C)$ ו- $|F(X)|=a$, דהיינו: $F(X)\in P_a(B\cup C)$ לכן $F\in P(C)\to P_a(B\cup C)$. נראה ש- F ח.ח.ע.. נניח אפוא, ש- $X_1,X_2\in P(C)$ ו- $F(X_1)=F(X_2)$. נראה, ש- $X_1=X_2$. נראה, למשל, ש- $X_1\subseteq X_2$ (ההוכחה ש- $X_2\subseteq X_1$ דומה). יהי אפוא $y\in X_1$. אז:

$$\begin{aligned}
& y\in X_1\cup B\\
\Rightarrow\ & y\in F(X_1)\\
\Rightarrow\ & y\in F(X_2) && (\text{כי } F(X_1)=F(X_2))\\
\Rightarrow\ & y\in X_2\cup B\\
\Rightarrow\ & y\in X_2 \ \vee\ y\in B
\end{aligned}$$

אבל כיוון ש- $B\cap C=\varnothing$ ו- $y\in X_1\subseteq C$, לא ייתכן ש- $y\in B$ ומכאן ש- $y\in X_2$.

קיבלנו ש- F הינה פונקציה ח.ח.ע. מ- $P(C)$ אל $P_a(B\cup C)$, ולכן

$$C(a,a)=|P_a(B\cup C)|\ge|P(C)|=2^{|C|}=2^a$$

מצד שני, $C(a,a)\le 2^a$ לכל a לפי (5)(ו) של טבלה ג.7.

סך-הכל קיבלנו: $C(a,a)=2^a$.

מסקנות: $C(\aleph_0,\aleph_0)=2^{\aleph_0}=$א ו- $C($א, א$)=2^{\text{א}}$.

הערה:

כיוון שלמעשה $a+a=a$ *לכל* a אינסופית (הגם שההוכחה חורגת מקורס זה), המשפט קובע ש- $C(a,a)=2^a$ לכל a אינסופית.

ד. קומבינטוריקה סופית

ד.1 עקרונות בסיסיים

קומבינטוריקה סופית היא ענף של הקומבינטוריקה הכללית, שלמדנו בחלק הקודם. ענף זה מטפל בקבוצות סופיות ובעוצמות סופיות. בחלק זה נניח, אפוא (אלא אם נאמר במפורש אחרת), שהקבוצות A , B וכו', שאנו עוסקים בהן, הינן סופיות, ושהעוצמות שאנו עוסקים בהן הינן סופיות (אנו נשתמש לכן עבורן במשתנים כמו n , m , k וכו'). מובן שכל העקרונות הכלליים, שלמדנו בפרק הקודם לגבי קבוצות כלשהן, נכונים גם לגבי קבוצות סופיות (בפרט אלו של טבלה ג.7).

הייחוד של קבוצות סופיות הוא בנקודות הבאות:

(1) אם A ו- B סופיות ו- $|A| = |B|$, אז כל פונקציה ח.ח.ע. מ- A ל- B או פונקציה מ- A על B היא פונקצית שקילות. בפרט:

$$P(n, n) = E(n)$$

(נוסחה זו נכונה בעצם גם לעוצמות אינסופיות, אך ההוכחה חורגת מקורס זה.)

(2) אי-שוויונות חלשים אפשר להכליל לאי-שוויונות חזקים. למשל:

$$m < n \wedge k \leq \ell \Rightarrow m + k < n + \ell$$

(3) פעולת החיסור, $b - a$, מוגדרת בכל מקרה ש- $a \leq b$, בפרט מתקיים ש-

$$|B - A| = |B| - |A|$$

כאשר $A \subseteq B$.

(4) גם בחילוק אפשר להשתמש. מכל שוויון מהצורה $a = b \cdot c$ נובע ש-

$$b = \frac{a}{c}$$

(כאשר $c \neq 0$).

טבלה מס' ד.1 מסכמת את העקרונות הקומבינטוריים הבסיסיים, שהינם מיוחדים לקומבינטוריקה סופית. ביחד, טבלאות ג.7 ו- ד.1 מספקות את כל העקרונות הבסיסיים של הקומבינטוריקה הסופית. כדאי לשים לב, שכמעט כל העקרונות בטבלה ד.1 קשורים בחיסור ובחילוק!

טבלה ד.1:
עקרונות בסיסיים המיוחדים לקומבינטוריקה סופית

(1)	(i)	$\lvert A-B\rvert = \lvert A\rvert - \lvert A\cap B\rvert$
	(ii)	אם $B\subseteq A$ אז $\lvert A-B\rvert = \lvert A\rvert - \lvert B\rvert$
	(iii)	$\lvert A\cup B\rvert = \lvert A\rvert + \lvert B\rvert - \lvert A\cap B\rvert$
(2)		אם R יחס שקילות על A ו- $\lvert X\rvert = n$ לכל $X\in A/R$ אז $\lvert A/R\rvert = \dfrac{\lvert A\rvert}{n}$
(3)		$p(n,n) = E(n) = n!$
(4)		$P(n,k) = \begin{cases} 0 & n<k \\ \dfrac{n!}{(n-k)!} & n\ge k \end{cases}$
(5)		$C(n,k) = \begin{cases} 0 & n<k \\ \dfrac{n!}{k!(n-k)!} & n\ge k \end{cases}$
(6)		$n>0 \Rightarrow S(n,k) = C(n+k-1,\, k) = \dfrac{(n+k-1)!}{k!(n-1)!}$
(7)		נניח: $\lvert B\rvert = \lvert A\rvert = n$, $\{A_1,\ldots,A_\ell\}$ פירוק של A , $\lvert A_i\rvert = q_i$ לכל $1\le i\le \ell$, ו- $R = \{\langle f,g\rangle \in (\mathrm{Eq}(A,B))^2 \mid \forall x\in B\forall i\in\{1,\ldots,\ell\}.\, f(x)\in A_i \Leftrightarrow g(x)\in A_i\}$ אז $\left\lvert \mathrm{Eq}(A,B)/R \right\rvert = \dfrac{n!}{q_1!\,q_2!\ldots q_\ell!}$

לרוב העקרונות המופיעים בטבלאות ג.7 ו- ד.1 מקובל (ועוזר) לתת ניסוחים אינטואיטיביים במונחים של מספר האפשרויות "לעשות" משהו מסוים (בדרך-כלל "לבחור" משהו):

(א) העיקרון ש- $\left|\biguplus_{i=1}^{n} A_i\right| = \sum_{i=1}^{n}|A_i|$ פירושו, שאם אפשר לפצל בעיית בחירה מסוימת למספר n של אפשרויות, ובאפשרות מס' i יש k_i בחירות אפשרויות, אז מספר הבחירות המבוקש הוא $k_1 + k_2 + \ldots + k_n$ (זאת בתנאי, שאין חפיפה בין שום שתיים מהאפשרויות, אליהן פוצלה הבעיה המקורית).

(ב) בעיקרון ש- $\left|\biguplus_{i=1}^{n} A_i\right| = n \cdot k$, אם $|A_i| = k$ לכל $1 \le i \le n$ (זהו עיקרון (2)(ב) מטבלה ג.7 בגרסתו למקרה הסופי), משתמשים בדרך-כלל באופן הבא: מפצלים בעיית בחירה מסוימת לבעיה של ביצוע שתי בחירות עוקבות, כאשר מה שנבחר בבחירה השנייה תלוי במה שנבחר בבחירה הראשונה. אם יש n אפשרויות לבחירה הראשונה, ועל כל אחת מהן יש k אפשרויות איך לבצע את הבחירה השנייה, אז סך-הכל מספר האפשרויות לביצוע שתי הבחירות הוא $n \cdot k$.

(ג) העקרונות ב- (3) של טבלה ג.7 הם עקרונות "שובך היונים", וכבר עמדנו על כך בפרק הקודם.

(ד) עיקרון (1)(ii) של טבלה ד.1 קובע אינטואיטיבית, שאם יש לנו מספר אפשרויות n לעשות דבר מסוים, ו- k מתוכן אינן "טובות" מבחינה מסוימת, אז מספר ה"טובות" הוא $n - k$.

(ה) לפונקציית שקילות מקבוצה סופית על עצמה קוראים, בדרך-כלל, "תמורה" של אותה קבוצה. $E(n)$ היא, אפוא, מספר התמורות של קבוצה בת n איברים. $E(n)$ הוא גם מספר הסידורים המלאים האפשריים של קבוצה A בת n איברים ("סידור" כזה, הקובע מי ראשון, מי שני, מי שלישי וכו', אינו אלא פונקציה מ- $\{1, \ldots, n\}$ על A).

(ו) ל- n^k, $P(n, k)$ ו- $C(n, k)$, אותם הגדרנו בפרק ג.6 לכל שתי עוצמות, יש מספר אינטרפרטציות אינטואיטיביות, המסוכמות בטבלה מס' ד.2. כך השאלה "כמה טורים אפשר למלא בטוטו?" היא דוגמה ל"בחירה" של 15 עצמים מתוך הקבוצה

$\{1, 2, X\}$ עם חשיבות לסדר ועם חזרות (ולכן מספר הטורים הוא 3^{15}. למעשה, כל טור הינו פונקציה מקבוצת המשחקים בה מדובר באותו שבוע אל הקבוצה $\{1, 2, X\}$). דירוג למצעד פזמונים בו יש לשלוח, נאמר, דירוג של חמישה פזמונים נבחרים מתוך 25 פזמונים אפשריים, הוא דוגמה לבעיה בה התשובה היא $p(25, 5)$, בעוד שהבעיה של בחירת חמישייה ראשונה למשחק כדורסל מתוך סגל של שנים-עשר שחקנים (ניתן לקרוא לה "בעיית מאמן הכדורסל") היא דוגמה לבעיה, שבה התשובה היא $C(12, 5)$ (או $\binom{12}{5}$, כמו שמקובל יותר לסמן במקרה הסופי).

טבלה ד.2:
הגדרות אינטואיטיביות

n^k	(1)	מספר האפשרויות "לבחור" k עצמים מתוך n, *עם* חזרות ו*עם* חשיבות לסדר.
	(2)	מספר האפשרויות לפזר k כדורים *שונים* ב- n תאים.
$P(n, k)$	(1)	מספר האפשרויות "לבחור" k עצמים מתוך n, *בלי* חזרות ו*עם* חשיבות לסדר.
	(2)	מספר האפשרויות לפזר k כדורים *שונים* ב- n תאים, לכל היותר כדור אחד בתא.
$C(n, k)$	(1)	מספר האפשרויות "לבחור" k עצמים מתוך n, *בלי* חזרות ו*בלי* חשיבות לסדר.
	(2)	מספר האפשרויות לפזר k כדורים *זהים* ב- n תאים, לכל היותר כדור אחד בתא.
$S(n, k)$	(1)	מספר האפשרויות "לבחור" k עצמים מתוך n, *עם* חזרות ו*בלי* חשיבות לסדר.
	(2)	מספר האפשרויות לפזר k כדורים *זהים* ב- n תאים (בלי הגבלות).
	(3)	מספר הפתרונות במספרים טבעיים של המשוואה: $x_1 + x_2 + \ldots + x_n = k$
$E(n)$	(1)	מספר התמורות של קבוצה בת n איברים.
	(2)	מספר הסידורים המלאים של קבוצה בת n איברים.

(ז) בטבלה מס' ד.1 מופיעה פעולה נוספת, S, אותה טרם הגדרנו. אינטואיטיבית היא משלימה את המשבצת החסרה במיון הבעיות של בחירת k עצמים מתוך n, לפי הקריטריונים של עם/בלי חזרות ועם/בלי חשיבות לסדר:

	בלי חזרות	עם חזרות
הסדר חשוב	$P(n, k)$	n^k
הסדר לא חשוב	$C(n, k)$	$S(n, k)$

דוגמה לבעיה כזו היא הבעיה הבאה: בכמה אופנים ניתן לחלק שמונה-עשרה משרות של סגן ראש-ממשלה בין ארבע מפלגות א', ב', ג', ד' (כולל האפשרות שמפלגה אחת תקבל את כל המשרות)? כאן צריך "לבחור" שמונה-עשר עצמים מתוך הקבוצה {"ד", "ג", "ב", "א"}, והתשובה לכן $S(4, 18)$.

הבה נגדיר את $S(n, k)$ באופן פורמלי:

הגדרה:

$$S(n,k) = \left|\left\{ f \in A \to \mathbf{N} \,\middle|\, \sum_{x \in A} f(x) = k \right\}\right|$$

כאשר A היא קבוצה שעוצמתה n.

הסבר:

הגדרנו כאן באופן מדויק מה פירוש "לבחור" k עצמים מתוך הקבוצה A עם אפשרות לחזרה ובלי חשיבות לסדר. לכל $x \in A$, $f(x)$ היא מספר הפעמים שהעצם x "נבחר", והדרישה $\sum_{x \in A} f(x) = k$ פירושה, שאכן סך-הכל "נבחרו" k איברים מ- A. $S(n, k)$ הוא מספר הפונקציות האלה.

הערות:

(1) לפונקציות $f: A \to \mathbf{N}$ קוראים *מולטי תת-קבוצות* של A (השם "bag" מקובל אף הוא במדעי המחשב). פונקציות כאלה הן הכללה של הפונקציות האופייניות χ_B^A של תת-קבוצות של A (השייכות, כזכור, ל- $A \to \{0, 1\}$).

(2) אם ניקח $A = \{1, \ldots, n\}$ ונסמן את $f(i)$ ב- x_i, נקבל ש- $S(n, k)$ הוא מספר הפתרונות ב*מספרים טבעיים* של המשוואה

$$x_1 + x_2 + \ldots + x_n = k$$

הבחנה זו יכולה לשמש לצורך הגדרה אלטרנטיבית של $S(n, k)$.

(3) כמו שקורה תמיד בעקבות (או לפני) הגדרה כמו זו של $S(n, k)$, עלינו להוכיח כאן:

משפט:

הפעולה S על עוצמות סופיות מוגדרת היטב. כלומר, אם $|A| = |A'|$ אז:

$$\left|\left\{ f \in A \to \mathbf{N} \,\middle|\, \sum_{x \in A} f(x) = k \right\}\right| = \left|\left\{ f \in A' \to \mathbf{N} \,\middle|\, \sum_{x \in A} f(x) = k \right\}\right|$$

את ההוכחה נשאיר לקורא.
נעבור להוכחת העקרונות של טבלה ד.1.

(1) (i) לכל A, B: $\quad A = (A - B) \uplus (A \cap B)$

$$\Rightarrow \quad |A| = |A - B| + |A \cap B|$$
$$\Rightarrow \quad |A - B| = |A| - |A \cap B|$$

(המעבר מהשורה השנייה לשלישית אפשרי רק בקבוצות סופיות!).

(ii) אם $B \subseteq A$ אז $A \cap B = B$, ולכן זהו מקרה פרטי של (i).

(iii) לכל A, B: $\quad A \cup B = (A - B) \uplus B$

$$\Rightarrow \quad |A \cup B| = |A - B| + |B|$$

לפי (i) $= (|A| - |A \cap B|) + |B|$

$$= |A| + |B| - |A \cap B|$$

(2) כזכור, אם R יחס שקילות על A, אז A/R היא *פירוק* של A, כלומר: אם $A/R = \{X_1, \ldots, X_\ell\}$, אז $A = \biguplus_{i=1}^{\ell} X_i$. עתה, אם ידוע ש- $|X_i| = n$ לכל $1 \le i \le \ell$,

אז לפי עיקרון הכפל ((2)(ב) של טבלה ג.7) $|A| = n \cdot \ell = n \cdot |A/R|$, לכן $|A/R| = \frac{|A|}{n}$.

הערה:

שוב, גם כאן כל הצעדים, *פרט לאחרון* (החילוק), נכונים גם במקרה ש- n ו- ℓ עוצמות אינסופיות.

(3) נוכיח ש- $E(n) = n!$ באינדוקציה על n. נזכיר שפעולת העצרת מוגדרת באופן הבא: $0! = 1$ ו- $n! = 1 \cdot 2 \cdot \ldots \cdot n$, כש- $n \geq 1$. התכונה היסודית של פעולת העצרת היא ש-

$$n! = (n-1)! \cdot n \quad \text{ל-} \quad n \geq 1 \quad (\text{או} \quad (n+1)! = n! \cdot (n+1))$$

שלב הבסיס: יש בדיוק פונקציית שקילות אחת מ- $\varnothing$ ל- $\varnothing$: $\varnothing$. לכן:

$$E(0) = |\mathrm{Eq}(\varnothing,\varnothing)| = |\{\varnothing\}| = 1 = 0!$$

שלב המעבר: נניח $E(n) = n!$. נוכיח $E(n+1) = (n+1)!$.

ההנחה ש- $E(n) = n!$ פירושה, שלכל שתי קבוצות, A ו- B, כך ש- $|A| = |B| = n$ מתקיים ש- $|\mathrm{Eq}(A,B)| = n!$.

עתה, לפי הגדרה:

$$E(n+1) = |\mathrm{Eq}(\{0, 1, \ldots, n\}, \{0, 1, \ldots, n\})|$$

נגדיר עתה ל- $0 \leq i \leq n$:

$$E_i = \{f \in \mathrm{Eq}(\{0, 1, \ldots, n\}, \{0, 1, \ldots, n\}) \mid f(n) = i\}$$

קל לראות שלכל $0 \leq i \leq n$ מתקיים:

$$E_i \sim \mathrm{Eq}(\{0, 1, \ldots, i-1, \; i+1, \ldots, n\}, \{0, 1, \ldots, n-1\}) \quad (\text{למה?})$$

(כלומר: $E_i \sim \mathrm{Eq}(\{0, 1, \ldots, n\} - \{i\}, \{0, 1, \ldots, n-1\})$).

לכן, לפי הנחת האינדוקציה, $|E_i| = n!$.

מזה ש- $|E_i| = n!$ לכל $0 \le i \le n$ וש- $A = \biguplus_{i=0}^{n} E_i$, נובע ש-

$$|A| = (n+1) \cdot n! = (n+1)!$$

הערה:
מה שכתבנו אינו אלא כתיבה בצורה מדויקת של השיקול הלא פורמלי הבא: כדי לבנות פונקציית שקילות f מ- $\{0, 1, \ldots, n\}$ על $\{0, 1, \ldots, n\}$, עלינו לבחור תחילה את הערך של $f(n)$, וזאת נוכל לעשות ב- $(n+1)$ אפשרויות. אחר-כך עלינו ליצור פונקציית שקילות בין $\{0, 1, \ldots, n-1\}$ לבין קבוצת הערכים שטרם נבחרו (שמספר איבריה n). זאת נוכל לעשות (לפי הנחת האינדוקציה) ב- $n!$ אפשרויות. סך-הכל יש לנו $(n+1) \cdot n!$ פונקציות שקילות מ- $\{0, 1, \ldots, n\}$ על עצמה.

בבית-ספר תיכון לא מנסחים זאת, בדרך-כלל, אפילו כך, וכביכול לא משתמשים כלל באינדוקציה. אומרים שם משהו כזה: "כדי לסדר קבוצה A בת n איברים, נבחר תחילה מי יהיה במקום הראשון: לכך יש n אפשרויות. אחר-כך מי יהיה במקום השני – לכך נותרו $n-1$ אפשרויות. אחר-כך מי יהיה במקום השלישי – $n-2$ אפשרויות, וכן הלאה, עד שנגיע למקום ה- n-י, שעבורו תישאר אפשרות אחת בלבד. סך-הכל נקבל $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 1$ אפשרויות, כלומר: $n!$.
יש להבין, אבל, שהן השימוש בצירוף המלים "וכן הלאה" והן בשלוש הנקודות "...", משמעותם האמיתית היא אחת: הוכחה ***באינדוקציה*** (וכמובן "סידור" הוא פונקציה מ- $\{1, \ldots, n\}$ ל- A, "המקום הראשון" פירושו $f(1)$, וכדומה). תרגום השיקול האינטואיטיבי מבית-הספר התיכון להוכחה במונחים מדויקים נותן בדיוק את מה שעשינו כאן (או דבר דומה מאוד).

(4) את הנוסחה עבור $P(n, k)$ אפשר להוכיח בצורה דומה לזו שבה הוכחנו את הנוסחה עבור E: באינדוקציה על n (למעשה, זה מה שעושים, אם כי לא אומרים זאת, בבית-ספר תיכון). יש, אבל, לשים לב, שהאינדוקציה כאן הינה מסובכת יותר מאשר במקרה של E: מה שאנו מוכיחים באינדוקציה על n הוא, שהנוסחה עבור $P(n, k)$ נכונה עבור *כל* k.
בשלב הבסיס אנו מוכיחים ש:

$$P(0, 0) = 1 \ \wedge\ \forall k \in \mathbf{N}^+ .\ P(0, k) = 0$$

בשלב המעבר אנו מוכיחים ש:

אם $\forall k \in \mathbf{N}.\ (k > n \Rightarrow P(n, k) = 0) \ \wedge\ (k \le n \Rightarrow P(n, k) = n!/(n-k)!)$

אז

$$\forall k \in \mathbf{N}.\ (k > n+1 \ \Rightarrow\ P(n+1, k) = 0)\ \wedge$$
$$(k \le n+1 \ \Rightarrow\ P(n+1, k) = \frac{(n+1)!}{(n+1-k)!})$$

כדי לגוון וכדי להדגים עקרונות אחרים, נוכיח כאן את (4) בדרך אחרת. נסתמך על כך ש- $P(n, k) = |\mathrm{In}(\{1, \ldots, n\}\ , \{1, \ldots, k\})|$.
ברור לכן, שאם $n < k$, אז קבוצה זו היא ריקה ולכן $P(n, k) = 0$.
נניח אפוא, ש- $k \le n$. נסמן $A = \{1, \ldots, k\}$, $B = \{1, \ldots, n\}$. אז $A \subseteq B$, ומתקיים:

$$(*)\qquad \mathrm{Eq}(B, B) = \biguplus_{f \in \mathrm{In}(B, A)} \{F \in \mathrm{Eq}(B, B) \mid F/A = f\}$$

(זאת, כיוון שאם $F \in \mathrm{Eq}(B,B)$ אז $F/A \in \mathrm{In}(B,A)$, וברור שהאיחוד זר, כי אם $F_1/A = f_1$, $F_2/A = f_2$ ו- $f_1 \neq f_2$, אז גם $F_1 \neq F_2$).

עתה, כמה איברים יש, בהינתן $f \in \mathrm{In}(B, A)$, בקבוצה $\{F \in \mathrm{Eq}(B,B) \mid F/A = f\}$?
הערכים של כל פונקציה F בקבוצה זו עבור $1, 2, \ldots, k$ ניתנים על-ידי f.
אינטואיטיבית, מה שנשאר עוד לעשות כדי להגדיר F כזאת, הוא להחליט על ערכיה עבור $k+1, k+2, \ldots, n$, ואת אלה צריכים לבחור מתוך $B - \{f(1), f(2), \ldots, f(k)\}$ (כלומר, מתוך $(B - f(A)$). ברור לכן, שלכל $f \in \mathrm{In}(B, A)$ מתקיים ש:

$$\{F \in \mathrm{Eq}(B,B) \mid F/A = f\} \sim \mathrm{Eq}(B - f(A), B - A)$$

(פורמלית, פונקציית השקילות כאן היא:

$$\lambda G \in \{F \in \mathrm{Eq}(B,B) \mid F/A = f\}.\ G/(B - A)$$

ויש להשלים את הפרטים הנחוצים כדי להראות, שאכן זו פונקציית שקילות בין שתי הקבוצות הנ"ל).

עתה, ב- $B - A$ ו- $B - f(A)$ יש $n - k$ איברים. לכן:

$$|\mathrm{Eq}(B - f(A), B - A)| = E(n - k) = (n - k)!$$

ומכאן שלכל $f \in \mathrm{In}(B, A)$, $\quad |\{F \in \mathrm{Eq}(B,B) \mid F/A = f\}| = (n - k)!$

מ- (*) למעלה נובע אפוא:

$$|E(B,B)| = |\mathrm{In}(B, A)| \cdot (n-k)!$$

כלומר:

$$n! = P(n, k) \cdot (n-k)!$$

ולכן

$$P(n,k) = \frac{n!}{(n-k)!}$$

(5) בפרק הקודם (טבלה ג.7 (5)(ג)) הוכחנו ש:

$$P(n, k) = C(n, k) \cdot E(k)$$

כיוון ש- $E(k) \geq 1$ לכל k, נובע מזה:

$$C(n,k) = \frac{P(n,k)}{E(k)}$$

ולכן (5) היא מסקנה מיידית מ- (3) ומ- (4).

לכל העקרונות עד כה ניתנו הוכחות מדויקות מאוד, כולם במונחים שלמדנו בתורת הקבוצות. מאחורי כל הוכחה כזו עומד, למעשה, רעיון אינטואיטיבי פשוט. לאחר שמתרגלים ומבינים כיצד מתרגמים רעיונות אינטואיטיביים כאלה להוכחות מלאות ומדויקות, מקובל, כדי למנוע סרבול יתר, להסתפק בהצגת ההסברים האינטואיטיביים (בלי לתרגמם להוכחות מן הסוג שהבאנו עד עתה). גם אנו נתחיל אפוא לנסח לעתים קרובות שיקולים קומבינטוריים באופן פחות פורמלי ונוקשה מכפי שעשינו עד עתה. אולם, יש לזכור תמיד, שהוכחות מלאות ומדויקות הן רק כאלה, המסתמכות במפורש על מושגי תורת הקבוצות ועל העקרונות המדויקים שלמדנו.

---

(6) במקרים רבים מוכיחים זהויות קומבינטוריות על-ידי שפותרים בעיה מסוימת בשתי דרכים שונות ומקבלים כך נוסחאות שונות לפתרונה. העובדה שמדובר ב*אותה* בעיה, גוררת את זהות שתי הנוסחאות. נדגים צורת הוכחה זו במקרה של הנוסחה עבור $S(n, k)$. הבעיה שנפתור בשתי דרכים תהיה: "בכמה אופנים ניתן להגיע במישור מהנקודה $(1,0)$ לנקודה (n, k) (כאשר $n \geq 1$, $k \geq 0$) במסלול שמורכב מצעדים באורך 1 כלפי מעלה או בכיוון ימינה?".

פתרון ראשון:

נייצג כל מסלול על-ידי n-יה $\langle x_1, \ldots, x_n\rangle \in \mathbf{N}^n$, שמשמעותה: המסלול מתחיל ב- x_1 צעדים כלפי מעלה $(0 \le x_1 \le k)$, אחר-כך צעד ימינה, אחר- כך x_2 צעדים כלפי מעלה $(0 \le x_2 \le k)$, אחר-כך צעד ימינה, וכן הלאה. סך-הכל יש לנו $n-1$ צעדים ימינה ו- $\sum_{i=1}^{n} x_i$ צעדים כלפי מעלה. ה- n-יה מתארת אפוא מסלול מותר אם"ם $\sum_{i=1}^{n} x_i = k$. ברור אפוא, שמספר המסלולים המותרים שווה למספר ה-n-יות מסוג זה, ומספר זה הינו $S(n, k)$, לפי אחת ההגדרות של $S(n, k)$.

פתרון שני:

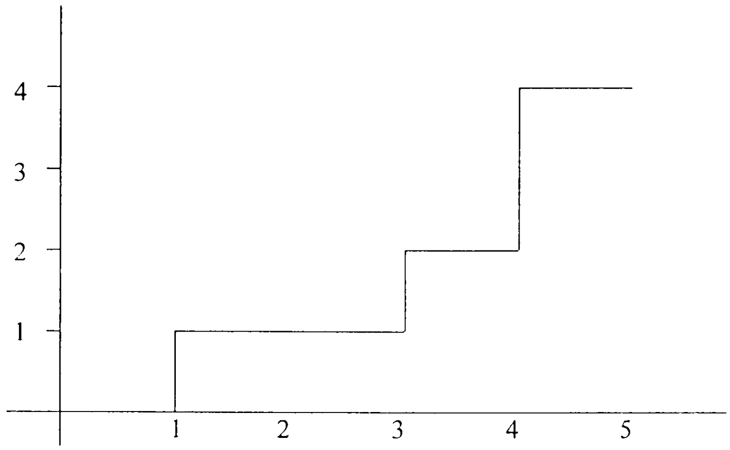
נייצג כל מסלול על-ידי סדרה באורך $k + n - 1$ של האותיות "U" ו- "R", שבה "U" מופיע k פעמים ו- "R" מופיע $n - 1$ פעמים. הופעת "U" פירושה: "עשה צעד אחד כלפי מעלה" (Up), והופעת "R" פירושה: "עשה צעד אחד ימינה" (Right). סך-הכל אנו עושים $n - 1$ צעדים ימינה (מ- 1 ועד n) ו- k צעדים כלפי מעלה (מ- 0 עד k).

עתה, כדי ליצור סדרה כזו עלינו לבחור את k המקומות בהם נכתוב "U" (בשאר ייכתב "R"). זאת נוכל לעשות ב- $C(k + n - 1, k)$ דרכים. לכן, מספר המסלולים הוא $C(n + k - 1, k)$.

מהעובדה, ששני הפתרונות נכונים, אנו מקבלים: $S(n, k) = C(n + k - 1, k)$.

ציור 3 מדגים שני תיאורים קונקרטיים של אותו מסלול במקרה בו $k = 4$, $n = 5$.

ציור 3: דוגמה להוכחת הנוסחה $S(n, k)$



תיאור מסלול, שיטה א': $\langle 1, 0, 1, 2, 0\rangle$

תיאור מסלול, שיטה ב': $\langle U, R, R, U, R, U, U, R\rangle$

הנה דוגמה נוספת לשיטת ההוכחה של זהויות קומבינטוריות, שהפעלנו בהוכחת $S(n, k)$:

בעיה:

מקבוצה של n סטודנטים יש לבחור משלחת לראש החוג, שתכלול שני יושבי-ראש. כמה אפשרויות יש להרכיב את המשלחת?

פתרון א':

את יושבי הראש אפשר לבחור ב- $\binom{n}{2}$ דרכים.* על כל אחת מהן אפשר לבחור את שאר חברי המשלחת ב- 2^{n-2} דרכים (כי כל קבוצה חלקית של $n-2$ הסטודנטים הנותרים אפשרית). סך-הכל יש לכן $\binom{n}{2}\cdot 2^{n-2}$.

פתרון ב':

נסמן ב- A_k את מספר המשלחות שיש בהן בדיוק k סטודנטים $(2 \le k \le n)$. כדי להרכיב משלחת כזו, עלינו לבחור תחילה את k חבריה. זאת נוכל לעשות ב- $\binom{n}{k}$ דרכים. אחר-כך עלינו לבחור את שני יושבי-הראש שלה, ונעשה זאת ב- $\binom{k}{2}$ דרכים. מכאן ש- $|A_k| = \binom{n}{k}\cdot\binom{k}{2}$. קבוצת המשלחות האפשריות היא $\biguplus_{k=2}^{n} A_k$,

לכן מספרן הוא $\sum_{k=2}^{n}\binom{n}{k}\cdot\binom{k}{2}$.

כיוון ששני הפתרונות נכונים, אנו מקבלים את הזהות:

$$\sum_{k=2}^{n}\binom{n}{k}\cdot\binom{k}{2} = \binom{n}{2}\cdot 2^{n-2}$$

או:

$$\sum_{k=2}^{n} k\cdot(k-1)\cdot\binom{n}{k} = n(n-1)\cdot 2^{n-2}$$

---

* מעתה נשתמש, בדרך-כלל, בסימן $\binom{n}{k}$ במקום $C(n, k)$. סימון זה מקובל יותר כשמדובר בקבוצות סופיות.

נעבור לבסוף לעיקרון (7) – העיקרון האחרון בטבלה ד.1. כיוון שניסוח העיקרון מסובך במידת-מה, נקדים להוכחתו דוגמה, שתבהיר למה הכוונה.

בעיה:

בכמה אופנים ניתן לסדר קבוצה של 5 כדורים אדומים, 4 שחורים, 3 ירוקים ו- 4 צהובים?

תשובה:

בקבוצה שבה עסקנו יש 16 כדורים. עקרונית, התשובה היא פשוט !16. ברם, "לא לכך התכוון המשורר". השואל כאן אינו רוצה להבדיל, למשל, בין שני סידורים, שכל ההבדל ביניהם הוא, שבאחד כדור צהוב א' הוא הראשון, וכדור צהוב ב' הוא השני, בעוד בשני המצב הפוך. אומנם, שני הסידורים אינם זהים באמת (על כדור צהוב א' יש אולי שריטה מיקרוסקופית, בעוד לכדור צהוב ב' אין שריטה כזו), אבל אנו איננו רוצים, בדרך-כלל, להבדיל ביניהם. לצרכינו (ברוב המקרים) הם *שקולים*. כללית, בבעיה מסוג זה אנו רוצים לזהות (אלא אם כן נאמר אחרת) כל שני סידורים, שאי-אפשר להבחין ביניהם לפי צבעי הכדורים.

המובן המתמטי של אמירה מהסוג של "אנו מזהים שני עצמים של קבוצה E בעלי תכונה מסוימת" בהקשר של בעיית ספירה מסוימת הוא, שאנו מגדירים יחס שקילות מסוים על E, ומה שאנו רוצים לדעת הוא את *מספר מחלקות השקילות* של E לפי יחס זה, לא את מספר איברי E (שני איברים באותה מחלקת שקילות הם "אותו דבר" לגבינו). בדוגמה שלפנינו, הקבוצה E היא אוסף הסידורים של 16 הכדורים (כלומר: קבוצת פונקציות השקילות בין $B=\{1,\ldots,16\}$ ובין A-קבוצת הכדורים). שני הסידורים הם שקולים לגבינו אם בכל מקום סידורי הם שמים כדורים בעלי צבע זהה. זהו בדיוק מקרה מהסוג, שעיקרון (7) של ד.1 מדבר עליו: A כאן היא קבוצת הכדורים. A_1 – היא קבוצת הכדורים האדומים, A_2 – קבוצת הכדורים השחורים, A_3 – קבוצת הכדורים הירוקים ו- A_4 – קבוצת הכדורים הצהובים. מתקיים אכן, ש- $A=\biguplus_{i=1}^{4} A_i$.

כמו כן, בבעיה זו: $n=16$, $\ell=4$, $q_1=5$, $q_2=4$, $q_3=3$ ו- $q_4=4$. לפי עיקרון (7), התשובה לשאלה היא אפוא $\frac{16!}{5!4!3!4!}$.

הוכחת (7):

כיוון ש- $|A| = |B| = n$, הרי ב- $\mathrm{Eq}(A,B)$ יש $E(n) = n!$. קל לברר, שאכן היחס R הוא יחס שקילות על $\mathrm{Eq}(A,B)$. נראה, שבכל מחלקת שקילות של R יש $q_1!\cdot q_2!\cdot\ldots\cdot q_\ell!$ איברים. מזה ומעיקרון (2) ינבע (7) באופן מיידי.

אינטואיטיבית, הסיבה לכך שבכל מחלקת שקילות יש $q_1!\cdot q_2!\cdot\ldots\cdot q_\ell!$ איברים ברורה: נניח $B = \{1, \ldots, n\}$, X מחלקת שקילות ו- $X = [f]_R$. f היא במקרה זה סידור של איברי A , ופונקציה g היא סידור שקול (כלומר $g \in [f]_R$), אם"ם היא מתקבלת מהסידור f על-ידי שעושים פרמוטציה (תמורה) כלשהי בין איברי A_1, פרמוטציה של איברי A_2 , וכן הלאה. מספר הפרמוטציות של איברי סוג i הוא $q_i!$, וסך-הכל נוכל לקבל כך $q_1!\cdot q_2!\cdot\ldots\cdot q_\ell!$ סידורים שקולים ל- f. מה שנכתוב עתה הוא פשוט ניסוח מדויק של שיקול אינטואיטיבי זה.

הוכחת (7) לפי העקרונות הבסיסיים של קומבינטוריקה כללית:

כיוון ש- $|A| = |B| = n$, הרי: $|\mathrm{Eq}(A,B)| = E(n) = n!$.

יהי $X \in \mathrm{Eq}(A,B)/R$. אז $X = [f]_R$ לאיזה $f \in \mathrm{Eq}(A,B)$.

נגדיר $F:\mathrm{Eq}(A_1,A_1)\times\ldots\times\mathrm{Eq}(A_\ell,A_\ell)\to X$ על-ידי:

$$F(g_1,\ldots,g_\ell) = \lambda y \in B.\begin{cases} g_1(f(y)) & f(y)\in A_1 \\ g_2(f(y)) & f(y)\in A_2 \\ \vdots & \vdots \\ g_\ell(f(y)) & f(y)\in A_\ell \end{cases}$$

F מוגדרת היטב, כיוון ש- $\{A_1,\ldots,A_\ell\}$ פירוק של A .

קל לברר שלכל $\langle g_1, \ldots, g_n\rangle$ בתחום, $F(g_1,\ldots,g_\ell)$ אכן פונקצית שקילות מ- B על A (כלומר $F(g_1,\ldots,g_\ell)\in\mathrm{Eq}(A,B)$). מהגדרת R נובע גם מיידית, ש- $F(g_1,\ldots,g_\ell)Rf$, דהיינו $F(g_1,\ldots,g_\ell)\in[f]_R = X$.

לבסוף, F עצמה היא פונקצית שקילות מ- $\mathrm{Eq}(A_1,A_1)\times\ldots\times\mathrm{Eq}(A_\ell,A_\ell)$ על X, כי לא קשה להראות ש- $\lambda g \in X.\left\langle g\circ f^{-1}/A_1,\ g\circ f^{-1}/A_2,\ldots,\ g\circ f^{-1}/A_\ell\right\rangle$ היא פונקציה הפוכה (הפרטים – תרגיל).

מכל זה נובע, שלכל $X \in \mathrm{Eq}(A,B)/R$ מתקיים:

$$|X| = |\mathrm{Eq}(A_1,A_1)\times\ldots\times\mathrm{Eq}(A_\ell,A_\ell)|$$

$$\Rightarrow \quad |X| = q_1!\cdot q_2!\cdot\ldots\cdot q_\ell!$$

מכאן

$$|\mathrm{Eq}(A,B)| = |\mathrm{Eq}(A,B)/R|\cdot q_1!\cdot q_2!\cdot\ldots\cdot q_\ell!$$

ולכן

$$|\mathrm{Eq}(A,B)/R| = \frac{n!}{q_1!\cdot q_2!\cdot\ldots\cdot q_\ell!}$$

הנה דוגמה נוספת לשימוש בעיקרון (7):
נניח שיש לחלק קבוצה B של 20 אנשים ל- 4 קבוצות, שלכל אחת מהן משימה אחרת. בקבוצה הראשונה צריכים להיות 7 אנשים, בשנייה 4, בשלישית 5 וברביעית 4. מהו מספר החלוקות האפשרי?

תשובה:
נבצע את החלוקה כך: נסדר את עשרים האנשים בשורה. שבעת הראשונים יהיו בקבוצה א', ארבעת הבאים יהיו בקבוצה ב', וכן הלאה. לרוע המזל, ברור ששני סידורים שונים יכולים לתת אותה חלוקה. ניתן אבל לראות כל סידור כפונקציה מ- B אל $A = \{1, 2, \ldots, 20\}$. אם ניקח $A_1 = \{1, 2, \ldots, 7\}$, $A_2 = \{8, 9, 10, 11\}$, $A_3 = \{12, \ldots, 16\}$ ו- $A_4 = \{17, \ldots, 20\}$, אז שני ה"סידורים", f ו- g , הם שקולים, אם $\forall x \in B\ \forall 1 \le i \le 4\quad f(x) \in A_i \Leftrightarrow g(x) \in A_i$. לכן עיקרון (7) ישים ומספר האפשרויות הוא $\frac{20!}{7!4!5!4!}$.

כאשר אין חשיבות לסדר בין חלקי קבוצה, אלא רק צריך לבצע חלוקה, אז יש לעשות זיהויים נוספים בתוך קבוצת החלוקות בהן *יש* חשיבות לסדר (כלומר, יש להגדיר יחס שקילות מתאים עליה ולהשתמש בעיקרון (2) בטבלה ד.1). העיקרון הכללי הוא, שאם $|A| = n$, ורוצים לחלק את A ל- k_1 קבוצות עם n_1 איברים, k_2 קבוצות עם n_2 איברים, ..., k_ℓ קבוצות עם n_ℓ איברים (כש- $n = \sum_{i=1}^{\ell} k_i n_i$), אז מספר האפשרויות הוא:

$$\frac{n!}{k_1!\cdot(n_1!)^{k_1}\cdot k_2!\cdot(n_2!)^{k_2}\cdot\ldots\cdot k_\ell!\cdot(n_\ell!)^{k_\ell}}$$

ד.2 עיקרון ההכלה וההפרדה

מציאת מספר האיברים ב- $\bigcup_{i=1}^{n} A_i$ כשהאיחוד אינו זר, וב- $\bigcap_{i=1}^{n} A_i$, אינה דבר פשוט.

בפרק הקודם ראינו נוסחה, שאפשר להיעזר בה עבור המקרה $n = 2$:

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|$$

בפרק זה נכליל נוסחה זו ל- n טבעי כלשהו. ההכללה נקראת עיקרון ההכלה וההפרדה (Inclusion-Exclusion Principle, ובקיצור In-Ex).

עיקרון ההכלה וההפרדה, נוסח א':

נניח $A_1, \ldots, A_n$ קבוצות סופיות. נסמן $E_n = P(\{1, 2, \ldots, n\}) - \{\varnothing\}$. אז:

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{X \in E_n} (-1)^{|X|-1} \cdot \left|\bigcap_{i \in X} A_i\right|$$

למרות שהניסוח הזה הוא צלול ובהיר כבדולח, הבה נדגים בכל זאת (רק ליתר ביטחון...) את מה שכתוב פה עבור המקרים $n = 2$ ו- $n = 3$:

(א) כש- $n = 2$ אז:

$$E_2 = P(\{1,2\}) - \{\varnothing\} = \{\varnothing, \{1\}, \{2\}, \{1,2\}\} - \{\varnothing\}$$
$$= \{\{1\}, \{2\}, \{1,2\}\}$$

לכן:

$$\sum_{X \in E_2} (-1)^{|X|-1} \cdot \left|\bigcap_{i \in X} A_i\right| = (-1)^{|\{1\}|-1} \cdot \left|\bigcap_{i \in \{1\}} A_i\right| + (-1)^{|\{2\}|-1} \cdot \left|\bigcap_{i \in \{2\}} A_i\right| + (-1)^{|\{1,2\}|-1} \cdot \left|\bigcap_{i \in \{1,2\}} A_i\right|$$

עתה:

$$|\{1\}| = 1 \quad , \quad |\{2\}| = 1 \quad , \quad |\{1,2\}| = 2$$

כמו-כן:

$$\bigcap_{i\in\{1\}} A_i = A_1 \qquad \bigcap_{i\in\{2\}} A_i = A_2 \qquad \bigcap_{i\in\{1,2\}} A_i = A_1 \cap A_2$$

סך-הכל מקבלים:

$$\sum_{X\in E_2} (-1)^{|X|-1} \cdot \left|\bigcap_{i\in X} A_i\right| = (-1)^{1-1} \cdot |A_1| + (-1)^{1-1} \cdot |A_2| + (-1)^{2-1} \cdot |A_1 \cap A_2| =$$

$$= |A_1| + |A_2| - |A_1 \cap A_2|$$

וזה אכן $|A_1 \cup A_2|$, לפי הזהות הבסיסית בה פתחנו את הפרק.

(ב) כש- $n = 3$ נקבל:

$$E_3 = P(\{1,2,3\}) - \{\varnothing\} = \{\{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$$

בסכום באגף ימין של עיקרון ההכלה וההפרדה יהיו כאן לכן $2^3 - 1 = 7$ מחוברים.
עתה כאן:

$$|\{1\}| = 1 \qquad \bigcap_{i\in\{1\}} A_i = A_1 \qquad |\{1,2\}| = 2 \qquad \bigcap_{i\in\{1,2\}} A_i = A_1 \cap A_2$$

$$|\{2\}| = 1 \qquad \bigcap_{i\in\{2\}} A_i = A_2 \qquad |\{1,3\}| = 2 \qquad \bigcap_{i\in\{1,3\}} A_i = A_1 \cap A_3$$

$$|\{3\}| = 1 \qquad \bigcap_{i\in\{3\}} A_i = A_3 \qquad |\{2,3\}| = 2 \qquad \bigcap_{i\in\{2,3\}} A_i = A_2 \cap A_3$$

$$|\{1,2,3\}| = 3 \qquad \bigcap_{i\in\{1,2,3\}} A_i = A_1 \cap A_2 \cap A_3$$

לכן במקרה זה:

$$\sum_{X\in E_3} (-1)^{|X|-1} \cdot \left|\bigcap_{i\in X} A_i\right| =$$

$$= |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$

הבה נראה, שזה אכן $|A_1 \cup A_2 \cup A_3|$. עבור זה נעזר בנוסחה עבור $n = 2$:

$$|A_1 \cup A_2 \cup A_3| = |(A_1 \cup A_2) \cup A_3|$$

$$= |A_1 \cup A_2| + |A_3| - |(A_1 \cup A_2) \cap A_3|$$

המקרה $n = 2$

$$= |A_1 \cup A_2| + |A_3| - |(A_1 \cap A_3) \cup (A_2 \cap A_3)|$$

חוק הפילוג

$$= (|A_1| + |A_2| - |A_1 \cap A_2|) + |A_3| - (|A_1 \cap A_3| + |A_2 \cap A_3| - |(A_1 \cap A_3) \cap (A_2 \cap A_3)|)$$

$n = 2$ פעמיים

$$= |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_1 \cap A_3| - |A_2 \cap A_3| + |A_1 \cap A_2 \cap A_3|$$

כדאי לשים לב, שהשתמשנו פה בנוסחה עבור $n = 2$ מספר פעמים, ועבור קומבינציות שונות, לא רק עבור A_1 ו- A_2 (גם עבור $A_1 \cap A_3$ ו- $A_2 \cap A_3$, למשל).

הוכחת עיקרון ההכלה וההפרדה נעשית באינדוקציה. הבסיס $n = 2$ הוכח בפרק הקודם. שלב המעבר מ- n ל- $n + 1$ דומה למה שעשינו כאן במעבר מ- $n = 2$ ל- $n = 3$, וגם בו צריך להשתמש בהנחת האינדוקציה בשני מקומות שונים, כמו גם במקרה הפרטי הידוע $n = 2$. ההוכחה מובאת במלואה בנספח לפרק זה.

הנוסח הראשון של עיקרון ההכלה וההפרדה מסייע (במקרים מסוימים) למצוא את מספר האיברים באיחוד של n קבוצות. נביא עתה נוסח שני, שימושי דווקא יותר, המסייע (שוב, במקרים מסוימים) למצוא את מספר האיברים ב*חיתוך* של n קבוצות.

עיקרון ההכלה וההפרדה, נוסח ב':

נניח U קבוצה, ותהיינה $A_1, \ldots, A_n$ קבוצות חלקיות של U. נסמן $\overline{A_i} = U - A_i$ (המשלים של A_i ביחס ל- U) ו- $\cap(\varnothing) = U$. אז:

$$\left|\bigcap_{i=1}^{n} \overline{A_i}\right| = \left|\overline{A_1} \cap \overline{A_2} \cap \ldots \cap \overline{A_n}\right| = \sum_{X \in P(\{1,\ldots,n\})} (-1)^{|X|} \cdot \left|\bigcap_{i \in X} A_i\right|$$

דוגמה: אם $n = 2$, אז $P(\{1, 2\}) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}$.

$\lvert\emptyset\rvert = 0$	$\bigcap\limits_{i\in\emptyset} A_i = U$	$\left\lvert\bigcap\limits_{i\in\emptyset} A_i\right\rvert = \lvert U\rvert$
$\lvert\{1\}\rvert = 1$	$\bigcap\limits_{i\in\{1\}} A_i = A_1$	$\left\lvert\bigcap\limits_{i\in\{1\}} A_i\right\rvert = \lvert A_1\rvert$
$\lvert\{2\}\rvert = 1$	$\bigcap\limits_{i\in\{2\}} A_i = A_2$	$\left\lvert\bigcap\limits_{i\in\{2\}} A_i\right\rvert = \lvert A_2\rvert$
$\lvert\{1,2\}\rvert = 2$	$\bigcap\limits_{i\in\{1,2\}} A_i = A_1 \cap A_2$	$\left\lvert\bigcap\limits_{i\in\{1,2\}} A_i\right\rvert = \lvert A_1 \cap A_2\rvert$

לכן, לפי נוסח זה של עיקרון ההכלה וההפרדה:

$$\lvert\overline{A}_1 \cap \overline{A}_2\rvert = (-1)^0\lvert U\rvert + (-1)^1\lvert A_1\rvert + (-1)^1\lvert A_2\rvert + (-1)^2\lvert A_1 \cap A_2\rvert =$$
$$= \lvert U\rvert - \lvert A_1\rvert - \lvert A_2\rvert + \lvert A_1 \cap A_2\rvert$$

הוכחת נוסח ב' של עיקרון ההכלה וההפרדה:

$$\lvert\overline{A}_1 \cap \ldots \cap \overline{A}_n\rvert = \lvert\overline{A_1 \cup \ldots \cup A_n}\rvert = \lvert U\rvert - \lvert A_1 \cup \ldots \cup A_n\rvert$$

חוק דה-מורגן

(לפי נוסח א')
$$= \lvert U\rvert - \sum_{X\in E_n} (-1)^{\lvert X\rvert - 1} \cdot \left\lvert\bigcap_{i\in X} A_i\right\rvert$$

$$= \lvert\cap\emptyset\rvert + \sum_{X\in E_n} (-1)^{\lvert X\rvert} \cdot \left\lvert\bigcap_{i\in X} A_i\right\rvert$$

(כי $E_n \cup \{\emptyset\} = P(\{1, \ldots, n\})$)
$$= \sum_{X\in P(\{1,\ldots,n\})} (-1)^{\lvert X\rvert} \cdot \left\lvert\bigcap_{i\in X} A_i\right\rvert$$

הערות:

(1) שני הנוסחים של עיקרון ההכלה וההפרדה הינם שקולים זה לזה: ממש כשם שהוכחנו כאן את נוסח ב' על-סמך נוסח א', יכולנו בקלות לעשות את ההיפך (ההוכחה דומה).

(2) בטקסטים רבים מובאת ההוכחה האינטואיטיבית הבאה לנוסח ב': נניח $a \in U$, ונניח ש- $A_{i_1},\dots,A_{i_k}$ הם איברי $\{A_1, \dots, A_n\}$ ש- a שייך אליהם (ייתכן ש- $k=0$).

אם $k>0$, אז a "תורם" לסכום באגף ימין של נוסח ב' $\sum_{j=1}^{k}(-1)^j\binom{k}{j}=0$ * (כל קבוצה של j קבוצות מתוך $A_{i_1},\dots,A_{i_k}$ תורמת משהו). אם $k=0$ אז a תורם 1 בלבד (רק $|U|$). סך-הכל, רק איברי $|\overline{A}_1 \cap \dots \cap \overline{A}_n|$ תורמים 1 כל אחד, ומכאן השוויון. הוכחה זו היא נכונה באופן בסיסי, אבל אין זו משימה של מה בכך להפוך אותה להוכחה מדויקת באמת!

צורת השימוש בנוסח השני של עיקרון ההכלה וההפרדה היא כזו: כאשר רוצים למצוא את מספר האיברים של קבוצה מסוימת U, המקיימים את התכונות $P_1, \dots, P_n$, אזי מגדירים את A_i בתור קבוצת האיברים ב- U, *שאינם* מקיימים את P_i:

$$A_i = \{x \in U \mid \neg P_i(x)\}$$

מחפשים אז את $|\overline{A}_1 \cap \overline{A}_2 \cap \dots \cap \overline{A}_n|$, ועיקרון ההכלה וההפרדה עשוי לעזור (בתנאי, כמובן, שמציאת מספר האיברים בחיתוכים השונים של קבוצות מתוך $A_1, \dots, A_n$ הוא קל יותר ממציאת $\left|\bigcap_{i=1}^{n}\overline{A}_i\right|$).

דוגמה 1:

מצא את מספר הטבעיים בין 1 ל- 1000, שאינם מתחלקים ב- 5, 6 ו- 8.

תשובה:

כאן $U=\{n \in \mathbf{N}^+ \mid n \le 1000\}$. לכן $|U| = 1000$.

נגדיר:

* זו נוסחה שנראה בפרק הבא.

$$A_1 = \{n \in U \mid n \text{ מתחלק ב- } 5\}$$
$$A_2 = \{n \in U \mid n \text{ מתחלק ב- } 6\}$$
$$A_3 = \{n \in U \mid n \text{ מתחלק ב- } 8\}$$

אז

$$|A_1| = \left[\frac{1000}{5}\right] = 200 \ , \ |A_2| = \left[\frac{1000}{6}\right] = 166 \ , \ |A_3| = \left[\frac{1000}{8}\right] = 125$$

כמו כן:

$$A_1 \cap A_2 = \{n \in U \mid n \text{ מתחלק ב- } 30\}$$
$$A_1 \cap A_3 = \{n \in U \mid n \text{ מתחלק ב- } 40\}$$
$$A_2 \cap A_3 = \{n \in U \mid n \text{ מתחלק ב- } 24\}$$
$$A_1 \cap A_2 \cap A_3 = \{n \in U \mid n \text{ מתחלק ב- } 120\}$$

לכן:

$$|A_1 \cap A_3| = \left[\frac{1000}{40}\right] = 25 \qquad |A_1 \cap A_2| = \left[\frac{1000}{30}\right] = 33$$

$$|A_1 \cap A_2 \cap A_3| = \left[\frac{1000}{120}\right] = 8 \qquad |A_2 \cap A_3| = \left[\frac{1000}{24}\right] = 41$$

לכן התשובה:

$$1000 - (200 + 166 + 125) + (33 + 25 + 41) - 8 = 600$$

מסקנה:

מספר הטבעיים בין 1 ל- 1000, שמתחלקים ב- 5, 6 או 8 הוא $1000 - 600 = 400$.

שימוש נפוץ במיוחד בעיקרון ההכלה וההפרדה נעשה במקרים, שמספר האיברים ב- $A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}$ תלוי ב- k בלבד (בתנאי שכל האינדקסים שונים זה מזה). אם נסמן מספר איברים זה ב- α_k (כש- $\alpha_0 = |U|$), נקבל מעיקרון ההכלה וההפרדה ש:

$$\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right| = \sum_{k=0}^{n} (-1)^k \alpha_k \cdot \binom{n}{k}$$

(זאת כיוון שמספר החיתוכים בנוסחה, שעוצמתם α_k , הוא כמספר האפשרויות לבחור k אינדקסים שונים $i_1, \dots, i_k$ מתוך הקבוצה $\{1, 2, \dots, n\}$, כלומר: $\binom{n}{k}$).

דוגמה 2:

מהו מספר המספרים בני n ספרות, שאפשר להרכיב מ- $\{1,2,3,4,5\}$, אם 1, 2 ו- 3 חייבים להופיע?

תשובה:

כאן U היא קבוצת המספרים בני n ספרות שאפשר להרכיב מ- $\{1,2,3,4,5\}$.

A_1 היא קבוצת המספרים ב- U בהם 1 אינו מופיע.

A_2 היא קבוצת המספרים ב- U בהם 2 אינו מופיע.

A_3 היא קבוצת המספרים ב- U בהם 3 אינו מופיע.

אנו מחפשים את $|\overline{A}_1 \cap \overline{A}_2 \cap \overline{A}_3|$. עתה:

$$|U| = 5^n \Rightarrow \alpha_0 = 5^n$$

$$|A_1| = |A_2| = |A_3| = 4^n \Rightarrow \alpha_1 = 4^n$$

$$|A_1 \cap A_2| = |A_1 \cap A_3| = |A_2 \cap A_3| = 3^n \Rightarrow \alpha_2 = 3^n$$

$$|A_1 \cap A_2 \cap A_3| = 2^n \Rightarrow \alpha_3 = 2^n$$

לכן התשובה היא:

$$\binom{3}{0} \cdot 5^n - \binom{3}{1} \cdot 4^n + \binom{3}{2} \cdot 3^n - \binom{3}{3} \cdot 2^n$$

כלומר:

$$5^n - 3 \cdot 4^n + 3 \cdot 3^n - 2^n$$

אזהרה:

הנוסחה $\sum_{k=0}^{n} (-1)^k \alpha_k \cdot \binom{n}{k}$ ישימה רק כאשר ל- $\alpha_1, \dots, \alpha_n$ יש מובן, כלומר: רק כשמתקיים התנאי, שמספר האיברים ב- $A_{i_1} \cap \dots \cap A_{i_k}$ תלוי רק ב- k (וזאת לכל k).

תנאי זה התקיים בדוגמה האחרונה. כך, למשל, עבור $k = 2$ היה לחיתוך של *כל* שתי קבוצות מתוך $\{A_1, A_2, A_3\}$ אותו מספר איברים: 3^n.

דוגמה 3:

בעיית המלצר ו- n הכובעים: n אורחים במסעדה מוסרים בעת הכניסה את כובעיהם למלצר. בשעת היציאה מחזיר להם המלצר את הכובעים באופן אקראי. כמה אפשרויות יש לו להחזיר את הכובעים, כך שאף אורח לא מקבל את הכובע של עצמו?

נוסח כללי יותר של הבעיה, במונחי תורת הקבוצות, הינו:
נניח כי A ו- B הן קבוצות כך ש- $|A| = |B| = n$, ונניח ש- g_0 היא פונקציית שקילות מ- B ל- A. מצא את $|\{f \in \mathrm{Eq}(A, B) \mid \forall x \in B\ f(x) \neq g_0(x)\}|$.
(בבעיה המקורית B היא קבוצת האורחים, A קבוצת הכובעים, g_0 היא הפונקציה המתאימה לכל אורח את הכובע שלו, ו- $\mathrm{Eq}(A, B)$ הינה קבוצת ההתאמות האפשריות בין קבוצת האורחים לקבוצת הכובעים).

פתרון:
נניח $B = \{b_1, \ldots, b_n\}$. תהי $U = \mathrm{Eq}(A, B)$.
נגדיר:

$$A_i = \{f \in \mathrm{Eq}(A, B) \mid f(b_i) = g_0(b_i)\}$$

אנו מחפשים את:

$$\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right|$$

עתה, קל לראות, שאם $i_1, \ldots, i_k$ הם k אינדקסים שונים מתוך $\{1, \ldots, n\}$, אז

$$\left|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right| = (n-k)!$$

(זה כולל את המקרה $k = 0$, כי $|\cap(\varnothing)| = |U| = |\mathrm{Eq}(A, B)| = n!$).

לכן, לפי In-Ex, התשובה היא:

$$\sum_{k=0}^{n}(-1)^k\binom{n}{k}\cdot(n-k)! = \sum_{k=0}^{n}(-1)^k\cdot\frac{n!}{k!} = n!\cdot\sum_{k=0}^{n}(-1)^k\cdot\frac{1}{k!}$$

הערה:

מנוסחה זו נובעת המסקנה המעניינת הבאה: אם נסמן את המספר שמצאנו ב- $D(n)$, אז $\frac{D(n)}{n!} = \sum_{k=0}^{n}(-1)^k\cdot\frac{1}{k!}$. מחשבון דיפרנציאלי ואינטגרלי נובע לכן ש-

$$\lim_{n\to\infty}\frac{D(n)}{n!}=\frac{1}{e}$$

עתה, $\frac{D(n)}{n!}$ היא *ההסתברות* שאף אורח לא יקבל את כובעו, ומכאן שהסתברות זו שואפת ל- $\frac{1}{e}$. יתר-על-כן, ידוע לנו מחשבון דיפרנציאלי ש-

$$\left|\frac{D(n)}{n!}-\frac{1}{e}\right|<\frac{1}{(n+1)!}$$

לכן, כבר עבור n-ים קטנים מאוד ההסתברות היא בערך $\frac{1}{e}$ (השגיאה עבור $n=7$, למשל, קטנה מ- 0.00003, והיא הולכת וקטנה). המסקנה המפתיעה היא, שבין אם יש 10 אורחים או 1,000,000, ההסתברות שאף אורח לא יקבל את כובעו בחזרה היא בכל המקרים כמעט זהה (וגדולה למדי): בערך $\frac{1}{e}$ (למעלה מ- $\frac{1}{3}$!), כשההבדלים בין המקרים השונים הם זניחים!

הערה:

במקרה הפרטי בו $A=B$ ו- $g_0=i_B$, נוהגים לקרוא לפונקציות שנספרו בדוגמה זו (כלומר, הפונקציות ב- $\{f\in \mathrm{Eq}(B,B)\mid \forall x\in B.\, f(x)\neq x\}$) "אי סדרים טוטליים על B".

מספר הפתרונות של משוואות בשלמים עם אילוצים

בעיה:

כמה פתרונות במספרים *שלמים* יש למשוואה

$$x_1+x_2+x_3=9$$

בתנאי ש- $1\le x_1\le 2$, $-2\le x_2\le 4$ ו- $2\le x_2\le 4$?

בעיה זו היא דוגמה לסוג כללי של שאלות, שניתן לפתור בעזרת עיקרון ההכלה וההפרדה: כמה פתרונות במספרים שלמים יש למשוואה

$$x_1+x_2+\ldots+x_n=k$$

בתנאי שמתקיימים האילוצים: $a_i \le x_i \le b_i$ (כאשר לכל $1 \le i < n$ a_i הוא מספר שלם, ו- b_i הוא שלם או $b_i = \infty$)?

הנוסחאות הבסיסיות של הקומבינטוריקה נותנות את התשובה במקרה הפרטי, בו לכל $1 \le i < n$ מתקיים ש- $a_i = 0$ ו- $b_i \ge k$. התשובה אז היא $S(n, k)$. עתה, אחת הדרכים העיקריות במתמטיקה להתמודדות עם בעיות חדשות היא לנסות לעשות להן רדוקציה לבעיות, אותן אנו כבר יודעים לפתור. שלב ראשון בפתרון בעיות מהסוג, שאנו דנים בו עכשיו, יהיה לכן ביצוע *נורמליזציה* שלהן, שתהפוך אותן לבעיות בהן $a_i = 0$ לכל n, כמו במקרה שפתרונו ידוע לנו. נוכל להשיג זאת באופן פשוט על-ידי שינוי המשתנים אתם אנו עובדים. לשם כך נגדיר:

$$y_i = x_i - a_i \qquad (i = 1, \ldots, n)$$

כיוון ש- $x_i = y_i + a_i$ לכל i, נוכל להציב $y_i + a_i$ במקום x_i במשוואה המקורית, ולקבל כך משוואה חדשה, עם אילוצים חדשים, אבל עם אותו מספר פתרונות כמו שהיה למשוואה המקורית. בדוגמה שלנו זה ייראה כך:

$$y_1 = x_1 - 1 \qquad y_2 = x_2 + 2 \qquad y_3 = x_3 - 2$$

המשוואה שנקבל תהיה לכן:

$$(y_1 + 1) + (y_2 - 2) + (y_3 + 2) = 9$$

כלומר:

$$y_1 + y_2 + y_3 = 8$$

האילוצים החדשים יהיו:

$$1 \le y_1 + 1 \le 2 \qquad -2 \le y_2 - 2 \le 4 \qquad 2 \le y_3 + 2 \le 4$$

כלומר:

$$0 \le y_1 \le 1 \qquad 0 \le y_2 \le 6 \qquad 0 \le y_3 \le 2$$

אם נעבור משימוש ב- $y_1, \ldots, y_3$ חזרה ל- $x_1, \ldots, x_3$ (צעד שאינו הכרחי, אך בוודאי אינו משנה את מספר הפתרונות), נקבל, שעלינו לפתור את המשוואה $x_1 + x_2 + x_3 = 8$ בהגבלות: $0 \le x_1 \le 1$, $0 \le x_2 \le 6$ ו- $0 \le x_3 \le 2$. במלים אחרות: עלינו שוב לפתור משוואה מהצורה $x_1 + x_2 + \ldots + x_n = k$ (לא אותו k שהיה במשוואה המקורית!) עם אילוצים מהצורה $0 \le x_i \le b_i$ (שוב, לא אותם b_i-ים שהיו במשוואה המקורית!). כדי לפתור בעיה זו בעזרת In-Ex נגדיר:

$$U = \{\langle x_1, \ldots, x_n\rangle \in \mathbf{N}^n \mid x_1 + x_2 + \ldots + x_n = k\}$$
$$A_i = \{\langle x_1, \ldots, x_n\rangle \in U \mid x_i \ge b_i + 1\}$$

ונמצא את $\left|\overline{A}_1 \cap \overline{A}_2 \cap \ldots \cap \overline{A}_n\right|$.

בדוגמה הספציפית שלנו:

$$U = \{\langle x_1, x_2, x_3\rangle \in \mathbf{N}^3 \mid x_1 + x_2 + x_3 = 8\}$$
$$A_1 = \{\langle x_1, x_2, x_3\rangle \in \mathbf{N}^3 \mid x_1 + x_2 + x_3 = 8 \wedge x_1 \ge 2\}$$
$$A_2 = \{\langle x_1, x_2, x_3\rangle \in \mathbf{N}^3 \mid x_1 + x_2 + x_3 = 8 \wedge x_2 \ge 7\}$$
$$A_3 = \{\langle x_1, x_2, x_3\rangle \in \mathbf{N}^3 \mid x_1 + x_2 + x_3 = 8 \wedge x_3 \ge 3\}$$

הפעלת In-Ex במצב אליו הגענו תדרוש למצוא את $\left|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right|$ בכל הצירופים האפשריים. משמעות הדבר היא שוב מציאת מספר הפתרונות של משוואות עם הגבלות (מהסוג בו אנו עוסקים). הפעם יוביל אבל נרמול הבעיות לכאלה, שפתרונן הוא $S(n, k)$ עבור k מתאים.

הבה נדגים את מציאת $|A_1 \cap A_3|$ בדוגמה הספציפית שלנו. המדובר, למעשה, במציאת מספר הפתרונות של $x_1 + x_2 + x_3 = 8$ עם האילוצים: $x_1 \ge 2$, $x_3 \ge 3$.

נגדיר אפוא:

$$y_3 = x_3 - 3 \qquad y_2 = x_2 \qquad y_1 = x_1 - 2$$

נקבל:

$$(y_1 + 2) + y_2 + (y_3 + 3) = 8$$

כלומר:

$$y_1 + y_2 + y_3 = 3$$

כש- y_1, y_2, y_3 חייבים להיות מספרים טבעיים (ואין הגבלות נוספות!), מספר הפתרונות של משוואה זו הוא $S(3, 3)$, כלומר $\binom{5}{3}$.

באופן דומה נקבל כאן:

$$|u| = S(3,8) = \binom{10}{8}$$

$$|A_1| = S(3,6) = \binom{8}{6} \qquad |A_2| = S(3,1) = \binom{3}{1} \qquad |A_3| = S(3,5) = \binom{7}{5}$$

כמו-כן, $A_1 \cap A_2 = A_2 \cap A_3 = A_1 \cap A_2 \cap A_3 = \varnothing$. לכן, מספר הפתרונות של הבעיה הוא סך-הכל:

$$\binom{10}{8} - \binom{8}{6} - \binom{3}{1} - \binom{7}{5} + \binom{5}{3} = 45 - 28 - 3 - 21 + 10 = 3$$

הערה:

בדוגמה ספציפית זו יכולנו פשוט לספור את הפתרונות ישירות, וזה היה מהיר יותר. עם זאת, זה אינו המצב בדרך-כלל!

נספח: הוכחת עיקרון ההכלה וההפרדה

סימונים:

$E_n = P(\{1,\ldots,n\}) - \{\varnothing\}$ $\qquad$ $E_n^* = \{X \in P(\{1,\ldots,n+1\}) \mid n+1 \in X\} - \{\{n+1\}\}$

למה 1

אם A ו- B סופיות, $F : A \to B$ היא פונקציית שקילות ו- $g : B \to \mathbf{R}$, אז:

$$\sum_{x\in A} g(F(x)) = \sum_{x\in B} g(x)$$

למה 2

(i) $E_{n+1} = E_n \cup \{\{n+1\}\} \cup E_n^*$.

(ii) $F_n = \lambda X \in E_n \,.\, X \cup \{n+1\}$ היא פונקציית שקילות מ- E_n על E_n^* ומתקיים:

$$\forall X \in E_n \,.\, |F_n(X)| - 1 = |X|$$

את הוכחת שתי הלמות נשאיר לקורא.

הוכחת העיקרון:

באינדוקציה. ל- $n = 2$ כבר הוכחנו.
נניח נכונות ל- n , נוכיח ל- $n+1$. יהיו אפוא $A_1, \ldots, A_{n+1}$ קבוצות.

$$\left|\bigcup_{i=1}^{n+1} A_i\right| = \left|\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right| = \left|\bigcup_{i=1}^{n} A_i\right| + |A_{n+1}| - \left|\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right|$$

$$= \left|\bigcup_{i=1}^{n} A_i\right| + |A_{n+1}| - \left|\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right| \qquad \text{לפי חוק הפילוג}$$

$$= \sum_{X\in E_n} (-1)^{|X|-1} \left|\bigcap_{i\in X} A_i\right| + (-1)^{|\{n+1\}|-1} |A_{n+1}| - \sum_{X\in E_n} (-1)^{|X|-1} \left|\bigcap_{i\in X} (A_i \cap A_{n+1})\right| \qquad \text{לפי הנחת האינדוקציה (פעמיים)}$$

$$= \sum_{X\in E_n} (-1)^{|X|-1} \left|\bigcap_{i\in X} A_i\right| + (-1)^{|\{n+1\}|-1} \left|\bigcap_{i\in\{n+1\}} A_i\right| + \sum_{X\in E_n} (-1)^{|X|} \left|\bigcap_{i\in X} (A_i \cap A_{n+1})\right|$$

ברור, אבל, ש-

$$\bigcap_{i\in X}\left(A_i \cap A_{n+1}\right) = \bigcap_{i\in F_n(X)} A_i$$

לכן נקבל לפי למה 2(ii):

$$\left|\bigcup_{i=1}^{n+1} A_i\right| = \sum_{X\in E_n}(-1)^{|X|-1}\left|\bigcap_{i\in X} A_i\right| + (-1)^{|\{n+1\}|-1}\left|\bigcap_{i\in\{n+1\}} A_i\right| + \sum_{X\in E_n}(-1)^{|F_n(X)|-1}\left|\bigcap_{i\in F_n(X)} A_i\right|$$

לפי למה 1 $$= \sum_{X\in E_n}(-1)^{|X|-1}\left|\bigcap_{i\in X} A_i\right| + (-1)^{|\{n+1\}|-1}\left|\bigcap_{i\in\{n+1\}} A_i\right| + \sum_{X\in E_n^*}(-1)^{|X|-1}\left|\bigcap_{i\in X} A_i\right|$$

$$= \sum_{X\in E_n\cup\{\{n+1\}\}\cup E_n^*}(-1)^{|X|-1}\left|\bigcap_{i\in X} A_i\right|$$

לפי למה 2 (i) $$= \sum_{X\in E_{n+1}}(-1)^{|X|-1}\left|\bigcap_{i\in X} A_i\right|$$

ד.3 תכונות המקדמים הבינומיאליים

המספרים $\binom{n}{k}$, כאשר $0 \le k < n$, הם בעלי חשיבות רבה במתמטיקה, מעבר לשימוש שנעשה בהם בקומבינטוריקה סופית. פרק זה יוקדש לכן לכמה מתכונותיהם. נתחיל בזהויות הפשוטות הבאות:

$$\binom{n}{k} = \binom{n}{n-k} \qquad (0 \le k \le n) \tag{1}$$

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \qquad (1 \le k \le n-1) \tag{2}$$

לרוב הטענות מסוג זה ניתן לתת שתי הוכחות שונות: אחת אלגברית והשנייה קומבינטורית. הוכחה אלגברית של (1), למשל, נראית כך:

$$\binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{(n-k)!k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}$$

הוכחה קומבינטורית, לעומת זאת, מתבססת על כך, שאם $|A| = n$, אז $\lambda X \in P_k(A).A - X$ היא פונקציית שקילות מ- $P_k(A)$ על $P_{n-k}(A)$, ולכן

$$\binom{n}{k} = |P_k(A)| = |P_{n-k}(A)| = \binom{n}{n-k}$$

(ניסוח יותר אינטואיטיבי: לבחור k עצמים מתוך n , או לבחור את $n-k$ העצמים הנותרים, זו בעצם אותה בעיה.)

לגבי (2) הוכחה אלגברית תיראה כך:

$$\binom{n-1}{k}+\binom{n-1}{k-1}=\frac{(n-1)!}{k!(n-1-k))!}+\frac{(n-1)!}{(k-1)!(n-k)!}=$$

$$=\frac{(n-1)!(n-k)+(n-1)!k}{k!(n-k)!}=\frac{(n-1)!(n-k+k)}{k!(n-k)!}=$$

$$=\frac{(n-1)!n}{k!(n-k)!}=\frac{n!}{k!(n-k)!}=\binom{n}{k}$$

(השתמשנו במהלך ההוכחה 3 פעמים בעובדה, שלכל $m \in \mathbf{N}^+$ מתקיים:
$m! = (m-1)! \cdot m$).

הוכחה קומבינטורית של אותה זהות:
את $P_k(\{1, \ldots, n\})$ אפשר לחלק לשתי תת-קבוצות זרות:

(1) אלה ש- n שייך אליהן. קבוצה זו אקויפוטנטית עם $P_{k-1}(\{1, \ldots, n-1\})$ (כי פרט ל- n עלינו לבחור בעוד $k-1$ מספרים מבין המספרים בין 1 ל- $n-1$).

(2) אלה ש- n לא שייך אליהן. תת-קבוצה זו היא *בדיוק* $P_k(\{1, \ldots, n-1\})$. לכן:

$$\binom{n}{k}=\left|P_k(\{1,\ldots,n\})\right|=\left|P_{k-1}(\{1,\ldots,n-1\})\right|+\left|P_k(\{1,\ldots,n-1\})\right|=\binom{n-1}{k-1}+\binom{n-1}{k}$$

שתי הזהויות שהוכחנו (בעיקר השנייה), יחד עם העובדה שלכל $n \in \mathbf{N}$ מתקיים

$$\binom{n}{0}=1 \qquad \binom{n}{n}=1$$

מאפשרות למצוא את המקדמים הבינומיאליים בעזרת מה שידוע בשם "משולש פסקל":

$$\begin{array}{c|ccccccccccc}
0 & & & & & & 1 & & & & & \\
1 & & & & & 1 & & 1 & & & & \\
2 & & & & 1 & & 2 & & 1 & & & \\
3 & & & 1 & & 3 & & 3 & & 1 & & \\
4 & & 1 & & 4 & & 6 & & 4 & & 1 & \\
5 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & \iddots & & & & \vdots & & & & \ddots &
\end{array}$$

האיבר ה- k-י בשורה מספר n (הספירה מתחילה *מאפס*) נותן את $\binom{n}{k}$. המשולש נבנה כך: 1-ים בקצוות (לפי $\binom{n}{0}=1, \binom{n}{n}=1$), וכל מספר אחר שווה לסכום השניים, שנמצאים מעליו (זה לפי זהות מס' (2) למעלה). זהות מס' (1) היא ה"אחראית" לכך שב"משולש" יש סימטריה בין צד ימין לצד שמאל.

משפט הבינום

את השם "מקדמים בינומיאליים" קיבלו המספרים $\binom{n}{k}$ ממשפט מפורסם אחר שהם מככבים בו, הלוא הוא *משפט הבינום*:

$$\boxed{(a+b)^n=\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k \quad \left(=a^n+\binom{n}{1}a^{n-1}b+\binom{n}{2}a^{n-2}b^2+\ldots+b^n\right)}$$

הסבר אינטואיטיבי למה נוסחה זו נכונה מתקבל מהעובדה ש-

$$(a+b)^n=\overbrace{(a+b)\cdot(a+b)\cdot\ldots\cdot(a+b)}^{n\text{ פעמים}}$$

אם נפתח את כל הסוגריים באגף ימין נקבל סכום של מונומים רבים מהצורה $a^{n-k}b^k$ (כש- $0\le k\le n$). עבור k ספציפי, מספר הפעמים שהמונום $a^{n-k}b^k$ מופיע הוא כמספר האפשרויות לבחור את k ה- $(a+b)$-ים (מתוך ה- n) שיתרמו את ה- b-ים של $a^{n-k}b^k$ (כל שאר ה- $(a+b)$-ים יתרמו a), לכן מספר זה הוא $\binom{n}{k}$. כינוס כל האיברים ייתן אפוא את נוסחת הבינום.

הוכחה מדויקת של משפט הבינום נעשית באינדוקציה. המשפט טריביאלי עבור $n=0$ (וגם עבור $n=1$). נניח נכונות ל- $n-1$. נוכיח ל- n.

לפי הנחת האינדוקציה:

$$(a+b)^n=(a+b)^{n-1}\cdot(a+b)=(a+b)\cdot\sum_{k=0}^{n-1}\binom{n-1}{k}a^{n-1-k}b^k$$

$$=\sum_{k=0}^{n-1}\binom{n-1}{k}a^{n-k}b^k+\sum_{k=0}^{n-1}\binom{n-1}{k}a^{n-k-1}b^{k+1}$$

$$=\binom{n-1}{0}a^n+\sum_{k=1}^{n-1}\binom{n-1}{k}a^{n-k}b^k+\sum_{k=0}^{n-2}\binom{n-1}{k}a^{n-(k+1)}b^{k+1}+\binom{n-1}{n-1}b^n$$

$$=a^n+\sum_{k=1}^{n-1}\binom{n-1}{k}a^{n-k}b^k+\sum_{k=1}^{n-1}\binom{n-1}{k-1}a^{n-k}b^k+b^n$$

$$=a^n+\sum_{k=1}^{n-1}\left(\binom{n-1}{k}+\binom{n-1}{k-1}\right)a^{n-k}b^k+b^n$$

לפי זהות (2)

$$=\binom{n}{0}a^n+\sum_{k=1}^{n-1}\binom{n}{k}a^{n-k}b^k+\binom{n}{n}b^n$$

$$=\sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k$$

מסקנות פשוטות ממשפט הבינום:

1. אם נציב במשפט הבינום $a=1$ ו- $b=1$, נקבל:

$$2^n=\sum_{k=0}^{n}\binom{n}{k}$$

מזה נובע שאם ב- A יש n איברים, אז:

$$|P(A)|=\sum_{k=0}^{n}|P_k(A)|$$

נוסחה אחרונה זו ברורה כמובן גם מהעובדה ש:

$$P(A) = \biguplus_{k=0}^{n} P_k(A)$$

2. אם נציב במשפט הבינום $a = 1$ ו- $b = -1$, אז עבור $n > 0$ נקבל:

$$0 = \sum_{k=0}^{n} (-1)^k \cdot \binom{n}{k}$$

(זה לא מה שנקבל עבור $n = 0$, כי בקומבינטוריקה $0^0 = 1$, וגם $(-1)^0 = 1$, $\binom{0}{0} = 1$).

הכללת המקדמים הבינומיאליים

נתחיל בהבחנה הבאה:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!}$$

את הביטוי $n(n-1) \ldots (n-k+1)$ מסמנים לעתים קרובות ב- $n^{\underline{k}}$ (או $(n)_k$). מסתבר, שבחשבון דיפרנציאלי ואינטגרלי יש חשיבות לצירוף זה,גם כשבמקום n מציבים לא רק מספר טבעי גדול מ- k, אלא מספר ממשי כלשהו. בהתאם מגדירים:

$$x^{\underline{k}} = \begin{cases} 1 & k = 0 \\ x(x-1)(x-2)\ldots(x-k+1) & k > 0 \end{cases} \qquad (k \in \mathbf{N},\ x \in \mathbf{R})$$

$$\binom{x}{k} = \frac{x^{\underline{k}}}{k!} \qquad (k \in \mathbf{N},\ x \in \mathbf{R})$$

דוגמאות:

(1) לכל $x \in \mathbf{R}$:
$$\binom{x}{2} = \frac{x(x-1)}{2}$$

(2) אם נציב $x = -n$ כש- $n \in \mathbf{N}$, נקבל עבור $k > 0$:

$$\binom{-n}{k} = \frac{(-n)(-n-1)(-n-2)\dots(-n-k+1)}{k!} =$$

$$= \frac{(-1)^k \cdot n(n+1)(n+2)\dots(n+k-1)}{k!} = \frac{(-1)^k \cdot (n+k-1)^{\underline{k}}}{k!} =$$

$$= (-1)^k \cdot \binom{n+k-1}{k} = (-1)^n \cdot S(n,k)$$

(3) לכל $x \in \mathbf{R}$: $\binom{x}{0} = 1$ ו- $\binom{x}{1} = x$.

(4) $\binom{n}{k}$ מוגדר עתה גם כש- $n \in \mathbf{N}$ ו- $n < k$, אבל ברור ש:

$$\forall k\, \forall n.\ k \in \mathbf{N} \wedge n \in \mathbf{N} \wedge n < k \Rightarrow \binom{n}{k} = 0$$

(כיוון שאחד הגורמים ב- $n(n-1)\dots(n-k+1)$ כאשר $n<k$ יהיה $n-n$).

הסיבה העיקרית להכללת המקדמים הבינומיאליים ל- x ממשי כלשהו הינה, שבחשבון דיפרנציאלי ואינטגרלי מוכיחים, שאם $\left|\frac{b}{a}\right| < 1$, אז משפט הבינום נכון גם כשהמעריך אינו מספר טבעי, אלא α ממשי כלשהו:

$$\forall a \in \mathbf{R}\, \forall b \in \mathbf{R}\, \forall \alpha \in \mathbf{R}.\ \left|\frac{b}{a}\right| < 1 \Rightarrow (a+b)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} a^{\alpha-k} b^k$$

הבהרות:

(1) הסכום באגף ימין של משפט בינום מוכלל זה הוא סכום *אינסופי*. הגדרה מדויקת של סכום כזה (הקובעת מתי הוא "מתכנס", דהיינו: מתי יש לו ערך ממשי) ניתנת בחשבון אינפיניטסימלי (אם כי הנוסחה של סכום טור הנדסי אינסופי מתכנס נלמדת כבר בתיכון).

(2) הנוסחה הנ"ל היא אכן הכללה של משפט הבינום שהוכחנו כאן, כיוון שאם n טבעי, אז מה שמתקבל מהנוסחה המוכללת הוא אכן:

$$(a+b)^n = \sum_{k=0}^{\infty} \binom{n}{k} a^{n-k} b^k$$

הסיבה היא, שכל המחוברים החל מ- $k+1$ והלאה הם 0 (כי $\binom{n}{k} = 0$ כש- $k > n$).

(3) המקרה השימושי ביותר (גם לצרכינו בהמשך) של משפט הבינום המוכלל, שראינו למעלה, הוא כאשר $a = 1$ ו- $b = x$ כך ש- $|x| < 1$. נקבל אז:

$$\forall x \in \mathbf{R}\ \forall \alpha \in \mathbf{R}.\ |x| < 1 \Rightarrow (1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} x^k$$

באופן מעשי, כדאי לציין, מוכיחים תחילה עובדה זו, וההכללה המלאה מתקבלת על-ידי ההצבה $x = \frac{b}{a}$ (ופישוט).

זהויות רבות על המקדמים הבינומיאליים אפשר להכליל אם מרשים x ממשי כלשהו בניסוחן במקום מספרים טבעיים בלבד (רק בחלק העליון של $\binom{-}{-}$, לא בתחתון!). כך, למשל, אפשר להכליל את (2) עבור $x \in \mathbf{R}$ כלשהו (ו- $k \in \mathbf{N}$):

$$\binom{x}{k} = \binom{x-1}{k} + \binom{x-1}{k-1}$$

אפשר כאן להוכיח עובדה זו בעזרת פיתוח אלגברי. ברם, אין צורך בכך. אפשר להסיק אותה ישירות מהעובדה הידועה לנו כבר, דהיינו: שהיא נכונה כאשר x הוא מספר טבעי גדול מ- k. השיקול הוא השיקול הבא: עבור k *קבוע*, x^k ו- $\binom{x}{k}$ הם פולינומים ב- x ממעלה k. לכן $\binom{x}{k} - \binom{x-1}{k} - \binom{x-1}{k-1}$ גם הוא פולינום ממעלה k (לכל היותר). לפולינום זה יש אינסוף שורשים: כל מספר טבעי גדול מ- k הוא שורש שלו, לפי מה שהוכחנו. ברם, לפולינום ממעלה $k \geq 1$ יש לכל היותר k שורשים, אלא אם כן הוא פולינום האפס. היוצא מכאן הוא ש- $\binom{x}{k} - \binom{x-1}{k} - \binom{x-1}{k-1}$ שווה זהותית לאפס, ולכן $\binom{x}{k} = \binom{x-1}{k} + \binom{x-1}{k-1}$ לכל $x \in \mathbf{R}$.

טכניקת ההוכחה, שתיארנו עתה, ידועה בשם "השיקול הפולינומיאלי". היא מאפשרת להוכיח זהויות בינומיאליות רבות עבור מספרים ממשיים כלשהם על-ידי הוכחתן למספרים טבעיים בלבד. כמובן יש להיזהר פה: לא כל זהות נוכל להכליל כך. לא נוכל, למשל, להכליל את זהות (1) ל- x ממשי כלשהו (במקום n), כיוון ש- $\binom{x}{x-k}$ אינו מוגדר כלל, כש- x אינו מספר טבעי $\leq k$.

טבלה מס' ד.3 כוללת את עשר הזהויות הבינומיאליות החשובות ביותר. בטבלה זו a, b, x, y, α הם מספרים ממשיים כלשהם; n, m, k – מספרים טבעיים. בצד ימין רשומים תנאים לנכונותן של הזהויות משמאל.

טבלה ד.3:
זהויות בינומיאליות חשובות

	תנאים	
(1)	$k \le n$	$\binom{n}{k} = \binom{n}{n-k}$
(2)	$k \ge 1$	$\binom{x}{k} = \binom{x-1}{k} + \binom{x-1}{k-1}$
(3)	$\alpha \in \mathbf{N} \vee \left\lvert \frac{b}{a} \right\rvert < 1$	$(a+b)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} a^{\alpha-k} b^k$
(4)		$\sum_{k=0}^{\infty} \binom{n}{k} = 2^n$
(5)	$n \ge 1$	$\sum_{k=0}^{\infty} (-1)^k \binom{n}{k} = 0$
(6)		$\binom{-x}{k} = (-1)^k \binom{x+k-1}{k}$
(7)	$k \le n$	$\binom{x}{n}\binom{n}{k} = \binom{x}{k}\binom{x-k}{n-k}$
(8)		$\sum_{k=0}^{n} \binom{x+k}{k} = \binom{x+n+1}{n}$
(9)	$m \le n$	$\sum_{k=m}^{n} \binom{k}{m} = \sum_{k=0}^{n} \binom{k}{m} = \binom{n+1}{m+1}$
(10)		$\sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k} = \binom{x+y}{n}$

נעבור להוכחת הזהויות. את (1)-(5) כבר הוכחנו במהלך פרק זה. את (6) הוכחנו במקרה $x = n \in \mathbf{N}$, וניתן להכליל ל- x כלשהו בעזרת השיקול הפולינומיאלי. בדומה, גם את נוסחאות (7), (8) ו-(10) די להוכיח למקרה ש- x ו- y הם מספרים טבעיים, ולהסתמך על השיקול הפולינומיאלי לצורך ההכללה למספרים ממשיים כלשהם.

הוכחת (7):

נניח $x \in \mathbf{N}$. נבחן את הבעיה הבאה: מתוך קבוצה A עם x איברים יש לבחור קבוצה B עם n איברים, ומתוך קבוצה B יש לבחור קבוצה C עם k איברים. כמה אפשרויות יש?

פתרון א': מספר האפשרויות לבחור את B הוא $\binom{x}{n}$. אחר-כך אפשר לבחור את C מתוכה ב- $\binom{n}{k}$ דרכים. סך-הכל $\binom{n}{k} \cdot \binom{x}{n}$.

פתרון ב': נבחר תחילה את הקבוצה C. אפשר לעשות זאת ב- $\binom{x}{k}$ דרכים. אחר-כך נבחר את שאר האיברים ב- B. עלינו לבחור $n-k$ איברים מתוך $x-k$ שנותרו. לכך יש $\binom{x-k}{n-k}$ דרכים. סך-הכל יש $\binom{x-k}{n-k}\binom{x}{k}$ אפשרויות.

משני הפתרונות מקבלים את השוויון ב- (7).

הוכחת (8): באינדוקציה על n.

כש- $n=0$, אז $\sum_{k=0}^{0}\binom{x+k}{k} = \binom{x}{0} = 1$ וגם $\binom{x+0+1}{0} = 1$.

נניח נכונות ל- $n-1$. נוכיח ל- n. ואכן:

$$\sum_{k=0}^{n}\binom{x+k}{k} \overset{(1)}{=} \sum_{k=0}^{n-1}\binom{x+k}{k} + \binom{x+n}{n} \overset{(2)}{=} \binom{x+(n-1)+1}{n-1} + \binom{x+n}{n} \overset{(3)}{=}$$

$$\overset{(3)}{=} \binom{x+n}{n-1} + \binom{x+n}{n} \overset{(4)}{=} \binom{x+n+1}{n}$$

השוויון (2) כאן הוא על-פי הנחת האינדוקציה. זה שב- (4) – לפי זהות (2) מטבלה ד.3.

הסבר קומבינטורי ל- (8):

נחשב בשתי דרכים את מספר הפתרונות במספרים טבעיים של האי-שוויון:

$$x_1 + x_2 + \ldots + x_m + x_{m+1} \le n \qquad (m \in \mathbf{N}\,,\; n \in \mathbf{N})$$

(א) המספר המבוקש זהה למספר הפתרונות בטבעיים של המשוואה:

$$x_1 + x_2 + \ldots + x_m + x_{m+1} + x_{m+2} = n$$

(כיוון ש- $\lambda x_1, x_2, \ldots, x_m, x_{m+1}.\langle x_1, \ldots, x_m, x_{m+1}, n - \sum_{i=1}^{m+1} x_1 \rangle$ היא פונקציית שקילות מקבוצת הפתרונות של האי-שוויון על זו של המשוואה).

מכאן נקבל שהמספר המבוקש הוא $S(m+2, n)$.

(ב) קבוצת הפתרונות של האי-שוויון הינה בבירור:

$$\biguplus_{k=0}^{n} \left\{ \langle x_1, \ldots, x_m, x_{m+1} \rangle \in \mathrm{N}^{n+1} \,\middle|\, \sum_{i=1}^{m+1} x_i \;=\; k \right\}$$

לכן מספרם הינו $\sum_{k=0}^{n} S(m+1, k)$.

משתי התשובות אנו מקבלים:

$$S(m+2, n) = \sum_{k=0}^{n} S(m+1, k)$$

כלומר:

$$\binom{m+n+1}{n} = \sum_{k=0}^{n} \binom{m+k}{k}$$

זהות (8) מתקבלת מזה בעזרת השיקול הפולינומיאלי.

הערה:
הוכחה זו מובילה לשיטה כללית למציאת מספר הפתרונות במספרים טבעיים של האי-שוויון: $x_1, x_2 + \ldots + x_n \le k$. מספר זה שווה למספר הפתרונות הטבעיים של *המשוואה:* $x_1 + \ldots + x_n + x_{n+1} = k$. מכאן שהוא שווה ל- $S(n+1, k)$.

הוכחת (9):

העובדה ש- $\sum_{k=0}^{n}\binom{k}{m} = \sum_{k=m}^{n}\binom{k}{m}$ נובעת מכך ש- $\binom{k}{m} = 0$ כאשר $k < m$.

עתה:

$$\sum_{k=m}^{n}\binom{k}{m} \underset{\text{זהות (1)}}{=} \sum_{k=m}^{n}\binom{k}{k-m} \underset{\text{הזזת אינדקסים}}{=} \sum_{k=m-m}^{n-m}\binom{k+m}{(k+m)-m} =$$

$$= \sum_{k=0}^{n-m}\binom{m+k}{k} \underset{\text{זהות (8)}}{=} \binom{m+(n-m)+1}{n-m} = \binom{n+1}{n-m} \underset{\text{זהות (1)}}{=} \binom{n+1}{m+1}$$

הוכחת (10):

נניח $x, y \in \mathbf{N}$ ו- $x + y \ge n$. נתבונן בבעיה הבאה: תהיינה A ו- B קבוצות כך ש- $|A| = x$, $|B| = y$ ו- $A \cap B = \varnothing$. כמה אפשרויות יש לבחור n עצמים מתוך $A \cup B$?

פתרון א': ב- $A \cup B$ יש $x + y$ איברים. לכן התשובה $\binom{x+y}{n}$.

פתרון ב': נסמן ב- k את מספר האיברים שנבחר מ- A . $n-k$ יהיה אז מספר האיברים הנבחרים מתוך B . לכן מספר האפשרויות לבחור בדיוק k איברים מ- A (והשאר מ- B) הוא $\binom{x}{k}\binom{y}{n-k}$. עתה k יכול להיות כל מספר בין 0 ל- n , לכן סך-הכל מספר האפשרויות הוא $\sum_{k=0}^{n}\binom{x}{k}\binom{y}{n-k}$.

מהשוואת שני הפתרונות נקבל את הזהות ב- (10).

דוגמה:

בבחירות לאגודת הסטודנטים יש עשרה מועמדים. לכל בוחר יש חמישה קולות, שהוא יכול לחלקם כרצונו: הוא יכול לתת את כולם לאותו מועמד, או לחלקם בין חמישה מועמדים שונים וכו'. הוא אינו חייב, כמו-כן, להשתמש בכל הקולות העומדים לרשותו. בבחירות השתתפו 3100 סטודנטים ו- 97 מהם הטילו פתקים לבנים. הוכח שבין הנותרים יש שניים שהצביעו באותה צורה.

פתרון:

נסמן ב- A_i את קבוצת פתקי ההצבעה האפשריים עם i קולות. ברור ש- $|A_i| = S(10, i)$, כי צריך לבחור i מועמדים מתוך 10, עם אפשרות חזרה וללא חשיבות לסדר. מכאן ש- $|A_i| = \binom{10+i-1}{i} = \binom{9+i}{i}$.

קבוצת פתקי ההצבעה האפשריים היא $\biguplus_{i=0}^{5} A_i$. לכן מספרם:

$$\sum_{i=0}^{5}\binom{9+i}{i} \underset{\uparrow}{=} \binom{9+5+1}{5} = \binom{15}{5} = 3003$$

לפי (8)

מכאן שמספר פתקי ההצבעה הלא-ריקים האפשריים: $3003 - 1 = 3002$.

מספר פתקי ההצבעה הלא-ריקים שנספרו: $3100 - 97 = 3003$.

לפי עיקרון שובך היונים הפשוט, יש לכן שני פתקי הצבעה לא-ריקים זהים.

הערה:

מציאת מספר הפתקים האפשריים היא בעצם מציאת מספר הפתרונות במספרים טבעיים של

$$x_1, x_2 + \ldots + x_{10} \le 5$$

כמו שראינו בהערה אחרי הוכחת זהות (8), זה שקול למציאת מספר הפתרונות של המשוואה:

$$x_1, x_2 + \ldots + x_{10} + x_{11} = 5$$

(נוסף בה כאילו מועמד נוסף, שהינו "חסר שם").

ולכן התשובה היא $S(11, 5)$, כלומר

$$\binom{11+5-1}{5} = \binom{15}{5}$$

ד.4. פונקציות יוצרות

הבה נתחיל בבעיה הבאה, אותה נפתור לקראת סוף הפרק.

בעיה

כמה אפשרויות יש לבחור 173 מספרים מתוך $\{0, 1, 2\}$, אם 0 חייב להיבחר מספר זוגי של פעמים?

אפשר לנסות לפתור בעיה זו באופן סבלני על-ידי בדיקת כל האפשרויות. זה יהפוך לסיוט אם במקום 173 יהיה מדובר ב- 1733. בשלב זה נעדיף, אולי, שמחשב יעשה את העבודה השחורה במקומנו. בשביל זה נצטרך כמובן לכתוב תכנית מתאימה. די ברור, אבל, שתכנית כזו תוכל (אולי בשינויים קלים) לפתור את הבעיה לא רק עבור 173 או 1733, אלא עבור n טבעי כלשהו. היוצא מזה הוא, שכל ניסיון לפתור בעיה זו (אפילו בעזרת מחשב) יוביל, כמעט בהכרח, לצורך במציאת פתרון של בעיה כללית יותר.
במתמטיקה, אפילו לא ממוחשבת, אחת הגישות העיקריות לפתרון בעיות היא אכן לנסות *להכליל* אותן. אולי זה נראה מוזר, אך לעתים קרובות קל יותר לפתור בעיה כללית, ואחר-כך ליישם את הפתרון למקרה פרטי, בו אנו מעוניינים, מאשר לנסות לפתור ישירות את המקרה הפרטי. בקומבינטוריקה זה מתבטא בכך, שלעתים נוח להחליף את אחד הנתונים של בעיה מסוימת ב*פרמטר* n (או k או כל אחד אחר) ואז לנסות לפתור את הבעיה עבור n כלשהו. פתרון הבעיה המוכללת יהיה צריך לספק דרך יעילה לחישוב ה*פונקציה*, המתאימה לכל n את פתרון הבעיה, כשערך הפרמטר הוא n. אם אכן נמצא דרך יעילה כזו (על-ידי נוסחה ישירה, או אולי בדרך אחרת), נוכל להפעיל אותה על המקרה הספציפי המעניין אותנו, וכך לפתור אותו.

בדוגמה שלנו, אם נסמן ב- a_n את מספר האפשרויות לבחור n מספרים מתוך $\{0, 1, 2\}$, כש- 0 חייב להופיע מספר זוגי של פעמים, הרי מה *שתשאלנו* עליו הוא a_{173}. מה שננסה *לעשות*, לעומת זאת, הוא למצוא דרך יעילה לחישוב הפונקציה $\lambda n \in \mathbf{N}.\, a_n$ (אנו נמצא בהמשך נוסחה מפורשת לפונקציה זו!). מציאת ערך הפונקציה עבור $n = 173$ יהיה אז עניין פשוט.

ל- $\lambda n \in \mathbf{N}.\, a_n$ אנו קוראים, כזכור, *סדרה* של מספרים (טבעיים, במקרה זה). אינטואיטיבית, אנו רואים פונקציה זו כמייצגת רשימה אינסופית של מספרים $\langle a_0, a_1, a_2, a_3, \ldots\rangle$. אינטואיציה זו יש בה עזר רב (אך יש לזכור שהיא אינטואיציה בלבד!). לסדרות של מספרים (טבעיים, בדרך-כלל, אך לא תמיד) יש אכן שימוש רב

בפתרון בעיות קומבינטוריות. בפרק הזה ובפרק הבא נפתח כלים לטיפול יעיל בסדרות כאלה*, וכמו-כן נראה, איך ליישם כלים אלה לצורך פתרון בעיות קומבינטוריות. נתחיל במה שהוא אולי הכלי החזק והכללי ביותר (בחלק גדול של המקרים): *פונקציות יוצרות*. לצורך הפעלת כלי זה יבוא לעזרתנו החשבון האינפיניטסימלי.

סימונים:

בפרק זה, המשתנים i, j, k, ℓ, m, n ישמשו כמשתנים עבור מספרים **טבעיים**. במקום לכתוב $\lambda n \in \mathbf{N}.\, t$ או $\forall n \in \mathbf{N}.\, \varphi$, נכתוב לכן פשוט $\lambda n.\, t$ ו- $\forall n.\, \varphi$ (וכנ"ל עם שאר המשתנים האלה). המשתנים $a, b, c, x, y, z, u, v, w, \alpha, \beta, \gamma$ ישמשו עבור מספרים ממשיים (ולפעמים אפילו מרוכבים). במקום לכתוב $\lambda x \in \mathbf{R}.\, t$ או $\forall x \in \mathbf{R}.\, y$, נכתוב לכן בקיצור $\lambda x.\, t$ ו- $\forall x.\, \varphi$ (וכנ"ל עם השאר). יש לזכור תמיד, שאלו רק קיצורים למה שהיינו צריכים לכתוב באמת!

הרעיון המרכזי של פרק זה הוא, שסדרות רבות של מספרים נוצרות באופן טבעי על-ידי פונקציות (חלקיות) מ- $\mathbf{R}$ ל- $\mathbf{R}$ באופן הבא:

הגדרה:

נאמר שפונקציה $F : \mathbf{R} \to \mathbf{R}$ *יוצרת* את הסדרה $\lambda n.\, a_n$ אם קיים $r > 0$ כך ש:

$$\forall x. |x| < r \Rightarrow F(x) = \sum_{n=0}^{\infty} a_n x^n$$

הערות:

(א) ב- $\sum_{n=0}^{\infty} a_n x^n$ הכוונה ל- $\lim_{k \to \infty} \sum_{n=0}^{k} a_n x^n$. הסכום ("טור") האינסופי מוגדר אפוא אם"ם הסדרה $\lambda k. \sum_{n=0}^{k} a_n x^n$ מתכנסת. במקרה זה נאמר שה"טור" *מתכנס*. הגדרה מדויקת זו של $\sum_{n=0}^{\infty} a_n x^n$ *לא* תהיה אבל חשובה לענייננו! שימושית הרבה יותר תהיה עבורנו האינטואיציה של ראיית טור זה כמעין סכום אינסופי, *הדומה בהתנהגותו לפולינום*:

$$F(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_n x^n + \ldots$$

---

* יש לשים לב שסדרות אצלנו מתחילות ב- a_0 ולא ב- a_1.

גם לגודלו של r לא תהייה לגבינו כל חשיבות. חשוב רק שהוא קיים וגדול מאפס.

(ב) יחס ה"יצירה" הוא יחס בין $\mathbf{R} \to \mathbf{R}$ לבין $\mathbf{N} \to \mathbf{R}$. יחס זה הוא חד-חד-ערכי במובן הבא: ברור, שלכל סדרה ב- $\mathbf{N} \to \mathbf{R}$ יש לכל היותר פונקציה רציפה אחת שיוצרת אותה,* ומיד נראה גם שכל פונקציה ב- $\mathbf{R} \to \mathbf{R}$ יוצרת לכל היותר סדרה אחת.

(ג) מה שנזדקק לו מחשבון דיפרנציאלי ואינטגרלי הוא רק מספר עובדות יסודיות על *יחס* ה"יצירה". המובן המדויק של מושג היצירה בעצם אינו כל כך חשוב לנו. חשיבות תהיה כאן רק לאותן עובדות יסודיות אודותיו, המרוכזות להלן בטבלאות ד.4 ו- ד.5. נדגיש עם זאת שוב, שכדאי בהחלט להיעזר ב*אינטואיציה*, לפיה הסדרה $\langle a_0, a_1, a_2, \ldots\rangle$ נוצרת על-ידי הפונקציה

$$\lambda x.\ a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots$$

דוגמאות

(1) בבית-ספר תיכון לומדים על סכום טור הנדסי אינסופי מתכנס.
לומדים שם שאם $|q| < 1$ אז $1 + q + q^2 + q^3 + \ldots = \frac{1}{1-q}$.
ממילא אם $|x| < 1$ אז $1+x+x^2+x^3+\ldots=\frac{1}{1-x}$.

במונחים של פרק זה פירוש הדבר ש- $\lambda x.\frac{1}{1-x}$ יוצרת את הסדרה $\langle 1,1,1,1,\ldots\rangle$, או ליתר דיוק: את הסדרה $\lambda n.\ 1$ (ערך r במקרה זה הוא 1, כיוון שהשוויון למעלה נכון עבור $|x| < 1$. עובדה זו, כאמור, אינה רלוונטית לענייננו).

(2) משפט הבינום המוכלל קובע, שהפונקציה $\lambda x.\ (1+x)^\alpha$ יוצרת את הסדרה $\lambda n.\binom{\alpha}{n}$ (כלומר: את הסדרה $\left\langle\binom{\alpha}{0},\binom{\alpha}{1},\binom{\alpha}{2},\ldots\right\rangle$). זה נכון לכל $\alpha \in \mathbf{R}$.

---

* ליתר דיוק, אם F_1 ו- F_2 יוצרות אותה סדרה, אז יש $\rho > 0$ כך ש- F_1 ו- F_2 מוגדרות שתיהן ב- $(-\rho, \rho)$ ו- $F_1(x) = F_2(x)$ לכל x בקטע זה. ה"יחידות" של הפונקציה הנוצרת, עליה מדובר כאן, היא לכן במובן קצת שונה מהרגיל.

משפט מרכזי על הקשר בין פונקציות וסדרות הוא המשפט הבא:

משפט

אם פונקציה F יוצרת את הסדרה $\lambda n.\, a_n$, אז

$$\boxed{a_n = \frac{F^{(n)}(0)}{n!}}$$

את המשפט הזה מוכיחים בחדו"א.

מסקנות:

(1) כל פונקציה (חלקית) מ- $\mathbf{R}$ ל- $\mathbf{R}$ יוצרת *לכל היותר* סדרה אחת.

(2) תנאי *הכרחי* לכך ש- F תיצור סדרה כלשהי הוא, שהיא תהיה מוגדרת ורציפה ב- 0, ויהיו לה שם כל הנגזרות (מכל סדר).

הערה:

אומנם זהו תנאי הכרחי בלבד, אך בכל המקרים הנורמליים (הצצים בשימושים) זהו גם תנאי מספיק.

דוגמה:

אם $F = \lambda x.\, e^x$, אז $F' = \lambda x.\, e^x$ ו- $F^{(n)} = \lambda x.\, e^x$ לכל n . לכן $F^{(n)}(0) = e^0 = 1$ לכל n . מהמשפט האחרון נובע לכן, שהסדרה היחידה שפונקציה זו יכולה ליצור היא $\lambda n.\, \frac{1}{n!}$. למעשה, היא אכן יוצרת אותה, וההתכנסות אפילו מתקיימת *לכל* x :

$$\forall x.\, e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

טבלה מס' ד.4 מסכמת את כל התכונות החשובות של יחס היצירה שנזדקק להן. חלק מהן נובעות מתכונות אחרות ברשימה. את האחרות מוכיחים במדויק בחדו"א, ואנו נשאיר הוכחות מדויקות אלה לקורס המתאים. פה ניתן במקום זאת *הסברים* אינטואיטיביים מדוע יש *לצפות* לנכונות התכונות.*

* ההוכחות המדויקות בחדו"א מבוססות למעשה על אינטואיציות אלה, והן שוות למה שנעשה פה "עד כדי ε".

טבלה ד.4: פונקציות יוצרות: התכונות היסודיות

נניח: F יוצרת את $\lambda n.\, a_n$.

G יוצרת את $\lambda n.\, b_n$.

אז:

(1)	$\lambda x.\ \alpha F(x) + \beta G(x)$	יוצרת את	$\lambda n.\ \alpha\, a_n + \beta\, b_n$
(2)	$\lambda x.\ x^m.\ F(x)$	יוצרת את	$\lambda n.\ \begin{cases} 0 & n < m \\ a_{n-m} & n \geq m \end{cases}$
(3)	$\lambda x. \dfrac{F(x) - a_0 - a_1 x - \ldots - a_{m-1}x^{m-1}}{x^m}$	יוצרת את	$\lambda n.\ a_{n+m}$
(4)	$\lambda x.\ F(cx)$	יוצרת את	$\lambda n.\ c^n a_n$
(5)	$\lambda x.\ F(x^m)$	יוצרת את	$\lambda n. \begin{cases} a_{\frac{n}{m}} & m \backslash n \\ 0 & \text{אחרת} \end{cases}$
(6)	$\lambda x.\ F(x) \cdot G(x)$	יוצרת את	$\lambda n\ .\ \sum_{k=0}^{n} a_k b_{n-k}$
(7)	$\lambda x. \dfrac{F(x)}{1-x}$	יוצרת את	$\lambda n\ .\ \sum_{k=0}^{n} a_k$
(8)	F'	יוצרת את	$\lambda n.\ (n+1)a_{n+1}$
(9)	$\lambda x.\ x \cdot F'(x)$	יוצרת את	$\lambda n.\ na_n$
(10)	$\lambda x. \int_0^x F(t)dt$	יוצרת את	$\lambda n\ .\ \begin{cases} 0 & n = 0 \\ \dfrac{a_{n-1}}{n} & n > 0 \end{cases}$

נעבור עתה על התכונות ברשימה, ונסביר למה הן אמורות להיות נכונות. נסתמך על כך שהמשמעות האינטואיטיבית של ההנחות שבראש הטבלה היא:

$$F(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots$$

$$G(x) = b_0 + b_1x + b_2x^2 + \ldots + b_nx^n + \ldots$$

(1) אינטואיטיבית,

$$\alpha F(x) = \alpha a_0 + \alpha a_1x + \alpha a_2x^2 + \ldots + \alpha a_nx^n + \ldots$$

$$\beta G(x) = \beta b_0 + \beta b_1x + \beta b_2x^2 + \ldots + \beta b_nx^n + \ldots$$

ולכן

$$\alpha F(x) + \beta G(x) = (\alpha a_0+\beta b_0) + (\alpha a_1+\beta b_1)x + (\alpha a_2+\beta b_2)x^2 + \ldots + (\alpha a_n+\beta b_n)x^n + \ldots$$

ולכן ברור ש- $\lambda x.\ \alpha F(x) + \beta F(x)$ יוצרת את $\lambda n.\ \alpha a_n + \beta b_n$.

צורה יותר קונקרטית של הטיעון הנ"ל היא:

$$\lambda x.\ \alpha F(x) + \beta G(x) = \lambda x. \alpha \sum_{n=0}^{\infty} a_n x^n + \beta \sum_{n=0}^{\infty} b_n x^n = \lambda x. \sum_{n=0}^{\infty} (\alpha a_n + \beta b_n) x^n$$

ולכן נקבל את (1) (על-פי המקדם של x^n במה שיצא).

צורת שיקול זו מניחה, שעל סכומים אינסופיים חלים אותם חוקים, שחלים על סכומים סופיים. בתנאים מסוימים (החלים כאן) זה אכן נכון.

(2) $$\lambda x.x^m F(x) = \lambda x.x^m \sum_{n=0}^{\infty} a_n x^n = \lambda x. \sum_{n=0}^{\infty} a_n x^{n+m} = \lambda x. \sum_{n=m}^{\infty} a_{n-m} x^n$$

חשוב מאוד להבין את השלב האחרון בחישוב זה. מה שאנו עושים בו הוא "הזזה" של האינדקסים, כמו שהדבר נעשה בסכומים סופיים: כיוון ה"הזזה" בגבולות של ה- Σ הפוך לזו הנעשית בפנים הביטוי. (הסבר פורמלי: מגדירים $k = n + m$. אז $n = k - m$. לכן כש- $n = 0$ אז $k = m$, וכש- "$n = \infty$" גם "$k = \infty$". מכאן אנו מקבלים ש- $\sum_{k=m}^{\infty} a_{k-m} x^k = \sum_{n=0}^{\infty} a_n x^{n+m}$.)

עתה ב- $\sum_{n=m}^{\infty} a_{n-m}x^n$ המקדם של x^n הוא a_{n-m} אם $n \geq m$. אם $n < m$, אז "אין" x^n, כלומר: המקדם שלו הוא 0. מכאן (2).

דוגמה:

$$\lambda x. x^2 e^x = \lambda x. x^2\left(\frac{1}{0!} + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots + \frac{x^{n-2}}{(n-2)!} + \frac{x^{n-1}}{(n-1)!} + \frac{x^n}{n!} + \ldots\right)$$

$$= \lambda x. \left(\frac{x^2}{0!} + \frac{x^3}{1!} + \frac{x^4}{2!} + \frac{x^5}{3!} + \ldots + \frac{x^n}{(n-2)!} + \frac{x^{n+1}}{(n-1)!} + \ldots\right)$$

ואנו רואים ש- $a_0 = a_1 = 0$, ושהמקדם של x^n הוא $\frac{1}{(n-2)!}$ כש- $n \geq 2$.

$$\lambda x. \frac{F(x) - a_0 - a_1 x - \ldots - a_{m-1}x^{m-1}}{x^m} = \lambda x. \frac{\sum_{n=0}^{\infty} a_n x^n - \sum_{n=0}^{m-1} a_n x^n}{x^m} = \lambda x. \frac{\sum_{n=m}^{\infty} a_n x^n}{x^m} \qquad (3)$$

$$= \lambda x. \frac{x^m \cdot \sum_{n=m}^{\infty} a_n x^{n-m}}{x^m} = \lambda x. \sum_{n=m}^{\infty} a_n x^{n-m} = \lambda x. \sum_{n=m-m}^{\infty} a_{n+m} x^{(n+m)-m} = \lambda x. \sum_{n=0}^{\infty} a_{n+m} x^n$$

דוגמה:

$$\lambda x. \frac{e^x - 1}{x} = \lambda x. \frac{(1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots) - 1}{x} = \lambda x. \frac{\frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots}{x} =$$

$$= \lambda x. \frac{1}{1!} + \frac{x}{2!} + \frac{x^2}{3!} + \ldots$$

ברור שהמקדם של x^n הוא $\frac{1}{(n+1)!}$. לכן $\lambda x. \frac{e^x - 1}{x}$ יוצרת את $\lambda n. \frac{1}{(n+1)!}$.

$$\lambda x. F(cx) = \lambda x. \sum_{n=0}^{\infty} a_n (cx)^n = \lambda x. \sum_{n=0}^{\infty} a_n c^n x^n \qquad (4)$$

ולכן $\lambda x. F(cx)$ יוצרת את $\lambda n. c^n a_n$.

דוגמה:

$\lambda x.\, e^x$ יוצרת את $\lambda n.\frac{1}{n!}$. לכן לכל $a \in \mathbf{R}$, $\lambda x.\, e^{ax}$ יוצרת את $\lambda n.\frac{a^n}{n!}$.

(5) $$\lambda x.\, F(x^m) = \sum_{n=0}^{\infty} a_n (x^m)^n = \sum_{n=0}^{\infty} a_n x^{m\cdot n}$$

אנו רואים, שאם החזקה של x אינה מתחלקת ב- m , אז המקדם שלה הוא 0.
לעומת זאת, אם החזקה של x היא מהצורה $m \cdot k$, אז המקדם שלה הוא a_k ,
כלומר $a_{\frac{m\cdot k}{m}}$.

דוגמה:

$$\lambda x.\, e^{x^3} = \lambda x.\, 1 + \frac{x^3}{1!} + \frac{x^6}{2!} + \frac{x^9}{3!} + \frac{x^{12}}{4!} + \ldots$$

ואנו רואים, שהמקדמים של x^{10} ו- x^{11} (למשל) הם 0, בעוד שהמקדם של x^{12}
הוא $\frac{1}{(\frac{12}{3})!}$.

(6) $$\lambda x.\, F(x)\cdot G(x) = \lambda x.\, (a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots)\cdot(b_0 + b_1x + b_2x^2 + b_3x^3 + \ldots)$$

אם "נפתח את הסוגריים" ונכנס, כמו שעושים לסכומים סופיים, נקבל:

$$\lambda x.\, a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + (a_0b_3 + a_1b_2 + a_2b_1 + a_3b_0)x^3 + \ldots$$

ולכן, אינטואיטיבית, המקדם של x^n הוא $\sum_{k=0}^{n} a_k b_{n-k}$.

דוגמה:

ניקח $F = \lambda x.\, e^{3x}$, $G = \lambda x.\, e^{2x}$. אז $F \cdot G = \lambda x.\, e^{3x} \cdot e^{2x} = \lambda x.\, e^{5x}$.

עתה, במקרה זה $a_n = \frac{3^n}{n!}$, $b_n = \frac{2^n}{n!}$. חישוב הנעזר במשפט הבינום נותן לכן:

$$\sum_{k=0}^{n} a_k b_{n-k} = \sum_{k=0}^{n} \frac{3^k}{k!} \cdot \frac{2^{n-k}}{(n-k)!} = \frac{1}{n!}\sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \cdot 2^{n-k} \cdot 3^k =$$

$$= \frac{1}{n!}\sum_{k=0}^{n} \binom{n}{k} \cdot 2^{n-k} \cdot 3^k = \frac{(2+3)^n}{n!} = \frac{5^n}{n!}$$

ואכן $\lambda x.\, e^{5x}$ יוצרת את $\lambda n.\frac{5^n}{n!}$ (כמו שהוכחנו באופן כללי למעלה).

(7) זהו מקרה פרטי חשוב במיוחד של (6), בו אנו לוקחים $G = \lambda x. \frac{1}{1-x}$. במקרה זה $b_n = 1$ לכל n, ו- (7) מתקבל.

בהמשך נראה כיצד משתמשים ב- (7) לחישוב נוסחאות של סכומים.

(8) $$F = \lambda x.\, a_0 + a_1x + a_2x^2 + a_3x^3 + \ldots + a_nx^n + a_{n+1}x^{n+1} + \ldots$$

ואם נגזור לפי נגזרת של סכום, נקבל:

$$F' = \lambda x.\, a_1 + 2a_2x + 3a_3x^2 + \ldots + na_nx^{n-1} + (n+1)a_{n+1}x^n \ldots$$

ולכן, אינטואיטיבית, F' יוצרת את $\lambda n.\,(n+1)a_{n+1}$.

דוגמה:

$\lambda x.\frac{1}{1-x} = \lambda x.\sum_{n=0}^{\infty} x^n$. אם נגזור נקבל:

$$\lambda x.\frac{1}{(1-x)^2} = \lambda x\sum_{n=1}^{\infty} nx^{n-1} = \lambda x\sum_{n=0}^{\infty}(n+1)x^n$$

מכאן ש- $\lambda x.\frac{1}{(1-x)^2}$ יוצרת את הסדרה: $\langle 1, 2, 3, 4, \ldots\rangle$.

(נשים לב אבל, ש- 1 נמצא כאן בְמקום האפס, לא בְמקום ה"אחד"!).

הערה:

עובדה אחרונה זו נובעת גם מנוסחת הבינום הכללית, והיא מקרה פרטי של נוסחה (7) בטבלה ד.5.

(9) נובע מיידית מ- (8) ומ- (2) ($m = 1$).

דוגמה:

לפי הדוגמה הקודמת ו- (8), $\lambda x.\dfrac{x}{(1-x)^2}$ יוצרת את $\lambda n.\, n$, כלומר את $\langle 0, 1, 2, 3, 4, \ldots\rangle$.

(10) השיקול האינטואיטיבי כאן דומה לזה של (8), ומושאר לקוראים.

דוגמה:

$$\lambda x.\frac{1}{1-x} = \lambda x.\ 1 + x + x^2 + x^3 + \ldots$$

אינטגרציה של שני הצדדים מ- 0 ועד x תיתן:

$$\lambda x. -\ln(1-x) = \lambda x.\ x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \ldots$$

כלומר, $\lambda x. -\ln(1-x)$ יוצרת את: $\lambda n.$ if $n = 0$ then 0 else $1/n$, כלומר את: $\left\langle 0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots\right\rangle$.

טבלה ד.5 כוללת רשימה של פונקציות יוצרות חשובות ואת הסדרות שהן יוצרות. מתוך הרשימה הזו, (2), (5), (8), (9) ו- (10) כבר הוסברו למעלה. (3) ו- (4) נובעים מ- (2) בעזרת הכלל על $\lambda x.\, F(cx)$ ((3), אגב, הוא מקרה פרטי של (4), בו $c = -1$). (1) הוא טריביאלי. נסביר אפוא עתה את (6), (7), (11) ו- (12).

הוכחת (6): מ- (5) נובע ש- $\lambda x.\,(1+x)^{-\alpha}$ יוצרת את $\lambda n.\binom{-\alpha}{n}$, דהיינו את $\lambda n.\,(-1)^n \cdot \binom{\alpha+n-1}{n}$ (לפי זהות (6) מטבלה ד.3). אם נפעיל עתה על $\lambda x.\,(1+x)^{-\alpha}$ את הכלל אודות $\lambda x.\, F(cx)$ (כש- $c = -1$), נקבל את (6).

(7) הוא מקרה פרטי של (6) עם $\alpha = m + 1$, תוך שימוש בעובדה ש- $\binom{m+n}{n} = \binom{m+n}{m}$.

טבלה ד.5: פונקציות יוצרות: דוגמאות חשובות

	פונקציה יוצרת	סדרה נוצרת	כתיבה לא פורמלית
(1)	$\lambda x.\, x^m$	$\lambda n.\begin{cases}1 & n=m\\ 0 & n\neq m\end{cases}$	$\langle 0, 0, \ldots, 1, 0, 0, 0, \ldots\rangle$
(2)	$\lambda x.\dfrac{1}{1-x}$	$\lambda n.\, 1$	$\langle 1, 1, 1, 1, \ldots\rangle$
(3)	$\lambda x.\dfrac{1}{1+x}$	$\lambda n.\, (-1)^n$	$\langle 1, -1, 1, -1, -1, 1, \ldots\rangle$
(4)	$\lambda x.\dfrac{1}{1-cx}$	$\lambda n.\, c^n$	$\langle 1, c, c^2, c^3, c^4, \ldots\rangle$
(5)	$\lambda x.\, (1+x)^{\alpha}$	$\lambda n.\dbinom{\alpha}{n}$	$\left\langle 1, \alpha, \dbinom{\alpha}{2}, \dbinom{\alpha}{3}, \ldots \right\rangle$
(6)	$\lambda x.\dfrac{1}{(1-x)^{\alpha}}$	$\lambda n.\dbinom{\alpha+n-1}{n}$	$\left\langle 1, \alpha, \dbinom{\alpha+1}{2}, \dbinom{\alpha+2}{3}, \ldots \right\rangle$
(7)	$\lambda x.\dfrac{1}{(1-x)^{m+1}}$	$\lambda n.\dbinom{m+n}{m}$	$\left\langle 1, \dbinom{m+1}{m}, \dbinom{m+2}{m}, \dbinom{m+3}{m}, \ldots \right\rangle$
(8)	$\lambda x.\dfrac{x}{(1-x)^2}$	$\lambda n.\, n$	$\langle 0, 1, 2, 3, 4, \ldots\rangle$
(9)	$\lambda x.\, -\ln(1-x)$	$\lambda n.\begin{cases}0 & n=0\\ 1/n & n>0\end{cases}$	$\left\langle 0, 1, \dfrac{1}{2}, \dfrac{1}{3}, \dfrac{1}{4}, \ldots \right\rangle$
(10)	$\lambda x.\, e^{ax}$	$\lambda n.\dfrac{a^n}{n!}$	$\left\langle 1, a, \dfrac{a^2}{2!}, \dfrac{a^3}{3!}, \ldots \right\rangle$
(11)	$\lambda x.\, \cosh(ax)$	$\lambda n.\begin{cases}\dfrac{a^n}{n!} & n \text{ זוגי}\\ 0 & n \text{ איזוגי}\end{cases}$	$\left\langle 1, 0, \dfrac{a^2}{2!}, 0, \dfrac{a^4}{4!}, \ldots \right\rangle$
(12)	$\lambda x.\, \sinh(ax)$	$\lambda n.\begin{cases}0 & n \text{ זוגי}\\ \dfrac{a^n}{n!} & n \text{ אי־זוגי}\end{cases}$	$\left\langle 0, a, 0, \dfrac{a^3}{3!}, 0, \dfrac{a^5}{5!}, 0 \ldots \right\rangle$

(11)-(12): cosh ("קוסינוס היפרבולי") ו- sinh ("סינוס היפרבולי") מוגדרות כך:

$$\cosh = \lambda x.\frac{e^x + e^{-x}}{2} \qquad\qquad \sinh = \lambda x.\frac{e^x - e^{-x}}{2}$$

נוסחאות (11) ו- (12) נובעות לכן מ- (10) של טבלה זו ומכללים (1) ו- (4) של טבלה ד.4.

נעבור עתה לשימושים. עיקר השימושים בקורס זה יהיו, באופן טבעי, עבור קומבינטוריקה סופית. נתחיל אבל דווקא בשימוש מתחום אחר:

דוגמה:

מצא נוסחה עבור $1^2 + 2^2 + \ldots + n^2 = \sum_{k=0}^{n} k^2 = \sum_{k=1}^{n} k^2$.

פתרון:

המפתח כאן הוא תכונה (7) של טבלה ד.4, לפיה אם הסדרה $\lambda n.\, a_n$ נוצרת על-ידי F, אז סדרת הסכומים החלקיים נוצרת על-ידי $\lambda x.\frac{F(x)}{1-x}$.

שימוש בעובדה זו לצורך חישוב $\lambda n.\sum_{i=0}^{n} a_i$ עבור סדרה $\lambda n.\, a_n$ מסוימת מתבצע בשני שלבים:

(א) מציאת פונקציה F, שיוצרת את $\lambda n.\, a_n$.

(ב) מציאת הסדרה ש- $\lambda x.\frac{F(x)}{1-x}$ יוצרת.

בדוגמה שלנו, $a_n = n^2$ לכל n. בשלב (א) עלינו למצוא לכן פונקציה יוצרת ל- $\lambda n.\, n^2$. מאין ניקח אותה? עיון בטבלה ד.4 מגלה, שעיקרון (9) שם נראה הקרוב ביותר. כיוון ש- $n^2 = n \cdot n$, הרי אם נמצא פונקציה יוצרת G עבור $\lambda n.\, n$, אזי $\lambda x.\, x \cdot G'(x)$ תהיה פונקציה יוצרת עבור $\lambda n.\, n^2$. הבעיה עתה היא, לכן, למצוא פונקציה יוצרת עבור $\lambda n.\, n$. יכולנו להפעיל אותו שיקול, ולמצוא תחילה פונקציה יוצרת עבור $\lambda n.\, 1$. ברם, כאן נוכל פשוט להסתמך על טבלה ד.5, שם נכתב במפורש (סעיף (8)), ש- $\lambda n.\, n$ נוצרת על-ידי $G = \lambda x.\frac{x}{(1-x)^2}$. מזה נקבל:

$$F = \lambda x. x \cdot G'(x) = \lambda x.\, x.\frac{1+x}{(1-x)^3} = \lambda x.\frac{x+x^2}{(1-x)^3}$$

לכן $\lambda x.\frac{x+x^2}{(1-x)^4} = \lambda x.\frac{F(x)}{1-x}$.

עתה, לפי נוסחה (7) של טבלה ד.5 $\lambda x.\frac{1}{(1-x)^4}$ יוצרת את $\lambda n.\binom{n+3}{3}$. לכן:

$$\lambda x.(x+x^2)\cdot\frac{1}{(1-x)^4}=\lambda x.(x+x^2)\cdot\sum_{n=0}^{\infty}\binom{n+3}{3}x^n=$$

$$=\lambda x.\sum_{n=0}^{\infty}\binom{n+3}{3}x^{n+1}+\sum_{n=0}^{\infty}\binom{n+3}{3}x^{n+2}=\lambda x.\sum_{n=1}^{\infty}\binom{n+2}{3}x^n+\sum_{n=2}^{\infty}\binom{n+1}{3}x^n=$$

$$=\lambda x.\sum_{n=0}^{\infty}\binom{n+2}{3}x^n+\sum_{n=0}^{\infty}\binom{n+1}{3}x^n=\lambda x.\sum_{n=0}^{\infty}\left[\binom{n+2}{3}.+\binom{n+1}{3}\right]x^n$$

מכאן שהסכום המבוקש הוא:

$$\binom{n+2}{3}+\binom{n+1}{3}=\frac{(n+2)(n+1)n}{6}+\frac{(n+1)n(n-1)}{6}=\frac{(n+1)n(n+2+n-1)}{6}=$$

$$=\frac{n(n+1)(2n+1)}{6}$$

הערות:

(1) במעבר שלפני האחרון בחישוב למעלה הסתמכנו על זה ש- $0=\binom{2}{3}=\binom{1}{3}$.

(2) בחישוב האחרון לא נעזרנו במבט ראשון בכללים פורמליים, אלא בחישובים מ"הסוג האינטואיטיבי". למעשה, מה שעשינו לא היה אלא שימוש בכללים (1) ו- (2) של טבלה ד.4. לפי (2), למשל, $\lambda x.\frac{x^2}{(1-x)^4}$ יוצרת את $\lambda n.\binom{n+1}{3}$, כיוון ש- $\lambda x.\frac{1}{(1-x)^4}$ יוצרת את $\lambda n.\binom{n+3}{3}$. שיקול דומה חל על $\lambda x.\frac{x}{(1-x)^4}$. האפשרות לחבר נובעת מכלל (1) של טבלה ד.4. אנו נמשיך, עם זאת, להסתמך על כללים אלה ואחרים באופן לא מפורש, ובפועל נעשה חישובים אינטואיטיביים בסכומים אינסופיים. את כל מה שנעשה ניתן יהיה לתרגם בקלות לשימוש בכללים של טבלה ד.4.

(3) את מה שעשינו בדוגמה האחרונה ניתן להכליל בקלות לשיטה למציאת נוסחה עבור $\sum_{n=1}^{k} n^p$, כש- p מספר טבעי כלשהו. כל שעלינו לעשות הוא למצוא פונקציה יוצרת H_p עבור $\lambda n.\, n^p$ ואז למצוא את הסדרה הנוצרת על-ידי $\lambda x.\frac{H_p(x)}{1-x}$. כמו בדוגמה, נוכל להיעזר ב- (9) מטבלה ד.4, ולקבל ש- $H_{p+1} = \lambda x.\, xH'_p(x)$. כשמתחילים ב- $H_0 = \lambda x.\frac{1}{1-x}$ אפשר לקבל כך בקלות את H_p לכל p מבוקש.

(4) בדוגמה האחרונה ראינו הדגמה של שתי הבעיות העיקריות, הקשורות בפונקציות יוצרות:

(א) בהינתן סדרה $\lambda n.\, a_n$, למצוא פונקציה אלמנטרית, שיוצרת אותה (אם קיימת).

(ב) בהינתן פונקציה אלמנטרית, למצוא איזו סדרה (אם בכלל) היא יוצרת.

העובדות בטבלאות ד.4 ו- ד.5 הן כלי העבודה העיקרי שלנו לצורך התמודדות עם שני סוגי הבעיות.

(5) בדוגמה האחרונה היינו צריכים למצוא, בשלב מסוים, את הסדרה, שיוצרת פונקציה *רציונלית*, דהיינו מנה של שני פולינומים (במקרה של הדוגמה היתה זו הפונקציה $\lambda x.\frac{x(1+x)}{(1-x)^4}$). זהו מצב החוזר הרבה בשימושים, בעיקר אלו הקומבינטוריים. למזלנו, בהנחה מסוימת זה *תמיד* אפשרי:

משפט:

אם $p(x)$ ו- $q(x)$ שני פולינומים, ו- 0 אינו שורש של q, אז $\lambda x.\frac{p(x)}{q(x)}$ יוצרת סדרה.

הוכחה מלאה של המשפט דורשת הרחבת התיאוריה למספרים מרוכבים, ואנו נוותר עליה כאן. באופן בסיסי היא דומה להוכחה, שלכל פונקציה רציונלית ניתן למצוא אינטגרל לא מסוים, שהוא פונקציה אלמנטרית. ההוכחה היא קונסטרוקטיבית, ומספקת למעשה את המתכון (אלגוריתם) הבא למציאת הסדרה המבוקשת:

מציאת הסדרה הנוצרת על-ידי $\lambda x.\dfrac{p(x)}{q(x)}$:

שלב 1:

חילוק פולינומים:

$$\frac{p(x)}{q(x)} = p_0(x) + \frac{p_1(x)}{q(x)}$$

כאשר המעלה של p_1 קטנה ממעלת q .

שלב 2:

למצוא את השורשים $x_1, \ldots, x_m$ של הפולינום q (m – המעלה של q). כלומר:
לפתור את המשוואה $q(x) = 0$.
($x_1, \ldots, x_m$ אינם בהכרח שונים זה מזה, והם יכולים להיות מספרים מרוכבים).

שלב 3:

לפרק את q לפי הנוסחה:

$$q(x) = b_0(1 - \frac{x}{x_1})(1 - \frac{x}{x_2})\ldots(1 - \frac{x}{x_m})$$

(b_0 – האיבר החופשי של q , ואנו מניחים ש- $b_0 \neq 0$).

קיבלנו פירוק של q מהצורה: $q(x) = b_0(1 - c_1x)^{m_1}(1 - c_2x)^{m_2}\ldots(1 - c_px)^{m_p}$

כאשר $\sum_{i=1}^{p} m_i = m$.

שלב 4:

מוצאים $A_1, \ldots, A_m$ כך ש-

$$\frac{p_1(x)}{q(x)} = \frac{1}{b_0}\left[\frac{A_1}{1-c_1x} + \frac{A_2}{(1-c_1x)^2} + \ldots + \frac{A_{m_1}}{(1-c_1x)^{m_1}} + \frac{A_{m_1+1}}{1-c_2x} + \ldots + \frac{A_m}{(1-c_px)^{m_p}}\right]$$

שלב 5:

מוצאים את הסדרה, שכל מחובר יוצר, ומחברים. לתוצאה מוסיפים את p_0 משלב 1.

דוגמה:

איזו סדרה נוצרת על-ידי הפונקציה $\lambda x.\dfrac{1-5x}{2x^2-7x+3}$?

שלב 1:

מיותר כאן, כי המעלה של $1-5x$ כבר קטנה מזו של $2x^2-7x+3$.

שלב 2:

פותרים: $2x^2-7x+3=0$. הפתרונות: $x_1=\dfrac{1}{2}$ $\quad x_2=3$.

שלב 3:

מפרקים את המכנה לפי הנוסחה:

$$b_m x^m + b_{m-1}x^{n-1} + \ldots + b_1 x + b_0 = b_0\left(1-\frac{x}{x_1}\right)\left(1-\frac{x}{x_2}\right)\ldots\left(1-\frac{x}{x_m}\right)$$

כאן נקבל:

$$2x^2-7x+3 = 3\left(1-\frac{x}{1/2}\right)\left(1-\frac{x}{3}\right) = 3(1-2x)\left(1-\frac{1}{3}x\right)$$

שלב 4:

מוצאים A ו- B כך ש:

$$\frac{1-5x}{(1-2x)(1-\frac{1}{3}x)} = \frac{A}{1-2x} + \frac{B}{1-\frac{1}{3}x}$$

כלומר: לכל x $\quad 1-5x = A(1-\frac{1}{3}x)+B(1-2x)$

נציב $x=\dfrac{1}{2}$. נקבל: $A=-\dfrac{9}{5}$.

נציב $x=3$. נקבל: $B=\dfrac{14}{5}$.

סיכום:

$$\lambda x.\frac{1-5x}{2x^2-7x+3} = \lambda x.\frac{1}{3}\left[\frac{-9/5}{1-2x}+\frac{14/5}{1-1/3\,x}\right]$$

$$= \lambda x.\frac{1}{15}\left[\frac{14}{1-1/3\,x}-\frac{9}{1-2x}\right]$$

שלב 5:

מה שמצאנו יוצר את הסדרה: $\lambda n.\dfrac{1}{15}\left[14\cdot\left(\dfrac{1}{3}\right)^n - 9\cdot 2^n\right]$

שימושים בקומבינטוריקה

השימושים העיקריים של פונקציות יוצרות בקומבינטוריקה מבוססים דווקא על עיקרון מספר (6) בטבלה ד.4, המטפל בהכפלת שתי פונקציות יוצרות. נביא תחילה דוגמאות בהן נשתמש בעיקרון זה באופן אינטואיטיבי. הביסוס התיאורטי המדויק יבוא אחר-כך.

א. בחירות עם הגבלות ללא חשיבות לסדר

נתחיל *בדוגמה:* כמה אפשרויות יש להרכיב סלסלה לט"ו בשבט, המורכבת מ- 44 תאנים, שקדים וצימוקים, אם חייבים להיות לפחות 2 שקדים, לא יותר מתמר אחד, ומספר הצימוקים חייב להיות זוגי?

אם נסמן את מספר השקדים ב- x_1, את מספר התמרים ב- x_2 ואת מספר הצימוקים ב- x_3, הרי מספר האפשרויות הוא כמו מספר הפתרונות של המשוואה $x_1+x_2+x_3=44$ בהגבלות הבאות:

$$x_1 \in \{n|\, n \geq 2\} = A_1 \qquad x_2 \in \{0, 1\} = A_2 \qquad x_3 \in \{2n|\, n \in \mathbf{N}\} = A_3$$

דרך אפשרית לפתור את הבעיה היא לעשות רשימה מסודרת של כל השלשות האפשריות. בשביל זה יש צורך בשיטת ארגון כלשהי. אבחנה מרכזית, שנעשתה פה (ושקל להיווכח בנכונותה), היא, שהמספר המבוקש יהיה בדיוק המקדם של x^{44} בפיתוח הפולינום: $(x^2+x^3+x^4+\ldots+x^{44})(1+x)(1+x^2+x^4+\ldots+x^{44})$.

פתיחת סוגריים בצורה השיטתית הרגילה וכינוס יתנו לכן את התשובה המבוקשת. בגישה זו הפולינומים מהווים רק כלי ארגוני ולא יותר. תועלתם כשמספר הפירות בסלסלה יהיה 4444 (ולא 44) מוטלת בספק רב. ברם, נוכל לצעוד צעד גדול קדימה אם נשים לב שאינטואיטיבית, התשובה המבוקשת עדיין תהיה המקדם של x^{44} ***גם אם לא נעצור ב-*** x^{44} ***בפולינום הראשון והשלישי במכפלה הנ"ל.*** כלומר, התשובה היא המקדם של x^{44} בפיתוח:

$$(x^2+x^3+x^4+\ldots+x^{44}+x^{45}+\ldots)(1+x)(1+x^2+x^4+\ldots+x^{44}+x^{46}+\ldots)$$

למעשה, קל לראות, *שלכל* n שנציב במקום 44 הנ"ל, התשובה לבעיה תהיה המקדם של x^n בפיתוח של

$$(x^2 + x^3 + x^4 + \ldots)(1 + x)(1 + x^2 + x^4 + \ldots)$$

אם נתייחס עתה למכפלה זו כמגדירה פונקציה, ולכל אחד מהגורמים בה נתייחס כפונקציה יוצרת של סדרת המקדמים באותו גורם, הרי עיקרון המכפלה קובע בפירוש, שהמקדם של x^n נותן אכן את התשובה המבוקשת. כלומר: המקדם יהיה בדיוק מספר הקומבינציות $\langle x_1, x_2, x_3\rangle$ כך ש- $x_1 + x_2 + x_3 = n$ וכך ש- $x_1 \in A_1$, $x_2 \in A_2$ ו- $x_3 \in A_3$.

עתה:

$$\lambda x. \sum_{n=2}^{\infty} x^n = \lambda x. \frac{x^2}{1-x} \qquad (= \lambda x.\ x^2 + x^3 + x^4 + \ldots)$$

$$\lambda x. \sum_{n=0}^{\infty} x^{2n} = \lambda x. \frac{1}{1-x^2} \qquad (= \lambda x.\ 1 + x^2 + x^4 + x^6 + \ldots)$$

לכן

$$\lambda x.\ (x^2 + x^3 + x^4 + \ldots)(1 + x)(1 + x^2 + x^4 + \ldots) = \lambda x. \frac{x^2}{1-x} \cdot (1 + x) \cdot \frac{1}{1-x^2}$$

כדי לפתור את הבעיה המקורית, עלינו למצוא איזו סדרה יוצרת פונקציה זו, ואז לראות מיהו המקדם של x^{44}! הבה נעשה זאת*:

$$\lambda x. \frac{x^2}{1-x} \cdot (1 + x) \cdot \frac{1}{1-x^2} = \lambda x. \frac{x^2(1+x)}{(1-x)(1-x)(1+x)} = \lambda x. \frac{x^2}{(1-x)^2} = \lambda x.\ x^2 \cdot \frac{1}{(1-x)^2}$$

מסקנה:

אם נסמן ב- a_k את התשובה לשאלה עבור סלסלה בת k פירות, הרי:

$$a_k = \begin{cases} 0 & k = 0 \\ k-1 & k > 0 \end{cases}$$

בפרט $a_{44} = 44 - 1 = 43$, וזוהי התשובה לבעיה.

---

* מכאן ולהבא נשתמש באופן חופשי בנוסחאות בטבלה ד.5 לצורך מציאת הסדרה, שפונקציה מסוימת יוצרת, או מציאת הפונקציה היוצרת סדרה נתונה. ברוב המקרים לא נטרח לציין באיזו נוסחה אנו משתמשים, או שבכלל אנו משתמשים בטבלה זו.

נשים לב שבבעיה זו:

$\lambda x.\ x^2 + x^3 + x^4 \ldots$ היא הפונקציה היוצרת של $\chi_{\{n|\,n\geq 2\}}$ $\left(= \lambda_k \cdot \begin{cases} 0 & k < 2 \\ 1 & k \geq 2 \end{cases}\right)$

$\lambda x.\ 1 + x$ היא הפונקציה היוצרת של $\chi_{\{0,1\}}$ $\left(= \lambda_k \cdot \begin{cases} 1 & k \in \{0,1\} \\ 0 & \text{אחרת} \end{cases}\right)$

$\lambda x.\ 1 + x^2 + x^4 + x^6 \ldots$ היא הפונקציה היוצרת של $\chi_{\mathrm{N_{even}}}$ $\left(= \lambda_k \cdot \begin{cases} 1 & \text{זוגי } k \\ 0 & \text{אי זוגי } k \end{cases}\right)$

(בכותבנו $\chi_{\mathrm{N_{even}}}$ וכו' כוונתנו, כמובן, ל- $\chi^{\mathrm{N}}_{\mathrm{N_{even}}}$).

אנו מצאנו אפוא את הסדרה הנוצרת על-ידי $\chi_{\mathrm{N_{even}}} \cdot \chi_{\{0,1\}} \cdot \chi_{\{n|n\geq 2\}}$.

הבה נכליל. הבעיה שעסקנו בה פה היא בעיה מהסוג הכללי הבא:

בחירות עם הגבלות וללא חשיבות לסדר:

בעיה:

נניח $B = \{t_1, \ldots, t_n\}$ קבוצה עם n עצמים ו- $A_1, \ldots, A_n \in P(\mathbf{N})$.
כמה אפשרויות יש לבחור k עצמים מתוך B (אולי עם חזרות, אך בלי חשיבות לסדר), אם מספר הפעמים ש- t_i נבחר חייב להיות שייך ל- A_i ?

(בדוגמה: $B =$ {צימוק, תמר, שקד} ו- $A_1 = \{n|\, n \geq 2\},\ A_2 = \{0, 1\},\ A_3 = \mathbf{N}_{\mathrm{even}}$).

ניסוח אקוויוולנטי:

אם x_i $(i = 1, \ldots, n)$ הוא מספר הפעמים ש- t_i נבחר, אז הבעיה היא: כמה פתרונות במספרים טבעיים יש למשוואה

$$x_1 + x_2 + \ldots + x_n = k$$

בהגבלות, שלכל $1 \leq i \leq n$ $x_i \in A_i$?

ניסוח אקוויוולנטי: $|\{\langle x_1, \ldots, x_n\rangle \in A_1 \times A_2 \times \ldots \times A_n|\, x_1 + x_2 + \ldots + x_n = k\}| = ?$

שיטת הפתרון (לפי אותו הסק שהפעלנו במקרה הפרטי של הדוגמה הקודמת):

אם a_k מסמן את מספר הפתרונות עבור k (ל- $A_1, \ldots, A_n$ נתונים), אז הפונקציה היוצרת של $\lambda k.\, a_k$ היא מכפלת הפונקציות היוצרות של χ_{A_i} $(i = 1, \ldots, n)$.

במלים אחרות: אם F_1 יוצרת את χ_{A_1},

F_2 יוצרת את χ_{A_2},

$\vdots$

F_n יוצרת את χ_{A_n},

אז $F_1 \cdot F_2 \cdot \ldots \cdot F_n$ יוצרת את $\lambda k.\, a_k$.

פתרון הבעיה דורש אפוא למצוא תחילה את $F_1, \ldots, F_n$ ואחר-כך את הסדרה הנוצרת על-ידי $F_1 \cdot F_2 \cdot \ldots \cdot F_n$ (בדרך נצטרך אפוא להתמודד עם שני סוגי הבעיות הטכניות, הקשורות בשימוש בפונקציות יוצרות!).

דוגמאות נוספות:

(1) $\boldsymbol{S(n, k)}$: כאן מדובר בבחירת k עצמים מתוך n ללא כל הגבלות. כלומר: $A_1 = A_2 = \ldots = A_n = \mathbf{N}$. עתה $\chi_{\mathbf{N}} = \lambda n.\, 1$, לכן $F_1 = F_2 = \ldots = F_n = \lambda x.\dfrac{1}{1-x}$.
הפונקציה היוצרת של $\lambda k.\, S(n, k)$ היא לכן $\lambda x.\dfrac{1}{(1-x)^n}$. הסדרה, שיוצרת פונקציה זו, היא (לפי (6) של טבלה ד.5):

$$\lambda k.\binom{n+k-1}{k}$$

לכן $S(n, k) = \binom{n+k-1}{k}$ לכל n ו- k.

הערה:

1. זוהי ***הוכחה חדשה*** של נוסחה זו!
2. חשוב להבין, שבדוגמה זו $a_k = S(n, k)$, ו- n משמש כאן לכן ***כפרמטר*** !

(2) $\boldsymbol{C(n, k)}$: כאן מדובר בבחירת k עצמים מתוך n לא חזרות. במלים אחרות: כל עצם יכול להיבחר פעם אחת לכל היותר. לכן $A_1 = A_2 = \ldots = A_n = \{0, 1\}$. מכאן שלכל $1 \le i \le n$ $\chi_{A_i} = \lambda x.1 + x$. הפונקציה היוצרת של $\lambda k.\, C(n, k)$ היא אפוא

$\lambda x.\,(1+x)^n$. הסדרה, שפונקציה זו יוצרת, היא $\lambda k.\binom{n}{k}$. (שוב, n הוא ***פרמטר*** כאן!). לכן $C(n,k)=\binom{n}{k}$ (וזה נכון גם ל- $k>n$!).

(3) נחזור לדוגמה בה התחלנו פרק זה. כמה אפשרויות יש לבחור 173 מספרים מתוך $\{0, 1, 2\}$, אם 0 חייב להיבחר מספר זוגי של פעמים?

כאן $n=3$, $k=173$, ו-

$A_1=\mathbf{N}_{\text{even}}$, והפונקציה היוצרת היא: $\lambda x.\,(1+x^2+x^4+\ldots)=\lambda x.\dfrac{1}{1-x^2}$.

$A_2=A_3=\mathbf{N}$, והפונקציה היוצרת היא: $\lambda x.\,(1+x+x^2+x^3+\ldots)=\lambda x.\dfrac{1}{1-x}$.

הפונקציה היוצרת את $\lambda k.\,a_k$ היא לכן: $\lambda x.\dfrac{1}{1-x^2}\cdot\dfrac{1}{(1-x)^2}$.

הבה נמצא איזו סדרה פונקציה זו יוצרת:

$$\lambda x.\frac{1}{(1-x^2)(1-x)^2}=\lambda x.\frac{1}{(1+x)(1-x)(1-x)^2}=\lambda x.\frac{1}{(1+x)(1-x)^3}$$

עלינו לחפש לכן A, B, C, D כך שמתקיימת הזהות:

$$\frac{1}{(1+x)(1-x)^3}=\frac{A}{1+x}+\frac{B}{1-x}+\frac{C}{(1-x)^2}+\frac{D}{(1-x)^3}$$

$$\Rightarrow\qquad A(1-x)^3+B(1+x)(1-x)^2+C(1+x)(1-x)+D(1+x)=1$$

נציב $x=1$. נקבל: $2D=1$

נציב $x=-1$. נקבל: $8A=1$

נציב $x=0$. נקבל: $A+B+C+D=1$

נציב $x=2$. נקבל: $-A+3B-3C+3D=1$

פתרון המשוואות נותן:

$$A=\frac{1}{8}\qquad B=\frac{1}{8}\qquad C=\frac{2}{8}\qquad D=\frac{1}{2}=\frac{4}{8}$$

לכן:

$$\lambda x.\frac{1}{(1+x)(1-x)^3}=\lambda x.\frac{1}{8}\left[\frac{1}{1+x}+\frac{1}{1-x}+\frac{2}{(1-x)^2}+\frac{4}{(1-x)^3}\right]$$

הסדרה הנוצרת היא לכן:

$$\lambda k.\frac{1}{8}\left[(-1)^k + 1 + 2(k+1) + 4\binom{k+2}{2}\right]$$

$$= \lambda k.\frac{2k^2 + 8k + 7 + (-1)^k}{8}$$

מספר האפשרויות עבור k מסוים הוא לכן $\frac{2k^2 + 8k + 7 + (-1)^k}{8}$. אם נציב $k = 173$, נקבל 7656, וזוהי התשובה המבוקשת.

הכללה: במקום משוואה מהצורה $x_1 + x_2 + \ldots + x_n = k$ עם הגבלות $A_1, \ldots, A_n$, נוכל לפתור דבר כללי יותר: משוואה מהצורה $m_1x_1 + \ldots + m_nx_n = k$ $(m_i \in \mathbf{N})$ עם הגבלות דומות.

ההצבה $m_ix_i = y_i$ תעביר אותנו כאן למשוואה מהצורה $y_1 + y_2 + \ldots + y_n = k$ עם ההגבלות $y_i \in B_i$ $(i = 1, \ldots, n)$, כאשר $B_i = \{m_ij \mid j \in A_i\}$, וזוהי בעיה מהסוג, שאנו כבר יודעים איך להתמודד עמו.

דוגמה:

כמה פתרונות במספרים טבעיים יש למשוואה $2x_1 + x_2 + x_3 = 173$?

תשובה:

ההצבה $y_1 = 2x_1$ תביא אותנו בדיוק לדוגמה הקודמת. התשובה היא לכן 7656.

(4) הבה נחזור עתה לבעיה, שפתרנו בפרק הקודם בעזרת ***עיקרון ההכלה וההפרדה***:
כמה פתרונות במספרים ***שלמים*** יש למשוואה

$$x_1 + x_2 + x_3 = 9$$

בהגבלות $1 \le x_1 \le 2$, $-2 \le x_2 \le 4$ ו- $2 \le x_3 \le 4$?

פתרון:

כמו בשיטה הקודמת, גם כאן כדאי להתחיל באותו תהליך נרמול המוביל לבעיה על מספרים טבעיים. כלומר, גם כאן נתחיל בהגדרת משתנים חדשים:
$y_1 = x_1 - 1$, $y_2 = x_2 + 2$ ו- $y_3 = x_3 - 2$, וכמו שם הבעיה צריכה להיות: מה מספר הפתרונות במספרים טבעיים של המשוואה $y_1 + y_2 + y_3 = 8$ בהגבלות:

$0 \le y_1 \le 1$, $0 \le y_2 \le 6$ ו- $0 \le y_3 \le 2$. זוהי כבר בעיה מהסוג, בו עסקנו בדוגמאות האחרונות. כאן: $A_1 = \{0,1\}$, $A_2 = \{0, 1, 2, 3, 4, 5, 6\}$ ו- $A_3 = \{0, 1, 2\}$. מכאן, שאם a_k מספר הפתרונות עבור $y_1 + y_2 + y_3 = k$ (בהגבלות הנ"ל), אז פונקציה יוצרת עבור $\lambda k.\, a_k$ היא $\lambda x.\, (1 + x)(1 + x + x^2 + \ldots + x^6)(1 + x + x^2)$. במקרה זה, במקום למצוא נוסחה כללית עבור a_k , עדיף אולי למצוא ישירות את a_8 , כלומר: את המקדם של x^8 בפיתוח המכפלה הנ"ל. בדוגמה פשוטה זו אפשר לפתוח פשוט את הסוגריים ולכנס. הבה נעשה זאת אבל בדרך, שתהיה עדיפה עם דוגמאות מסובכות יותר. המפתח הוא שימוש בנוסחת הסכום של סדרה הנדסית:

$$\lambda x.\, (1 + x)(1 + x + \ldots + x^6)(1 + x + x^2) = \lambda x.\, \frac{1 - x^2}{1 - x} \cdot \frac{1 - x^7}{1 - x} \cdot \frac{1 - x^3}{1 - x}$$

$$= \lambda x.\, (1 - x^2)(1 - x^3)(1 - x^7) \cdot \frac{1}{(1 - x)^3}$$

$$= \lambda x. (1 - x^2 - x^3 + x^5 - x^7 + \text{חזקות גדולות מ- } 8) \cdot \sum_{n=0}^{\infty} \binom{n+2}{n} x^n$$

המקדם של x^8 יהיה לכן: $3 = \binom{3}{1} - \binom{5}{3} + \binom{7}{5} - \binom{8}{6} - \binom{10}{8}$.

(למשל: את $-x^3$ שבסוגריים נצטרך להכפיל ב- x^5 מתוך הסכום האינסופי. לכן הוא יתרום $-\binom{7}{5} = \binom{5+2}{5} \cdot (-1)$ למקדם של x^8).

השוואה עם התשובה שקיבלנו לפי עיקרון ההכלה וההפרדה תראה, שקיבלנו שם אותו סכום בדיוק! זהו מצב אופייני: שימוש בשיטה זו, שעשינו כאן בפרוטרוט, ייתן אכן תמיד אותו סכום, שנותן השימוש בעיקרון ההכלה וההפרדה. עם זאת, פה מגוון האפשרויות שלנו גדול יותר. בין פתיחה כללית של הסוגריים (הדרך הכי פחות מתוחכמת) ובין הדרך של שימוש בלתי מתפשר בנוסחת הסכום של סדרה הנדסית, קיימות דרכי ביניים. כאן יכולנו, למשל, לעשות גם את החישוב הבא:

$$\lambda x.\, (1 + x)(1 + x + x^2)(1 + x + \ldots + x^6) = \lambda x.\, (1 + x) \cdot \frac{1 - x^3}{1 - x} \cdot \frac{1 - x^7}{1 - x}$$

$$= \lambda x.\, (1 + x)(1 - x^3 - x^7 + x^{10}) \cdot \frac{1}{1 - x^2}$$

$$= \lambda x.\, (1 + x - x^3 - x^4 - x^7 - x^8 + \text{חזקות גדולות מ- } 8) \cdot \sum_{n=0}^{\infty} \binom{n+1}{n} x^n$$

והמקדם של x^8 יהיה לכן: $3 = 1 - 2 - 5 - 6 - 8 + 9$.

ב. בחירות עם הגבלות ועם חשיבות לסדר

נתחיל גם כאן בדוגמה: מה מספר האפשרויות לסדר בשורה 11 כדורים כחולים, אדומים וירוקים, אם מספר הכדורים הכחולים והאדומים חייב להיות חזקה של 3 וחייב גם להיות לפחות כדור ירוק אחד?

פתרון:

נסמן את ההתפלגות האפשרית של מספר הכדורים:

ירוקים	כחולים	אדומים
9	1	1
7	3	1
7	1	3
5	3	3
1	1	9
1	9	1

לאור מה שלמדנו בעבר על סידור כדורים בשורה, נקבל לכן:

$$a_{11} = \frac{11!}{9!1!1!} + \frac{11!}{7!3!1!} + \frac{11!}{7!1!3!} + \frac{11!}{5!3!3!} + \frac{11!}{1!1!9!} + \frac{11!}{1!9!1!}$$

כדי לראות איך ניתן למצוא פתרון כללי נוח, נשים לב ש-

$$\frac{a_{11}}{11!} = \frac{1}{9!1!1!} + \frac{1}{7!3!1!} + \frac{1}{7!1!3!} + \frac{1}{5!3!3!} + \frac{1}{1!1!9!} + \frac{1}{1!9!1!}$$

נקודת המפתח היא עתה, שלפי עיקרון הכפל של פונקציות יוצרות, אגף ימין של המשוואה האחרונה הוא בדיוק המקדם של x^{11} בפיתוח של:

$$\lambda x.\left(\frac{x^1}{1!}+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\ldots\right)\left(\frac{x^1}{1!}+\frac{x^3}{3!}+\frac{x^9}{9!}+\frac{x^{27}}{27!}+\ldots\right)\left(\frac{x^1}{1!}+\frac{x^3}{3!}+\frac{x^9}{9!}+\frac{x^{27}}{27!}+\ldots\right)$$

ברור אבל, שבמקום 11 יכולנו לקחת כאן כל k אחר. כמו כן ברור, שאם נגדיר:

$$A_2 = A_3 = \{3^n \mid n \in \mathbf{N}\} \quad , \quad A_1 = \{n \mid n \geq 1\}$$

אז הפונקציה למעלה היא בדיוק מכפלת הפונקציות היוצרות של $\lambda n.\dfrac{\chi_{A_i}(n)}{n!}$.

הבה נכליל לעיקרון כללי:

בעיה:

נניח $B = \{t_1, \ldots, t_n\}$ קבוצה עם n עצמים, ו- $A_1, \ldots, A_n \in P(\mathbf{N})$. כמה אפשרויות יש לבחור k עצמים מתוך B עם חשיבות לסדר (ואולי עם חזרות), אם מספר הפעמים ש- t_i נבחר חייב להיות שייך ל- A_i?

ניסוח מדויק יותר: מה המספר של הרשימות באורך k של איברים מ- B, שבהן מספר הפעמים ש- t_i מופיע שייך ל- A_i? דהיינו:

$$|\{f \in \{1, \ldots, k\} \to B \mid \forall 1 \le i \le n.\ |f^{-1}[\{t_i\}]| \in A_i\}| = ?$$

שיטת הפתרון: אם a_k מסמן את מספר הפתרונות של הבעיה (עבור $A_n, \ldots, A_1$ נתונים), אז הפונקציה היוצרת של $\lambda k.\dfrac{a_k}{k!}$ היא מכפלת הפונקציות היוצרות* של $\lambda k.\dfrac{\chi_{A_i}(k)}{k!}$ $(i = 1, \ldots, n)$.

הסבר: נניח ש- k_i הוא מספר הפעמים ש- t_i נבחר. ברור ש- $\sum_{i=1}^{n} k_i = k$. עתה, מספר הסידורים, שה- n-יה $\langle k_1, \ldots, k_n\rangle$ מספקת, הוא $\dfrac{k!}{k_1!\,k_2!\ldots k_n!}$.

ברור לכן, שאם נסמן:

$$I_k = \left\{\langle k_1, \ldots, k_n\rangle \,\middle|\, \sum_{i=1}^{n} k_i = k\right\}$$

$$J_k = \{\langle k_1, \ldots, k_n\rangle \in I_k \mid k_1 \in A_1 \wedge k_2 \in A_2 \wedge \ldots \wedge k_n \in A_n\}$$

אז:

$$a_k = \sum_{\langle k_1,\ldots k_n\rangle \in J_k} \frac{k!}{k_1!\,k_2!\ldots k_n!} = k! \cdot \sum_{\langle k_1,\ldots k_n\rangle \in J_k} \frac{1}{k_1!\ldots k_n!}$$

---

* ממשפטים בחדו"א נובע בקלות, שכל הפונקציות היוצרות האלה אכן קיימות. בדוגמאות הספציפיות זה ממילא יהיה ברור ממה שאנו כבר יודעים.

ולכן:

$$\frac{a_k}{k!}=\sum_{\langle k_1,\ldots k_n\rangle\in J_k}\frac{1}{k_1!}\cdot\frac{1}{k_2!}\cdots\frac{1}{k_n!}=\sum_{\langle k_1,\ldots k_n\rangle\in I_k}\frac{\chi_{A_1}(k_1)}{k_1!}\cdot\frac{\chi_{A_2}(k_2)}{k_2!}\cdots\frac{\chi_{A_n}(k_n)}{k_n!}$$

מעיקרון הכפל (טבלה ד.4, (6)) נובע לכן, שאם G_i יוצרת את $\lambda x.\dfrac{\chi_{A_i}(k)}{k\,!}$ $(i=1,\ldots,n)$,

אז $G_1\cdot G_2\cdot\ldots\cdot G_n$ יוצרת את $\lambda k.\dfrac{a_k}{k!}$.

הערה:

כדי לפשט את הניסוח נהוג לקרוא לפונקציה יוצרת של הסדרה $\lambda k.\dfrac{b_k}{k!}$ בשם *פונקציה יוצרת מעריכית* של הסדרה $\lambda k.\ b_k$. בטרמינולוגיה זו העיקרון שהסברנו זה עתה קובע, שהפונקציה היוצרת מעריכית את $\lambda k.\ a_k$ מהבעיה היא מכפלת הפונקציות היוצרות מעריכית את χ_{A_i}. כדאי לציין, שמהמשפט היסודי על פונקציות יוצרות נובע, שאם F יוצרת בכלל איזו סדרה, אז היא גם פונקציה יוצרת מעריכית של $\lambda n.\ F^{(n)}(0)$, ושלה בלבד. כמו כן, אם F יוצרת את $\lambda k.\ a_k$, אז F יוצרת מעריכית את $\lambda k.\ k!a_k$.

דוגמאות

(1) מהו מספר האפשרויות לבחור k עצמים מתוך n עם חשיבות לסדר (ועם חזרות)?

פתרון:

כאן אין הגבלות, דהיינו $A_1=A_2=\ldots=A_n=\mathbf{N}$. לכן $\chi_{A_i}(k)=1$ לכל $k\in\mathbf{N}$ ולכל $1\le i\le n$. מכאן שהפונקציה היוצרת של $\lambda k.\dfrac{\chi_{A_i}(k)}{k!}$ היא כאן הפונקציה היוצרת של $\lambda k.\dfrac{1}{k!}$, דהיינו: $\lambda x.\ e^x$. מכפלת n הפונקציות היוצרות תהיה לכן $\lambda x.\ (e^x)^n=\lambda x.\ e^{nx}$ (n כאן הוא שוב *פרמטר*!).

עתה, $\lambda x.\ e^{nx}$ יוצרת את $\lambda x.\dfrac{n^k}{k!}$, ולכן היא יוצרת מעריכית את $\lambda k.\ n^k$. מכאן שהתשובה היא n^k (תשובה זו ידועה לנו כבר, כמובן).

(2) $P(n, k)$: זהו מספר האפשרויות לבחור k עצמים מתוך n עם חשיבות לסדר, כשמספר הפעמים שכל עצם יכול להיבחר הוא לכל היותר 1. מכאן, שבבעיה זו $A_1 = A_2 = \ldots = A_n = \{0, 1\}$. לכן פונקציה יוצרת עבור $\lambda k.\frac{\chi_{A_i}(k)}{k!}$ היא כאן $\lambda x.1 + \frac{x}{1!} = \lambda x.1 + x$ (לכל $1 \le i \le n$). ההכפלה של n פונקציות כאלו נותנת את $\lambda x.\ (1 + x)^n$. פונקציה זו יוצרת את $\lambda k.\binom{n}{k}$, ולכן היא יוצרת מעריכית את $\lambda k.k!\binom{n}{k}$. מכאן: $P(n,k) = k!\binom{n}{k}$, דהיינו:

$$P(n,k) = \begin{cases} 0 & k > n \\ \frac{n!}{(n-k)!} & 0 \le k \le n \end{cases}$$

הערה:

הנוסחה $P(n,k) = \binom{n}{k} \cdot k!$ אינה אלא מקרה פרטי של הזהות: $P(a, b) = C(a, b) \cdot E(b)$.

(3) הבה נפתור עתה בעזרת השיטה האחרונה בעיה, שפתרנו בעבר בעזרת עיקרון ההכלה וההפרדה: כמה מספרים בני k ספרות אפשר להרכיב מהספרות 1, 2, 3, 4 ו- 5, אם 1, 2, 3 חייבים להופיע?

כאן $n = 5$ ו-

$$A_1 = A_2 = A_3 = \{n |\ n \ge 1\} \qquad A_4 = A_5 = \mathbf{N}$$

לכן פונקציה יוצרת מעריכית של χ_{A_i} $(i = 1, 2, 3)$ היא

$$\lambda x.\left(\frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\right) = \lambda x.e^x - 1$$

בעוד פונקציה יוצרת מעריכית של χ_{A_4} ו- χ_{A_5} היא $\lambda x.\ e^x$.

מכאן שפונקציה יוצרת מעריכית של $\lambda k.\, a_k$ היא:

$$\lambda x.\,(e^x-1)^3\cdot(e^x)^2=\lambda x.\,(e^{3x}-3e^{2x}+3e^x-1)e^{2x}=$$
$$=\lambda x.\, e^{5x}-3\cdot e^{4x}+3\cdot e^{3x}-e^{2x}$$

פונקציה זו יוצרת את $\lambda k.\dfrac{5^k-3\cdot 4^k+3\cdot 3^k-2^k}{k!}$.

לכן $a_k=5^k-3\cdot 4^k+3\cdot 3^k-2^k$. (זו גם התשובה, שקיבלנו בפרק ההוא, כמובן).

(4) מהו מספר הרשימות באורך n, שאפשר להרכיב מ- $\{0, 1, 2, 3\}$, ושיש בהן מספר זוגי של אפסים?

פתרון:

כאן $n=4$. $A_0=\{2n|\, n\in \mathbf{N}\}$, $A_1=A_2=A_3=\mathbf{N}$.

פונקציה יוצרת מעריכית של χ_{A_1}, χ_{A_2} ו- χ_{A_3} היא כאן שוב $\lambda x.\, e^x$.

לעומת זאת, פונקציה יוצרת מעריכית של χ_{A_0} היא כאן:

$$\lambda x.\left(1+\frac{x^2}{2!}+\frac{x^4}{4!}+\ldots\right)=\lambda x.\cosh(x)=\lambda x.\frac{e^x+e^{-x}}{2}$$

(ראה טבלה ד.5, (11)). לכן, אם a_n מסמן את הפתרון, אז פונקציה מעריכית יוצרת עבור $\lambda n.\, a_n$ היא: $\lambda x.\dfrac{e^x+e^{-x}}{2}\cdot(e^x)^3=\lambda x.\dfrac{e^{4x}+e^{2x}}{2}$. עתה:

$$\lambda x.\frac{e^{4x}+e^{2x}}{2}=\lambda x.\frac{1}{2}\left(\sum_{n=0}^{\infty}\frac{4^n}{n!}x^n+\sum_{n=0}^{\infty}\frac{2^n}{n!}x^n\right)=\lambda x.\sum_{n=0}^{\infty}\frac{4^n+2^n}{2}\cdot\frac{x^n}{n!}$$

לכן $a_n=\dfrac{4^n+2^n}{2}$ לכל n.

הערה:

יכולנו כמובן לקרוא ישירות תשובה זו מ- $\lambda x.\frac{1}{2}e^{4x}+\frac{1}{2}e^{2x}$, כמו שעשינו בדוגמה הקודמת.

ד.5. נוסחאות נסיגה

הקדמה

בפרק זה נציג שיטת טיפול בסדרות, שיש לה חשיבות עצומה בתכנות. כאן ניישם אותה בעיקר לצורך פתרון בעיות קומבינטוריות.
נפתח, כרגיל, בדוגמה:

בעיה

כמה מחרוזות (או רשימות) של "0" ו- "1" שאורכן 30 ניתן להרכיב, בתנאי שלא יהיו בהן שני אפסים רצופים?

ניתוח:
עקרונית, זוהי בעיה מהסוג של בחירת k עצמים מתוך n (כאשר כאן $n = 2$) עם חשיבות לסדר ועם הגבלות. לרוע המזל, ההגבלה כאן היא מסוג שונה מאלו, בהן טיפלנו בפרק הקודם. כך אין כאן הגבלה על מספר הפעמים ש- "0" יכול להופיע: הוא *יכול* להופיע, למשל, בדיוק פעמיים, אך לא פעמיים ברציפות! השיטות של הפרק הקודם לא יועילו לכן כאן.

כיצד בכל זאת נתמודד עם הבעיה? ראשית, ננסה *להכליל* אותה (כבר ראינו יעילות גישה זו בעבר!). נסמן אפוא ב- a_n את מספר הפתרונות עבור n כלשהו, לאו דווקא 30. שנית, ננסה לעשות ניתוח למקרים. ובכן, אם $n > 0$ ונשים בראש המחרוזת "1", אז אחר-כך נצטרך לשים מחרוזת באורך $n - 1$ מאותו סוג בדיוק. לעומת זאת, אם נשים בראש "0", אז אחריו נהיה חייבים (אם $n > 1$) לשים 1, ואחר-כך נהיה חופשיים לשים מחרוזת כלשהי מאותו סוג שאורכה $n - 2$. מספר האפשרויות במקרה הראשון הוא a_{n-1} ובשני הינו a_{n-2} (בהנחה ש- $n \geq 2$). אנו מקבלים אפוא, שאם $n \geq 2$, אז מתקיים הקשר:

$$\boxed{a_n = a_{n-1} + a_{n-2}}$$

אנו נוכל למצוא אפוא את a_{30} בתנאי שנמצא תחילה את a_{29} ו- a_{28}, ואותם נוכל למצוא, בתנאי שנמצא את אלו שקודמים להם, וכן הלאה. אם נמשיך כך לסגת, נגיע

לבסוף לצורך למצוא את a_0 ו- a_1. אותם, למרבה השמחה, קל למצוא ישירות. $a_0 = 1$ (המחרוזת הריקה עומדת בדרישות!) ו- $a_1 = 2$. מנקודת מוצא זו נוכל לצאת בדרך חזרה: $a_2 = 2 + 1 = 3$, $a_3 = 3 + 2 = 5$, $a_4 = 5 + 3 = 8$, וכן הלאה. תוך פרק זמן של מספר דקות נגיע (אפילו ידנית!) ל- a_{30}. התהליך אבל קל במיוחד לתכנות, ובעזרת תוכנות מחשב נוכל בקלות ובמהירות למצוא את a_n גם ל- n-ים גדולים.

אם נסכם: מהותית, פתרנו את הבעיה ל- n כלשהו על-ידי גילוי התנאים:

$$\begin{cases} a_n = a_{n-1} + a_{n-2} \\ a_0 = 1 \\ a_1 = 2 \end{cases}$$

תנאים אלו מאפשרים מציאת הערך המספרי של a_{30} ביעילות ובקלות. הסדרה $\lambda n.\, a_n$ המתקבלת: $\langle 1, 2, 3, 5, 8, 13, 21, 34, \ldots \rangle$ היא, אגב, סדרה מפורסמת מאוד, הידועה בשם "סדרת פיבונאצ'י".*

שתי נקודות מהותיות יש להעיר בקשר למה שעשינו:

(א) תיאור הסדרה בצורה שעשינו למעלה הוא, באופן עקרוני ו*מעשי*, פתרון לא רע בכלל של הבעיה המקורית, כיוון שהוא מאפשר חישוב יעיל של ערכי הסדרה – וזה עיקר מטרתנו.

(ב) עם זאת, זוהי בהחלט צורת תיאור/ייצוג חדשה של פונקציות, שהתחום שלהן הוא **N**, בה לא השתמשנו עד כה. עד עתה תיארנו את כל הפונקציות על-ידי תיאור *מפורש*, כלומר: על-ידי ביטוי-λ. חשוב להבין אבל, שאת הכתוב למעלה *אי-אפשר* לכתוב כביטוי λ! בפרט, אם נסמן $f = \lambda n.\, a_n$, אז ה"הגדרה" הבאה של f *אינה* הגדרה על-ידי ביטוי-λ:

$$(**)\quad f = \lambda n \in \mathbf{N}\,.\ \text{If } n = 0 \text{ then } 1 \text{ else if } n = 1 \text{ then } 2 \text{ else } f(n-1) + f(n-2)$$

הסיבה לכך היא, שהביטוי ה"מוגדר", f, מופיע בהגדרה שלו עצמו. באופן תקין, מחשב שהיה בא ליישם הגדרה זו בחישוב $f(10)$, למשל, היה מגיע לצורך בחישוב $f(9) + f(8)$, ואז הוא היה מפעיל לצורך זה את הערך של f, שהיה לו בזיכרון

* למעשה, סדרת פיבונאצ'י היא לא בדיוק זו אלא $\langle 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots \rangle$. כלומר, אם $\lambda n.\, F_n$ היא סדרת פיבונאצ'י ה"רשמית", אז הסדרה שמצאנו היא בדיוק $\lambda n.\, F_{n+2}$, ולא $\lambda n.\, F_n$.

לפני הכנסת ה"הגדרה" הזו! (אם אין ערך כזה, אז מה שהוא היה עושה תלוי באימפלמנטציה: הוא יכול לעצור עם הודעת "error" או להפעיל ברירת-מחדל כמו $f = \lambda n.\ 0$). המחשב *לא* היה מקשר בין ה- "f"-ים המופיעים באגף ימין ובין ה- "f" המופיע בצד שמאל של השוויון (באופן דומה במקצת לכך, שבביצוע פקודת ההשמה $x := x + 1$, המחשב משתמש בערך של x *לפני* ביצוע פקודת ההשמה לצורך חישוב ערכו *אחרי* ביצוע הפקודה).
כדי שהנקודה תובן טוב יותר, הבה ננסה לעשות משהו אנלוגי עם צורת ההגדרה המפורשת, המקובלת עלינו עבור קבוצות. ובכן, ה"הגדרה" הבאה גם היא אינה נחשבת כשימוש תקין ב- $\{...|\_\_\}$ לצורך הגדרת קבוצה A :

$$(***) \quad A = \{n \in \mathbf{N} \mid n = 0 \vee (n \geq 2 \wedge (n-2)) \in A\}$$

העובדה שתיאורים של סדרה כמו ב- (*) למעלה אינם ניתנים לתרגום לביטויי-λ, אין פירושה, שאי-אפשר להשתמש בהם לצורך הגדרת פונקציות. תיאורים כאלה ידועים בשם *הגדרות רקורסיביות* של פונקציות, והם נשמת-אפן של שפות תכנות פונקציונליות (כמו SCHEME). האימפלמנטציה שלהן פשוט לא נעשית דרך ביטויי-λ, אלא דרך אופרטורים מסובכים יותר, שלא כאן המקום והזמן לתאר אותם. (גם דברים כמו (***) מתורגמים לעתים להגדרות רקורסיביות של *קבוצות*, אך גם לכך לא ניכנס כרגע).

אם אי-אפשר להסתכל על (**) ו- (***) בתור הגדרות, אפשר בהחלט להסתכל עליהם בתור *משוואות*. כך, אם נראה ב- (***) משוואה, ש- A הוא הנעלם שלה, הרי למשוואה זו יש פתרון יחיד: $A = \mathbf{N}_{\text{even}}$ (נא לשים לב: מדובר במשוואה, שהנעלם בה הוא *קבוצה*, לא מספר!). בדומה גם את (**) אפשר לראות *כמשוואה*, שסדרת פיבונאצ'י היא הפתרון היחיד שלה (כאן הנעלם הוא *פונקציה* מ- $\mathbf{N}$ ל- $\mathbf{N}$!). למעשה, אנו אומרים, שפונקציה מוגדרת בצורה רקורסיבית, אם היא מוגדרת על-ידי משוואה כמו ב- (**), שיש לה פתרון יחיד, ופתרון זה ניתן לחישוב מתוך המשוואה (כשמדובר בפונקציה f "ניתנות לחישוב" פירושה האפשרות לחשב את ערכי $f(x)$ לכל x בתחום הפונקציה). לא תמיד, אומנם, כשרושמים משוואה בסגנון (**) ברור אם זה המצב או לא. לכן, לא לכל תכנית שכותבים ב- SCHEME אפשר תמיד להכריע, אם היא מגדירה פונקציה מ- $\mathbf{N}$ ל- $\mathbf{N}$, או לא. ברם, במקרים רבים זה אפשרי ואפילו קל, ומיד נציין מקרה אחד כזה, שהוא חשוב במיוחד!

נתחיל אבל בלימוד של סימונים מקובלים בנושא. את (**) אפשר לרשום גם בצורה:

$$X = \lambda n \in \mathbf{N}.\begin{cases} 1 & n = 0 \\ 2 & n = 1 \\ X(n-1) + X(n-2) & n \geq 2 \end{cases}$$

(צורה זו מחזקת את האינטואיציה, שמדובר למעשה ב*משוואה*).

במקום זאת מקובל יותר לכתוב:

$$\begin{cases} n \geq 2 \Rightarrow & X(n) = X(n-2) + X(n-1) \\ & X(0) = 1 \\ & X(1) = 2 \end{cases}$$

(צורה זו הופכת למעשה את המערכת המקורית ל*מערכת* של משוואות, שאחת מהן היא בבירור העיקרית – והשאר פשוטות ביותר. עוד נשוב לכך).

עתה, ברוב המקרים מקובל לכתוב a_n (או b_n , או c_n ...) במקום $X(n)$, וכותבים, כמו ב- (*) למעלה:

$$\begin{cases} a_n = a_{n-2} + a_{n-1} & (n \geq 2) \\ a_0 = 1 \\ a_1 = 2 \end{cases}$$

בצורה זו מְטשטשת במקצת המשוואה, ומודגשת יותר פתירותה, דהיינו: האפשרות לחשב $\lambda n.\ a_n$ מתוך המשוואות הללו.

חשוב לזכור שכל הצורות הללו הן שקולות!*

העובדה, שלמשוואה (או מערכת משוואות) כנ"ל יש פתרון יחיד, היא מקרה פרטי של המשפט הבא:

* צורה שקולה נוספת היא לכתוב $a_{n+2} = a_{n+1} + a_{n}$ במקום "$a_n = a_{n-1} + a_{n-2}$ עבור $n \geq 2$".

משפט:

תהי A קבוצה. יהיו $c_0, \ldots, c_{k-1}$ איברים של A, ותהי F פונקציה $(k+1)$-מקומית מ- $\mathbf{N} \times A \times A \times \ldots \times A$ ל- A. אז ***קיימת*** פונקציה X ***יחידה*** מ- $\mathbf{N}$ ל- A כך ש-

(א) $X(n) = c_n$ ל- $n = 0, \ldots, k-1$

(ב) $X(n) = F(n, X(n-1), \ldots, X(n-k))$ ל- $n \geq k$.

בדוגמה שלנו, למשל, $A = \mathbf{N}$ (זה יהיה המצב ברוב הדוגמאות), $c_0 = 1$, $c_1 = 2$ ו- $F = \lambda n \in \mathbf{N}, x \in \mathbf{N}, y \in \mathbf{N}.\ x + y$.

הגדרה

משוואה כללית מהצורה $X(n) = F(n, X(n-1), \ldots, X(n-k))$ נקראת ***נוסחת נסיגה***, או "משוואת נסיגה", כי היא מאפשרת חישוב ערכי $X(n)$ ל- $n \geq k$ על-ידי "נסיגה" לערכים קודמים של X. המספר k נקרא ***הסדר*** של נוסחת הנסיגה (או "משוואת הנסיגה"). תנאים מהצורה $X(n) = c_n$ (עבור n ו- c_n ***קונקרטיים***) נקראים ***תנאי התחלה***.

הערה:
למעשה, משוואת נסיגה היא משוואה מהצורה:

$$X / \{n \in \mathbf{N} \mid n \geq k\} = \lambda n \in \{n \in \mathbf{N} \mid n \geq k\}.\ F(n, X(n-1), \ldots, X(n-k))$$

בה הפונקציה X מ- $\mathbf{N}$ ל- A היא ***הנעלם***.

דוגמאות נוספות

(1) ניקח $A = \mathbf{N}$, $c_0 = 1$, $F = \lambda n \in \mathbf{N}, x \in \mathbf{N}.\ nx$. המשפט אומר שיש $X: \mathbf{N} \to \mathbf{N}$ יחידה כך ש-:

$$\begin{cases} X(0) = 1 \\ X(n) = n \cdot X(n-1) \quad (n \geq 1) \end{cases}$$

הפונקציה X היא כמובן $\lambda n.\ n!$. נשים לב, שכאן החישוב בעזרת נוסחת הנסיגה הוא הדרך היחידה, המוכרת לנו לחישוב פונקציה זו. כך לצורך חישוב $7!$ חייבים אנו לחשב תחילה את $6!$, ולהכפיל את התוצאה ב- 7.

(2) ניקח $A = R$, $c_0 = 3.5$, $F = \lambda n \in N,\, x \in \mathbf{R}^+ .\dfrac{x + 3/x}{2}$. לפי המשפט קיימת סדרה יחידה $\lambda n.\, a_n$, כך ש-

$$\begin{cases} a_0 = 3.5 \\ a_n = \dfrac{a_{n-1} + 3/a_{n-1}}{2} \qquad (n \geq 1) \end{cases}$$

המיוחד בסדרה זו הוא ש- $\lim_{n\to\infty} a_n = \sqrt{3}$ (כמו שמוכיחים בחדו"א).

הוכחת המשפט:

הוכחת היחידות היא קלה: אם G_1 ו- G_2 הם שני פתרונות, אז מוכיחים באינדוקציה על n ש- $G_1(n) = G_2(n)$ לכל n. מכאן ש- $G_1 = G_2$.

ניסוח *מדויק* של הוכחת הקיום מסובך יותר. נסתפק פה ב*אינטואיציה* הברורה, שנוסחת הנסיגה יחד עם תנאי ההתחלה מאפשרים את חישוב $G(n)$ לכל n, ומכאן ש- G קיימת.

השימוש בקומבינטוריקה

לעתים קרובות, כשיש לנו בעיה קומבינטורית של מציאת מספר מסוים a_n (התלוי בפרמטר n), קשה לנו למצוא באופן ישיר נוסחה ל- a_n, או אפילו לחשב את ערכו באופן ספציפי. במקרה כזה ניתן לנסות לתאר את $\lambda n.\, a_n$ באופן רקורסיבי, כלומר: על-ידי נוסחת נסיגה עם תנאי התחלה. הצלחה בכך פירושה, בדרך-כלל, שמצאנו דרך אפקטיבית לחישוב a_n לכל n שנחפוץ (בעיקר אם עומד לרשותנו מחשב). במלים אחרות: פתרנו, למעשה, את הבעיה.

דוגמאות:

(1) הדוגמה של סדרת פיבונאצ'י בה התחלנו פרק זה.

(2) הדוגמה מהפרק הקודם של מספר המחרוזות באורך n שאפשר לבנות מ- $\{0, 1, 2, 3\}$, שיש בהן מספר זוגי של אפסים:

פתרון רקורסיבי:

אם נשים בהתחלה 1, 2 או 3, הרי אחר-כך נוכל להשלים את המחרוזות בכל מקרה ב- a_{n-1} דרכים. אם נשים 0 בהתחלה, הרי אחר-כך נצטרך לשים מחרוזת באורך $n-1$ שיש בה מספר ***אי-זוגי*** של אפסים. מספר האפשרויות הוא $4^{n-1} - a_{n-1}$ (המספר הכללי של מחרוזות באורך $n-1$, המורכבות מ- $\{0, 1, 2, 3\}$, פחות מספר אלה מתוכן, שיש בהן מספר זוגי של אפסים). לכן, אם $n \geq 1$, אז

$$a_n = 3a_{n-1} + (4^{n-1} - a_{n-1})$$

כיוון ש- $a_0 = 1$ (המחרוזת הריקה היא כשרה), נקבל:

$$\begin{cases} a_n = 2a_{n-1} + 4^{n-1} & (n \geq 1) \\ a_0 = 1 \end{cases}$$

אם נרצה את התשובה באופן ספציפי כש- $n = 5$, נקבל:

$a_3 = 2 \cdot 10 + 4^2 = 36$ $\quad a_2 = 2 \cdot 3 + 4^1 = 10$ $\quad a_1 = 2 \cdot 1 + 4^0 = 3$

$a_5 = 2 \cdot 136 + 4^4 = 528$ $\quad a_4 = 2 \cdot 36 + 4^3 = 136$

(3) **מגדלי-האנוי**: זוהי בעיית תכנות קלאסית. הסיפור הוא, שיש לנו שלושה מוטות: א', ב' ו-ג'. על מוט א' מסודרות n טבעות שונות בגודלן, כשכל טבעת מונחת על טבעת גדולה ממנה. צעד חוקי הוא העברת הטבעת העליונה ממוט מסוים x אל מוט אחר y, בתנאי שהטבעת העליונה במוט y גדולה יותר מזו, שאנו מעבירים לשם, ובתנאי שהטבעת המועברת תושם על הטבעת העליונה הקודמת של y. הבעיה היא להעביר, בצעדים חוקיים בלבד, את כל הטבעות ממוט א' למוט ב' (כשאפשר, כמובן, להשתמש במוט ג' כעזר).

הפתרון הקלאסי הוא העברת $n-1$ טבעות ממוט א' למוט ג' בצעדים חוקיים בלבד (כאשר $n > 1$ ובהנחה שעבור $n-1$ טבעות אנו כבר יודעים לפתור). אחר-כך העברת הטבעת שנשארה ממוט א' למוט ב' (זה יהיה הצעד היחיד כש- $n = 1$), ולבסוף העברת $n-1$ הטבעות ממוט ג' למוט ב' בצעדים חוקיים בלבד. הבעיה הקומבינטורית: כמה צעדים שיטה זו דורשת? (זה יותר בעיה באנליזה של אלגוריתמים/תכניות מאשר בקומבינטוריקה, אבל הגבול הינו מטושטש למדי).

מהתיאור למעלה ברור ש-:

$$\begin{cases} a_1 = 1 \\ a_n = 2a_{n-1} + 1 \qquad (n > 1) \end{cases}$$

(במקום להתחיל ב- $a_1 = 1$ אפשר להתחיל ב- $a_0 = 0$, ונוסחת הנסיגה תישאר נכונה!).

(4) בכמה אופנים ניתן להרכיב מחרוזות באורך n של "0" ו- "1", שלא מופיע בהן הצירוף "101"?

פתרון:

נסמן ב- a_n את המספר המבוקש, ובאופן זמני ב- b_n את מספר המחרוזות המותרות שמתחילות ב- "01".

ברור שאם $n \geq 1$, אז $a_n = a_{n-1} + (a_{n-1} - b_{n-1})$ (או שהמחרוזת מתחילה ב- 0 (a_{n-1} אפשרויות), או שהיא מתחילה ב- 1, ואז אפשר לשים אחריה כל מחרוזת אחרת מותרת באורך $n-1$, בתנאי ***שאינה*** מתחילה ב- 01 ($a_{n-1} - b_{n-1}$ אפשרויות!).

כמו-כן, אם $n \geq 2$, אז

$$b_n = a_{n-2} - b_{n-2} \quad \text{(למה?)}$$

סך-הכל:

$$\begin{cases} n \geq 1 \;\Rightarrow\; a_n = 2a_{n-1} - b_{n-1} \qquad (1) \\ n \geq 2 \;\Rightarrow\; b_n = a_{n-2} - b_{n-2} \qquad (2) \end{cases}$$

מ- (1) נובע ש- $a_{n+1} = 2a_n - b_n$ לכל n, ולכן $b_n = 2a_n - a_{n+1}$ לכל n.

מכאן שאם $n \geq 2$, אז $b_{n-2} = 2a_{n-2} - a_{n-1}$. הצבת הנוסחאות ל- b_n ול- b_{n-2} ב- (2) נותנת: $2a_n - a_{n+1} = a_{n-2} - (2a_{n-2} - a_{n-1})$ עבור $n \geq 2$, ומזה נקבל:

$$a_{n+1} = 2a_n - a_{n-1} + a_{n-2} \quad \text{עבור } n \geq 2$$

לכן: $a_n = 2a_{n-1} - a_{n-2} + a_{n-3}$ לכל $n \geq 3$, וזוהי נוסחת נסיגה עבור $\lambda n.\, a_n$.

עבור $n = 0, 1, 2$ אפשר לראות ישירות ש- $a_2 = 4$, $a_1 = 2$, $a_0 = 1$. מפה נקבל בקלות, למשל, ש- $a_6 = 37$.

פתרון משוואות נסיגה בעזרת פונקציות יוצרות

כאמור למעלה, את ההגדרה של פונקציה בעזרת נוסחת נסיגה עם תנאי התחלה, אפשר לראות גם *כמשוואה*. למרות שההצגה הרקורסיבית היא הצגה אפקטיבית של הפונקציה (ומספקת דרך אפקטיבית לחישובה), מעוניינים אנו לפעמים בהצגה *מפורשת*; דהיינו *הצגה בעזרת ביטוי-λ* . בז'רגון, המקובל בטקסטים בנושא, קוראים אכן בשם "פתרון" של משוואה רקורסיבית רק להצגה מפורשת, בעזרת ביטוי-λ , של הסדרה המספקת משוואה זו (רק לא משתמשים בהם בסימון-λ!).

עם זאת, העניינים אינם כה פשוטים. לא די לדרוש "ביטוי-λ". השאלה הנשאלת היא: באילו פונקציות נתיר להשתמש בביטוי-λ זה? התשובה היא, שאנו נחפש ביטוי-λ, המשתמש בפונקציות האלמנטריות הידועות מחדו"א, בפעולות המוכרות מחדו"א (כמו חיבור, כפל וכו'), בהרכבת פונקציות, וכן בפונקציה $\lambda n.\, n!$ – שהיא התוספת המיוחדת כאן. למעשה, כמו שציינו למעלה, פונקצית העצרת בעצמה מוגדרת על-ידי נוסחת נסיגה – אך לצורך המטרה של "פתרון" משוואות נסיגה נתייחס אליה כאל פונקציה בסיסית.

אחד מכלי הנשק העיקריים שלנו למציאת פתרון כזה הוא חיפוש פונקציה יוצרת עבור הסדרה $\lambda n.\, a_n$, המוגדרת על-ידי נוסחת הנסיגה ותנאי ההתחלה. חיפוש זה מתחיל בהנחה, ש- F היא פונקציה יוצרת כזו, ואחר-כך בא שימוש במה שידוע על $\lambda n.\, a_n$, בשביל למצוא *מועמדת* עבור F . אם הצלחנו, הרי בדרך-כלל אפשר להראות בקלות, ש- F זו שמצאנו אכן יוצרת סדרה, המקיימת את המשוואה (כולל תנאי ההתחלה). כיוון שמשפט הקיום והיחידות למעלה קובע, שלכל משוואה כזו יש רק סדרה אחת, שמקיימת אותה, הרי בהכרח F יוצרת את $\lambda n.\, a_n$ המבוקש. מה שנותר עוד לעשות הוא למצוא נוסחה מפורשת עבור הסדרה, ש- F יוצרת. בעיות מסוג זה פתרנו כבר בעבר.

נביא עתה מספר דוגמאות, שמהן יהיה ברור התהליך.

1) נתחיל בדוגמה של מגדלי האנוי. הסדרה $\lambda n.\, a_n$ שאנו מחפשים מקיימת:

$$a_n = \begin{cases} 0 & n = 0 \\ 2a_{n-1} + 1 & n > 0 \end{cases}$$

נניח ש- F פונקציה יוצרת של $\lambda n.\, a_n$, אז $F(x)=\sum_{n=0}^{\infty}a_nx^n$.

נציב את תנאי ההתחלה $a_0 = 0$. נקבל:

$$F(x)=0+\sum_{n=1}^{\infty}a_nx^n=\sum_{n=1}^{\infty}a_nx^n$$

נציב עתה את נוסחת הנסיגה $a_n = 2a_{n-1} + 1$ (ל- $n \geq 1$). נקבל, של- x בתחום ההתכנסות מתקיים:

$$F(x)=\sum_{n=1}^{\infty}(2a_{n-1}+1)x^n$$

$$\Leftrightarrow \qquad F(x)=2\sum_{n=1}^{\infty}a_{n-1}x^n+\sum_{n=1}^{\infty}x^n$$

$$\Leftrightarrow \qquad F(x)=2x\sum_{n=1}^{\infty}a_{n-1}x^{n-1}+x\sum_{n=1}^{\infty}x^{n-1}$$

$$\Leftrightarrow \qquad F(x)=2x\sum_{n=0}^{\infty}a_nx^n+x\sum_{n=0}^{\infty}x^n$$ (שינוי גבולות הסכום)

$$\Leftrightarrow \qquad F(x)=2x\cdot F(x)+\frac{x}{1-x}$$ (בהנחה ש- $|x|<1$)

$$\Leftrightarrow \qquad F(x)(1-2x)=\frac{x}{1-x}$$

ולכן בהכרח

$$F=\lambda x.\frac{x}{(1-2x)(1-x)}$$

מה שמצאנו עד כה הוא, ש- $\lambda x.\dfrac{x}{(1-2x)(1-x)}$ היא המועמדת האפשרית היחידה להיות פונקציה יוצרת של $\lambda n.\, a_n$. עתה, מכיוון שזו פונקציה רציונלית, אנו יודעים שהיא יוצרת סדרה כלשהי $\lambda n.\, b_n$. אם נלך אחורה בחישוב שעשינו, נקבל

שבהכרח $b_0 = 0$ וש- $b_n = 2b_{n-1} + 1$ ל- $n > 0$, דהיינו: $\lambda n.\, b_n = \lambda n.\, a_n$. לכן F זו אכן יוצרת את הסדרה, שאנו מחפשים. הבה נמצא עתה סדרה זו. בשלב ראשון עלינו למצוא, כזכור, A ו- B כך ש-

$$\frac{x}{(1-2x)(1-x)} = \frac{A}{1-2x} + \frac{B}{1-x}$$

$$x = A(1-x) + B(1-2x)$$

קל לברר שכאן $A = 1$ ו- $B = -1$. לכן

$$\lambda x.\frac{x}{(1-2x)(1-x)} = \lambda x.\frac{1}{1-2x} - \frac{1}{1-x}$$

ממה שלמדנו בפרק הקודם ברור, שפונקציה זו יוצרת את הסדרה $\lambda n.\, 2^n - 1$. מספר הצעדים הדרוש להעברת n טבעות בשיטה (אלגוריתם) שהוצעה למעלה הוא לכן $2^n - 1$ (לא קשה, אגב, להוכיח, שזהו מספר *אופטימלי*, דהיינו: אי-אפשר לעשות זאת בפחות צעדים, ולא משנה באיזו שיטה ננקוט).

הערה חשובה:

השיקול, שהפעלנו בדוגמה האחרונה, מדוע F שמצאנו אכן יוצרת את הפונקציה המבוקשת, עובד באופן כללי. אומנם, כאשר אנו *מתחילים*, אנו אומרים *שנניח* ש- F יוצרת את $\lambda n.\, a_n$. אין לנו אז ביטחון, ש- F כזו אכן קיימת. ברם, אם הצלחנו למצוא מועמדת, ואנו יודעים שמועמדת זו אכן יוצרת סדרה כלשהי, אז בדרך-כלל היפוך השיקולים, שהביאו למציאת המועמדת, מראה, שהסדרה הנוצרת על ידה אכן מקיימת את נוסחת הנסיגה ואת תנאי ההתחלה. לכן סדרה זו היא הסדרה המבוקשת. *בדוגמאות הבאות לא נטרח לחזור על שיקול זה.*

2) מצא בעזרת פונקציות יוצרות נוסחה מפורשת לסדרה $\lambda n.\, a_n$ המוגדרת על-ידי:

$$\begin{cases} a_n = 5a_{n-1} - 6a_{n-2} & n \ge 2 \\ a_0 = 1 \quad a_1 = -2 \end{cases}$$

פתרון:

שוב, נניח ש- F פונקציה יוצרת עבור $\lambda n.\, a_n$, כלומר ש-

$$F = \lambda x. \sum_{n=0}^{\infty} a_n x^n$$

אז ל- x בתחום ההתכנסות:

$$F(x) = \sum_{n=0}^{\infty} a_n x^n = 1 - 2x + \sum_{n=2}^{\infty} a_n x^n$$

$$F(x) = 1 - 2x + \sum_{n=2}^{\infty} (5a_{n-1} - 6a_{n-2}) x^n$$

$$F(x) = 1 - 2x + 5x \sum_{n=2}^{\infty} a_{n-1} x^{n-1} - 6x^2 \sum_{n=2}^{\infty} a_{n-2} x^{n-2}$$

$$F(x) = 1 - 2x + 5x \sum_{n=1}^{\infty} a_n x^n - 6x^2 \sum_{n=0}^{\infty} a_n x^n$$

כיוון ש- $\sum_{n=1}^{\infty} a_n x^n = \left(\sum_{n=0}^{\infty} a_n x^n\right) - a_0$ ו- $a_0 = 1$, נקבל:

$$F(x) = 1 - 2x + 5x(F(x) - 1) - 6x^2 F(x)$$

$$(1 - 5x + 6x^2)\, F(x) = 1 - 2x - 5x$$

$$F(x) = \frac{1 - 7x}{1 - 5x + 6x^2}$$

מכאן ש- $\lambda x. \frac{1 - 7x}{1 - 5x + 6x^2}$ פונקציה יוצרת עבור $\lambda n.\, a_n$. בשביל לראות מה פונקציה זו יוצרת, נפרק תחילה את המכנה:

$$1 - 5x + 6x^2 = 0$$

$$x_1 = \frac{1}{2} \qquad x_2 = \frac{1}{3}$$

ואם נלך לפי הנוסחה:

$$ax^2+bx+c=c(1-\frac{1}{x_1}x)(1-\frac{1}{x_2}x)$$

נקבל:

$$1-5x+6x^2=(1-2x)(1-3x)$$

בשלב הבא עלינו למצוא A ו- B כך שלכל x

$$\frac{1-7x}{1-5x+6x^2}=\frac{A}{1-2x}+\frac{B}{1-3x}$$

נקבל $A=5$, $B=-4$. לכן $\lambda x.\frac{5}{1-2x}-\frac{4}{1-3x}$ פונקציה יוצרת עבור $\lambda n.\,a_n$,
ולכן לכל n :

$$a_n=5\cdot 2^n-4\cdot 3^n$$

(3) פתור:

$$\begin{cases} a_n=4(a_{n-1}-a_{n-2}) & n\geq 2 \\ a_0=3 \quad a_1=8 \end{cases}$$

פתרון:

נסמן $F(x)=\sum_{n=0}^{\infty}a_nx^n$. אז:

$$F(x)=3+8x+\sum_{n=2}^{\infty}a_nx^n$$

$$F(x)=3+8x+\sum_{n=2}^{\infty}4(a_{n-1}-a_{n-2})x^n$$

$$F(x)=3+8x+4x\sum_{n=2}^{\infty}a_{n-1}x^{n-1}-4x^2\sum_{n=2}^{\infty}a_{n-2}x^{n-2}$$

$$F(x) = 3 + 8x + 4x\sum_{n=1}^{\infty} a_n x^n - 4x^2 \sum_{n=0}^{\infty} a_n x^n$$

$$F(x) = 3 + 8x + 4x(F(x) - 3) - 4x^2 F(x)$$

$$(1 - 4x + 4x^2) \cdot F(x) = 3 - 4x$$

$$F(x) = \frac{3-4x}{(1-2x)^2}$$

כאן עלינו לחפש A ו- B כך ש-

$$\frac{3-4x}{(1-2x)^2} = \frac{A}{1-2x} + \frac{B}{(1-2x)^2}$$

כלומר,

$$3 - 4x = A(1 - 2x) + B$$

נקבל ש- $A = 2$ ו- $B = 1$.
לכן, פונקציה יוצרת כאן היא

$$\lambda x.\ \frac{2}{1-2x} + \frac{1}{(1-2x)^2}$$

והסדרה הנוצרת היא:

$$\lambda n.\ 2 \cdot 2^n + (n + 1) \cdot 2^n$$

כלומר:

$$\lambda n.\ 2^n(n + 3)$$

(4) נפתור לבסוף את הדוגמה על מספר המחרוזות, שאפשר להרכיב מ- $\{0,1,2,3\}$, שיש בהן מספר זוגי של אפסים. כזכור, אם a_n הוא התשובה עבור מחרוזת באורך n, אז:

$$\begin{cases} a_0 = 1 \\ a_n = 2a_{n-1} + 4^{n-1} \quad (n > 0) \end{cases}$$

נניח אפוא ש- F פונקציה יוצרת. אז $F(x) = \sum_{n=0}^{\infty} a_n x^n$, ולכן:

$$F(x) = 1 + \sum_{n=1}^{\infty} (2a_{n-1} + 4^{n-1}) x^n$$

$$F(x) = 1 + 2x \sum_{n=1}^{\infty} a_{n-1} x^{n-1} + x \sum_{n=1}^{\infty} (4x)^{n-1}$$

$$F(x) = 1 + 2x \sum_{n=0}^{\infty} a_n x^n + x \sum_{n=0}^{\infty} (4x)^n$$

$$F(x) = 1 + 2x \cdot F(x) + \frac{x}{1-4x}$$

לכן:

$$(1-2x)F(x) = 1 + \frac{x}{1-4x} = \frac{1-3x}{1-4x}$$

ו-

$$F = \lambda x. \frac{1-3x}{(1-4x)(1-2x)} = \lambda x. \frac{1}{2} \cdot \frac{1}{1-2x} + \frac{1}{2} \cdot \frac{1}{1-4x}$$

F זו יוצרת את הסדרה $\lambda n. \frac{2^n + 4^n}{2}$, ולכן $a_n = \frac{2^n + 4^n}{2}$ (כמו שמצאנו בעבר בדרך אחרת).

6.ד משוואות נסיגה ליניאריות עם מקדמים קבועים

לצורך מתן מוטיבציה למה שייעשה בפרק זה, הבה נחזור לדוגמה כלשהי מאלה שפתרנו בסעיף הקודם, למשל ל-

$$\begin{cases} a_n = 4(a_{n-1} - a_{n-2}) & n \geq 2 \\ a_0 = 3 \quad a_1 = 8 \end{cases}$$

נניח שבמקום תנאי ההתחלה $a_0 = 3$ ו- $a_1 = 8$, היינו מקבלים, למחרת היום, תרגיל דומה עם תנאי התחלה אחרים, למשל $a_0 = -1$ ו- $a_2 = \sqrt{3}$. על מנת לפתור בעיה חדשה זו היה עלינו לחזור על כל העבודה מחדש – וזהו בזבוז רב של משאבים ואנרגיה! יתר על כן, ישנן אכן בשימוש משוואות, שחוזרות על עצמן במקרים רבים, כאשר כל השינוי הוא בתנאי ההתחלה. לפתור כל פעם מחדש אינו רעיון משובב נפש.

אפשר, כמובן, להתגבר על הבעיה בצורה, שנעשה הדבר בזמנו לגבי משוואות ריבועיות בבית-ספר תיכון: נפתור את המשוואה $a_n = 4a_{n-1} - 4a_{n-2}$ עם *פרמטרים* כמו a ו- b במקום ה- 3 ו- 8 למעלה, ובאותה שיטה (פונקציות יוצרות), בה נעשה הדבר למעלה. זה אפשרי בהחלט ולא לגמרי בלתי מומלץ. עם זאת, החישוב כאן עם פרמטרים לא כל-כך נוח.

בסעיף זה נציג שיטה אלטרנטיבית, שהולכת בכיוון הפוך. היא מאפשרת (במקרים רבים) מציאה נוחה של פתרון *כללי* למשוואת הנסיגה (*בלא* שיינתנו תנאי התחלה), בעוד שמציאת פתרון לבעיה ספציפית, *עם* תנאי התחלה, דורשת עבודה נוספת של פתרון משוואות ממעלה ראשונה.

כדי להבין למה אנו חותרים, הבה ניקח דוגמה מתחום אחר.

בעיה

מצא פונקציה F כך ש- $F' = \lambda x.\, x^2$ ו- $F(2) = 7$.

פתרון:

תחילה נפתור פתרון כללי את הבעיה של מציאת פונקציה F כך ש- $F' = \lambda x.\, x^2$ (זה מה שנקרא מציאת "אינטגרל לא מסוים" של $\lambda x.\, x^2$). התשובה היא, כידוע, $\lambda x.\frac{x^3}{3}+c$ (בעצם, התשובה באמת היא *קבוצה* של פונקציות: הקבוצה $\{\lambda x.\frac{x^3}{3}+c \mid c\in \mathrm{R}\}$!). כדי למצוא את הפתרון של הבעיה הספציפית שלנו, נציב את תנאי ההתחלה $F(2) = 7$ ונקבל $\frac{2^3}{3}+c=7$ ולכן $c=\frac{13}{3}$. מכאן שפתרון הבעיה הוא $F=\lambda x.\frac{x^3+13}{3}$.

לצורך התיאוריה אותה נפתח כאן, יהיה נוח להציג את משוואת הנסיגה בצורה שונה במקצת מזו, שבה עבדנו עד כה (אם כי שקולה לחלוטין). במקום לכתוב אותה בצורה:

$$a_n = F(n, a_{n-1}, a_{n-2}, \ldots, a_{n-k}) \qquad \text{ל-}\ n \geq k$$

או

$$X(n) = F(n, X(n-1), X(n-2), \ldots, X(n-k)) \qquad \text{ל-}\ n \geq k$$

נעדיף כאן את הצורה:

$$a_{n+k} = F(n, a_{n+k-1}, a_{n+k-2}, \ldots, a_n) \qquad (\text{לא אותה } F\ !)$$

או

$$X(n+k) = F(n, X(n+k-1), X(n+k-2), \ldots, X(n))$$

כך, לדוגמה, משוואת פיבונאצ'י תיכתב עכשיו בצורה $a_{n+2} = a_{n+1} + a_n$ (במקום $a_n = a_{n-1} + a_{n-2}$), והמשוואה שמגדירה את פעולת העצרת תיכתב בצורה $a_{n+1} = (n+1)a_n$ (במקום $a_n = n \cdot a_{n-1}$. נשים לב, שבמקרה זה ה- F המקורי הוא $\lambda n,x.\, nx$, בעוד שה- F בצורה החדשה הינו $\lambda n,x.\,(n+1)x$).

היתרון הוא, שבצורה החדשה המשוואה אמורה להיות נכונה *לכל* n , וזה מועיל לצרכים מסוימים.

בעוד שאת טכניקת הפתרון עם פונקציות יוצרות ניתן לנסות להפעיל עבור כל משוואת נסיגה (אם כי לא תמיד היא תישא פירות), הרי את התיאוריה של סעיף זה נקדיש רק לסוג מיוחד, חשוב ביותר, של משוואות: משוואות לינאריות עם מקדמים

קבועים. הכוונה למשוואות כמו שתיארנו למעלה, בהן יש ל- F הצורה המיוחדת הבאה:

$$F(n, a_{n+k-1}, a_{n+k-2}, \ldots, a_n) = d_{k-1}a_{n+k-1} + d_{k-2}a_{n+k-2} + \ldots + d_0a_n + h(n)$$

כאשר $d_0, \ldots, d_{k-1} \in \mathbf{R}$ הם קבועים, ו- $h : \mathbf{N} \to \mathbf{R}$.

כשל- F צורה מיוחדת זו, משוואת הנסיגה נראית כך:

$$X(n+k) = d_{k-1}X(n+k-1) + \ldots + d_0X(n) + h(n) \qquad (\text{לכל } n)$$

ואפשר לכתוב אותה גם כך:

$$X(n+k) - d_{k-1}X(n+k-1) - \ldots - d_0X(n) = h(n) \qquad (\text{לכל } n)$$

(לדוגמה: משוואת פיבונאצ'י תיראה אז כך: $X(n+2) - X(n+1) - X(n) = 0$, ומשוואת מגדלי האנוי: $X(n+1) - 2X(n) = 1$. בשני המקרים זה צריך להיות נכון לכל n).

הכללה נוחה, שאינה מוסיפה ואינה גורעת, היא להרשות גם ל- $X(n+k)$ מקדם כלשהו *שונה מאפס* (ולא רק המקדם 1). כך, למשל, את המשוואה

$$a_{n+2} = \frac{1}{2}a_{n+1} - \frac{1}{3}a_n + n^2$$

נוח להעביר לצורה:

$$6a_{n+2} - 3a_{n+1} + 2a_n = 6n^2$$

(או: $6\cdot X(n+2) - 3\cdot X(n+1) + 2\cdot X(n) = 6n^2$).

נקבל אז את הצורה הכללית:

$$c_kX(n+k) + c_{k-1}X(n+k-1) + \ldots + c_0X(n) = h(n) \qquad \text{לכל } n \quad (c_k \neq 0)$$

(כיוון ש- $c_k \neq 0$, נוכל תמיד לחלק בו ולחזור לצורה הקודמת!).
המספר k נקרא *הסדר* של משוואה מסוג זה.

עתה, במקום לציין בכל מקום "לכל n", נוכל גם להציג את המשוואה כך:

$$\lambda n.\ c_k X(n+k) + c_{k-1} X(n+k-1) + \ldots + c_0 X(n) = h$$

צורת כתיבה זו מאפשרת לנו להסתכל על המשוואה מזווית אחרת:
בהינתן $\vec{c} = \langle c_0, c_1, \ldots, c_k \rangle \in \mathrm{R}^{k+1}$, נוכל להגדיר פונקציה $L^{\vec{c}} : (\mathrm{N} \to \mathrm{R}) \to (\mathrm{N} \to \mathrm{R})$
על-ידי:

$$L^{\vec{c}} = \lambda g \in \mathbf{N} \to \mathbf{R}.\ \lambda n \in \mathbf{N}. \sum_{i=0}^{k} c_{k-i}\, g(n+k-i)$$

מה שאנו מחפשים הוא פונקציה $X : \mathbf{N} \to \mathbf{R}$ כך ש-

$$L^{\vec{c}}(X) = h$$

(כאשר h היא פונקציה נתונה מ- $\mathbf{N}$ ל- $\mathbf{R}$).

כדוגמה לפעולת $L^{\vec{c}}$ הבה ניקח $k = 2$ ו- $\vec{c} = \langle \frac{1}{2}, 3, -1 \rangle$. אז ל- $g : \mathbf{N} \to \mathbf{R}$ מתקיים:

$$L^{\vec{c}}(g) = \lambda n \in \mathrm{N}.\ -g(n+2) + 3g(n+1) + \frac{1}{2} g(n)$$

בפרט

$$L^{\vec{c}}(\lambda n.2^n) = \lambda n \in \mathrm{N}.\ -2^{n+2} + 3 \cdot 2^{n+1} + \frac{1}{2} \cdot 2^n$$

וזוהי אכן פונקציה חדשה מ- $\mathbf{N}$ ל- $\mathbf{R}$.

הערות:

(1) הסימון $L^{\vec{c}}$ בא להדגיש, כי זהותה של הפונקציה L תלויה בערכם של k ושל $c_0, \ldots, c_k$. עם זאת, לצורך מניעת סרבול בכתיבה, נכתוב מעתה בדרך-כלל פשוט "L" במקום "$L^{\vec{c}}$".

(2) כאשר במקום קבועים $c_0, c_1, \ldots, c_k$ מרשים להשתמש כמקדמים ב*פונקציות* $h_0, \ldots, h_k$ מ- $\mathbf{N}$ ל- $\mathbf{R}$ (בתנאי ש- $h_k(n) \neq 0$ לכל n), אז המשוואות $L(X) = h$ המתקבלות נקראות *משוואות ליניאריות*. משוואות ליניאריות עם מקדמים קבועים

הן מקרה פרטי, חשוב במיוחד, של משוואות לינאריות; זהו, בעצם, המקרה בו המקדמים הם פונקציות קבועות.

(3) עד כה הנחנו כל הזמן, שהמקדמים הקבועים הם מספרים *ממשיים*, ש- h (ב- $L(X) = h$) היא סדרה של מספרים ממשיים, וש- $L : (\mathbf{N} \to \mathbf{R}) \to (\mathbf{N} \to \mathbf{R})$. למעשה, ניתן ורצוי לפתח וליישם את התיאוריה, כאשר במקום סדרות של מספרים ממשיים ומקדמים ממשיים נדבר על סדרות של מספרים *מרוכבים* ונרשה גם מקדמים מרוכבים. מכאן ואילך נניח אפוא, ש- $c_0, \ldots, c_k \in \mathbf{C}$ ו- $h : \mathbf{N} \to \mathbf{C}$. על L נסתכל אז כפונקציה מ- $\mathbf{N} \to \mathbf{C}$ אל $\mathbf{N} \to \mathbf{C}$.

עם כל-כך הרבה טרמינולוגיה חדשה וסימונים חדשים כדאי אולי לנסח מחדש את המשפט מהפרק הקודם על קיום ויחידות פתרונות. בשביל זה נצטרך אבל להכניס סימון נוסף, שנשתמש בו הרבה בהמשך.

הגדרה:

אם $f : \mathbf{N} \to \mathbf{C}$ היא סדרה של מספרים מרוכבים, אז ה- k-רישא שלה, $f[k]$, מוגדרת על-ידי:

$$f[k] = \langle f(0), f(1), \ldots, f(k-1) \rangle \in \mathbf{C}^k$$

(בצורה אחרת: ה- k-רישא של סדרה $\lambda n.\, a_n$ היא הווקטור $\langle a_0, a_1, \ldots, a_{k-1} \rangle$).

משפט 1 (משפט הקיום והיחידות):

תהי $L(X) = h$ משוואת נסיגה לינארית מסדר k. אז לכל $\vec{v} = \langle v_0, \ldots, v_{k-1} \rangle \in \mathbf{C}^k$ קיים פתרון יחיד $X_{\vec{v}}$ למשוואה $L(X) = h$, המקיים:

$$X_{\vec{v}}[k] = \vec{v}$$

(במלים פשוטות יותר: יש סדרה יחידה $X_{\vec{v}} : \mathrm{N} \to \mathrm{C}$ כך ש- $L(X_{\vec{v}}) = h$ וכך ש- $X_{\vec{v}}(i) = v_i$ ל- $i = 0, 1, \ldots, k-1$).

מסקנה 1:

אם f ו- g שני פתרונות של משוואה לינארית $L(X) = h$ מסדר k, כך ש- $f[k] = g[k]$, אז $f = g$.

לצורך ההמשך נגייס לעזרתנו את האלגברה הלינארית (לא ייתכן לקפח קורס חשוב זה, אחרי שבנושא הפונקציות היוצרות הסתמכנו כה הרבה על חשבון דיפרנציאלי ואינטגרלי!). ההבחנה הראשונית לה אנו זקוקים היא, שאם נגדיר חיבור סדרות ומכפלת סדרה בקבוע בצורה הרגילה:

$$f+g=\lambda n.\, f(n)+g(n) \qquad \alpha\cdot f=\lambda n.\ \alpha f(n)$$

אז $\mathbf{N}\to\mathbf{C}$ היא מרחב וקטורי מעל $\mathbf{C}$ עם פעולות אלו. וקטור האפס של מרחב זה $(0_{\mathbf{N}\to\mathbf{C}})$ הוא הסדרה $\lambda n.\ 0$.

הבחנה שנייה, זו שמכניסה את האלגברה הלינארית לתמונה, היא הטענה הבאה:

משפט 2:

בהינתן $\vec{c}=c_0,\ldots,c_k\in \mathrm{C}$, הפונקציה $L^{\vec{c}}$ מ- $\mathbf{N}\to\mathbf{C}$ אל $\mathbf{N}\to\mathbf{C}$, שהוגדרה לעיל, היא העתקה לינארית מהמרחב הווקטורי $\mathbf{N}\to\mathbf{C}$ אל עצמו.

הוכחה:

אם $f,g\in\mathbf{N}\to\mathbf{C}$ ו- $\alpha,\beta\in\mathbf{C}$, אז:

$$\begin{aligned}
L^{\vec{c}}(\alpha f+\beta g)&=\lambda n.\ \sum_{i=0}^{k}c_{k-i}(\alpha f+\beta g)(n+k-i)\\
&=\lambda n.\ \sum_{i=0}^{k}c_{k-i}(\alpha f(n+k-i)+\beta g(n+k-i))\\
&=\lambda n.\ \alpha\cdot\sum_{i=0}^{k}c_{k-i}f(n+k-i)+\beta\cdot\sum_{i=0}^{k}c_{k-i}g(n+k-i)\\
&=\lambda n.\ \alpha\cdot(L^{\vec{c}}(f))(n)+\beta\cdot(L^{\vec{c}}(g))(n)\\
&=\alpha L^{\vec{c}}(f)+\beta L^{\vec{c}}(g)
\end{aligned}$$

מסקנה 2:

לכל $\alpha_1,\ldots,\alpha_\ell\in \mathrm{C}$ ו- $g_1,\ldots,g_\ell\in \mathrm{N}\to \mathrm{C}$ מתקיים:

$$L\left(\sum_{i=1}^{\ell}\alpha_i g_i\right)=\sum_{i=1}^{\ell}\alpha_i L(g_i)$$

מסקנה 3:

$\ker(L)$, הגרעין של L (כלומר: $\{g \in \mathbf{N} \to \mathbf{C} \mid L(g) = \lambda n.\ 0\}$) הינו תת-מרחב של $\mathbf{N} \to \mathbf{C}$.

הוכחה:

זהו מקרה פרטי של משפט כללי על העתקות לינאריות (וגם עניין של שתי שורות להוכיח באופן ישיר).

המסקנה האחרונה מובילה לעניין מיוחד בסוג המשוואות הבא:

הגדרה:

משוואת נסיגה לינארית נקראת *הומוגנית* אם היא מהצורה $L(X) = \lambda n.\ 0$. אחרת היא *לא הומוגנית.*

הערה:

בצורה יותר סטנדרטית נכתבת משוואה לינארית הומוגנית מסדר k עם מקדמים קבועים בצורה הבאה:

$$c_k a_{n+k} + c_{k-1} a_{n+k-1} + \ldots + c_1 a_{n+1} + c_0 a_n = 0$$

יש לזכור, כשכותבים את המשוואה כך, שהנעלם בה הוא *סדרות* $\lambda n.\ a_n$, שמקיימות את המשוואה המספרית *לכל* n (k הינו קבוע כאן). מקובל גם לכתוב $L(X) = 0$ במקום $L(X) = \lambda n.\ 0$. יש לזכור אבל, שבצורת כתיבה זו "0" אינו המספר 0, אלא וקטור האפס של $\mathbf{N} \to \mathbf{C}$ (כלומר: $\lambda n.\ 0$).

את מסקנה 3 נוכל עכשיו לנסח כך:

משפט 3:

אוסף הפתרונות של המשוואה ההומוגנית $L(X) = \lambda n.\ 0$ מהווה תת-מרחב של $\mathbf{N} \to \mathbf{C}$.

מסקנה 4:

אם $X_1,\ldots,X_\ell$ פתרונות של משוואה הומוגנית כנ"ל, אז כך גם כל צירוף לינארי $\sum_{i=1}^{\ell}\alpha_i X_i$ שלהם.

אנו יכולים עתה לנסח ולהוכיח את המשפט המרכזי על משוואות לינאריות הומוגניות:

משפט 4:

תהי $L(X) = \lambda n.\,0$ משוואה לינארית הומוגנית מסדר k, ויהיו $X_1, \ldots, X_k$ פתרונות שלה, כך שהווקטורים $X_1[k], \ldots, X_k[k]$ מהווים בסיס של $\mathbf{C}^k$. אז $X_1, \ldots, X_k$ מהווים בסיס למרחב הפתרונות של $L(X) = \lambda n.\,0$.

הערה:

המשפט יהיה נכון, גם אם במקום $\mathbf{C}$ היינו מדברים לכל אורך הדרך על $\mathbf{R}$.

הוכחה:

עלינו להראות, ש- $X_1, \ldots, X_k$ בלתי-תלויים לינארית, ושהם פורשים את מרחב הפתרונות.

אי-תלות:

נניח $\sum_{i=1}^{k}\alpha_i X_i = \lambda n.0$ $(\alpha_1, \ldots, \alpha_k \in \mathbf{C})$. עלינו להראות שלכל i $\alpha_i = 0$.

עתה, פירוש ההנחה הוא, שלכל $n \in \mathbf{N}$, $\sum_{i=1}^{k}\alpha_i X_i(n) = 0$. בפרט נכון הדבר ל- $0 \le n \le k-1$. מזה נובע מיידית ש- $\sum_{i=1}^{k}\alpha_i X_i[k] = \vec{0}_k$ (כאשר $\vec{0}_k = \langle 0,0,\ldots,0\rangle$ הוא וקטור האפס של $\mathbf{C}^k$). נתון אבל, שהווקטורים $X_1[k], \ldots, X_k[k]$ הם בלתי תלויים לינארית (כי הם בסיס של $\mathbf{C}^k$). מכאן, שאכן $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.

פרישה:

יהי Y פתרון כלשהו של המשוואה. כיוון ש- $X_1[k], \ldots, X_k[k]$ פורשים את $\mathbf{C}^k$, קיימים $\alpha_1, \ldots, \alpha_k \in \mathbf{C}$ כך ש- $Y[k]=\sum_{i=1}^{k}\alpha_i X_i[k]$.

עתה, לפי מסקנה 4, $Z=\sum_{i=1}^{k}\alpha_i X_i$ גם הוא פתרון של המשוואה. יתר-על-כן, לכל $0 \le j \le k-1$ מתקיים ש- $Z(j)=\sum_{i=1}^{k}\alpha_i X_i(j)=Y(j)$ (כיוון שזו בדיוק משמעות הטענה ש- $Y[k]=\sum_{i=1}^{k}\alpha_i X_i[k]$). לכן $Y[k] = Z[k]$, וממילא $Y=Z$ לפי מסקנה 1 למעלה. מכאן $Y=\sum_{i=1}^{k}\alpha_i X_i$ (בניסוח אחר: Y ו- Z הם שני פתרונות של המשוואה, שמקיימים אותם תנאי התחלה, ולכן הם שווים).

שים לב: הטענה ש- $Y[k]=\sum_{i=1}^{k}\alpha_i X_i[k]$ פירושה ש-

$$\begin{bmatrix} Y(0) \\ Y(1) \\ \vdots \\ Y(k-1) \end{bmatrix} = \alpha_1 \begin{bmatrix} X_1(0) \\ X_1(1) \\ \vdots \\ X_1(k-1) \end{bmatrix} + \alpha_2 \begin{bmatrix} X_2(0) \\ X_2(1) \\ \vdots \\ X_2(k-1) \end{bmatrix} + \ldots + \alpha_k \begin{bmatrix} X_k(0) \\ X_k(1) \\ \vdots \\ X_k(k-1) \end{bmatrix}$$

היעזרות בכך תסייע בהבנת ההוכחה האחרונה.

מסקנה 5:

הממד של מרחב הפתרונות של משוואה לינארית הומוגנית מסדר k הוא k.

הוכחה:

יהי $\vec{e}_1,\ldots,\vec{e}_k$ בסיס כלשהו של $\mathbf{C}^k$ (למשל, הבסיס הקנוני). לפי משפט 1, קיים לכל $1 \le i \le k$ פתרון X_i למשוואה, המקיים $X_i[k]=\vec{e}_i$. לפי המשפט האחרון $X_1, \ldots, X_k$ אלו מהווים בסיס של מרחב הפתרונות, ולכן הממד שלו הוא k.

מסקנה 6:

א. אם $X_1, \ldots, X_k$ הם k פתרונות בלתי תלויים של משוואה לינארית הומוגנית מסדר k, אז הם מהווים בסיס למרחב הפתרונות של המשוואה.

ב. אם $X_1, \ldots, X_k$ הם k פתרונות של משוואה כנ"ל, כך ש- $X_1[k], \ldots, X_k[k]$ הם בלתי תלויים (ב- $\mathbf{C}^k$), אז $X_1, \ldots, X_k$ מהווים בסיס למרחב הפתרונות של המשוואה.

הוכחה

א. מיידי ממסקנה 5.

ב. נובע ממשפט 4 או מהתרגיל הבא:

תרגיל

אם $X_1,\ldots,X_\ell$ הם פתרונות של משוואה הומוגנית מסדר k, אז $X_1,\ldots,X_\ell$ בלתי תלויים (ב- $\mathbf{C} \to \mathbf{N}$) אס"ם $X_1[k],\ldots,X_\ell[k]$ בלתי תלויים (ב- $\mathbf{C}^k$).

מסקנה 7:

אם $X_1, \ldots, X_k$ כמו במסקנה 6 (א או ב), אז כל פתרון של המשוואה הוא מהצורה $\alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k$ עבור $\alpha_1, \ldots, \alpha_k \in \mathbf{C}$ יחידים.

במלים אחרות: כל פתרון הוא מהצורה $\lambda n. \sum_{i=1}^{k} \alpha_i X_i(n)$.

הגדרה:

אם $X_1, \ldots, X_k$ הם פתרונות בלתי תלויים של משוואה לינארית הומוגנית מסוימת מסדר k, אז לביטוי $\lambda n.\ \alpha_1 X_1(n) + \ldots + \alpha_k X_k(n)$ (כש- $\alpha_1, \ldots, \alpha_k$ משתנים חופשיים) קוראים ***פתרון כללי של המשוואה***.

ממה שראינו נובע, שכל פתרון ספציפי מתקבל מפתרון כללי על-ידי הצבת ערכים קונקרטיים במקום $\alpha_1, \ldots, \alpha_k$.

דוגמה:

נתבונן במשוואה הלינארית $a_n = 5a_{n-1} - 6a_{n-2}$, או:

$$a_{n+2} - 5a_{n+1} + 6a_n = 0$$

קל לברר ש- $X_1 = \lambda n.\ 2^n$ ו- $X_2 = \lambda n.\ 3^n$ הם פתרונות שלה. למשל, אם נציב את X_2 באגף השמאלי נקבל לכל n:

$$3^{n+2} - 5 \cdot 3^{n+1} + 6 \cdot 3^n = 3^n(9 - 15 + 6) = 3^n \cdot 0 = 0$$

בדוגמה זו $k = 2$ ו-

$$X_1[2] = \langle X_1(0), X_1(0) \rangle = \langle 2^0, 2^1 \rangle = \langle 1, 2 \rangle \qquad X_2[2] = \langle 3^0, 3^1 \rangle = \langle 1, 3 \rangle$$

עתה $\langle 1, 3 \rangle$ ו- $\langle 1, 2 \rangle$ הם בלתי-תלויים לינארית, כי $\begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} = 3 - 2 = 1 \neq 0$ ולכן, לפי מסקנה 6ב, X_1 ו- X_2 אלו מהווים בסיס למרחב הפתרונות של המשוואה הנ"ל.

הפתרון הכללי הוא אפוא:

$$\lambda n.\ \alpha_1 \cdot 2^n + \alpha_2 \cdot 3^n \qquad (\alpha_1, \alpha_2 \in \mathbf{C})$$

נניח עתה שנתונים תנאי התחלה כמו בפרק הקודם (שם פתרנו משוואה זו עם תנאי התחלה קונקרטיים בעזרת פונקציות יוצרות): $a_0 = 1$, $a_1 = -2$. עלינו למצוא את ה- α_1 וה- α_2 המתאימים. מה ש- α_1 ו- α_2 אלו צריכים לקיים הוא:

$$\begin{cases} \alpha_1 \cdot 2^0 + \alpha_2 \cdot 3^0 = 1 \\ \alpha_1 \cdot 2^1 + \alpha_2 \cdot 3^1 = -2 \end{cases}$$

במלים אחרות, עלינו לפתור את מערכת המשוואות:

$$\alpha_1 + \alpha_2 = 1$$
$$2\alpha_1 + 3\alpha_2 = -2$$

הפתרון של מערכת זו הוא $\alpha_1 = 5$, $\alpha_2 = -4$. מכאן שפתרון המשוואה עם תנאי ההתחלה הנ"ל הוא $\lambda n.\ 5 \cdot 2^n - 3 \cdot 4^n$, כמו שקיבלנו בסעיף הקודם.

טוב ויפה, אך איך מוצאים בדוגמה זו את הפתרונות הבסיסיים $\lambda n.\ 2^n$ ו- $\lambda n.\ 3^n$? ובאופן כללי יותר: איך מוצאים בסיס $X_1, \ldots, X_k$ למרחב הפתרונות של משוואת נסיגה הומוגנית לינארית מסדר k ? זהו הנושא הבא שלנו.

ההגדרה הבאה מכילה את המפתח לפתרון הבעיה.

הגדרה:

תהי $\lambda n.\sum_{i=0}^{k} c_{k-i}X(n+k-i)=\lambda n.0$ (או $\sum_{i=0}^{k} c_{k-i}a_{n+k-i}=0$) משוואה לינארית הומוגנית. הפולינום האופייני של משוואה זו הוא הפולינום:

$$c_kx^k + c_{k-1}x^{k-1} + \ldots + c_1x + c_0$$

דוגמה:

הפולינום האופייני של המשוואה $a_{n+2} - 5a_{n+1} + 6a_n = 0$ מהדוגמה האחרונה הוא $x^2 - 5x + 6$.

משפט 5:

אם α הוא שורש של הפולינום האופייני של משוואה לינארית הומוגנית מסוימת עם מקדמים קבועים, אז $\lambda n.\ \alpha^n$ הוא פתרון של המשוואה.

הוכחה:

נניח שהמשוואה היא:

$$\lambda n.\ c_kX(n+k) + c_{k-1}X(n+k-1) + \ldots + c_1X(n+1) + c_0X(n) = \lambda n.\ 0$$

הנתון ש- α שורש של הפולינום האופייני פירושו ש:

$$c_k\alpha^k + c_{k-1}\alpha^{k-1} + \ldots + c_1\alpha + c_0 = 0$$

נציב עתה $\lambda n.\ \alpha^n$ באגף השמאלי של משוואת הנסיגה. נקבל:

$$\lambda n.\ c_k\alpha^{n+k} + c_{k-1}\alpha^{n+k-1} + \ldots + c_1\alpha^{n+1} + c_0\alpha^n =$$

$$= \lambda n.\ \alpha^n(c_k\alpha^k + c_{k-1}\alpha^{k-1} + \ldots + c_1\alpha + c_0) \underset{\downarrow}{=} \lambda n.\ \alpha^n \cdot 0 = \lambda n.\ 0$$

כי α שורש של הפולינום האופייני

דוגמה:

השורשים של הפולינום $x^2 - 5x + 6$ (שהינו הפולינום האופייני של המשוואה בדוגמה, שפתרנו למעלה) הם 2 ו- 3. לכן, לפי משפט 5, $\lambda n.\ 2^n$ ו- $\lambda n.\ 3^n$ הם פתרונות של משוואה זו (כלומר, של $a_{n+2} - 5a_{n+1} + 6a_n = 0$).

עתה, בדוגמה זו שני פתרונות אלו יצאו בלתי תלויים ופורשים את מרחב הפתרונות. המשפט הבא מראה, שזה איננו מקרי.

משפט 6:

נניח $x_1, \ldots, x_k$ הם k שורשים שונים של הפולינום האופייני של משוואה לינארית הומוגנית מסדר k עם מקדמים קבועים. אז $\lambda n.\ \alpha_1 x_1{}^n + \ldots + \alpha_k x_k{}^n$ הוא פתרון כללי של המשוואה.

הוכחה:

אנו יודעים כבר ש- $\lambda n.\ x_1{}^n, \ldots, \lambda n.\ x_k{}^n$ הם k פתרונות שונים של המשוואה. נותר לכן לראות שהם בלתי-תלויים, כלומר: שה- k-רישות שלהן הן בלתי-תלויות. k-רישות אלה יוצרות את המטריצה:

$$\begin{bmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_k \\ x_1^2 & x_2^2 & \ldots & x_k^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{k-1} & x_2^{k-1} & \ldots & x_k^{k-1} \end{bmatrix}$$

מטריצה זו ידועה בשם מטריצת ואנ-דר-מונדה של $x_1, \ldots, x_k$. הדטרמיננטה שלה שווה ל-* $(x_2 - x_1) \cdot (x_3 - x_1) \ldots (x_k - x_1) \cdot (x_3 - x_2) \ldots (x_k - x_2) \ldots (x_k - x_{k-1})$, והיא שונה לכן מאפס אמ"ם $x_1, \ldots, x_k$ שונים כולם זה מזה. כיוון שזוהי ההנחה שלנו, פירוש הדבר, שה- k-רישות אכן בלתי-תלויות לינארית.

דוגמאות נוספות:

(1) סדרת פיבונאצ'י:

$$a_{n+2} = a_{n+1} + a_n$$

או

$$a_{n+2} - a_{n+1} - a_n = 0$$

* זהו תרגיל סטנדרטי באלגברה לינארית.

הפולינום האופייני כאן הוא $x^2 - x - 1$. השורשים שלו (פתרונות המשוואה $x^2 - x - 1 = 0$) הם $x_1 = \frac{1+\sqrt{5}}{2}$, $x_2 = \frac{1-\sqrt{5}}{2}$, והם שונים זה מזה. סדר המשוואה הוא 2. לכן פתרון כללי של המשוואה הוא:

$$\lambda n.\alpha_1\left(\frac{1+\sqrt{5}}{2}\right)^n + \alpha_2\left(\frac{1-\sqrt{5}}{2}\right)^n$$

עתה, בסדרת פיבונאצ'י הקלאסית $a_0 = 0$ ו- $a_1 = 1$. אם נציב, נקבל:

$$\begin{cases} \alpha_1 + \alpha_2 = 0 \\ \alpha_1.\frac{1+\sqrt{5}}{2} + \alpha_2.\frac{1-\sqrt{5}}{2} = 1 \end{cases}$$

פתרון המערכת הוא: $\alpha_1 = \frac{1}{\sqrt{5}}$, $\alpha_2 = -\frac{1}{\sqrt{5}}$.

לכן נוסחת האיבר הכללי של סדרת פיבונאצ'י הקלאסית היא:

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right]$$

הערות:

1. בתרגיל בו התחלנו סעיף זה היו תנאי ההתחלה $a_0 = 1$ ו- $a_1 = 2$, ויכולנו למצוא α_1 ו- α_2 מתאימים באותה שיטה. אפשר גם לשים לב,שהסדרה שם היא בדיוק $\lambda n.\, F_{n+2}$, ולכן התשובה שם היא:

$$\frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+2} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+2}\right]$$

2. כדאי לשים לב, שלמרות שסדרת פיבונאצ'י היא סדרה של מספרים טבעיים, הנוסחה המפורשת עבורה משתמשת במספר אי-רציונלי $(\sqrt{5})$!

(2) סדרות הנדסיות:

נמצא, למשל, נוסחה לסדרה הנדסית, שבה $q = 3$ ו- $a_1 = 7$. הסדרה מוגדרת על-ידי:

$$\begin{cases} a_1 = 7 \\ a_{n+1} = 3a_n \end{cases}$$

משוואת הנסיגה כאן היא: $\lambda n.\, X(n+1) - 3X(n) = \lambda n.\, 0$.

הפולינום האופייני שלה הוא $x - 3$. השורש היחיד שלו הוא 3.

לכן פתרון כללי של המשוואה הינו: $\lambda n.\, \alpha \cdot 3^n$.

כדי למצוא את α נציב $n = 1$ ונקבל:

$$\alpha \cdot 3 = 7$$

$$\alpha = \frac{7}{3}$$

ולכן הסדרה היא $\lambda n. \frac{7}{3} \cdot 3^n = \lambda n. 7 \cdot 3^{n-1}$

הערה:

אנו רואים בדוגמה, שתנאי התחלה לא חייב להיות דווקא מהצורה $X(0) = c_0$!

מה קורה כאשר מספר השורשים של הפולינום האופייני קטן ממעלת הפולינום? לפני שניגש לטפל במקרה זה להלן קצת חומר רקע: לפי מה שידוע כ"משפט היסודי של האלגברה" (אותו מוכיחים רק בשנה ב), כל פולינום p (אפילו כזה שמקדמיו מרוכבים) ניתן לפרק בצורה הבאה:

$$p(x) = a_k (x - x_1)^{r_1} (x - x_2)^{r_2} \dots (x - x_\ell)^{r_\ell} \qquad r_1, \dots, r_k \ge 1$$

כאשר מעלת הפולינום היא $k = \sum_{i=1}^{\ell} r_i$, a_k הוא המקדם של x^k $(a_k \neq 0)$ ו- $x_1, \dots, x_\ell$ הם שורשיו של הפולינום (שיכולים, כמובן, להיות מספרים מרוכבים!). המספר הטבעי r_i נקרא *הריבוי* של השורש x_i. המשפט הבא עונה על השאלה, איך לפתור את המשוואה, כאשר הריבוי של חלק מהשורשים גדול מאחד. משפט זה מהווה הכללה של המשפט הקודם, ועל ההוכחה שלו נוותר כאן.

משפט 7:

נניח שהפולינום האופייני של משוואת נסיגה הומוגנית מסדר k הוא:

$$p(x) = a(x-x_1)^{r_1}(x-x_2)^{r_2}\ldots(x-x_\ell)^{r_\ell} \quad \left(k = \sum_{i=1}^{\ell} r_i\right)$$

אז בסיס למרחב הפתרונות ניתן על-ידי:

$$\{\lambda n.x_1^n, \lambda n.nx_1^n, \ldots, \lambda n.n^{r_1-1}x_1^n, \lambda n.x_2^n, \ldots, \lambda n.n^{r_2-1}x_2^n, \ldots, \lambda n.x_\ell^n, \ldots, \lambda n.n^{r_\ell-1}x_\ell^n\}$$

במלים אחרות: פתרון כללי של המשוואה הוא:

$$\lambda n.\alpha_1 x_1^n + \alpha_2 n\cdot x_1^n + \ldots + \alpha_{r_1} n^{r_1-1}x_1^n + \alpha_{r_1+1}x_2^n + \alpha_{r_1+2}n\cdot x_2^n + \ldots +$$

$$+ \ldots + \alpha_{r_1+r_2} n^{r_2-1}\cdot x_2^n + \ldots + \alpha_k n^{r_\ell-1}\cdot x_\ell^n$$

הערה:

הרעיון הוא, שכל שורש x_i תורם r_i פתרונות לבסיס הנ"ל:

$$\lambda n.x_i^n,\ \lambda n.nx_i^n,\ \lambda n.n^2x_i^n, \ldots, \lambda n.n^{r_i-1}x_i^n$$

דוגמה:

נפתור את המשוואה:

$$a_{n+2} = 4(a_{n+1} - a_n)$$

כלומר:

$$a_{n+2} - 4a_{n+1} + 4a_n = 0$$

הפולינום האופייני הינו:

$$p(x) = x^2 - 4x + 4 = (x-2)^2$$

יש פה שורש אחד $(x_1 = 2)$, והריבוי שלו הוא 2. לכן פתרון כללי למשוואה הוא:

$$\lambda n.\ \alpha_1 \cdot 2^n + \alpha_2 n \cdot 2^n$$

או

$$\lambda n.\ 2^n(\alpha_1 + \alpha_2 n)$$

בפרט אם ניקח את תנאי ההתחלה $a_0 = 3$ ו- $a_1 = 8$ (כמו בפרק הקודם), נקבל:

$$\alpha_1 = 3$$

$$2(\alpha_1 + \alpha_2) = 8 \quad \Rightarrow \quad \alpha_2 = 1$$

והפתרון לכן הוא:

$$\lambda n.\ 2^n(3 + n)$$

אין ספק, שבמקרה זה הפתרון נמצא באופן קל ומהיר יותר מאופן מציאת הפתרון בעזרת פונקציות יוצרות (אותו ביצענו בפרק הקודם)!

טבלה ד.6 מסכמת את העובדות המרכזיות הקשורות בפתרון משוואות נסיגה לינאריות הומוגניות עם מקדמים קבועים.

טבלה ד.6
משוואות נסיגה לינאריות הומוגניות עם מקדמים קבועים

נניח כי $L(X) = h$ היא משוואה כנ"ל מסדר k.

משפט: לכל $\vec{V} \in \mathbf{C}^k$ קיים פתרון $X_{\vec{v}}$ יחיד למשוואה, המקיים $X_{\vec{v}}[k] = \vec{v}$.

מסקנה: אם X_1 ו- X_2 שני פתרונות של המשוואה כך ש- $X_1[k] = X_2[k]$, אז $X_1 = X_2$.

משפט: א. L היא העתקה לינארית מ- $\mathbf{N} \to \mathbf{C}$ אל $\mathbf{N} \to \mathbf{C}$.
ב. קבוצת הפתרונות של $L(X) = \lambda n.\ 0$ היא תת-מרחב של $\mathbf{N} \to \mathbf{C}$.

משפט: אם $X_1, \ldots, X_k$ פתרונות של $L(X) = \lambda n.\ 0$, ו- $X_1[k], \ldots, X_k[k]$ בסיס של $\mathbf{C}^k$, אז: $X_1, \ldots, X_k$ בסיס למרחב הפתרונות של $L(X) = \lambda n.\ 0$.

מסקנה: ממד מרחב הפתרונות של $L(X) = \lambda n.\ 0$ הוא k.

משפט: אם α שורש של הפולינום האופייני של משוואת נסיגה לינארית הומוגנית עם מקדמים קבועים, אז $\lambda n.\ \alpha^n$ פתרון של המשוואה הנ"ל.

משפט: אם הפולינום האופייני של משוואה כנ"ל הוא:

$$p(x) = a(x-\alpha_1)^{r_1}(x-\alpha_2)^{r_2}\ldots(x-\alpha_\ell)^{r_\ell} \qquad \left(\sum_{i=1}^{\ell} r_i = k\right)$$

אז פתרון כללי למשוואה הוא:

$$\lambda n. c_1\alpha_1^n + c_2 n\alpha_1^n + \ldots + c_{r_1} n^{r_1-1}\alpha_1^n + \ldots + c_{k-r_\ell+1}\alpha_\ell^n + \ldots \quad + c_k n^{r_\ell-1}\alpha_\ell^n$$

%%

בעיה:

מה נעשה אם כל מקדמי המשוואה הם מספרים ממשיים, תנאי ההתחלה קשורים רק במספרים ממשיים, אנו מתעניינים רק בפתרונות, שהם סדרות של מספרים ממשיים (בקיצור: כאשר מספרים מרוכבים לא מעניינים אותנו בכלל), אבל לפולינום האופייני יש שורשים לא-ממשיים?

תשובה:

עקרונית, אין בזה כל רע. כבר ראינו למעלה, שבסדרת פיבונאצ'י, בה כל המקדמים הם טבעיים, תנאי ההתחלה קשורים במספרים טבעיים, והפתרון הינו סדרה של טבעיים – הנוסחה עבור פתרון זה מתבססת על חזקות של מספרים, שלא רק שאינם טבעיים – הם אפילו אינם מספרים רציונליים ($\frac{1-\sqrt{5}}{2}$ ו- $\frac{1+\sqrt{5}}{2}$). אין אפוא שום פסול בכך, שנוסחה עבור סדרות של מספרים ממשיים (ואפילו טבעיים) תשתמש במספרים מרוכבים. עם זאת, אם מתעקשים, אפשר כאן לעשות רדוקציה לשימוש בפתרונות ממשיים על-סמך העובדה הבאה: אם מקדמיו של פולינום הם ממשיים, ו- $z=\alpha+\beta i$ הוא שורש של אותו פולינום, אז כך גם $\bar{z}=\alpha-\beta i$, והריבוי של שניהם הינו שווה. עתה, אם $\alpha+\beta i=r(\cos\theta+i\sin\theta)$, אז

$$(\alpha+\beta i)^n=r^n(\cos n\theta+i\sin n\theta)$$
$$(\alpha-\beta i)^n=r^n(\cos n\theta-i\sin n\theta)$$

$$r^n\cos n\theta=\frac{(\alpha+\beta i)^n+(\alpha-\beta i)^n}{2}$$

$$r^n\sin n\theta=\frac{(\alpha+\beta i)^n-(\alpha-\beta i)^n}{2i}$$

מכאן ש- $\{\lambda n.\,(\alpha+i\beta)^n,\ \lambda n.\,(\alpha-i\beta)^n\}$ ו- $\{\lambda n.\ r^n\cos n\theta,\ \lambda n.\ r^n\sin n\theta\}$ פורשים אותו מרחב, ואפשר להשתמש בזוג השני במקום בזוג הראשון. בזוג זה אכן אין כל התייחסות למספרים מרוכבים.

דוגמה:

$$a_{n+2}-2a_{n+1}+2a_n=0$$

הפולינום האופייני הוא:

$$x^2-2x+2$$

שורשיו:

$$1 \pm i = \sqrt{2}\left(\cos\frac{\pi}{4} \pm i\sin\frac{\pi}{4}\right)$$

לכן פתרון כללי של המשוואה הינו:

$$\lambda n. \sqrt{2}^n\left(\alpha_1 \cos\frac{\pi}{4}n + \alpha_2 \sin\frac{\pi}{4}n\right)$$

%%

פתרון משוואות לא הומוגניות

נעבור עתה לדיון במשוואות מהצורה $L(X) = f$ כש- $f \neq \lambda n.\ 0$, דהיינו: במשוואות לא-הומוגניות. משפט המפתח כאן הוא המשפט הבא:

משפט 8:

יהי X_p פתרון כלשהו של המשוואה הלינארית $L(X) = f$ (L – כמו למעלה), ותהי H קבוצת הפתרונות של המשוואה ההומוגנית המתאימה $L(X) = \lambda n.\ 0$, אז:

$$\{X \in \mathbf{N} \to \mathbf{C} \mid L(X) = f\} = \{Y + X_p \mid \ Y \in H\}$$

הוכחה:

נסמן את אגף שמאל של השוויון למעלה ב- A, ואת אגף ימין ב- B.

הוכחה ש- $\boldsymbol{A \subseteq B}$:

נניח $X \in A$. אז $L(X) = f$. נגדיר $Y_0 = X - X_p$. אז:

$$L(Y_0) = L(X) - L(X_p) = f - f = \lambda n.\ 0$$

(כי L העתקה לינארית). לכן $Y_0 \in H$. כמו-כן $X = Y_0 + X_p$. לכן $X \in B$.

הוכחה ש- $\boldsymbol{B \subseteq A}$:

נניח $Z \in B$. אז קיים $Y \in H$ (כלומר: $L(Y) = \lambda n.\ 0$) כך ש- $Z = Y + X_p$. לכן:

$$L(Z) = L(Y) + L(X_p) = \lambda n.\ 0 + f = f$$

מכאן ש- Z פתרון של $L(X) = f$, ולכן $Z \in A$.

משמעות המשפט האחרון היא, שכדי למצוא פתרון כללי למשוואה הלא-הומוגנית $L(X) = f$, עלינו לעשות שני דברים:

(א) למצוא פתרון *כללי* למשוואה ההומוגנית המתאימה $L(X) = \lambda n.\ 0$.

(ב) למצוא פתרון *פרטי* כלשהו למשוואה $L(X) = f$.

הסכום של מה שמוצאים ב-(א) וב-(ב) נותן, לפי המשפט, פתרון כללי ל- $L(X) = f$.

בסעיפים הקודמים טיפלנו כבר בשאלה, כיצד לבצע את שלב (א). אשר לשלב (ב) – כאן אנו חוזרים בעצם לבעיה של פתירת $L(X) = f$. החידוש הוא, שכעת מסתפקים אנו במציאת פתרון *אחד*, בכל דרך שהיא. בקשר לעניין זה יש חדשות טובות וחדשות רעות:

החדשות *הרעות* הן, שאין שיטה כוללת לעשות זאת, שתעבוד עבור כל f.

החדשות *הטובות* הן, שעבור מספר סוגים חשובים ונפוצים של פונקציות f *יש* שיטה איך למצוא פתרון פרטי יחיד כלשהו למשוואה $L(X) = f$. אנו נסתפק כאן במחלקה רחבה אחת: קבוצת הפונקציות מהצורה:

$$\lambda n.\ a^n P(n)$$

כאשר $a \in \mathbf{C}$ (ייתכן גם ש- $a = 1$!) ו- P פולינום (אולי ממעלה 0, כלומר קבוע).

העיקרון (אותו לא נוכיח) הוא:

עיקרון:

אם P פולינום ממעלה ℓ ו- a שורש מריבוי i של הפולינום האופייני (כולל המקרה $i = 0$, כלומר: כש- a *אינו* שורש של הפולינום האופייני), אז יש פתרון למשוואה $L(X) = \lambda n.\ a^n P(n)$ מהצורה $f = \lambda n.\ n^i a^n Q(n)$, כאשר Q פולינום ממעלה ℓ לכל היותר. את המקדמים של הפולינום Q מוצאים על-ידי הצבה במשוואה וחילוץ.

דוגמה 1: משוואת מגדלי האנוי:

$$a_{n+1} - 2a_n = 1$$

כאן $a = 1$, ו- P הוא הפולינום $P(n) = 1$, שמעלתו $\ell = 0$. הפולינום האופייני של המשוואה הוא $x - 2$. לכן פתרון כללי למשוואה ההומוגנית המתאימה הוא: $\lambda n.\ \alpha \cdot 2^n$. עתה a כאן אינו שורש של הפולינום האופייני. לכן יש פתרון פרטי מהצורה $\lambda n.\ n^0 \cdot 1^n \cdot d$ כש- $d \in \mathbf{C}$ (מעלת Q היא 0, כמו זו של P !). במלים אחרות, יש קבוע d כך ש- $\lambda n.\ d$ פתרון של המשוואה.

נציב במשוואה הרקורסיבית (דהיינו, במשוואות הנסיגה). נקבל:

$$d - 2d = 1$$
$$-d = 1$$
$$d = -1$$

לכן פתרון כללי של משוואת הנסיגה כאן הוא $\lambda n.\ \alpha \cdot 2^n - 1$. במקרה של מגדלי האנוי אנו יודעים ש- $a_0 = 0$. לכן:

$$\alpha \cdot 2^0 - 1 = 0 \ \Rightarrow \ \alpha = 1$$

והתשובה הסופית היא לכן:

$$a_n = 2^n - 1$$

הערה:

תשובה זהה קיבלנו קודם בעזרת פונקציות יוצרות. נזכיר בהזדמנות זו, שהשימוש בפונקציות יוצרות הוא כלי נשק אלטרנטיבי להתקפה על משוואות לא-הומוגניות.

דוגמה 2:

נחזור לדוגמה של מספר המחרוזות באורך n שאפשר לבנות מ- $\{0, 1, 2, 3\}$, שיש בהן מספר זוגי של אפסים. משוואת הנסיגה ותנאי ההתחלה היו, כזכור:

$$\begin{cases} a_{n+1} = 2a_n + 4^n \\ a_0 = 1 \end{cases}$$

כלומר:

$$\begin{cases} a_{n+1} - 2a_n = 4^n \\ a_0 = 1 \end{cases}$$

שוב, הפולינום האופייני הוא $x - 2$, 2 הוא השורש היחיד שלו, והפתרון הכללי של המשוואה ההומוגנית הוא $\lambda n.\ 2^n\alpha$. לעומת זאת, $a = 4$ כאן, והוא שוב אינו שורש

של הפולינום האופייני. לפי העיקרון המרכזי אנו יודעים שיש d כך ש- $\lambda n.\, 4^n d$ פתרון של המשוואה. נציב, ונקבל שלכל n צריך להתקיים:

$$d \cdot 4^{n+1} - 2d \cdot 4^n = 4^n$$

ואם נחלק ב- 4^n, נקבל:

$$d = ½ \quad \Leftarrow \quad 4d - 2d = 1$$

פתרון כללי של המשוואה הוא לכן: $\lambda n.\alpha \cdot 2^n + \frac{1}{2} \cdot 4^n$. אם נציב $a_0 = 1$ נקבל ש- $\alpha = ½$, והפתרון לכן הינו $\lambda n.\frac{2^n + 4^n}{2}$.

דוגמה 3:

$$a_{n+1} - 2a_n = 4^n (n+1)$$

הפולינום האופייני נשאר כמו בדוגמאות הקודמות. לעומת זאת, בדוגמה זו $a = 4$, ו- $P(n) = n + 1$ הוא פולינום ממעלה ראשונה. לפי העיקרון יש לכן A ו- B כך ש- $\lambda n.\, 4^n(An + B)$ פתרון של המשוואה. הצבה תיתן:

$$4^{n+1}(A(n+1) + B) - 2 \cdot 4^n(An + B) = 4^n(n+1)$$

נחלק ב- 4^n:

$$4(A(n+1) + B) - 2(An + B) = n + 1$$

זו צריכה להיות *זהות*, ולכן מהשוואת מקדמים נובע:

$$\begin{cases} 2A = 1 \\ 4A + 2B = 1 \end{cases}$$

ונקבל לכן ש- $A = ½$ ו- $B = -½$.

מכאן, שפתרון פרטי למשוואה הוא: $\lambda n.\frac{1}{2}4^n(n-1)$.

ופתרון כללי: $\lambda n.\alpha \cdot 2^n + \frac{1}{2}4^n(n-1)$.

דוגמה 4:

$$a_{n+1} - 2a_n = 2^n\,(2n + 3)$$

ההבדל בין דוגמה זו ובין הקודמות הוא, שכאן $a = 2$, ולכן a *הינו* פתרון של הפולינום האופייני. הריבוי של a הוא 1. אנו מחפשים לכן עתה A ו- B , כך ש- $\lambda n.\ n \cdot 2^n\,(An + B)$ הינו פתרון של המשוואה. הצבה תיתן כאן:

$$(n + 1) \cdot 2^{n+1}(A(n + 1) + B) - 2n \cdot 2^n(An + B) = 2^n(2n + 3)$$

$$\Rightarrow \qquad 2(n + 1)(An + A + B) - 2n(An + B) = 2n + 3$$

אפשר להמשיך כאן כמו בדוגמה הקודמת: לפתוח סוגריים, לכנס ולהשוות מקדמים (כדאי לשים לב, שהגורם עם n^2 מתבטל!). אלטרנטיבית, אפשר להציב ערכים:

$$\begin{array}{ll} n = 0 & \Rightarrow \left\{ \begin{array}{c} 2(A + B) = 3 \\ \\ 4(2A + B) - 2(A + B) = 5 \end{array} \right. \\ n = 1 & \Rightarrow \end{array}$$

פתרון המשוואות נותן $A = ½$, $B = 1$. לכן פתרון כללי למשוואה הוא:

$$\lambda n.\alpha \cdot 2^n + n \cdot 2^n\left(\frac{1}{2}n + 1\right) = \lambda n.\ 2^n\left(\frac{1}{2}n^2 + n + \alpha\right)$$

הערה:
נוכל להרחיב במידה ניכרת את אוסף הפונקציות f , עבורן יש לנו דרך ישירה לפתור את $L(X) = f$, בעזרת האבחנה הבאה:
אם $f = f_1 + f_2 + \ldots + f_j$ ו- $X_1, \ldots, X_j$ פתרונות של $L(X_i) = f_i$ $(i = 1, \ldots, j)$, אז $X_1 + X_2 + \ldots + X_j$ פתרון של $L(X) = f$.

הוכחה:

$$L(X_1 + \ldots + X_j) = L(X_1) + \ldots + L(X_j) = f_1 + \ldots + f_j = f$$

(יש כאן שימוש חזק בעובדה, ש- L הינה העתקה לינארית!).

דוגמה 5:

יש לפתור:

$$a_{n+1} - 2a_n = 4^n(n + 1) + 2^n(2n + 3)$$

משתי הדוגמאות האחרונות נובע, שפתרון משוואה זו הוא:

$$\lambda n.\alpha \cdot 2^n + \frac{1}{2}4^n(n-1) + n \cdot 2^n\left(\frac{1}{2}n+1\right) =$$

$$= \lambda n.\frac{1}{2}4^n(n-1) + 2^n\left(\frac{1}{2}n^2 + n + \alpha\right)$$

דוגמה 6:

נחזור לסיום לבעיה של מציאת נוסחה ל- $a_n = 1^2 + 2^2 + \ldots + n^2$, הפעם בעזרת משוואות נסיגה. מתקיים כאן:

$$a_{n+1} - a_n = (n + 1)^2 = 1^n(n^2 + 2n + 1)$$

זוהי משוואה לא הומוגנית מהסוג שאנו *יודעים* לפתור! כאן $a = 1$ ו- $\ell = 2$. הפולינום האופייני הוא $x - 1$, והשורש שלו, 1, הוא מריבוי 1 (ושווה ל- a !). לכן יש פתרון פרטי מהצורה $\lambda n.\ 1^n \cdot n^1(An^2 + Bn + C)$, כלומר $\lambda n.\ n(An^2 + Bn + C)$.
הצבה במשוואה תיתן:

$$(n + 1)(A(n + 1)^2 + B(n + 1) + C) - n(An^2 + Bn + C) = (n + 1)^2$$

זה אמור להיות נכון לכל n טבעי, ולכן (השיקול הפולינומיאלי!) לכל $x \in R$. נציב את הערכים $-1, 0, 1$ ונקבל:

$$\begin{aligned} n = -1 &\Rightarrow A - B + C = 0 \\ n = 0 &\Rightarrow A + B + C = 1 \\ n = 1 &\Rightarrow 2(4A + 2B + C) - (A + B + C) = 4 \end{aligned}$$

פתרון מערכת זו הוא: $C = \frac{1}{6}$, $B = \frac{1}{2}$, $A = \frac{1}{3}$.

לכן פתרון פרטי של המשוואה הוא:

$$\lambda n.n\left(\frac{1}{3}n^2+\frac{1}{2}n+\frac{1}{6}\right)=\lambda n.\frac{n(2n^2+3n+1)}{6}$$

$$=\lambda n.\frac{n(n+1)(2n+1)}{6}$$

פתרון כללי של המשוואה ההומוגנית $a_{n+1}-a_n=0$ הוא $\lambda n.\ \alpha$.
לכן, פתרון כללי של המשוואה הלא-הומוגנית הוא:

$$\lambda n.\alpha+\frac{n(n+1)(2n+1)}{6}$$

הצבת תנאי ההתחלה $a_1=1$ (או אפילו $a_0=0$) תראה ש- $\alpha=0$. לכן התשובה הסופית היא:

$$\lambda n.\frac{n(n+1)(2n+1)}{6}$$

ה. יסודות תורת הגרפים

ה.1 מושגי יסוד

תורת הגרפים היא מכשיר מתמטי רב-עוצמה, המשמש לפתרון סוגים רבים של בעיות. הסיבה לכך הינה, שמצבים ובעיות רבים ניתן לייצג בעזרת גרפים סופיים.

פירוש מושג ה"גרף" בתורת הגרפים שונה מאוד מהמושג של "גרף של פונקציה", המוכר לקוראים. בתורת הגרפים גרף הינו קבוצה של נקודות (הנקראות "קדקודי הגרף") יחד עם קבוצה של קווים המחברים ביניהן (ונקראים "קשתות הגרף"). מפות עירוניות מספקות דוגמה טובה לגרפים במובן זה. במפה כזו הרחובות (ליתר דיוק: קטעים של רחובות) משמשים כקשתות, והצמתים מתפקדים כקדקודים. כל קטע של רחוב *מחבר* שני צמתים. שני צמתים אלו הינם קדקודי הקצה (או פשוט: הקצוות) של אותו קטע, והם נחשבים *סמוכים* זה לזה. כל אלה הינן דוגמאות של מושגים יסודיים של תורת הגרפים. בדומה, בעיות כגון "האם אפשר לצעוד דרך כל הרחובות של רמת אביב ג' בלי לחזור על אף קטע-רחוב פעמיים?", או "מה המסלול הקצר ביותר מהאוניברסיטה לתחנת הרכבת?" הן בעיות אופייניות של תורת הגרפים.

נשים לב שמפות עירוניות נחלקות לשני סוגים עיקריים: אלו שנועדו לשימושם של הולכי רגל, ואלו שנועדו לשימושם של נהגי מכוניות. במפות, שנועדו לנהגים, מצוין ביחס לכל קטע-רחוב, באיזה כיוון מותר לנסוע בו. הקשתות במפות אלו הינן אפוא *מכוונות*. בקשתות כאלה אנו מבחינים בין קדקוד היציאה של הקשת (הקדקוד בו היא מתחילה) וקדקוד הכניסה שלה. מפות להולכי רגל, לעומת זאת, הן דוגמה *לגרפים לא מכוונים*. בגרפים כאלה לכל קשת יש עדיין שני קצוות, אך לשניהם יש מעמד זהה.

מספר הערות

(1) כיוון שבגרף מכוון לכל קשת יש כיוון מוגדר, הרי לרחובות דו-סטריים (במפות עירוניות לנהגים) מתייחסים בתורת הגרפים כמייצגים שתי קשתות שונות, עם כיוונים הפוכים. כל צד של רחוב כזה הוא בעצם קשת נפרדת (לזוג קשתות כאלה קוראים קשתות *אנטי-מקבילות*).

(2) קורה לעתים, ששני צמתים מחוברים על-ידי שתי קשתות שונות (או יותר). קשתות כאלה, שיש להן אותן קצוות, נקראות *מקבילות* (כשמדובר בגרף מכוון, הרי ב"אותן קצוות" הכוונה, שלשתי הקשתות יש אותו קדקוד יציאה ואותו קדקוד כניסה).[1]

(3) ייתכן גם מצב, שבו קשת מחברת קדקוד מסוים עם עצמו (כיכרות עם מוצא יחיד, למשל). קשתות כאלה נקראות *לולאות*. גרפים, שאין בהם לולאות או קשתות מקבילות, נקראים גרפים *פשוטים*.

(4) במפות עירוניות ייתכנו גם מדרחובים. נשים לב, שמבחינת הנהגים מדרחובים אינם מייצגים קשתות, בעוד שמבחינת הולכי הרגל – כן. מפות, שנועדו הן לשימושם של הולכי רגל והן לשימושם של נהגים, מהוות לכן דוגמה למבנים מתמטיים, שהם מסובכים יותר מאשר גרפים.

הן בהערה האחרונה והן בהקדמה התייחסנו לגרפים כאל "מבנים מתמטיים". עד כה דנו אבל בגרפים מנקודת מבט אינטואיטיבית בלבד. כדי שנוכל אכן לראותם כמבנים מתמטיים ולטפל בהם באופן מתמטי, עלינו לספק למושגים השונים הגדרות מתמטיות מדויקות. כרגיל במתמטיקה המודרנית, הגדרות אלו ניתנות במסגרת (השפה של) תורת הקבוצות. עתה, בחינה מדוקדקת של ההסברים, שהבאנו למעלה, תראה, שאנו מכירים גרף לחלוטין, אם אנו יודעים (א) מי הם הקדקודים; (ב) מי הן הקשתות; (ג) אילו קדקודים מקשרת כל קשת. בהתאם נגדיר:

גרף מכוון הוא שלשה $G = \langle V, E, F \rangle$, שבה V ו- E הן קבוצות זרות, $V \neq \varnothing$, ו- $F : E \to V^2$.

V נקראת קבוצת הקדקודים של הגרף G.

E נקראת קבוצת הקשתות של הגרף G.

אם $e \in E$ ו- $F(e) = \langle v_1, v_2 \rangle$, אז אומרים ש-$e$ מחברת את v_1 ל-v_2, ש-v_1 היא קדקוד היציאה (או ההתחלה) של e, ש-v_2 הוא קדקוד הכניסה (או הסוף) שלה, ש-v_1 ו-v_2 נמצאים (או חלים) על e, וש-v_1 ו-v_2 הם *סמוכים*.

אם $e_1 \in E$, $e_2 \in E$, ו- $F(e_1) = F(e_2)$, אז e_1 ו-e_2 נקראות קשתות מקבילות.

אם $e \in E$ ו- $\pi_1(F(e)) = \pi_2(F(e))$, אז e נקראת לולאה.

גרף מכוון $\langle V, E, F \rangle$ נקרא *פשוט* אם F חח"ע ו- $\pi_1(F(e)) \neq \pi_2(F(e))$ לכל $e \in E$ (דהיינו: $F(E) \cap i_V = \varnothing$).

[1] בטקסטים מסוימים, גרפים עם קשתות מקבילות נקראים "מולטיגרפים", בעוד מושג ה"גרף" נשמר לגרפים, שאין בהם קשתות מקבילות. גם אנו ננהג כך לפעמים.

גרף לא מכוון הוא שלשה $G = <V, E, F>$, שבה V ו- E הן קבוצות זרות, $V \neq \varnothing$, ו- $F : E \to P_1(V) \cup P_2(V)$ (ראו פרק ג.6).
V נקראת קבוצת הקדקודים של הגרף G.
E נקראת קבוצת הקשתות של הגרף G.

אם $e \in E$ ו- $F(e) = \{v_1, v_2\}$ (כולל המקרה, שבו $v_1 = v_2$), אז v_1 ו-v_2 נקראים הקצוות של e, ואומרים ש-e מחברת את v_1 ל-v_2, ש-v_1 ו-v_2 *נמצאים* (או חלים) על e, וש-v_1 ו-v_2 הם *סמוכים*.
אם $e_1 \in E$, $e_2 \in E$ ו- $F(e_1) = F(e_2)$, אז e_1 ו-e_2 נקראות קשתות מקבילות.
אם $e \in E$ ו- $F(e) \in P_1(V)$, אז e נקראת לולאה.
גרף לא מכוון $<V, E, F>$ נקרא *פשוט*, אם F חח"ע, ו- $F : V \to P_2(E)$.

הגדרה, שבה גרף הוא שלשה, היא מעט מסורבלת. הגדרה כזאת היא הכרחית למעשה, רק אם מרשים קשתות מקבילות (מה שקוראים לפעמים מולטיגרף). כאשר עוסקים בגרף ללא קשתות מקבילות, אפשר לזהות קשת עם זוג הקצוות שלה (זוג סדור, במקרה שהגרף מכוון, זוג לא סדור, שעשוי להיות בעצם סינגלטון, אם הגרף אינו מכוון). זיהוי זה מאפשר את ההגדרות החילופיות הבאות:

גרף מכוון ללא קשתות מקבילות הוא זוג $G = <V, E>$, כך ש- $V \neq \varnothing$, ו- $E \subseteq V^2$ (כלומר: E הוא יחס על V). גרף כזה נקרא *פשוט* אם E יחס אי-רפלקסיבי.

גרף לא מכוון ללא קשתות מקבילות הוא זוג $G = <V, E>$, כך ש- $V \neq \varnothing$, ו- $E \subseteq P_1(V) \cup P_2(V)$. גרף כזה נקרא פשוט אם $E \subseteq P_2(V)$.

כמו בהגדרות הקודמות, איברי V נקראים קדקודי הגרף, ואיברי E – הקשתות שלו. במקרה הלא מכוון, אם $\{v_1, v_2\}$ היא קשת של G (כלומר $\{v_1, v_2\} \in E$), אז v_1 ו- v_2 הם קדקודים סמוכים, המהווים את *קצות* הקשת הזו. קשת מהצורה $\{v\}$ (דהיינו: סינגלטון) נקראת *לולאה*. כרגיל, הגרף הינו פשוט אם ורק אם אין בו לולאות.

במקרה המכוון, אם $<v_1, v_2>$ הוא קשת של G (כלומר: $<v_1, v_2> \in E$), אז v_1 הוא קדקוד היציאה, ו- v_2 הוא קדקוד הכניסה של קשת זו. קשת מהצורה $<v, v>$ נקראת *לולאה*. שוב, גרף כזה הינו פשוט אם ורק אם אין בו לולאות.

הבאנו למעלה שתי הגדרות שונות למושג הגרף. מה הקשר ביניהן? התשובה פשוטה: אם $\langle V, E\rangle$ הינו גרף לפי ההגדרה השנייה, הרי $\langle V, E, i_E\rangle$ הינו גרף (ללא קשתות מקבילות) לפי ההגדרה הראשונה. לפי הגדרותינו, אין כל הבדל ממשי בין "שני" הגרפים האלה, כיוון שלא רק שיש להם אותן קשתות ואותם קדקודים, אלא גם ההגדרות ה"שונות" נותנות תשובות זהות לשאלות הבסיסיות, כגון: מי הם קדקודי הקצה של קשת נתונה? אילו קשתות הן לולאות? האם הגרף הוא פשוט? וכו'. כל עוד אין קשתות מקבילות, הבחירה עם איזו הגדרה לעבוד היא שאלה של נוחות וטעם. בספר זה נשתמש בהגדרה הראשונה (גרף כשלשה), כאשר המשפטים בהם נעסוק יהיו תקפים ובעלי עניין גם למולטיגרפים. לעומת זאת, נעדיף לעתים את ההגדרה השנייה (גרף כזוג), כאשר התיאוריה תתרכז בגרפים ללא קשתות מקבילות (ובפרט אם היא תעסוק בגרפים פשוטים). מושגים וטענות, שינוסחו על-פי ההגדרה הראשונה, אפשר תמיד לתרגם לגרפים המאופיינים כזוג $\langle V, E\rangle$ על-ידי כך שמסתכלים על "$\langle V, E\rangle$" כקיצור של "$\langle V, E, i_E\rangle$".

תיארנו למעלה גם שני סוגים שונים של גרפים: גרף מכוון וגרף לא מכוון. גם כאן יש קשרים ברורים: בהינתן גרף מכוון, אפשר לקבל ממנו באופן טבעי גרף לא מכוון על-ידי "מחיקת" הכיוונים של הקשתות. ולהפך, מכל גרף לא מכוון ניתן לקבל באופן טבעי הרבה גרפים מכוונים על-ידי בחירת כיוון לכל קשת, שאינה לולאה.[2] ניסוח מתמטי מדויק של רעיונות אלה מוביל להגדרה הבאה:

הגדרה:

יהיו $G = \langle V, E, F\rangle$ גרף לא מכוון ו- $G' = \langle V', E', F'\rangle$ גרף מכוון. נאמר ש-G' *מבוסס* על G, וש- G *מושרה* על-ידי G', אם:

(א) $V = V'$

(ב) $E = E'$

(ג) $F = \lambda e \in E.\ \{\pi_1(F'(e))\ ,\ \pi_2(F'(e))\}$

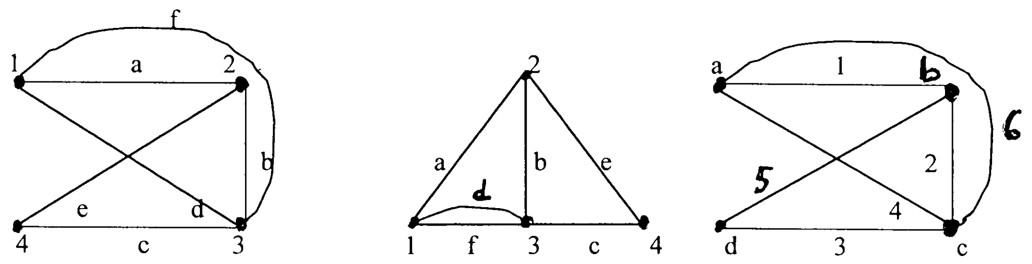
גרף מכוון מבוסס אפוא על גרף לא מכוון, אם לשני הגרפים אותם קדקודים, אותן קשתות, ולכל קשת יש בשניהם אותם קדקודי קצה.

מעתה נשתמש במונח "גרף" לציין גרפים *לא* מכוונים (אלא אם כן נאמר בפירוש אחרת). בדרך-כלל נטפל רק בגרפים עם מספר סופי של קדקודים וקשתות (גרפים כאלה נקראים *סופיים*).

---

[2] מבחינת תורת הגרפים אין הבדל בין "עם כיוון השעון" ו"נגד כיוון השעון". ללולאה $\{x\}$ של גרף לא סדור G אפשר להתאים בגרף המבוסס על G רק את הקשת $\langle x, x\rangle$.

המחשות חזותיות ("גרפיות") הן בדרך-כלל אמצעי מועיל מאוד לצורך רכישת הבנה ואינטואיציה בתחום מסוים. הדבר נכון במיוחד בתורת הגרפים. כאן מקובל מאוד להמחיש גרף סופי בעזרת דיאגרמה. בדיאגרמה כזו מייצגים קדקוד על-ידי נקודה במישור (עם תווית), וקשת – על-ידי קו רציף (עם תווית), המחבר את הנקודות המייצגות את הקצוות של אותה קשת (בגרף מכוון מוסיפים לקשת ראש חץ מכיוון קדקוד היציאה אל קדקוד הכניסה). הנה שלוש דוגמאות:



(א) (ב) (ג)

נתבונן תחילה בדיאגרמה (א). היא מייצגת את הגרף $G_1 = <V_1, E_1, F_1>$ הבא:

$$V_1 = \{1, 2, 3, 4\}$$

$$E_1 = \{a, b, c, d, e, f\}$$

$$F_1(a) = \{1, 2\},\ F_1(b) = \{2, 3\},\ F_1(c) = \{3, 4\},\ F_1(d) = \{1, 3\},$$
$$F_1(e) = \{2, 4\},\ F_1(f) = \{1, 3\}$$

(כלומר:

$$(F_1 = \{<a, \{1, 2\}>, <b, \{2, 3\}>, <c, \{3, 4\}>, <d, \{1, 3\}>, <e, \{2, 4\}>, <f, \{1, 3\}>\}$$

ואיזה גרף מייצגת דיאגרמה (ב)? ובכן, בדיקה תגלה ללא קושי, שלמרות הצורה השונה, דיאגרמה (ב) אף היא מייצגת את G_1 (כלומר: דיאגרמות (א) ו-(ב) מייצגות אותו גרף בדיוק). אנו רואים אפוא, שאותו גרף ניתן לייצג על-ידי דיאגרמות רבות, שיכולות להיות שונות מאוד זו מזו. כך, למשל, בין דיאגרמה (א) לבין דיאגרמה (ב) קיים השוני המהותי הבא: בעוד בדיאגרמה (ב) אין שתי קשתות הנחתכות זו עם זו (בנקודה פנימית שלהן), בדיאגרמה (א) המצב שונה: הקווים המייצגים את הקשתות d ו-e נחתכים. נציין שגרף, שניתן לתת לו ייצוג כמו בדיאגרמה (ב) (כלומר ייצוג, שבו שום שתי קשתות אינן נחתכות), נקרא גרף *מישורי*. הגרף G_1, לדוגמה, הוא מישורי, כיוון שדיאגרמה (ב) היא ייצוג שלו מהסוג הנדרש. דבר זה אינו עומד בסתירה לכך, שדיאגרמה (א), שגם

היא מייצגת את G_1, אינה עומדת בדרישת המישוריות. גרף הינו מישורי, אם יש לו *איזשהו* ייצוג,המקיים את תנאי המישוריות. התשובה לשאלה,אם גרף נתון הוא מישורי או לא,עלולה לכן להיות קשה.[3]

נעבור עתה לדיאגרמה (ג). היא מייצגת את הגרף $G_2 = <V_2, E_2, F_2>$ הבא:

$$V_2 = \{a, b, c, d\}$$

$$E_1 = \{1, 2, 3, 4, 5, 6\}$$

$$F_2(1) = \{a, b\},\ F_2(2) = \{b, c\}, F_2(3) = \{c, d\},\ F_2(4) = \{a, c\},\ F_2(5) = \{b, d\},$$
$$F_2(6) = \{a, c\}$$

ברור שהגרפים G_1 ו-G_2 הינם שונים (די לציין, למשל, שקבוצות הקדקודים שלהם הינן שונות). עם זאת, הדמיון בצורתן של הדיאגרמות (א) ו-(ג) מרמז על קשר הדוק בין שני הגרפים. קשר זה מבוטא בעזרת מושג ה*איזומורפיזם*. בקצרה, איזומורפיזם של שני גרפים (מכוונים שניהם או לא מכוונים) הוא פונקציית שקילות המתאימה לכל קדקוד של האחד קדקוד של השני ולכל קשת של האחד קשת של השני, כך שקדקודי הקצה של כל קשת מותאמים לקדקודי הקצה של תמונתה. ההגדרה הרשמית הבאה מנסחת זאת באופן מדויק יותר:

הגדרה:

(א) יהיו $G_1 = <V_1, E_1>$ גרפים (שניהם מכוונים או שניהם לא מכוונים).
איזומורפיזם בין G_1 ל- G_2 הוא פונקציית שקילות g מ- V_1 ל- V_2, עבורה מתקיים:

(1) במקרה הלא מכוון:

$$\forall a, b \in V_1.\ \{g(a), g(b)\} \in E_2 \Leftrightarrow \{a, b\} \in E_1$$

(2) במקרה המכוון:

$$\forall a, b \in V_1.\ <g(a), g(b)> \in E_2 \Leftrightarrow <a, b> \in E_1$$

(ב) יהיו $G_1 = <V_1, E_1, F_1>$, $G_2 = <V_2, E_2, F_2>$.

איזומורפיזם בין G_1 ו- G_2 הוא פונקציית שקילות g בין $V_1 \cup E_1$ ובין $V_2 \cup E_2$ כך שמתקיים:[4]

[3] הנושא של גרפים מישוריים הוא נושא חשוב, עם תוצאות מעניינות רבות, אך בקורס זה לא ניכנס אליו (מעבר להזכרת עצם המושג).

$$g(V_1) = V_2 \quad \text{(i)}$$
$$g(E_1) = E_2 \quad \text{(ii)}$$
$$\forall e \in E_1.\ F_2(g(e)) = g(F_1(e)) \quad \text{(iii)}$$

תרגיל:

בדקו, שההגדרות ב-(א) וב-(ב) תואמות זו את זו, כאשר מסתכלים על "$G = <V, E>$" כעל קיצור של "$G = <V, E, i_E>$".

שני גרפים נקראים *איזומורפיים*, אם קיים איזומורפיזם ביניהם. כך למשל G_1 ו- G_2, המיוצגים בדיאגרמות למעלה, הם איזומורפיים, כיוון שהפונקציה g הבאה מהווה איזומורפיזם ביניהם:

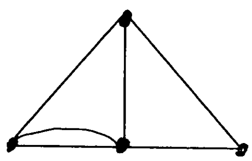
$g(a) = 1$	$g(b) = 2$	$g(c) = 3$	$g(d) = 4$	$g(e) = 5$	$g(f) = 6$
$g(1) = a$	$g(2) = b$	$g(3) = c$	$g(4) = d$		

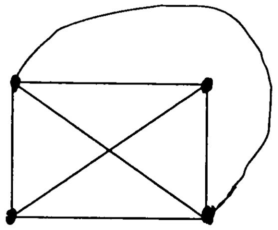
שני גרפים איזומורפיים יכולים להיות שונים זה מזה מאוד בזהותם של הקדקודים והקשתות שלהם (ב- G_1 הקדקודים הם אותיות, והקשתות הם מספרים, בעוד שב- G_2 המצב הפוך), אך מבחינת תורת הגרפים אין לזה שום חשיבות. כל מה שיש לתורת הגרפים לומר באשר לגרף מסוים G, תקף באותה מידה לגבי כל גרף שהוא איזומורפי ל- G. לפעמים אומרים אפילו, שמדובר ב"אותו הגרף". זה אינו מדויק, כמובן. מה שנכון לומר הוא, *שטיפוס הגרף* שלהם זהה. "טיפוס הגרף" הוא כאן מושג, המתקבל ממושג הגרף ומיחס האיזומורפיות בין גרפים בעזרת אותו תהליך הפשטה, שתואר בפרק ג.1, והוביל שם למושג ה"עוצמה" (קל לברר, שיחס האיזומורפיות הוא אכן יחס שקילות במובן הרחב – ראו דיון בפרק ג.1).[5] גם את מושג "טיפוס הגרף" קל להמחיש על-ידי דיאגרמות: פשוט מוחקים את התוויות מהקדקודים והקשתות של הדיאגרמות המייצגות גרפים מטיפוס זה. כך למשל, את טיפוס הגרף של הגרפים G_1 ו- G_2 מהדיאגרמות למעלה ניתן לייצג בשתי הצורות הבאות (כמו גם בצורות רבות אחרות, כמובן):

---

[4] נזכיר שאם $f: A \to B$ ו- $X \subseteq A$, אז $f(X)$ מסמן את התמונה של X לפי f (ולכן $f(X) \subseteq B$). בפרט: אם $a_1, a_2 \in A$, אז $f(\{a_1, a_2\}) = \{f(a_1), f(a_2)\}$. בדומה, מגדירים $f(<a_1, a_2>) = <f(a_1), f(a_2)>$. עובדות אלו חשובות לצורך הבנת התנאים (i)-(iii) (במיוחד (iii)).

[5] גם כאן מקובל בחלק מהטקסטים להגדיר את "טיפוס הגרף" של גרף G כאוסף כל הגרפים שאיזומורפיים ל- G, וגם כאן הגדרה זו הינה בעייתית, כי אוסף זה אינו מהווה קבוצה. כמו במקרה של עוצמות, הגדרה מדויקת, אך מסובכת יותר, ניתנת במסגרת תורת הקבוצות האקסיומטית.





להלן שתי דוגמאות של טיפוסי גרפים חשובים:

הגדרה:

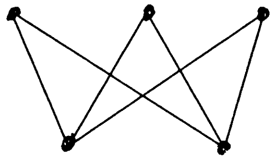
(א) טיפוס הגרף השלם K_n הוא הטיפוס של כל הגרפים הפשוטים, שיש להם n קדקודים, וכל שני קדקודים (שונים) שלהם הם סמוכים.

(ב) טיפוס הגרף הדו-צדדי השלם $K_{m,n}$ הוא הטיפוס של כל הגרפים הפשוטים, שניתן לפרק את קבוצת הקדקודים שלהם לאיחוד זר של שתי קבוצות V_1 ו-V_2, כך ש-$|V_1| = m$, $|V_2| = n$ ושני קדקודים הם סמוכים, אם"ם אחד מהם שייך ל-V_1, והשני ל-V_2.

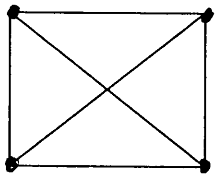
תרגיל

הוכיחו, שכל שני גרפים ב- K_n הם אכן איזומורפיים, וכן שכל שני גרפים ב- $K_{n,m}$ הם איזומורפיים (לכל n ו- m).

להלן ייצוגים גרפיים של K_4 *ושל* $K_{3,2}$:



$K_{3,2}$



K_4

כדי לפשט ניסוח טענות, נהוג לעתים קרובות שלא להקפיד על ההבחנה בין גרף ובין הטיפוס שלו, ולדבר למשל על *הגרף* K_5 (במקום על טיפוס הגרף K_5). מעתה גם אנו ננהג כך לפעמים (אך לא במקומות, בהם עלול להיווצר בלבול!).

כאמור, הבעיות, שתורת הגרפים עוסקת בהן, הן כאלה, שכדי לפתור אותן לגבי גרף מסוים G, כל מה שנחוץ לדעת על G הוא טיפוס-הגרף שלו. דוגמה פשוטה (מאוד)

לבעיה כזו היא הבעיה הבאה: "כמה קשתות יש בגרפים מטיפוס K_n?". התשובה היא כמובן $\binom{n}{2}$, כיוון שכל זוג קדקודים (מתוך ה-n שיש לגרף) מגדיר קשת (יחידה). לזהותם של הקדקודים אין כמובן שום רלוונטיות עבור בעיה זו. כדאי אגב לשים לב, שעצם *ניסוח* הבעיה מניח למעשה מראש, שהתשובה זהה לכל הגרפים ב-K_n! מעתה ננסח לכן בעיות כאלה פשוט כך: "מה מספר הקשתות ב-K_n?".

נעבור עתה למושג בסיסי נוסף של תורת הגרפים.

הגדרה:

גרף $<V', E'>$ הוא תת-גרף של הגרף $<V, E>$ אם $V' \subseteq V$ ו-$E' \subseteq E$.

הערות:

(1) ההגדרה תופסת במידה שווה הן לגרפים מכוונים והן לגרפים לא מכוונים.

(2) בהגדרה מניחים במפורש, ש-$<V', E'>$ *גרף* (דהיינו ש-$E' \subseteq (V')^2$, אם הגרף מכוון, ו-$E' \subseteq P_1(V') \cup P_2(V')$, אם הוא גרף לא מכוון). בלי קיום תנאי זה אין הזוג $<V', E'>$ מהווה תת-גרף של $<V, E>$, אפילו אם $V' \subseteq V$ ו-$E' \subseteq E$.

תרגיל

הגדירו מתי הגרף $<V', E', F'>$ מהווה תת-גרף של הגרף $<V, E, F>$.

הגדרה:

(1) אם $<V, E>$ גרף מכוון ו-$V' \subseteq V$, אז $<V', E \cap V'^2>$ הוא תת-הגרף של $<V, E>$ *המושרה* על-ידי V'.

(2) אם $<V, E>$ גרף לא מכוון ו-$V' \subseteq V$, אז $<V', E \cap (P_1(V') \cup P_2(V'))>$ הוא תת-הגרף של $<V, E>$ *המושרה* על-ידי V'.

תת-הגרף המושרה על-ידי V' הוא תת-הגרף המקסימלי של $<V, E>$, שקבוצת הקדקודים שלו היא V' (על הקורא לוודא לעצמו עובדה זו, ובפרט שתת-הגרף המושרה על-ידי V' הוא אכן תת-גרף של $<V, E>$!).

באופן אנלוגי נגדיר:

הגדרה:

יהי $<V, E>$ גרף, ותהי $E' \subseteq E$. תת-הגרף של $<V, E>$ *המושרה* על-ידי E' הוא:

$$<\{x \in V \mid \exists y \in V.\ \{x, y\} \in E'\},\ E'>$$

שוב, תת-הגרף של $<V, E>$ המוגדר על-ידי E' הוא תת-הגרף המקסימלי של $<V, E>$, ש- E' הינה קבוצת הקשתות שלו.

תרגיל

(1) הגדירו את מושג תת-הגרף המוגדר על-ידי קבוצה של קשתות במקרה של גרפים מכוונים.

(2) הגדירו את המושגים השונים הקשורים במושג תת-הגרף במקרה של מולטיגרפים (מהצורה $<V, E, F>$).

לסיום הפרק נביא דוגמה פשוטה של תרגום בעיה מ"החיים" לבעיה של תורת הגרפים.

בעיה

הראו, שמבין כל שישה בני-אדם יש שלושה, שכל שניים מהם מכירים זה את זה, או שיש ביניהם שלושה, שאף אחד מהם אינו מכיר איש מהשניים האחרים.

פתרון:

יהי a אחד מהשישה. את האחרים נחלק לשתי קבוצות: אלו שמכירים את a ואלו שלא. כיוון שיש חמישה "אחרים", באחת משתי קבוצות אלו יש שלושה איברים או יותר (ובשנייה – שני איברים או פחות). נניח למשל, שיש לפחות שלושה אנשים מהחמישה, שמכירים את a (במקרה, שיש שלושה שלא מכירים אותו, מטפלים באופן זהה לחלוטין). יהיו לכן b_1, b_2 ו- b_3 שלושה אנשים מהחמישה, המכירים את a. אם אף אחד משלושה אנשים אלו אינו מכיר את האחר, הרי מצאנו שלושה אנשים בקבוצה, שאף אחד מהם אינו מכיר איש מהשניים האחרים. אחרת שניים לפחות מבין b_1, b_2 ו- b_3 מכירים זה את זה. נניח למשל שאלו b_1 ו- b_2. אז b_1, b_2 ו- a הם שלושה אנשים מהקבוצה, שכל שניים מהם מכירים זה את זה.

הבה נציג עתה את הבעיה במונחים של תורת הגרפים. ובכן, "היכרות", כפי השימוש שנעשה במושג זה בטענה ובהוכחתה, הינה יחס סימטרי בין חברי השישייה. נוח לכן להציג אותה באמצעות גרף לא מכוון $<V, E_1>$, בו קבוצת הקדקודים V היא שישיית האנשים, ויש קשת ב- E_1 בין שני אנשים בקבוצה, אם"ם הם מכירים זה את זה. מצד שני, ליחס האי-היכרות יש בטענה מקום לא פחות חשוב מיחס ההיכרות. גם יחס האי-היכרות הוא סימטרי (בטיעון למעלה) וגם אותו נוכל לייצג על-ידי גרף לא מכוון $<V, E_2>$. קבוצת הקדקודים של גרף זה, V, היא אותה קבוצת הקדקודים של הגרף הראשון, אבל הפעם יש קשת בין שני אנשים אם"ם הם אינם מכירים זה את זה.

מה הקשר בין $<V, E_1>$ ו- $<V, E_2>$? התשובה היא, שקבוצת כל הקשתות, המחברות שני איברים שונים של V ($P_2(V)$), היא האיחוד הזר של E_1 ו- E_2. במלים אחרות, $<V, E_1 \uplus E_2>$ הוא גרף שלם בן 6 איברים. את הטענה, שהוכחנו למעלה, נוכל לכן לנסח במונחי תורת הגרפים באופן הבא:

טענה:

נניח ש-$<V, E> \in K_6$ וש- $E = E_1 \uplus E_2$. אז לפחות לאחד מהגרפים $<V, E_1>$ ו- $<V, E_2>$ יש תת-גרף הנמצא ב- K_3 (ובניסוח פחות פדנטי: אם $<V, E>$ הוא הגרף השלם K_6, ו- $E = E_1 \uplus E_2$, אז K_3 הוא תת-גרף של $<V, E_1>$, או של $<V, E_2>$).

הוכחה:

יהי $a \in V$, ותהי E_a קבוצת הקשתות, ש- a נמצא עליהן. אז $|E_a| = 5$. יתר על כן, כיוון ש- $E_a \subseteq E$, ו- $E = E_1 \uplus E_2$, אז

$$E_a = E_a \cap (E_1 \uplus E_2) = (E_a \cap E_1) \uplus (E_a \cap E_2)$$

ולכן

$$5 = |E_a \cap E_1| + |E_a \cap E_2|$$

מכאן, שאו ש- $|E_a \cap E_1| \geq 3$, או ש- $|E_a \cap E_2| \geq 3$. נניח למשל, ש- $|E_a \cap E_1| \geq 3$. יהיו אפוא b_1, b_2, b_3 שלושה איברים שונים של V, כך ש- $\{\{a, b_1\}, \{a, b_2\}, \{a, b_3\}\} \subseteq E_1$. אם גם $\{b_1, b_2\} \notin E_1$, גם $\{b_1, b_3\} \notin E_1$ וגם $\{b_2, b_3\} \notin E_1$, אז כל שלוש הקשתות הללו נמצאות ב- E_2, ולכן $<\{b_1, b_2, b_3\}, \{\{b_1, b_2\}, \{b_1, b_3\}, \{b_2, b_3\}\}>$ הוא תת-גרף של $<V, E_2>$, השייך ל- K_3. אחרת, אחת משלוש הקשתות הללו נמצאות ב- E_1. נניח למשל, שזו $\{b_1, b_2\}$. במקרה זה נקבל ש- $<\{a, b_1, b_2\}, \{\{a, b_1\}, \{a, b_2\}, \{b_1, b_2\}\}>$ הוא תת-גרף של $<V, E_1>$, השייך ל- K_3.

ההוכחה, שהבאנו זה עתה, היא למעשה תרגום מדויק ללשון תורת הגרפים של ההוכחה, שהבאנו תחילה (במונחי "היכרות" ו"אי-היכרות"). אין להכחיש, עם זאת, שהיא הרבה יוותר קשה לקריאה ולהבנה מאשר ההוכחה המקורית (שהיא בעצם זהה!). עלול אפוא להיווצר הרושם, שתרגום כזה לא יוסיף דבר, ולהיפך: רק מסבך ללא צורך דברים פשוטים. הבה נסביר אפוא, למה תרגום כזה אינו רק מועיל, אלא שברוב המקרים הוא כמעט הכרחי.

ראשית, אותה בעיה עצמה יכולה לצוץ בצורות שונות ומשונות, לאו דווקא במונחים של "היכרות" ו"אי-היכרות" (דוגמה לכך, במונחים של "צביעה", נראה עוד מעט). ללא ניסוח אחיד, אבסטרקטי, במונחים מתמטיים, נהיה נידונים לפתור אותה כל פעם מחדש. הניסוח המתמטי האבסטרקטי מאפשר גם להבדיל בצורה ברורה, מה רלוונטי ומה לא, ולכן הוא מאפשר את יישום הטענה בקשת רחבה של מקרים (כך למשל, העובדה, שבבעיה המקורית דובר בבני-אדם דווקא, היא לא רלוונטית לחלוטין כאן, בעוד שכל מה שרלוונטי לגבי יחס ההיכרות הינו רק הסימטריות שלו).

שנית, את הבעיה הזו ניתן היה לפתור ישירות בקלות כזו, רק בגלל שהיא *פשוטה* מאוד, ואף עוסקת במספרים קטנים מאוד. גם "בעיות מילוליות" פשוטות באלגברה תיכונית ניתן לפעמים לפתור ביתר קלות ישירות (במקום לתרגמן ללשון המשוואות) – אך זה הופך בלתי אפשרי, כשהבעיות מסתבכות קצת יותר! בדומה, גם כאן, בעיות קצת יותר מסובכות מאותו סוג דורשות כבר טיפול מתמטי של ממש, וטיפול כזה ניתן להיעשות רק במסגרת תורה מתמטית, המנוסחת במונחים מתמטיים מדויקים (כבר בפרק הבא נראה דוגמה לכך!). בניגוד לרושם, שעוררה אולי הדוגמה האחרונה, במציאות, הצעד המכריע בפתרון בעיות רבות מ"החיים" הוא ייצוגן בעזרת גרפים והמונחים של תורת הגרפים. ייצוג כזה מאפשר ליישם את התיאוריה העשירה, הטכניקות והאלגוריתמים, שפותחו במסגרת תורת הגרפים (ורק קצת מהם נראה בקורס זה).

עם זאת, כיוון ששימוש עקבי במונחים הטכניים של תורת הקבוצות עלול להיות מכביד, מקובל לעתים בתורת הגרפים להשתמש עבור אותם מושגים עצמם בשמות, שנראים פחות טכניים ויותר אינטואיטיביים. דוגמה טובה לכך היא מושג *הצביעה*. כשמדובר בתורת הגרפים על "צביעה" של איברי קבוצה מסוימת (בדרך-כלל קבוצת הקדקודים או קבוצת הקשתות של איזה גרף) ב- k צבעים, הכוונה היא בעצם לפרוק אותה קבוצה ל- k תת-קבוצות או פחות.[6] הרעיון האינטואיטיבי הוא, שההבדלה בין תת-הקבוצות בפירוק יכולה להיעשות על-ידי ש"צובעים" את איברי כל אחת מהן בצבע משלה, המייחד אותה. כך למשל, מקובל לתאר את הבעיה האחרונה שפתרנו במונחים של צביעה על-ידי שני צבעים (לדוגמה, קשתות בצבע לבן מייצגות היכרות בין שני אנשים, קשתות בצבע שחור מייצגות אי-היכרות). במונחים אלו ינוסח המשפט האחרון כך: "בכל צביעה של קשתות הגרף השלם K_6 על-ידי שני צבעים, יתקבל תת-גרף מטיפוס K_3, שלכל קשתותיו יש אותו צבע" (תת-גרף כזה נקרא "מונוכרומטי"). נשים

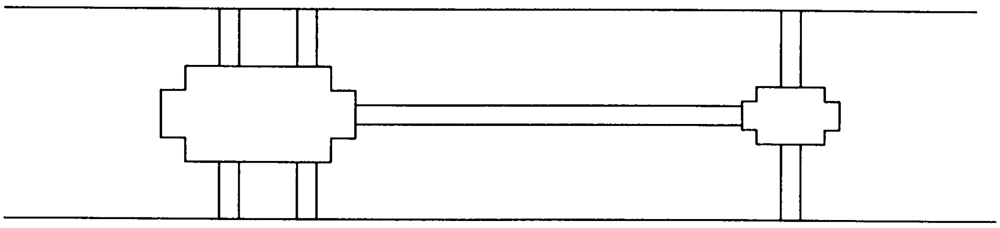
[6] לפי הגדרת "פירוק" בפרק ב.5 (IV), תת-הקבוצות בפירוק הינן לא ריקות. לעומת זאת, כשמדובר בתורת הגרפים בצביעה על-ידי k צבעים, אין הכוונה, שחייבים להשתמש *בכל* הצבעים. לכן התוספת "או פחות" כאן.

לב אבל, שלמרות שמושג הצביעה הוא בעל שם "צבעוני" יותר ממושג הפירוק, הוא בעצם מושג טכני לא פחות (וכמעט זהה במשמעותו למושג הפירוק)!

נציין לבסוף, שהמשפט, שהוכחנו בסוף פרק זה הוא מקרה פרטי של משפט יסודי חשוב (בתורת הגרפים בפרט, ובקומבינטוריקה בכלל), הידוע כמשפט *רמזי*. נוסח רחב יותר שלו, אף כי לא הכללי ביותר, הוא:
לכל גרף G ולכל מספר טבעי k, קיים מספר טבעי n, כך שבכל צביעה של קשתות הגרף השלם K_n ב- k צבעים, יש ב- K_n תת-גרף איזומורפי ל- G, שכל הקשתות שלו באותו צבע.
ההוכחה של משפט זה חורגת ממסגרת קורס זה.

ה.2 מסילות ומעגלים

המסורת מייחסת את הולדתה של תורת הגרפים לפתרונה של בעיה קונקרטית, הידועה כ"בעיית הגשרים של קניגסברג". מעשה שהיה כך היה: בעיר קניגסברג שבגרמניה (עירו של קאנט) זורם נהר, ובנהר יש שני איים. כל אחת משתי גדות הנהר מחוברת אל האי הגדול בין השניים באמצעות שני גשרים, ועם הקטן ביניהם – באמצעות גשר אחד. גשר נוסף מחבר את שני האיים (ראו ציור):



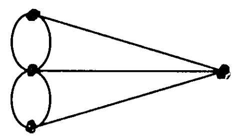
ציור 1

משימה ידועה, שהוצבה בפני המבקרים בעיר, היתה: לתכנן מסלול טיולים, שיעבור דרך כל שבעת הגשרים, אך מבלי לעבור אף אחד מהגשרים יותר מפעם אחת. רבים ניסו את כוחם, אך כולם נכשלו. עד שבא ב-1736 ליאונרד אוילר (מגדולי המתמטיקאים של כל הזמנים) והראה באופן מתמטי, שסיבת הכישלון היא היות המשימה בלתי ניתנת לביצוע באופן עקרוני. בפרק זה אכן נוכיח זאת. מובן, שמה שמעניין אותנו באמת כאן, אינו הפתרון של בעיה ספציפית זו (עם כל חשיבותה ההיסטורית), אלא התיאוריה, המאפשרת את פתרונה של בעיה זו, כמו גם הצורה, שבה מתמודדים מתמטיקאים עם בעיות קונקרטיות כאלו מ"החיים".

נתחיל בסוגייה היותר כללית. ובכן, פתרון מתמטי של בעיות כגון זו של גשרי קניגסברג, כרוך כמעט תמיד *בשתי* הפשטות. ראשית, כדי לתת פתרון מתמטי לבעיה, יש להציג אותה באופן מתמטי, דהיינו: על-ידי הצגתה בעזרת מבנים מתמטיים מוגדרים היטב, שיפשיטו את הבעיה מכל המרכיבים, שאינם באמת רלוונטיים (כמו: שהיא קשורה בנהר, איים, גשרים וכו'). שנית, המפתח לפתרון בעיה כזו הוא כמעט תמיד *בהכללתה* ומציאת פתרון, שמטפל בהצלחה לא רק בה, אלא בקבוצה שלמה של

בעיות דומות (במיוחד נכון היה הדבר אם הבעיה היתה עוסקת ב- 700,000 גשרים, במקום רק בשבעה).

בדוגמה של גשרי קניגסברג, כל מה שרלוונטי במצב המתואר בבעיה, מיוצג בצורה שלמה על-ידי גרף מהטיפוס הבא (בו הקדקוד העליון והתחתון מייצגים את שתי גדות הנהר, אלה שבאמצע מייצגים את שני האיים, והקשתות מייצגות את הגשרים).



ציור 2

ייצוג זה הופך את הבעיה לבעיה של תורת הגרפים (שתוגדר בהמשך באופן מדויק). כמו שרמזנו, פתרונה ייעשה על-ידי כך שנמצא *לכל* גרף סופי, אם מסלול מהסוג המתואר בבעיה אפשרי בו או לא.

נתחיל את התמודדותנו עם בעיית גשרי קניגסברג על-ידי הכנסת מושגים, שיאפשרו לנסח אותה באופן מדויק.

הגדרה:

(1) יהי G גרף (מכוון או לא מכוון). *טיול* ב- G הוא רשימה סופית מהצורה

$$p = \langle v_0, e_1, v_1, e_2, v_2, \ldots, v_{n-1}, e_n, v_n \rangle$$

שבה $v_0, \ldots, v_n$ $(n \geq 0)$ הם קדקודים של G, $e_1, \ldots, e_n$ הן קשתות של G, ולכל $1 \leq i \leq n$, e_i מחברת את v_{i-1} ל- v_i. לגבי טיול p כנ"ל אומרים, שהוא *מחבר* את v_0 ל- v_n (או שהוא *בין* v_0 ל- v_n), וכמו כן ש- v_0 ו- v_n הם *הקצוות* שלו, בעוד $v_1, \ldots, v_{n-1}$ הם הקדקודים *פנימיים* שלו. על איברי p אומרים, שהם *על* הטיול p.

(2) טיול p, שבו אף קשת לא חוזרת יותר מפעם אחת, נקרא *מסלול* של G. אם הקצוות שלו זהים (כלומר אם $v_0 = v_n$), אז הטיול נקרא *מעגלי*. אם אף קדקוד פנימי של p לא חוזר ב-p יותר מפעם אחת, אז p נקרא *פשוט*.

(3) אורך של טיול הוא מספר הקשתות שבו.

הערות:

(1) המקרה שבו $n = 0$ (כלומר $p = <v_0>$) אפשרי לפי ההגדרה. טיול כזה נקרא *מנוון* (או *טריביאלי*), והוא כמובן מסלול פשוט ומעגלי.

(2) טיול פשוט הוא מסלול (כי אם אף קדקוד פנימי לא חוזר יותר מפעם אחת, אז גם שום קשת לא חוזרת יותר מפעם אחת!).

(3) אם אחד הקדקודים הפנימיים של מסלול זהה לאחד הקצוות, אז המסלול אינו פשוט (כי אותו קדקוד פנימי חוזר פעמיים). לעומת זאת, מסלול מעגלי, שבו כל הקדקודים הפנימיים שונים זה מזה, והם שונים גם מנקודת הקצה, כן נחשב לפשוט.

(4) מהגדרת טיול נובע, שקשת יכולה להימצא עליו רק אם (אך לא בהכרח אם ורק אם!) שני הקצוות שלה נמצאים עליו.

(5) אורך של מסלול ב- $<V, E, F>$ קטן או שווה ל- $|E|$.

(6) נשים לב, שאם p מסלול בגרף מכוון G, אז p הוא גם מסלול בגרף הלא-מכוון, ש-G מבוסס עליו (ההפך *אינו* נכון!).

הגדרה:

1. *מסלול אוילר* בגרף הוא מסלול שעובר דרך כל הקשתות וכל הקדקודים של הגרף. *מעגל-אוילר* הוא מסלול אוילר מעגלי.

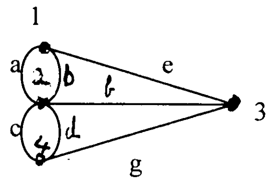
2. *מסילה* היא גרף, שיש לו מסלול אוילר. *מעגל* הוא גרף, שיש לו מעגל אוילר. מסילה נקראת *פשוטה* אם יש לה מסלול אוילר פשוט. מעגל נקרא *פשוט* אם יש לו מעגל אוילר פשוט.

3. יהי G גרף. *מסילה ב-* G (או "מסילה של G") היא תת-גרף של G המהווה מסילה. *מעגל של* G הוא תת-גרף של G המהווה מעגל.

הערות:

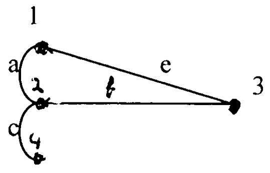
(1) בטקסטים רבים "מסלול אוילר" ב-G מוגדר רק בתור מסלול, שעובר דרך כל הקשתות של G. אם ב- G יש נקודות *מבודדות* (דהיינו: נקודות שאינן נמצאות על שום קשת), אז יש הבדל בין שתי ההגדרות. ב- G כזה לא ייתכן כלל מסלול אוילר לפי הגדרתנו (אלא אם כן יש בו קדקוד אחד בדיוק), אך עדיין ייתכן מסלול אוילר לפי ההגדרה הדורשת פחות. בחירתנו בהגדרה למעלה מאפשרת ניסוח פשוט ו"נקי" יותר של המשפטים. אין אבל כל בעיה לנסח את המשפטים הללו מחדש עם ההגדרה האחרת.

(2) "מסלול בגרף G" ו"מסילה בגרף G" הם מושגים, שיש ביניהם קשר הדוק, אך הם אינם זהים. בעוד מסלול של G הינו רשימה (מסודרת!) של קדקודים וקשתות של G, "מסילה בגרף G" הינה תת-גרף של G. כדי להדגים את ההבדל, נתבונן בגרף G_0 הבא (שהוא מטיפוס הגרף שמופיע בציור 1):



ציור 3: הגרף G_0

תת-ה-גרף הבא של G_0 הינו *מסילה* של G_0:



ציור 4

כדי להיווכח, שאכן כך הדבר, עלינו להציג *מסלול*, שעובר דרך כל הקשתות והקדקודים של גרף זה. למעשה, יש (בדיוק!) ארבעה מסלולים כאלה:

$$<4, c, 2, a, 1, e, 3, f, 2>$$
$$<4, c, 2, f, 3, e, 1, a, 2>$$
$$<2, a, 1, e, 3, f, 2, c, 4>$$
$$<2, f, 3, e, 1, a, 2, c, 4>$$

(3) מרבית הטקסטים משתמשים במונח "מעגל" הן במובן שהגדרנו למעלה, והן במקום המונח "מסלול מעגלי". בהמשך ננהג כך גם אנו, ועל הקוראים יהיה להחליט למה הכוונה בכל מקרה. (למעשה כבר נהגנו כך, עת הגדרנו "מעגל אוילר". מעגל כזה אינו מעגל כלל לפי הגדרותינו, אלא מסלול מעגלי!).

כפי שראינו בדוגמה האחרונה, למסילה בגרף G יכולים להתאים מספר מסלולים של G. ברור, לעומת זאת, שכל מסלול ב- G קובע מסילה יחידה של G (שקדקודיה וקשתותיה הן אלה הנמצאים על מסלול זה). כדאי לציין גם, שבעוד קל לברר לפי ההגדרה, אם רשימה מסוימת מהווה מסלול של G, אין זה כה קל לברר לפי ההגדרה, אם תת-גרף מסוים של G מהווה מסילה של G. למעשה, את בעיית גשרי קניגסברג אנו יכולים עתה לנסח (סוף סוף) בצורה המדויקת הבאה:
האם G_0 של ציור 3 (או כל גרף אחר מהטיפוס המתואר בציור 2) מהווה מסילה? (דהיינו: האם יש בו מסלול אוילר?).

אפשר לפתור בעיה ספציפית זו על-ידי עריכת רשימה מלאה של כל המסלולים, שיש ב- G_0, ובדיקה אם אחד מהם הוא מסלול אוילר. הרבה יותר מעניין (ומעשי!) למצוא תנאים כלליים, קלים לבדיקה, שעבור *כל* גרף יהוו ביחד תנאים הכרחיים ומספיקים לקיום מסלול אוילר בו. כדי לנסח תנאים כאלה נצטרך להכניס כמה מושגים יסודיים נוספים של תורת הגרפים. תחילה נסקור אבל מספר פעולות על טיולים ומסלולים, שנזדקק להם בהמשך.

(א) הוצאת תת-טיול: אם $\langle v_0, e_1, v_1, \ldots, e_n, v_n\rangle$ טיול ו- $0 \le i \le j \le n$, אז הרשימה $\langle v_i, e_{i+1}, \ldots, e_j, v_j\rangle$ נקראת תת-טיול שלו. נשים לב שתת-טיול של טיול הוא בעצמו טיול, ותת-טיול של מסלול הוא בעצמו מסלול (ולכן נדבר פשוט על תת-מסלול במקרה זה), והוא פשוט, אם המסלול המקורי פשוט.

(ב) היפוך: אם רשימה מסוימת ($\langle v_0, e_1, v_1, \ldots, v_{n-1}, e_n, v_n\rangle$) מהווה טיול, אז הרשימה $\langle v_n, e_n, v_{n-1}, \ldots, v_1, e_1, v_0\rangle$ מהווה טיול גם היא. יתר על כן: היפוך של מסלול מהווה בעצמו מסלול (והוא פשוט, אם המסלול המקורי פשוט).

(ג) השמטת קטע מעגלי (ממסלול לא פשוט):
אם $\langle v_0, e_1, \ldots, v_i, e_{i+1}, v_{i+1}, \ldots, e_j, v_j, e_{j+1}, \ldots, v_n\rangle$ טיול, $0 \le i, j \le n$ ו- $v_i = v_j$, אז הרשימה המתקבלת מטיול זה על-ידי השמטת הקטע בין v_i ל- v_j (דהיינו, הרשימה $\langle v_0, e_1, \ldots, v_i, e_{j+1}, v_{j+1}, \ldots, v_n\rangle$ גם היא טיול בין אותם שני קדקודים. אם הרשימה המקורית היא מסלול, אז כך גם הרשימה החדשה.

(ד) חיבור טיולים: אם $<v_0, \ldots, v_n>$ ו- $<v'_0, e'_1, v'_1, \ldots, v'_k>$ הן טיולים ו- $v_n = v'_0$ (כלומר: הסיום של טיול אחד הוא ההתחלה של הטיול השני), אז חיבורן הוא הרשימה $<v_0, \ldots, v_n, e'_1, v'_1, \ldots, v'_k>$,[1] המהווה טיול בעצמה.

(ה) הזזת מעגלים: אם $<v_0, e_1, \ldots, v_i, \ldots, e_n, v_0>$ הוא טיול (מסלול/מסלול פשוט) מעגלי, אז כך גם $<v_i, e_{i+1}, v_{i+1}, \ldots, e_n, v_0, e_1, v_1, \ldots, e_i, v_i>$. לטיול זה יש כמובן בדיוק אותם קשתות וקדקודים כמו של הטיול המקורי.

את הבדיקה הקלה של כל הטענות, שנכללו בתיאור הפעולות (א)-(ה) למעלה, נשאיר לקוראים. נעיר רק, שהפעולה האחרונה היא למעשה צירוף של שתיים קודמות: תחילה יוצרים את תתי-הטיולים $<v_0, \ldots, v_i>$ ו- $<v_i, e_{i+1}, \ldots, v_0>$ של הטיול המקורי, ואז מחברים אותם מחדש, בסדר הפוך.

הערה:
משמעות הטענה ב-(ה) הינה, שכל קדקוד של מעגל מהווה את נקודת ההתחלה והסיום של איזשהו מסלול מעגלי, שעובר דרך כל הקשתות והקדקודים של אותו מעגל.

נעבור עתה למושגים החדשים, שנזדקק להם לצורך פתרון בעייתנו.

הגדרה:
שני קדקודים של גרף G נקראים *קשורים* ב- G, אם יש טיול, שמחבר ביניהם. גרף נקרא *קשיר*, אם כל שני קדקודים שלו הינם קשורים.

דוגמה: הגרפים בציורים 3 ו-4 הם קשירים.

הטענה הבאה מספקת הגדרות אלטרנטיביות למושג הקשירות.

טענה 1:
(1) שני קדקודים של גרף G הם קשורים, אם"ם יש מסלול ב- G המחבר אותם.
(2) שני קדקודים של גרף G הם קשורים, אם"ם יש מסלול פשוט ב- G המחבר אותם.

---

[1] בטקסטים מסוימים משתמשים במונח "שרשור" במקום ה"חיבור" שלנו. מונח זה לא מתאים להגדרות בספר *זה*, כיוון שבשרשורן של שתי הרשימות (לפי השימוש הרגיל של מושג זה) יופיע הקדקוד v_n פעמיים רצוף (פעם בתור v_n, פעם בתור v'_0).

הוכחה:
כיוון שכל מסלול פשוט הוא מסלול, וכל מסלול הוא טיול, מספיק להוכיח, שאם x ו- y קשורים ב- G, אז יש מסלול פשוט, שמחבר אותם. יהיו אפוא x ו- y קשורים, ויהי p הטיול הקצר ביותר המחבר אותם. p זה הינו בהכרח מסלול פשוט, משום שאם קדקוד v חוזר בו פעמיים, אז נוכל להשמיט מ-p את הקטע המעגלי המחבר את שני המופעים השונים הראשונים של v, ולקבל טיול חדש בין x ל- y, שהוא קצר יותר מ- p (בסתירה לבחירת p).

העובדה הבסיסית החשובה ביותר על יחס הקשירות ניתנת בטענה הבאה:

טענה 2:
יחס הקשירות הינו יחס שקילות.

הוכחה:
(א) רפלקסיביות: כל קדקוד v_0 בגרף G מחובר לעצמו על-ידי הטיול $<v_0>$.
(ב) סימטריות: אם הטיול p מחבר את הקדקוד x ל- y, אז היפוכו של p הוא טיול המחבר את y ל- x.
(ג) טרנזיטיביות: נניח ש-x קשור ל- y, ו- y קשור ל- z. נניח שהטיול p_1 מחבר את x ל- y, והטיול p_2 מחבר את y ל- z. אז החיבור של p_1 ו- p_2 מוגדר (כי נקודת הסיום של p_1 היא נקודת ההתחלה של p_2), והוא מחבר את x ל- z.

הגדרה:
תת-הגרפים של G, המושרים על-ידי מחלקות השקילות של יחס הקשירות על קבוצת הקדקודים של G, נקראים *רכיבי הקשירות* של G. גרף G הינו קשיר, אם"ם יש לו רכיב קשירות יחיד.

תרגיל
נניח שהגרף G' מתקבל מהגרף G על-ידי תוספת קשת אחת, המחברת קדקודים מרכיבי קשירות שונים. הוכח כי שני רכיבי קשירות אלה של G מתאחדים ב- G' לרכיב קשירות אחד, בעוד שאר הרכיבים נשארים ללא שינוי. (מה קורה אם מוסיפים ל- G קשת בין שני קדקודים באותו רכיב קשירות?).

קשירות היא, באופן כללי, תכונה חשובה מאוד של גרף. חשיבותה לפרק ספציפי זה נובעת מהטענה הבאה:

טענה 3:
כל מסילה היא גרף קשיר.

הוכחה:
אם G מסילה, אז יש מסלול p שעובר דרך כל הקדקודים של G. בפרט כל שני קדקודים של G מחוברים על-ידי הקטע (או תת-המסלול) של p הנמצא ביניהם.

קשירות היא רק תנאי ***הכרחי*** לקיום מסלול אוילר בגרף. כך למשל, הגרף G_0 של ציור 3 מהווה (כמו שנראה) דוגמה לגרף קשיר שאינו מסילה. כדי לנסח תנאי, שיהיה הכרחי ומספיק עבור כל גרף סופי, עלינו להכניס מושג בסיסי נוסף: מושג ***הדרגה***. קל אבל יותר להכניס מושג זה תחילה לגרפים מכוונים דווקא.

מעתה והלאה נניח, שאנו עוסקים בגרפים ***סופיים*** בלבד.

הגדרה:
יהי G גרף מכוון. ***דרגת הכניסה*** של קדקוד v בגרף, $d_G^+(v)$, היא מספר הקשתות המסתיימות ב- v, בעוד ***דרגת היציאה*** של v, $d_G^-(v)$, היא מספר הקשתות המתחילות בו. הדרגה של v, $d_G(v)$, מוגדרת בתור $d_G^+(v)+d_G^-(v)$.

טענה 4:
בגרף מכוון $<V, E, F>$ מתקיים:

$$\sum_{v\in V} d_G^+(v) = \sum_{v\in V} d_G^-(v) = |E|$$

הוכחה:
תהי E_v קבוצת הקדקודים המסתיימים ב- v. אז $|E_v| = d_G^+(v)$ ו- $E = \biguplus_{v\in V} E_v$. לכן

$\sum_{v\in V} d_G^+(v) = |E|$. באופן דומה מראים ש- $\sum_{v\in V} d_G^-(v) = |E|$.

כבסיס להגדרת דרגה של קדקוד בגרף לא מכוון תשמש הטענה הבאה.

טענה 5:
אם G גרף לא מכוון, ו- G' הוא גרף מכוון המבוסס עליו, אז הדרגה ב- G' של קדקוד v שווה לסכום של מספר הקשתות (כולל לולאות) ב- G, ש- v נמצא עליהן, ושל מספר הלולאות, ש- v נמצא עליהן (במלים אחרות: הדרגה שווה למספר הקשתות ש- v נמצא עליהן, כשלולאות נספרות פעמיים).

הוכחה:

יהי v קדקוד. נסמן ב- O_v את קבוצת הקשתות ב- G, שמתחילות ב- v בגרף המכוון G', ב- I_v את קבוצת הקשתות של G, שמסתיימות ב-v בגרף G', וב- E_v את קבוצת הקשתות ב- G, ש- v נמצא עליהן. אז $E_v = O_v \cup I_v$, ולכן:

$$|E_v| = |O_v| + |I_v| - |I_v \cap O_v|$$

$$|O_v| + |I_v| = |E_v| + |I_v \cap O_v|$$

אבל $|O_v| = d^-_{G'}(v)$, $|I_v| = d^+_{G'}(v)$, ולכן $|O_v| + |I_v| = d_{G'}(v)$. מצד שני, $I_v \cap O_v$ היא בדיוק קבוצת הלולאות, ש- v נמצא עליהן (דהיינו: קבוצת הקשתות ב- G, שב- G' הן גם מתחילות ב- v וגם נגמרות בו). מכאן שהשוויון האחרון מבטא בדיוק את מה שצריך להוכיח.

מסקנה 1:

לקדקוד של גרף לא מכוון G יש אותה דרגה בכל הגרפים המכוונים, המבוססים על G.

הדרגה של קדקוד בגרף לא מכוון G מוגדרת כך, שתהיה שווה לדרגה המשותפת, שיש לו בכל הגרפים המכוונים המבוססים על G. לאור הטענה האחרונה, ההגדרה הפורמלית היא:

הגדרה:

הדרגה $d_G(v)$ של קדקוד v בגרף מוגדרת כסכום של מספר הקשתות, ש- v נמצא עליהן, ומספר הלולאות, שהוא נמצא עליהן (במלים אחרות: כל לולאה "נספרת פעמיים").

טענה 6:

סכום הדרגות של כל הקדקודים בגרף שווה לפעמיים מספר הקשתות.

הוכחה:

נניח ש- $G = \langle V, E, F \rangle$. יהי $G' = \langle V, E', F' \rangle$ גרף מכוון המבוסס עליו. אז $|E'| = |E|$,

$$\sum_{v\in V} d_G(v) = \sum_{v\in V} d_{G'}(v) = \sum_{v\in V}(d^+_{G'}(v) + d^-_{G'}(v)) = \sum_{v\in V} d^+_{G'}(v) + \sum_{v\in V} d^-_{G'}(v) =$$

$$= |E'| + |E'| = 2|E'| = 2|E|$$

הערה:
הסבר אינטואיטיבי לטענה האחרונה הוא, שכשמחברים את הדרגות של כל הקדקודים, אז כל קשת נספרת פעמיים. (זה כולל, לפי הגדרה, את הלולאות!). מכאן שסכום זה שווה לפעמיים מספר הקשתות.

%%
זהו הסבר אינטואיטיבי בלבד, משום שהמושג של "ספירת קשת פעמיים" דורש הבהרה, ובמיוחד באיזה מובן, כשאנו מחברים מספרים, אנו בעצם סופרים קשתות (ועוד פעמיים!). ניתן אבל בהחלט להפוך הסבר זה להוכחה, שמשתמשת במושגים מוגדרים היטב ובעקרונות שהכרנו. הדבר יכול להיעשות, למשל, כך:
יהי $G = <V, E, F>$ גרף. נגדיר:

$$A = \{<v, e, 1> \mid e \in E \wedge v \in V \wedge v \in F(e)\} \cup \{<v, e, 2> \mid e \in E \wedge v \in V \wedge F(e) = \{v\}\}$$

נספור עתה את מספר איברי A בשתי דרכים: אחת לפי הקדקודים, אחת לפי הקשתות.

ספירה לפי הקדקודים: נגדיר עבור קדקוד v:

$$A_v =< x \in A \mid \pi_1^3(x) = v >$$

קל לראות ש- $|A_v| = d_G(v)$. כמו-כן, $A = \biguplus_{v \in V} A_v$. לכן

$$|A| = \sum_{v \in V} d_G(v)$$

ספירה לפי הקשתות: נגדיר עבור קשת e:

$$A_e =< x \in A \mid \pi_2^3(x) = e >$$

אז $A = \biguplus_{e \in E} A_e$. הפעם אבל $|A_e| = 2$ לכל $e \in E$ ($A_e = \{<v_1, e, 1>, <v_2, e, 1>\}$, אם e מחברת את v_1 ו- v_2 ו- $v_1 \neq v_2$, $A_e = \{<v_0, e, 1>, <v_0, e, 2>\}$ אם e היא לולאה ב- v_0).
לכן $|A| = 2|E|$.

בסך-הכל קיבלנו:

$$\sum_{v \in V} d_G(v) = 2|E|$$

%%

הטענה הבאה על דרגות אינה חשובה לצורך מטרתו העיקרית של הפרק, אך היא מדגימה את צורת החשיבה הקשורה במושג הדרגה, והוכחתה תשמש לנו תזכורת של עיקרון שובך היונים, אותו הכרנו בפרק ג.6.

טענה 7:

בכל גרף פשוט בעל שני קדקודים לפחות, יש שני קדקודים, שדרגותיהם שוות.

הוכחה:

נניח שמספר הקדקודים הוא n, וש- $n \geq 2$ (ולכן $n-1>0$). כיוון שהגרף פשוט, דרגת כל קדקוד היא $n-1$ לכל היותר (אפשר לחבר קדקוד ל- $n-1$ הקדקודים הנותרים). עתה לא ייתכן, שיש גם קדקוד שדרגתו 0 (דהיינו: קדקוד, שאינו מחובר בקשת לאף קדקוד אחר), וגם קדקוד, שדרגתו $n-1$ (דהיינו: קדקוד, המחובר בקשת לכל קדקוד אחר). מכאן שאו שדרגות כל הקדקודים הן בין 0 ל- $n-2$, או שדרגות כולם הן בין 1 ל- $n-1$. בשני המקרים יש $n-1$ ערכים אפשריים של הדרגות של n הקדקודים, ולכן לפי עיקרון שובך היונים יש שני קדקודים, שדרגתם זהה.

נחזור עתה למסלולי אוילר, ולמקום המרכזי של מושג הדרגה בחקירתם.

טענה 8:

(1) אם G מעגל (כלומר ב- G יש מסלול אוילר מעגלי), אז דרגת כל קדקוד של G הינה זוגית.

(2) אם G גרף שיש בו מסלול אוילר שאינו מעגלי, אז יש ב- G בדיוק שני קדקודים שדרגתם אי-זוגית (והם מהווים את קצות המסלול הנ"ל).

הוכחה:

(1) יהי $c = <v_0, e_1, v_1, \ldots, e_n, v_n>$ מסלול אוילר ב- G. יהי G' הגרף המכוון המבוסס על G, שבו הכיוון של הקשת e_i הוא מ- v_{i-1} ל- v_i (כיוון שמדובר במסלול אוילר, כל קשת של G מופיעה בו פעם אחת ויחידה, ולכן זה מגדיר כיוון יחיד *לכל* קשת). ברור ש- c הוא גם מסלול אוילר ב- G'. עתה, אם c הוא מסלול מעגלי (כלומר $v_0 = v_n$), אז לכל קדקוד v, מספר הקשתות ב- c, הנכנסות ל- v, שווה למספר הקשתות בו, היוצאות מ- v. (תרגיל: הוכח זאת באופן פורמלי, על-ידי בניית פונקציית שקילות מתאימה). כיוון ש- c מכיל את כל הקשתות של G, פירוש הדבר ש- $d^+_{G'}(v) = d^-_{G'}(v)$ לכל קדקוד v. כיוון ש- $d_G(v) = d_{G'}(v) = d^+_{G'}(v) + d^-_{G'}(v)$, נקבל ש- $d_G(v) = 2d^+_{G'}(v)$ לכל v. מכאן ש- $d_G(v)$ הינו זוגי.

(2) סעיף זה יושאר כתרגיל לקוראים.

מסקנה 2:

במסילה מספר הקדקודים, שיש להם דרגה אי-זוגית, הוא 0 או 2. יתר על כן:

(i) אם דרגת כל הקדקודים הינה זוגית, אז כל מסלול אוילר דרכה חייב להיות מעגל אוילר (והמסילה הינה מעגל).

(ii) אם יש במסילה שני קדקודים שדרגתם אי-זוגית, אז כל מסלול אוילר דרכה מחבר שני קדקודים אלו (קדקודים אלו נקראים לכן *קצות המסילה*).

הערה:

ממסקנה 2 נובע, שלמרות שייתכנו במסילה מסלולי אוילר רבים, זהותם של הקצוות (כאשר יש כאלה) אינה תלויה בבחירה של המסלול (הגרף בציור 4 מהווה דוגמה טובה לכך). לעומת זאת מסתבר, שכאשר מסילה היא מעגל, אז *כל* מסלול אוילר דרכה הינו מעגלי (הגדרת מעגל דרשה רק שיש *איזשהו* מסלול מעגלי!). מעתה נוכל לכן להבדיל בין מעגל ובין מסילה שאינה מעגל. בסוג הראשון, כל מסלולי אוילר הם מעגליים. בסוג השני – אף אחד מהם אינו מעגלי.

טענה 3 ומסקנה 2 מספקות, כל אחת, תנאי *הכרחי* לקיום מסלול אוילר בגרף, אבל אף אחד מהתנאים אינו כשלעצמו מספיק (הראו זאת!). המשפט הבא (שהוא המשפט המרכזי של הפרק) מראה, ש*הצירוף* שלהם אכן מהווה תנאי הכרחי ומספיק.

משפט:

בגרף יש מסלול אוילר (כלומר, הגרף הוא מסילה) אם"ם הוא קשיר, ומספר קדקודיו, שדרגתם היא אי-זוגית, הוא 0 או 2.

הוכחה:

הכרחיות התנאים הוכחה בטענה 3 ומסקנה 2.

בשביל המספיקות, נניח תחילה ש- G הוא גרף קשיר, שדרגת כל קדקודיו היא זוגית, ונראה שיש בו מעגל אוילר. בשביל זה די להראות שכל מסלול ב- G, שאורכו מקסימלי (כלומר: אין שום מסלול אחר ב- G, שהוא ארוך יותר), חייב להיות מעגל אוילר ב- G. יהי אפוא c מסלול מקסימלי כזה. נראה תחילה ש- c הוא מעגלי. נניח שלא. יהי v_0 קדקוד הראשית של c. כיוון ש- c אינו מעגל, יש מספר אי-זוגי של קשתות על c, ש- v_0 הוא קצה שלהן (כשלולאות נספרות פעמיים). כיוון שדרגת v_0 זוגית, פירוש הדבר שיש קשת e, ש- v_0 נמצא עליה, אך אינה נמצאת על c. אם e מחברת את v_0 לקדקוד x, אז חיבורם של המסלול $<x, e, v_0>$ ושל c נותן מסלול ארוך מ- c, בסתירה למקסימליות c.

נראה עתה, של- c התכונה הבסיסית הבאה: אם v_0 קדקוד על c, ו- e קשת ש- v_0 נמצאת עליה, אזי e (וממילא גם הקצה השני שלה) נמצאת על c. נניח בשלילה, שאין

הדבר כך. כיוון ש- c מעגלי, אפשר לבצע בו הזזה, ולקבל מסלול מעגלי c' בעל אותו אורך, ועם אותם קשתות וקדקודים, המתחיל (ונגמר) ב- v_0. אפשר עתה לצרף ל- c' את הקשת e (עם הקצה השני שלה) ולקבל מסלול חדש, שהינו ארוך ב- 1 מ- c. זוהי אבל שוב סתירה למקסימליות c.

בעזרת התכונה הבסיסית הזו של c ניתן להוכיח בקלות, שאם p הינו מסלול ב- G המתחיל בנקודה על c, אז כל הקשתות והקדקודים של p נמצאים על c (הוכחה מלאה נעשית באינדוקציה על האורך k של p. שלב הבסיס, בו $k = 0$, הוא טריביאלי, כי אז $p = \langle v_0 \rangle$. שלב המעבר נעשה בעזרת התכונה הבסיסית של c, שהוכחנו).
יהי עתה v קדקוד כלשהו של G, ויהי v_0 קדקוד על c. כיוון ש- G קשיר, יש מסלול המחבר את v_0 ל- v, ולכן לפי מה שזה עתה הראינו, v חייב להיות על c.
תהי לבסוף e קשת של G. כיוון שהקצוות של e הם על c (כי *כל* הקדקודים של G הם על c!), אז e בעצמה הינה על c (שוב – לפי התכונה הבסיסית של c, שהראינו קודם).

הראינו ש- c מסלול מעגלי, שכל הקדקודים והקשתות של G נמצאים עליו. מכאן ש- c הוא מעגל אוילר ב- G (ו- G הינו מעגל).

הראינו שאם G גרף קשיר, שמספר הקדקודים בעלי דרגה אי-זוגית בו הוא 0, אז G מעגל. נניח עתה ש- G הוא גרף קשיר, שיש בו בדיוק שני קדקודים, x ו- y, שדרגתם אי-זוגית. נוסיף עתה ל- G קשת חדשה e בין x ל- y, ונקבל כך גרף חדש G'. G' הוא עדיין קשיר כמובן, ודרגת *כל* קדקודיו היא זוגית. לפי מה שהראינו יש לכן ב- G' מעגל אוילר. על-ידי הזזתו (ואולי היפוכו) ניתן להניח, שמעגל זה מתחיל ב- x ובקשת e (כלומר: הוא מהצורה $\langle x, e, y, \ldots, x \rangle$. אז תת-המסלול של מעגל אוילר זה בין y ל- x (כלומר, המסלול ללא הקשת e) הוא מסלול אוילר בגרף G, שעובר דרך כל הקדקודים והקשתות של e. במלים אחרות: הוא מסלול אוילר ב- G (שלא במפתיע, הוא מחבר כמובן את שני הקדקודים ב- G, שדרגתם אי-זוגית).

תרגיל

הראה, שאם נגדיר "מסלול-אוילר" כמסלול המכיל את כל הקשתות של G (אך לא בהכרח את כל הקדקודים), אז ב- G יש מסלול כזה אם"ם מספר הקדקודים שדרגתם אי-זוגית הוא 0 או 2, וכל שני קדקודים, שדרגתם שונה מאפס, הם קשורים.

בשלב זה נוכל סוף-סוף לספק פתרון מלא לבעיית הגשרים של קניגסברג.

דוגמאות:

(1) בגרף G_0 (ציור 3) אין מסלול אוילר, ולכן לבעיית גשרי קניגסברג אין פתרון.

(2) אם לגרף G_0 נוסיף קשת כלשהי (שאינה לולאה) בין שניים מקדקודיו, נקבל גרף שיש בו מסלול אוילר.

הסבר:
בגרף G_0 הדרגה של כל הקדקודים היא אי-זוגית. כיוון שמספר הקדקודים הוא 4, לא ייתכן, לפי המשפט האחרון, שיש בו מסלול אוילר. לעומת זאת, אם נוסיף לו קשת, שאינה לולאה, נקבל גרף, שבו לשני קדקודים (קצוות הקשת שהוספנו) יש דרגה זוגית, בעוד לשאר – דרגה אי-זוגית. לכן יש מסלול אוילר (שאינו מעגלי) בגרף זה.

הערות:

(1) למעשה, על מנת להראות, שב- G_0 *אין* מסלול אוילר, אין צורך במשפט האחרון. מסקנה 2 לפניו (שהוכחתה קלה בהרבה) מספיקה. לעומת זאת, על מנת להראות, שעל-ידי הוספה ל- G_0 של קשת כלשהי (שאיננה לולאה) מקבלים מסילה, השתמשנו אכן במשפט האחרון. יכולנו כמובן להראות זאת גם על-ידי בדיקת כל האפשרויות להוסיף קשת ל- G_0, ומציאת מסלולי אוילר בגרפים המתקבלים. זו דרך ארוכה הרבה יותר, ואינה פרקטית, כשמספר הקדקודים בגרף גדול.

(2) המשפט האחרון נוסח כמשפט קיום טהור: הוא קובע, שבתנאים מסוימים (קלים לבדיקה) יש מסלול אוילר. כדאי לשים לב אבל, שההוכחה שלו היתה *קונסטרוקטיבית*, ולכן אפשר להוציא ממנה *אלגוריתם* למציאה של מסלול אוילר בגרף, כאשר זה קיים (כלומר: כשהתנאים המתוארים במשפט מתקיימים). ואכן, ההוכחה מאפיינת מסלול אוילר במסילה כמסלול שלה, שיש לו אורך מקסימלי. בחינה לעומק של ההוכחה מראה עוד, שהיא כוללת למעשה פרוצדורה איך אפשר, בהינתן מסלול שאינו מסלול מקסימלי, למצוא מסלול ארוך יותר. אם נתחיל אפוא במסלול בעל קשת יחידה (או אפילו ממסלול מנוון, ללא קשתות!), נוכל להפעיל פרוצדורה זו שוב ושוב, וכשהגרף הנ"ל סופי, הרי בהכרח נגיע אחרי מספר פעמים למסלול בעל אורך מקסימלי. אם הגרף הוא מסילה, אז לפי הוכחת המשפט, מסלול זה יהווה מסלול אוילר באותו גרף (ניתן לכתוב בקלות תכנית מחשב, שתיישם אלגוריתם זה!).

תרגילים

הוכח:

1. כל מסילה לא פשוטה מכילה תת-גרף, שהוא מעגל שאינו מנוון.
2. כל שני קדקודים שונים במסילה p הם קדקודי הקצה של מסילה p', שהיא תת-גרף של p. אם p פשוטה ואינה מעגל, אז p' זו היא יחידה.

ה.3 עצים ויערות

פרק זה מוקדש לסוג מיוחד של גרפים, שיש לו חשיבות רבה במתמטיקה, ועוד יותר – במדעי המחשב: עצים.

בטרם נתחיל, נזכיר לקוראים, שבחלק זה במונח "גרפים" כוונתנו לגרפים סופיים. עוד קונוונציה טרמינולוגית נוחה, שנאמץ בפרק זה, היא, שבמונח "מעגל" נתכוון ל*מעגל לא מנוון* (דהיינו: מעגל, שיש בו לפחות קשת אחת). הסכם זה מאפשר ניסוח קצר ובהיר יותר של ההגדרות והמשפטים, והוא תואם את המקובל בספרות.

הגדרה 1:

(א) לגרף, שאין בו מעגלים, קוראים *יער*.

(ב) לגרף קשיר, שאין בו מעגלים, קוראים *עץ*.

עץ הינו, אפוא, *יער קשיר*. יער, מצד שני, הינו גרף, שכל רכיבי הקשירות שלו הינם עצים.

תרגיל

הוכח את הטענה האחרונה.

עיקר המשפטים בפרק זה יוקדש לאפיונם של יערות ועצים. נפתח במשפט המאפיין יערות. לצורך הוכחתו נזדקק לטענה הבאה, שיש לה חשיבות בפני עצמה.

משפט 1:

בכל גרף G, שדרגת כל קדקודיו היא 2 לפחות, יש מעגל.

הוכחה:

יהי $p = <v_0, e_1, v_1, \ldots, e_n, v_n>$ מסלול ב- G, שאורכו מקסימלי (כיוון ש- G סופי, מסלול כזה קיים, כי אורך כל מסלול אינו יכול לעלות על מספר הקשתות של G). המסלול p חייב לכלול את כל הקשתות, ש- v_n נמצא עליהן (לו היתה קשת e מחוץ ל- p, ש- v_n נמצא עליה, היינו יכולים להוסיף אותה ואת הקצה השני שלה ל- p, ולקבל

כך מסלול ארוך מ- p). כיוון ש- $d_G(v_n) \geq 2$, פירוש הדבר הוא, שחייב להיות $i < n$, כך ש- $v_i = v_n$. מכאן, שעבור i זה $<v_i, e_{i-1}, \ldots, e_n, v_n>$ הינו מעגל.[1]

מסקנה 1:

אם G יער, אז יש ב- G קדקוד, שדרגתו 0 או 1.

משפט 2:

גרף G הינו יער, אם"ם בכל תת-גרף שלו יש קדקוד, שדרגתו 0 או 1.

הוכחה:

כיוון שברור, שכל תת-גרף של יער הוא בעצמו יער, ה"רק אם" נובע ממסקנה 1. לכיוון ההפוך, נניח שבכל תת-גרף של G יש קדקוד, שדרגתו 0 או 1. כיוון שבמעגל דרגת כל קדקוד 2 לפחות, פירוש הדבר שאין ב- G מעגלים.

הגדרה 2:

(א) קדקוד בגרף, שדרגתו 0, נקרא קדקוד ***מבודד***.

(ב) קדקוד בגרף, שדרגתו 1, נקרא ***עלה***.

את משפט 2 נוכל לכן לנסח כך: גרף G הינו יער, אם"ם בכל תת-גרף שלו יש קדקוד מבודד או עלה.

מסקנה 2:

גרף G הינו עץ אם"ם הוא קשיר, ובכל תת-גרף שלו יש קדקוד, שדרגתו 0 או 1.

תרגיל

בכל עץ, שמספר קדקודיו גדול מ- 1, יש לפחות שני עלים. (רמז: ניקח מסלול p בעל אורך מקסימלי. p אינו יכול להיות מעגלי או מנוון, ושני קצותיו חייבים להיות עלים.)

משפט 3:

יהי $G = <V, E, F>$.

(א) אם G יער, אז $|E| \leq |V| - 1$.

(ב) אם G קשיר, אז $|E| \geq |V| - 1$.

(ג) אם G הינו עץ, אז $|E| = |V| - 1$.

---

[1] ייתכן ש- $i = n - 1$, ואז e_n הינה לולאה. לולאה היא אכן מעגל, שאורכו 1.

הוכחה:

(א) באינדוקציה על $|V|$.

בסיס האינדוקציה: נניח $|V| = 1$. במקרה זה כל קשת היא בהכרח לולאה. כיוון שכל לולאה היא מעגל, ואילו G הוא יער, הרי מספר הקשתות חייב להיות 0 (ולכן $|V| - 1 = |E| = 0$ במקרה זה).

שלב המעבר: נניח נכונות הטענה כאשר $|V| = n$, ונניח $|V| = n + 1$. כיוון ש- G יער, יש בו לפי מסקנה 1 קדקוד v_0, שדרגתו 0 או 1. יהי $V' = V - \{v_0\}$, ויהי $G' = <V', E', F'>$ תת-הגרף של G, המושרה על-ידי V'. כיוון ש- $d_G(v_0) \le 1$, E' מכיל את כל הקשתות של G, פרט אולי לאחת. לכן $|E| \le |E'| + 1$. מצד שני, G' הוא יער עם n קדקודים, ולכן $|E'| \le |V'| - 1$ לפי הנחת האינדוקציה. מכאן ש- $|E'| \le |V| - 2$, ולכן:

$$|E| \le |E'| + 1 \le (|V| - 2) + 1 = |V| - 1$$

(ב) נניח G קשיר. יהי $x_0 \in V$. נבנה פונקציה $f : (V - \{x_0\}) \to E$ באופן הבא: לכל קדקוד y, השונה מ- x_0, נבחר מסלול בעל אורך מינימלי מ- x_0 אל y (קיים, כי G קשיר), ונתאים ל- y את הקשת האחרונה במסלול זה. נוכיח עתה, ש- f זו הינה ח.ח.ע., ומזה ינבע ש- $|E| \ge |V - \{x_0\}| = |V| - 1$.

נניח בשלילה, ש- f לא ח.ח.ע.. אז קיימים $y_1, y_2 \in V$ ו- $e \in E$, כך ש- $f(y_1) = f(y_2) = e$, אבל $y_1 \ne y_2$. כיוון שמהגדרת f ברור, שכל קדקוד y הוא נקודת קצה של $f(y)$, y_1 ו- y_2 הם שני הקצוות של e. מהגדרת f נובע גם, שקיימים מסלולים p_1 ו-p_2, כך ש- p_1 הוא מסלול באורך מינימלי מ- x_0 ל- y_1, p_2 הוא מסלול באורך מינימלי מ- x_0 ל- y_2, ו- e היא הקשת האחרונה בשניהם. לכן ל- p_1 ול- p_2 הצורה הבאה:

$$p_1 = <x_0, e_1, \ldots, e_n, y_2, e, y_1>$$

$$p_2 = <x_0, e'_1, \ldots, e'_k, y_1, e, y_2>$$

עתה, $<x_0, e'_1, \ldots, e'_k, y_1>$ הוא מסלול מ- x_0 אל y_1. לכן אורכו גדול או שווה לזה של p_1, דהיינו: $k \ge n + 1$ (ולכן $k > n$). מצד שני $<x_0, e_1, \ldots, e_n, y_2>$ הוא מסלול מ- x_0 אל y_2, ולכן אורכו גדול או שווה לזה של p_2, דהיינו $n \ge k + 1$ (ולכן $n > k$). קיבלנו סתירה, כי לא ייתכן,שגם $n > k$ וגם $k > n$.

(ג) סעיף זה נובע מיידית משני קודמיו (והגדרת עץ כיער קשיר).

מסקנה 3:

התנאים הבאים שקולים עבור גרף $G = <V, E, F>$:

(א) G עץ.

(ב) G חסר מעגלים ו- $|E| = |V| - 1$.

(ג) G חסר מעגלים מקסימלי, כלומר: G חסר מעגלים, אך על-ידי הוספת קשת כלשהי ל- E (בין קדקודים של V) נקבל גרף שיש בו מעגלים.[2]

הוכחה:

(א) ⇐ (ב): מיידי מהגדרת עץ וסעיף (ג) של משפט 3.

(ב) ⇐ (ג): מיידי מסעיף (א) של משפט 3.

(ג) ⇐ (א): נניח ש- G הוא חסר מעגלים מקסימלי. כדי להוכיח, ש- G הינו עץ, עלינו להראות, שכל שני קדקודים v ו- w של G הינם קשורים. ברור, שזהו המצב כאשר יש קשת ב- E בין v ל- w. נניח אפוא, שאין ב- E קשת כזו. ניצור גרף חדש, G', על V על-ידי שנוסיף ל- E קשת חדשה, e^*, בין v ל- w. מתכונת המקסימליות של G נובע, ש- G' מכיל מעגל. הקשת e^* חייבת להימצא על מעגל זה (כי אחרת היה מדובר במעגל ב- G, אך G חסר מעגלים!). מכאן שיש ב- G' מסלול מעגלי מהצורה:

$$<v, e^*, w, e_1, \ldots, e_n, v>$$

כאשר $e_1, \ldots, e_n$ הן קשתות ב- E. עתה, $<w, e_1, \ldots, e_n, v>$ הינו מסלול ב- G בין w ל- v, ולכן v ו- w אכן קשורים ב- G.

עבור האפיון הבא של עצים נזדקק ללמה:

למה 1:

יהי $G = <V, E, F>$ גרף קשיר, ונניח ש- $e \in E$ היא קשת המונחת על איזשהו מעגל ב- G. אז $G' = <V, E - \{e\}, F/(E - \{e\})>$ הינו גרף קשיר (ובניסוח פחות מדויק: הסרת קשת ממעגל בגרף קשיר אינה פוגעת בקשירות).

הוכחה:

נניח שהקצוות של e הם a ו- b. כיוון ש- e נמצאת על מעגל ב- G, קיים ב- G מסלול מעגלי מהצורה: $<a, e, b, e_1, \ldots, e_k, a>$, כאשר $e_1, \ldots, e_k \in E - \{e\}$. מכאן ש- $p = <b, e_1, \ldots, e_k, a>$ הוא מסלול ב- G' בין b ל- a.

---

[2] ניסוח מדויק יותר: G חסר מעגלים, אך בכל גרף מהצורה $<V, E', F'>$, כך ש- $E \subset E'$, יש מעגלים.

יהיו עתה $v_1, v_2 \in V$. נראה שהם קשורים ב- G'. כיוון ש- G קשיר, הרי יש ב- G מסלול q בין v_1 ל- v_2. אם e אינה נמצאת על מסלול זה, אז q עצמו הינו מסלול ב- G' בין v_1 ל- v_2. אם e כן נמצאת על q, אז אחד מהקטעים $<a, e, b>$ או $<b, e, a>$ הוא קטע של q. במקרה הראשון נחליף ב- q קטע זה עם p. במקרה השני נחליף אותו בהיפוכו של p. בשני המקרים נקבל כך מ- q מסלול חדש בין v_1 ל- v_2, שנמצא כולו ב- G'.

מסקנה 4:

התנאים הבאים שקולים עבור גרף $G = <V, E, F>$:

(א) G עץ.

(ב) G קשיר ו- $|E| = |V| - 1$.

(ג) G קשיר מינימלי, דהיינו: G קשיר, אך אם נסלק קשת כלשהי מ- E, נקבל גרף, שאינו קשיר.

הוכחה:

(א) $\Leftarrow$ (ב): מיידי מהגדרת עץ וסעיף (ג) של משפט 3.

(ב) $\Leftarrow$ (ג): מיידי מסעיף (ב) של משפט 3.

(ג) $\Leftarrow$ (א): נניח ש- G קשיר מינימלי. כדי להראות ש- G הינו עץ, יש להוכיח רק, שב- G אין מעגלים. זה נובע מלמה 1, כיוון שלו היה ב- G מעגל, היינו יכולים, לפי למה זו, להסיר מ- G את אחת הקשתות של מעגל זה, ולקבל כך גרף קשיר. זה סותר את תכונת המינימליות של G.

ללמה 1 יש מסקנה חשובה נוספת. כדי לנסח אותה, עלינו להכניס מושג חדש:

הגדרה 3:

(א) תת-גרף G' של גרף G *פורש* את G, אם ל- G ול- G' יש אותם קדקודים.

(ב) *עץ פורש* של גרף G הוא תת-גרף של G, הפורש את G ומהווה עץ.

משפט 4:

לגרף $G = <V, E, F>$ יש עץ פורש אם"ם הוא קשיר.

הוכחה:

ברור שאם ל- G יש תת-גרף קשיר, אז G עצמו קשיר. לכן אם ל- G יש עץ פורש, אז G קשיר.

לכיוון ההפוך נוכיח באינדוקציה על $|E|$, שאם G קשיר, אז יש לו עץ פורש.

אם $|E| = 0$, אז כיוון ש- G קשיר, $|V| = 1$. לכן, במקרה זה G עצמו הוא עץ, וממילא הוא עץ פורש עבור עצמו.

נניח את נכונות הטענה כאשר $|E| = n$, ונניח $|E| = n + 1$. אם אין ב- G מעגלים, אז G הוא עץ, והוא פורש את עצמו. אחרת יש ב- G מעגל, ולכן, לפי למה 1, נוכל להסיר מ- G את אחת הקשתות של מעגל זה ולקבל גרף קשיר G' עם n קשתות. לפי הנחת האינדוקציה יש ל- G' עץ פורש T. כיוון של- G ול- G' יש אותם קדקודים, T הוא גם עץ פורש של G (שימו לב, שאנו מסתמכים כאן על העובדה הברורה, שאם T הינו תת-גרף של G', ו- G' הוא תת-גרף של G, אז T הוא תת-גרף גם של G).

הערה:

ההוכחה מקפלת למעשה בחובה אלגוריתם למציאת עץ פורש עבור גרף נתון G על-ידי סילוק חוזר של קשתות, עד שנשארים עם עץ פורש.

נחזור עתה לנושא של אפיוני עצים. תחילה למה חשובה נוספת:

למה 2:

אם קיימים בגרף G שני קדקודים (לא בהכרח שונים), שיש ביניהם שני מסלולים שונים, אז יש ב- G מעגל.

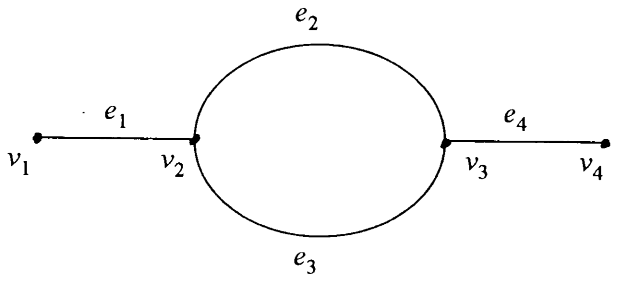
הוכחה:

נניח, שיש ב- G שני מסלולים שונים עם אותם קדקודי התחלה וסוף. נבחר שני מסלולים p ו- p' כאלה, שסכום אורכיהם הוא מינימלי. ל- p ול- p' יש אפוא הצורה הבאה:

$$p = <x, e_1, z_1, \ldots, z_n, e_n, y> \quad , \quad p' = <x, e'_1, z'_1, \ldots, z'_k, e'_k, y>$$

נסמן ב- q את תת-המסלול של p בין z_1 ל- y, וב- q' את תת-המסלול של p' בין z'_1 ל- y. כיוון ש- $p \neq p'$, גם $e_1 \neq e'_1$, כי אחרת היינו מקבלים ש- $z_1 = z'_1$, ולכן q ו- q' היו אז שני מסלולים שונים עם אותם קדקודי התחלה וסוף, שסכום אורכיהם קטן מזה של p ו- p' (בסתירה לבחירת p ו- p'). עתה, אם x מופיע על q או על q', אז p או p' מכיל מעגל, וסיימנו. נניח אפוא שלא. אז e_1 ו- e'_1 גם הן אינן נמצאות על q או על q'. לכן החיבור של q עם היפוכו של q' מהווה טיול בין z_1 ל- z'_1, שאינו מכיל את e_1 או את e'_1. מטיול זה ניתן להוציא מסלול r בין z_1 ל- z'_1, שאף הוא אינו מכיל כמובן את e_1 או את e'_1 (השווה להוכחת טענה 1 של הפרק הקודם). מכאן, שחיבורם של $<x, e_1, z_1>$, r ו- $<z'_1, e'_1, x>$ נותן מעגל (מהצורה: $<x, e_1, z_1, \ldots, z'_1, e'_1, x>$).

הערה:
יש לשים לב, שקיומם של שני מסלולים שונים בין שני קדקודים *אין* פירושו, ששני קדקודים אלו, או אפילו רק אחד מהם, חייבים להיות מונחים על מעגל. כך למשל, יש שני מסלולים שונים בין הקדקודים v_1 ו- v_4 בגרף של ציור 5, אך גם v_1 וגם v_4 אינם נמצאים על שום מעגל בגרף זה.



ציור 5

מסקנה 5:
גרף G הינו עץ אם"ם בין כל שני קדקודים שלו יש מסלול (פשוט) יחיד.

הוכחה:
נניח ש- G הינו עץ. אז G קשיר, ולכן יש מסלול בין כל שני קדקודים שלו. מסלול זה הוא יחיד לפי למה 2, כיוון ש- G הוא חסר מעגלים (והוא גם פשוט מאותה סיבה בדיוק!).
לכיוון ההפוך, נניח כי בין כל שני קדקודים של G יש מסלול יחיד. אז G הוא בוודאי קשיר. כמו-כן, לא ייתכנו ב- G מעגלים, כי כל שני קדקודים על מעגל קשורים זה לזה בשני מסלולים שונים.[3]

המשפט הבא מסכם את כל האפיונים, שמצאנו עבור מושג ה"עץ":

משפט 5:
התנאים הבאים לגבי גרף $G = <V, E, F>$ הם שקולים:
(א) G עץ.
(ב) G קשיר, ובכל תת-גרף שלו יש קדקוד, שדרגתו 0 או 1.
(ג) G קשיר, ו- $|E| = |V| - 1$.
(ד) G קשיר מינימלי.

---

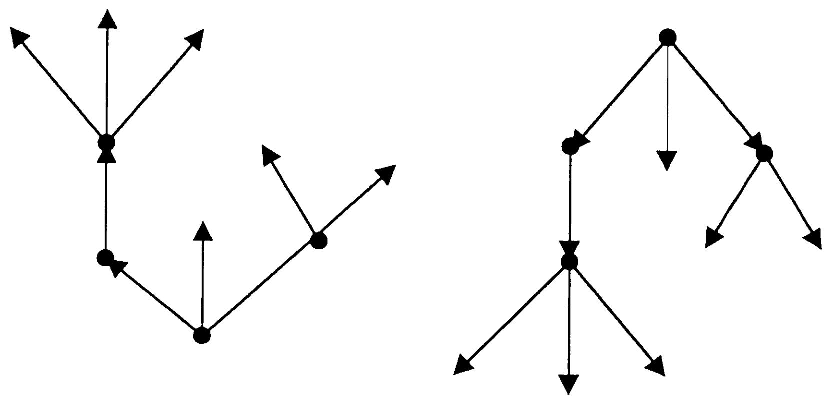
[3] כדאי לשים לב שזה נכון גם כאשר שני הקדקודים הם זהים: כל קדקוד v על מעגל p קשור לעצמו בשני מסלולים שונים: p עצמו, והמסלול הטריביאלי $<v>$.

(ה) בין כל שני קדקודים של G יש מסלול (פשוט) יחיד.

(ו) G חסר מעגלים, ו- $|G| = |V| - 1$.

(ז) G חסר מעגלים מקסימלי.

מצאנו עד כה אפיונים שונים ומשונים למושג ה"עץ", אבל אף אחד מהם לא מבהיר, בעצם, למה קוראים אנו לגרף (לא-מכוון) קשיר וחסר מעגלים בשם "עץ". אינטואיטיבית, עץ אמור דווקא להיות גרף מכוון, שיש לו "שורש", והוא מתפצל ל"ענפים". זהו אכן מושג העץ, כפי שהוא מופיע במרבית השימושים במדעי המחשב. דוגמה לגרף כזה ניתן למצוא בציור 6:[4]



ציור 6

מטרתנו הבאה תהיה להגדיר מושג אינטואיטיבי זה של עץ באופן מדויק, ולהבהיר את הקשר בינו ובין מושג העץ, כפי שהגדרנו למעלה.

הגדרה 4:

(א) ***עץ מכוון*** הוא גרף מכוון G, שיש בו קדקוד a, הקשור לכל קדקוד של הגרף על-ידי מסלול יחיד. לקדקוד a קוראים ***השורש*** של G.

(ב) אם G הינו עץ מכוון ו- v_1, v_2 הם קדקודים שלו, אז v_1 נקרא ***הורה*** של v_2, ו- v_2 ***ילד*** של v_1, אם יש קשת ב- G מ- v_1 אל v_2. v_1 נקרא ***אב קדמון*** של v_2, ו- v_2 ***צאצא*** של v_1, אם יש מסלול ב- G בין v_1 ל- v_2, שאורכו גדול מאפס.

(ג) קדקוד ללא ילדים בעץ מכוון נקרא ***עלה*** של אותו עץ.

[4] בשני חלקי הציור מתואר אותו גרף מכוון עצמו, כשפעם הוא מצויר כשכיוון הסתעפותו כלפי מעלה, ופעם – כלפי מטה. מתמטית אין לכך כל חשיבות כמובן (במדעי המחשב, אגב, מרבית העצים "גדלים" כלפי מטה, משום מה).

למה 3:

יהי G עץ מכוון עם שורש a, ויהיו v ו-w קדקודים שלו. נניח ש- p_1 הוא המסלול היחיד מ- a ל- v, ונניח ש- p_2 הוא מסלול מ- v ל- w. אז חיבורם של p_1 ו- p_2 הוא מסלול מ- a ל- w.

הוכחה:

החיבור של p_1 ו- p_2 הוא כמובן טיול מ- a ל- w. כדי להראות, שהוא גם מסלול, די להוכיח, של- p_1 ול- p_2 אין שום קשת משותפת. נניח בשלילה, שיש להם קשת כזו. תהי e הקשת הראשונה של p_2, הנמצאת גם על p_1. יהי u קדקוד היציאה של e. נסמן ב- p'_2 את קטע המסלול של p_2 עד e (לא כולל e). p'_2 הוא מסלול מ- v ל- u, שאין לו שום קשת משותפת עם p_1. לכן החיבור של p_1 ו- p'_2 הוא מסלול מ- a ל- u. אורך מסלול זה הינו גדול או שווה לזה של p_1. מצד שני, הקטע של p_1 עד e (לא כולל e) הוא מסלול מ- a ל- u, שאורכו קטן מזה של p_1. קיבלנו שני מסלולים שונים מהשורש a לקדקוד u, בסתירה להגדרת עץ מכוון.

משפט 6:

נניח ש- G הינו עץ מכוון עם שורש a. אזי:

(א) $d_G^+(a) = 0$

(ב) $d_G^+(x) = 1$ לכל x השונה מ- a.

הוכחה:

(א) נניח בשלילה, שיש קשת e הנכנסת ל- a. נניח שקשת זו מתחילה בקדקוד w. כיוון ש- a שורש של העץ, יש מסלול בעץ מ- a ל- w. אם נחבר למסלול זה את $\langle w, e, a\rangle$, נקבל, לפי למה 3, מסלול לא טריביאלי מ- a ל- a. בין a ל- a קיים אבל גם המסלול הטריביאלי $\langle a\rangle$. מכאן שבין a לעצמו יש שני מסלולים שונים, בסתירה לכך, שבין a לכל קדקוד של העץ יש, לפי הגדרה, מסלול יחיד.

(ב) נניח ש- x הוא קדקוד השונה מ- a. אז קיים מסלול בין a ל- x. מסלול זה אינו טריביאלי (כי $x \neq a$) ולכן הוא חייב להסתיים בקשת, הנכנסת ל- x. מכאן ש- $d_G^+(x) \geq 1$. נניח בשלילה, ש- $d_G^+(x) > 1$. מכאן שיש שתי קשתות שונות, e_1 ו- e_2, הנכנסות ל- x. נניח שקשתות אלו יוצאות מהקדקודים v_1 ו- v_2 (בהתאמה). כיוון ש- a הוא שורש, יש מסלולים p_1 ו- p_2 (בהתאמה) מ- a אל v_1 ואל v_2. על-ידי חיבור p_1 ל- $\langle v_1, e_1, x\rangle$ וחיבור p_2 עם $\langle v_2, e_2, x\rangle$ נקבל לפי למה 3 שני מסלולים שונים מ- a ל- x (הם שונים, כי $e_1 \neq e_2$). זוהי סתירה לכך, שמ- a יש מסלול יחיד אל x. לכן לא ייתכן, ש- $d_G^+(x) > 1$, ומכאן ש- $d_G^+(x) = 1$.

מסקנה 6:

לכל עץ מכוון יש רק שורש אחד.

הוכחה:

ממשפט 6 נובע מצד אחד, שלכל שורש יש דרגת כניסה 0, ומצד שני שיש רק קדקוד אחד עם דרגת כניסה כזו.

הערה:

למעשה רק עכשיו, במסקנה 6, הצדקנו את השימוש בה"א הידוע בהגדרת "השורש של עץ מכוון" למעלה (הגדרה (4)(א)).

מסקנה 7:

אם $G = <V, E, F>$ הוא עץ מכוון, אז $|E| = |V| - 1$.

הוכחה:

זה מיידי ממשפט 6, ומהעובדה ש- $|E| = \sum_{x \in V} d_G^+(x)$ (טענה 4 בפרק הקודם).

למה 4:

(א) בין כל שני קדקודים של עץ מכוון יש לכל היותר מסלול אחד.

(ב) בעץ מכוון אין מעגלים.

הוכחה:

(א) נניח שבין הקדקודים v ו- w יש שני מסלולים שונים. חיבורם של מסלולים אלו אל המסלול, שקיים בין השורש a ל- v, ייתן, לפי למה 3, שני מסלולים שונים בין a ל- w, בסתירה להגדרת השורש.

(ב) כל מעגל מספק מסלול לא טריביאלי בין כל קדקוד v עליו אל עצמו. כיוון שתמיד קיים גם המסלול הטריביאלי $<v>$, חלק (ב) הוא מקרה פרטי של חלק (א).

משפט 6, מסקנותיו ולמה 4 מספקים יחד תמונה טובה של המבנה של עץ מכוון: בכל עץ כזה שום קדקוד אינו אב קדמון של עצמו. השורש הוא אב קדמון של כל הקדקודים האחרים. לשורש אין הורה, ולכל קדקוד אחר יש הורה יחיד. בין הורה לכל אחד מילדיו אין מסלול אחר, פרט לקשת המחברת אותם. בין כל קדקוד לבין צאצא שלו יש מסלול יחיד. האבות הקדמונים של כל קדקוד v הם הקדקודים הנמצאים על המסלול היחיד מהשורש ל- v. ענף בעץ הוא מסלול, המתחיל בשורש ונגמר באיזה עלה, וכל מסלול

חלקי לאיזשהו ענף. אם v_1 ו- v_2 הם שניהם אבות קדמונים שונים של קדקוד w, אז אחד מהם הינו אב קדמון של השני.

נשאיר לקוראים לבדוק, שכל הטענות הללו על מבנה עץ מכוון נובעים אכן בקלות ממה שכבר הוכחנו, ונעבור עתה לקשרים בין עצים מכוונים ועצים (לא מכוונים).

משפט 7:

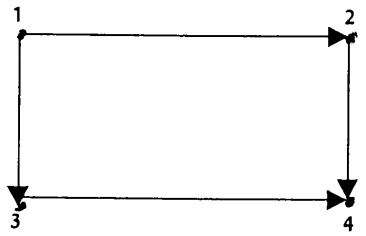
אם G' הינו עץ מכוון, אז הגרף (הלא מכוון), ש- G' מבוסס עליו, הוא עץ (לא מכוון).

הוכחה:

נניח ש- G' מבוסס על G. כיוון שכל מסלול ב- G' הוא גם מסלול של G[5], השורש של G' קשור ב- G לכל קדקוד אחר. מזה נובע, בגלל הסימטריות והטרנזיטיביות של יחס הקשירות, שכל קדקוד של G קשור לכל קדקוד אחר. דהיינו: G קשיר. מזה, מסעיף (ג) של משפט 5 וממסקנה 7 נובע, ש- G הינו עץ.

אזהרה:

הסתמכות על למה 4 על מנת לטעון ישירות, ש- G הוא חסר מעגלים, לא היתה מספיקה כאן: מהעובדה, שבגרף מכוון מסוים G' אין מעגלים, ***לא*** נובע בהכרח, שגם בגרף הלא מכוון, ש- G' מבוסס עליו, אין מעגלים. כך למשל, אין מעגלים בגרף המכוון של ציור 7, אך הגרף הלא מכוון, ש- G' מבוסס עליו, הינו מעגל.



ציור 7

המשפט הבא הולך בכיוון ההפוך מזה של משפט 7. הוא מראה, כיצד ניתן להפוך כל עץ לא-מכוון לעץ מכוון, על-ידי שבוחרים, איזה קדקוד יהווה את השורש.

משפט 8:

יהי $G = \langle V, E, F\rangle$ עץ (לא מכוון), ויהי $a \in V$. אז קיים עץ מכוון יחיד G' המבוסס על G, כך שהקדקוד a הוא השורש שלו.

[5] ההיפך, נזכור, אינו נכון.

כדי להוכיח את משפט 8, נזדקק לשתי הלמות הבאות:

למה 5:

יהי $G = <V, E, F>$ עץ (לא מכוון), ויהיו $a \in V$, $e \in E$. נניח ש- e מחברת את הקדקודים v_1 ו- v_2, ויהיו p_1 ו- p_2 המסלולים היחידים המחברים את a ל- v_1 ול- v_2 (בהתאמה)[6]. אז או ש- e הינה הקשת האחרונה ב- p_1 ואינה נמצאת על p_2, או ש- e הינה הקשת האחרונה ב- p_2 ואינה נמצאת על p_1.

הוכחת למה 5:

אם e נמצאת על p_1, אז כיוון שלפי מסקנה 5 p_1 הינו פשוט, ל- p_1 יש הצורה $<a, \ldots, v_2, e, v_1>$, ו- e הינה אפוא הקשת האחרונה ב- p_1. כמו-כן, הקטע של p_1 בין a ל- v_2 הוא מסלול בין a ל- v_2, ולכן שווה ל- p_2. מכאן ש- e אינה נמצאת על p_2 במקרה זה.

אם e אינה נמצאת על p_1, אז נוכל לחבר את $<v_1, e, v_2>$ ל- p_1 ולקבל כך מסלול בין a ל- v_2, דהיינו: את p_2. מכאן, שבמקרה זה e הינה הקשת האחרונה ב- p_2.

למה 6:

יהי $G = <V, E, F>$ עץ, $a \in V$, ויהי G' עץ מכוון המבוסס על G, ש- a הינו השורש שלו. נניח ש- e היא קשת ב- G המחברת את הקדקודים v ו- w, וש- e נמצאת בסופו של המסלול היחיד ב- G מ- a ל- w. אז e מחברת ב- G' את v ל- w (ולא להיפך).

הוכחת למה 6:

יהי p המסלול היחיד ב- G' מ- a ל- w. p הינו גם מסלול (יחיד בהכרח) ב- G בין a ל- w [7], ולכן e נמצאת בסופו של p. מכאן של- p הצורה: $<a, \ldots, v, e, w>$. כיוון ש- p מסלול של G', הכיוון של e ב- G' חייב להיות מ- v אל w.

הוכחת משפט 8:

יחידות: נניח ש- $G_1 = <V, E, F_1>$ ו- $G_2 = <V, E, F_2>$ הם שני עצים מכוונים המבוססים על G, ונניח ש- $a \in V$ הוא השורש של שניהם. יהי $e \in E$, ונניח ש- e מחברת ב- G את v_1 ו- v_2 (כלומר: $F(e) = \{v_1, v_2\}$). יהיו p_1 ו- p_2 המסלולים היחידים ב- G בין a ל- v_1 ו- v_2 (בהתאמה). לפי למה 6, אם e היא הקשת האחרונה ב- p_1, אז הכיוון שלה גם ב- G_1 וגם ב- G_2 הוא מ- v_2 אל v_1 (כלומר: $F_1(e) = F_2(e) = <v_2, v_1>$), בעוד שאם e היא הקשת האחרונה ב- p_2, אז הכיוון שלה גם ב- G_1 וגם ב- G_2 הוא

---

[6] ראה סעיף (ה) של משפט 5 או מסקנה 5.

[7] ראה הערה (6) אחרי הגדרת "טיול" בפרק הקודם. ההערה נכונה כמובן גם למסלולים.

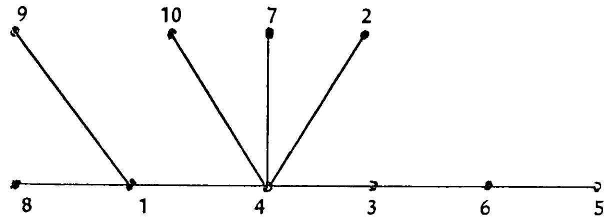
מ- v_1 אל v_2 (כלומר: $F_1(e) = F_2(e) = <v_1, v_2>$). לאור למה 5, פירוש הדבר, שבכל מקרה $F_1(e) = F_2(e)$. זה נכון לכל $e \in E$, ולכן $F_1 = F_2$, וממילא $G_1 = G_2$.

קיום: נגדיר גרף מכוון $G' = <V, E, F'>$, המבוסס על G, באופן הבא: אם $e \in E$, ו- $F(e) = \{v_1, v_2\}$, אז $F'(e) = <v_1, v_2>$, אם e היא הקשת האחרונה על p_2, ו- $F'(e) = <v_2, v_1>$, אם e היא הקשת האחרונה על p_1 (כאשר p_1 ו- p_2 מוגדרים כמו בלמה 5). F' מוגדר היטב לאור למה 5. נראה, ש- G' הוא עץ מכוון עם הקדקוד a כשורש, דהיינו: לכל $v \in V$ יש ב- G' מסלול יחיד בין a ל- v. עתה כל מסלול ב- G' הוא גם מסלול ב- G, ולכל $v \in V$ יש ב- G בדיוק מסלול אחד בין a ל- v. לכן מספיק להוכיח, שכל מסלול p ב- G המתחיל ב- a, הוא גם מסלול ב- G'. נראה זאת באינדוקציה על n – אורך המסלול p.
אם $n = 0$, אז $p = <a>$, וזהו כמובן מסלול גם ב- G'.

נניח את נכונות הטענה לגבי מסלולים באורך n, ויהי p מסלול ב- G, שאורכו $n + 1$. נניח $p = <a, e_1, \ldots, e_n, v_n, e_{n+1}, v_{n+1}>$. אז $<a, e_1, \ldots, e_n, v_n>$ מסלול באורך n, המתחיל ב- a. לכן, לפי הנחת האינדוקציה, זה גם מסלול ב- G'. כיוון ש- p הינו מסלול ב- G המתחיל ב- a, נובע מהגדרת G', ש- $F(e_{n+1}) = <v_n, v_{n+1}>$. משתי עובדות אלו נובע, ש- p הוא אכן מסלול גם ב- G'.

הערה:
נשים לב, שהדרך, בה הוכחנו את היחידות במשפט הקודם, כפתה למעשה את הגדרת G', כפי שניתנה בהוכחת חלק הקיום של משפט זה.

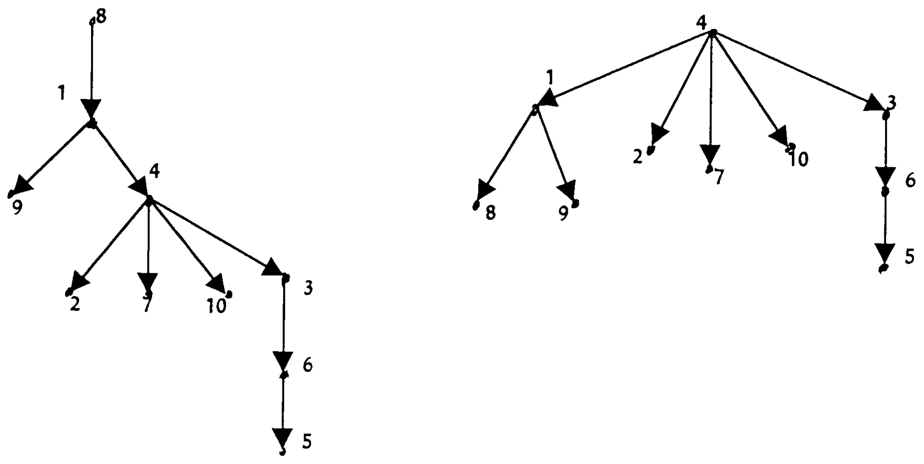
דוגמה:
יהי $G = <V, E>$ העץ[8] המתואר בציור 8:



ציור 8

[8] כיוון שבעצים אין לולאות ואין קשתות מקבילות, אפשר בדרך-כלל להתייחס לעץ כאל זוג $<V, E>$, וכך אכן מקובל לעשות. ראו דיון בפרק ה.1.

בחירת 4 כשורש תיתן את העץ המכוון, המתואר בצד ימין של ציור 9, בעוד בחירת 8 כשורש תיתן את העץ המכוון, המתואר בצד שמאל של ציור זה.



ציור 9

תרגיל:

מה הקשר בין העלים של עץ (לא מכוון) והעלים של עץ מכוון המבוסס עליו? (ראה הגדרות (2)(ב) ו- (4)(ג)).

התרגיל הבא נותן אפיון מסוג אחר למושג של "עץ".

תרגיל:

(א) יהי $T = <V, E, F>$ עץ מכוון. נגדיר יחס $\leq_T$ על V באופן הבא: $v \leq_T w$ אם"ם $v = w$, או w הינו צאצא של v.

הראה שליחס $\leq_T$ התכונות הבאות:

(1) $\leq_T$ הוא יחס סדר חלקי.

(2) קיים ב- V איבר קטן ביותר לפי $\leq_T$, דהיינו איבר a, כך ש- $\forall x \in V.\ a \leq_T x$.

(3) לכל $v \in V$ הקבוצה $\{x \in V \mid x \leq_T v\}$ סדורה מלא על-ידי $\leq_T$.

(ב) נניח ש- V קבוצה סופית לא ריקה, ו- $\leq$ יחס על V, המקיים את תכונות (1)-(3) מחלק (א). נגדיר גרף מכוון $<V, E>$ על V, על-ידי שניצור קשת בין קדקוד v לקדקוד w אם"ם w הינו עוקב של v לפי $\leq$ (דהיינו: $v \leq w$, $w \neq v$, ואם $v \leq x \leq w$, אז $x = v$ או $x = w$). הראה ש- $T = <V, E>$ הוא עץ מכוון, ו- $\leq_T = \leq$.

ה.4 נוסחת קיילי

הפרקים הקודמים הוקדשו להכרת המושגים היסודיים של תורת הגרפים. לא היה בהם למעשה יותר מאשר מבוא מתמטי מדויק לתורה עשירה זו. כדי לתת לקורא רמז כלשהו לעושר זה, נקדיש פרק זה למשפט לא טריביאלי אחד של תורת הגרפים. את השאר נשאיר לקורסים מתקדמים יותר.

השאלה המרכזית, שנעסוק בה בפרק זה, הינה: בהינתן קבוצה סופית V עם שני איברים לפחות, כמה עצים מהצורה $<V, E>$ (כש- $E \subseteq V^2$) קיימים? במלים אחרות: בכמה אופנים ניתן להפוך קבוצה סופית נתונה לעץ? התשובה ניתנת במשפט הבא:

נוסחת קיילי:

תהי V קבוצה סופית כך ש- $|V| \geq 2$. אז מספר העצים מהצורה $<V, E>$ הוא $|V|^{|V|-2}$. ובסימונים:

$$|V| \geq 2 \;\;\Rightarrow\;\; |T(V)| = |V|^{|V|-2}$$

כאשר:

$$T(V) = \{E \in P(P_2(V)) \mid \text{הינו עץ } <V, E>\}$$

כצעד ראשון בהוכחת משפט זה נציין, שדי להוכיח אותו עבור המקרה בו V הינה קבוצה סופית של מספרים טבעיים. זה מתבסס על התרגיל הפשוט הבא, שאת הוכחתו אנו משאירים לקורא:

תרגיל:

אם $|V_1| = |V_2|$, אז $|T(V_1)| = |T(V_2)|$.

המשפט שנוכיח הוא לכן:

משפט קיילי:

$$\forall V \in P_{\text{fin}}(\mathbf{N}).\ |V| \geq 2 \;\Rightarrow\; |T(V)| = |V|^{|V|-2}$$

הדרך הטבעית ביותר להוכחת נוסחה כמו נוסחת קיילי, היא למצוא קבוצה S הקשורה ב- V, עליה ידוע כבר ש- $|S| = |V|^{|V|-2}$, ולהראות ש- S ו- $T(V)$ הינן שוות-עוצמה. החלק הראשון בתכנית זו הינו קל לביצוע. מהעקרונות הבסיסיים, שלמדנו בפרקים קודמים, אנו יודעים, ש- $|V|^{|V|-2}$ היא העוצמה של $V^{|V|-2}$ (דהיינו: עוצמת קבוצת הרשימות מאורך $|V| - 2$ של איברי V, שזה מספר "האפשרויות לבחור $|V| - 2$ עצמים מתוך V עם חזרות ועם חשיבות לסדר"). אשר לחלק השני – הוא קשה יותר. העקרונות הקומבינטוריים שלמדנו אינם מסייעים כאן. אנו נאלץ לכן לנקוט בדרך הישירה: בניית פונקציית שקילות בין $T(V)$ ל- $V^{|V|-2}$. מה שאנו מחפשים, אם כך, היא דרך להתאים לכל $E \in T(V)$ רשימה באורך $|V| - 2$ של איברי V, כך שההתאמה תגדיר פונקציה הפיכה. מעשית, אנו מחפשים אפוא שיטה, כיצד ***לקודד*** כל $E \in T(V)$ על-ידי רשימה כנ"ל, כך שנוכל לשחזר כל E כזה מהקוד עבורו. הקוד עבור E צריך לכן לכלול בתוכו את כל האינפורמציה הדרושה לשם כך. זה מציג בעיה מסוימת: מתוך התבוננות ברשימה באורך $|V| - 2$ של איברי V ובה בלבד, לא נוכל לדעת אפילו מי הם כל איברי V: יהיו לפחות שניים, שאותם לא תהיה לנו כל דרך לנחש. כדי להתגבר על בעיה זו יצטרך תהליך השחזור לקבל ***שני*** נתונים: קבוצת הקדקודים V והקוד s של E (זוג כזה יהיה תקין, אם כל מספר המופיע ברשימה s שייך לקבוצה V).

הבה נסכם. מטרתנו היא, בהינתן קבוצה סופית V של מספרים טבעיים, למצוא פונקציות $H_V : V^{|V|-2} \to T(V)$ ו- $F_V : T(V) \to V^{|V|-2}$, שתהיינה הפוכות זו לזו. את זאת נשיג על-ידי שנבנה פונקציות כלליות F ו-H, כך ש- F מספקת ***לכל*** V כנ"ל ולכל E, כך ש- $E \in T(V)$, קוד מתאים ב- $V^{|V|-2}$, בעוד H מתאימה לכל V ו- s, כך ש- $s \in V^{|V|-2}$, קבוצה E, כך ש- $<V, E>$ עץ. הפעלת פונקציית קורי Cu על F תיתן פונקציה, שלכל V תספק F_V מתאים, בעוד הפעלת Cu על H תיתן פונקציה, שלכל V תספק H_V מתאים.

נתחיל בקביעת התחומים של F ו- H. לשם כך נשתמש בסימונים הבאים:

$$\mathbf{E} = P_{\text{fin}}(P_2(\mathbf{N}))$$

E היא קבוצת כל הקבוצות הסופיות של קשתות, שקדקודיהן ב- **N**.

$$\mathbf{T} = \{<V, E> \in P_{\text{fin}}(\mathbf{N}) \times \mathbf{E} \mid |V| \geq 2 \wedge E \in T(V)\}$$

T היא קבוצת כל העצים הלא-טריביאליים הסופיים, שקדקודיהם שייכים ל- **N**.

$$\mathbf{S} = \bigcup_{n=0}^{\infty} \mathbf{N}^n$$

S היא קבוצת כל הרשימות הסופיות של מספרים טבעיים (כאן $\mathbf{N}^0 = \{\varnothing\}$).

$$\mathbf{B} = \{<V, s> \in P_{\text{fin}}(\mathbf{N}) \times \mathbf{S} \mid |V| \geq 2 \wedge s \in V^{|V|-2}\}$$

B הינה קבוצת כל הקודים המלאים של איברי **T**.

אנו עומדים אפוא להגדיר פונקציות $F : \mathbf{T} \to \mathbf{S}$ ו- $H : \mathbf{B} \to \mathbf{E}$. ההגדרה של F ו- H תהיה ***רקורסיבית*** במובן הבא: בעוד הערך של F (למשל) עבור העצים הלא-טריביאליים הפשוטים ביותר יינתן ישירות, הרי כדי לחשב את $F(V, E)$ עבור עץ מורכב יותר נצטרך לחשב תחילה את ערך F עבור עץ עם מספר קשתות קטן יותר. העיקרון פה הוא אותו עיקרון שהפעלנו, בעת שהגדרנו פונקציות בעזרת נוסחאות נסיגה. רעיון ההגדרה ברקורסיה הוא אכן רעיון כללי, שמשתמשים בו רבות בתכנות (ובמיוחד בשפות תכנות פונקציונליות, דוגמת SCHEME).

לצורך ההגדרה המדויקת של F ו- H, נשתמש במספר סימונים מקובלים, הקשורים בטיפול ברשימות: אם $s = <a_1, \ldots, a_n>$ היא רשימה לא ריקה, אז $\text{hd}(s)$ יסמן את a_1 ("ראש" הרשימה s), **בעוד** $\text{tail}(s)$ יסמן את שאר הרשימה, דהיינו: הרשימה $<a_2, \ldots, a_n>$ (רשימה זו היא ריקה כאשר $n = 1$). כמו-כן, אם $k \in \mathbf{N}$, אז $k\hat{\ }s$ יסמן את הרשימה $<k, a_1, \ldots, a_n>$.[1,2]

הגדרה 1:

הפונקציה $F : \mathbf{T} \to \mathbf{S}$ מוגדרת באופן הבא:

$$F(V, E) = \begin{cases} \varnothing & |E| = 1 \\ n_1 \hat{\ } F(V - \{n_0\}, E - \{a\}) & |E| > 1 \end{cases}$$

כאשר: n_0 הוא האיבר הקטן ביותר ב- V, שהוא עלה של $<V, E>$.
a היא הקשת ביחידה ב- E, ש- n_0 נמצא עליה.
n_1 הוא הקצה השני של a (זה השונה מ- n_0).

---

1 בשפת התכנות LISP כותבים $\text{CAR}(s)$ במקום $\text{hd}(s)$, $\text{CDR}(s)$ במקום $\text{tail}(s)$, ו- $\text{cons}(k, s)$ במקום $k\hat{\ }s$.

2 נשים לב, שאם $<a_1, \ldots, a_n>$ מוגדר (כמו בפרק ב.3) בתור $<a_1, <\ldots <a_{n-1}, a_n> \ldots>$, אז $\text{hd}(s) = \pi_1(s)$, $\text{tail}(s) = \pi_2(s)$ ו- $n\hat{\ }s = <n, s>$.

הגדרה רקורסיבית מסוג זה של F יש להצדיק, כלומר: עלינו להראות ש- $F(V, E)$ אכן מוגדר לכל $<V, E> \in \mathbf{T}$, ו- $F(V, E)$ אכן איבר של **S**. למעשה, נוכיח יותר:

למה 1:

הפונקציה F מוגדרת היטב לכל $<V, E> \in \mathbf{T}$, ו- $s = F(V, E)$ היא רשימה באורך $|E| - 1$ (שזה $|V| - 2$, כי $<V, E>$ עץ) של איברים של V, שאף אחד מהם אינו עלה של $<V, E>$. יתר על כן: כל איבר של V, שאינו עלה, מופיע ברשימה s.

הוכחת למה 1:

באינדוקציה על $|E|$.

אם $|E| = 1$ אז $|V| = 2$, ושני קדקודי V הינם עלים. $F(V, E)$ הינו $\varnothing$, שהיא אכן רשימה ריקה באורך $|E| - 1 = 0$. מובן, שאף עלה אינו מופיע ברשימה זו, אך כל איבר של V, שאינו עלה, כן מופיע בה (פשוט משום שאין איבר כזה!).

נניח נכונות הטענה כאשר $|E| = n \ge 1$. נניח ש- $|E| = n + 1$. אז $|V| = n + 2 \ge 3$. כיוון ש- $<V, E>$ הינו עץ, קבוצת העלים של עץ זה אינה ריקה. זוהי קבוצה של מספרים טבעיים, ולכן יש בה איבר קטן ביותר. מכאן, ש- n_0 קיים, ולכן גם a ו- n_1 קיימים. יתר על כן: n_1 אינו עלה, כי שני עלים בעץ יכולים להיות מחוברים על-ידי קשת רק אם מספר קדקודי העץ הוא 2 (הוכח/י!), ואילו $|V| \ge 3$. עתה כל קשת ב- $E - \{a\}$ מחברת שני קדקודים של $V - \{n_0\}$ (כי a הינה הקשת היחידה, ש- n_0 נמצא עליה). לכן $G' = <V - \{n_0\}, E - \{a\}>$ הינו תת-גרף של $<V, E>$. ברור ש- G' חסר מעגלים, כי $<V, E>$ הינו חסר מעגלים. כמו-כן, $|E - \{a\}| = |V - \{n_0\}| - 1$ (כי $|E| = |V| - 1$). מכאן ש- G' הינו עץ (לפי משפט (5)(ו) בפרק הקודם). מספר הקשתות של עץ זה הוא $(n + 1) - 1 = n$. ניתן לכן להפעיל לגביו את הנחת האינדוקציה, ולקבל ש- $F(V - \{n_0\}, E - \{a\})$ היא רשימה באורך $n - 1$ $(|E - \{a\}| - 1)$ של איברי $V - \{n_0\}$, שאף אחד מהם אינו עלה ב- G', וממילא אף אחד מהם אינו עלה של $<V, E>$. כיוון שגם n_1 אינו עלה של $<V, E>$, $F(V, E)$ הוא רשימה באורך n (דהיינו: $|E| - 1$) של איברים ב- V, שאף אחד מהם אינו עלה של $<V, E>$. נראה עוד, שכל איברי V, שאינם עלים של $<V, E>$ נמצאים ברשימה זו. יהי אפוא w איבר של V, שאינו עלה של $<V, E>$. אז $w \ne n_0$. אם $w = n_1$, אז הוא מופיע ב- $F(V, E)$ (ואפילו בראש הרשימה). אם $w \ne n_1$, אז w לא נמצא על a. לכן דרגתו ב- $<V, E>$ זהה לדרגתו ב- G'. מכאן ש- w גם אינו עלה של G', ולכן, לפי הנחת האינדוקציה, w נמצא על $F(V - \{n_0\}, E - \{a\})$, וממילא w נמצא גם על $F(V, E)$.

מסקנה 1:

לכל $<V, E> \in \mathbf{T}$, $F(V, E) \in V^{|V|-2}$, ו- $<V, F(V, E)> \in \mathbf{B}$.

דוגמה 1:

יהי $<V, E>$ העץ מציור 8 בפרק הקודם. אז $V = \{1, 2, \ldots, 10\}$. כדי לקבל את $F(V, E)$ עלינו למצוא את n_0. זהו האיבר הקטן ביותר בקבוצת העלים של $<V, E>$. קבוצה זו הינה $\{2, 5, 7, 8, 9, 10\}$. לכן $n_0 = 2$. מכאן ש- $a = \{2, 4\}$ ו- $n_1 = 4$. $F(V, E)$ היא לכן רשימה באורך 8, המתחילה ב- 4. כדי למצוא את שאר הרשימה עלינו למצוא את $F(V - \{2\}, E - \{\{2, 4\}\})$, דהיינו: את הקוד של העץ המתקבל מ- $<V, E>$ על-ידי מחיקת הקדקוד 2 והקשת, המחברת אותו אל הקדקוד 4. אם נמשיך את התהליך נקבל ש:

$$F(V, E) = <4, 6, 3, 4, 4, 1, 1, 4>$$

הקדקודים המופיעים ב- $F(V, E)$ הם 1, 3, 4, 6. אלה הינם אכן הקדקודים של V, שאינם עלים של $<V, E>$.

הגדרה 2:

הפונקציה $H : \mathbf{B} \to \mathbf{E}$ מוגדרת ברקורסיה כך:

$$H(V, s) = \begin{cases} \{V\} & s = \varnothing \\ \{\{k_0, \mathrm{hd}(s)\}\} \cup H(V - \{k_0\}, \mathrm{tail}(s)) & s \neq \varnothing \end{cases}$$

כאשר k_0 הוא האיבר הקטן ביותר של V, שאינו מופיע ב- s.

למה 2:

הפונקציה H מוגדרת היטב לכל $<V, s> \in \mathbf{E}$, ו- $H(V, s) \in T(V)$. יתר על כן: $<V, H(V, s)>$ הינו עץ, שהעלים שלו הם בדיוק אותם איברים של V, שאינם מופיעים ברשימה s (בפרט: $<V, H(V, s)> \in \mathbf{T}$).

הוכחת למה 2:

באינדוקציה על $\ell(s)$ – אורך הרשימה s.

אם $\ell(s) = 0$ אז $s = \varnothing$. כיוון ש- $<V, s> \in B$, פירוש הדבר ש- $|V| = 2$. נניח למשל, ש- $V = \{n_1, n_2\}$. אז $<V, H(V, s)> = <\{n_1, n_2\}, \{\{n_1, n_2\}\}>$, וזהו עץ עם שני קדקודים וקשת אחת ($\{n_1, n_2\}$), המחברת אותם. שני הקדקודים הם עלים במקרה זה, ולכן הטענה מתקיימת בו.

נניח נכונות הטענה כאשר $\ell(s) = n$, ונניח $\ell(s) = n + 1$. אז $|V| = n + 3 > 2$. עתה k_0 קיים, כיוון ש- $\ell(s) < |V|$. כיוון ש- k_0 אינו מופיע ב- S, $k_0 \neq \mathrm{hd}(s)$, וכמו-כן $\mathrm{tail}(s)$ הוא רשימה באורך n של איברים מ- $V - \{k_0\}$, דהיינו: $\mathrm{tail}(s) \in (V - \{k_0\})^{|V - \{k_0\}|-2}$. מכאן ש- $M = \,<V - \{k_0\}, \mathrm{tail}(s)> \,\in \mathbf{B}$. כיוון ש- $\ell(\mathrm{tail}(s)) = n$, ניתן להפעיל את הנחת האינדוקציה לגבי M, ולקבל ש- $H(M)$ מוגדר היטב, ו- $G' = \,<V - \{k_0\}, H(M)>$ הינו עץ, שהעלים שלו הם בדיוק האיברים ב- $V - \{k_0\}$, שאינם מופיעים ב- $\mathrm{tail}(s)$. עתה $<V, H(V, s)>$ מתקבל מהעץ G' על-ידי הוספת קדקוד חדש (k_0) וקשת חדשה ($\{k_0, \mathrm{hd}(s)\}$), שמחברת קדקוד זה לאחד הקדקודים של G' (הלא הוא $\mathrm{hd}(s)$, שאכן שייך ל- $V - \{k_0\}$). ממשפט (5) (ג) בפרק הקודם נובע לכן בקלות, ש- $<V, H(V, s)>$ הינו עץ. העלים של עץ זה הם k_0, ואותם עלים של G', השונים מ- $\mathrm{hd}(s)$ ($\mathrm{hd}(s)$ אינו עלה של $<V, H(V, S)>$, כיוון שדרגתו בעץ זה היא לפחות 2). מזה ומההנחת האינדוקציה לגבי $\mathrm{tail}(s)$ נובע, שהעלים של $<V, H(V, s)>$ הם בדיוק איברי V, שאינם נמצאים על s (זאת כיוון ש- k_0 לא נמצא על s, בעוד $\mathrm{hd}(s)$ נמצא ב- s).

דוגמה 2:

יהי $V = \{1, \ldots, 10\}$, ו- s הרשימה שנבנתה על-ידי F בדוגמה 1 (כלומר: $s = \,<4, 6, 3, 4, 4, 1, 1, 4>$). כדי למצוא את $H(V, s)$ עלינו למצוא תחילה את k_0. זהו האיבר הראשון של V, שאינו מופיע ב- s, ומכאן ש- $k_0 = 2$. $\mathrm{hd}(s)$ היא כאן 4. מהגדרת H, כדי למצוא את $H(V, s)$ עלינו למצוא תחילה את

$$H(\{1, 3, 4, \ldots, 10\}\,,\ <6, 3, 4, 4, 1, 1, 4>)$$

ולהוסיף לקבוצת קשתות זו על V את הקשת $\{2, 4\}$. אם נבצע זאת, נקבל ש- $<V, H(V, s)>$ הוא בדיוק הגרף בציור 8 בפרק הקודם.

הלמה הבאה מראה, שמה שקיבלנו בדוגמה 2 אינו מקרי:

למה 3:

(א) אם $<V, E> \,\in \mathbf{T}$ אז $H(V, F(V, E)) = E$;

(ב) אם $<V, s> \,\in \mathbf{B}$ אז $F(V, H(V, s)) = s$.

הוכחת למה 3:

(א) באינדוקציה על $|E|$.

אם $|E| = 1$, אז $F(V, E) = \varnothing$. לכן $H(V, F(V, E)) = H(V, \varnothing) = \{V\}$. אבל כאשר $|E| = 1$, אז $|V| = 2$ ו- $E = \{V\}$. לכן $H(V, F(V, E)) = E$ במקרה זה.

נניח נכונות הטענה עבור $|E| = n$, ונניח $|E| = n + 1$. אז לפי הגדרה 1, $F(V, E) = n_1{}^\wedge F(V - \{n_0\}, E - \{a\})$ כש- n_0, n_1 ו- a הם כמו בהגדרה הנ"ל. לכן $\text{hd}(F(V, E)) = n_1$, ו- $\text{tail}(F(V, E)) = F(V - \{n_0\}, E - \{a\})$. כדי למצוא את $H(V, F(V, E))$ עלינו למצוא (לפי הגדרה 2) את k_0 – האיבר הראשון של V, שאינו מופיע ב- $F(V, E)$. לפי למה 1 זהו האיבר הקטן ביותר בקבוצת העלים של $<V, E>$. לכן $k_0 = n_0$. מכל זה נובע, לפי הגדרה 2, ש:

$$H(V, F(V, E)) = \{\{n_0, n_1\}\} \cup H(V - \{n_0\}, F(V - \{n_0\}, E - \{a\}))$$

אבל לפי הנחת האינדוקציה $H(V - \{n_0\}, F(V - \{n_0\}, E - \{a\})) = E - \{a\}$. כמו-כן, $a = \{n_0, n_1\}$. לכן:

$$H(V, F(V, E)) = \{a\} \cup (E - \{a\}) = E$$

(ב) באינדוקציה על $\ell(s)$.

אם $\ell(s) = 0$, אז $s = \varnothing$ ו- $H(V, s) = \{V\}$. לכן $F(V, H(V, s)) = F(V, \{V\}) = \varnothing$ (כי $|\{V\}| = 1$), דהיינו $F(V, H(V, s)) = s$ במקרה זה.

נניח נכונות הטענה כאשר $\ell(s) = n$, ונניח $\ell(s) = n + 1$. אז לפי הגדרה 2, $H(V, s) = \{k_0, \text{hd}(s)\} \cup H(V - \{k_0\}, \text{tail}(s))$, כאשר k_0 הינו האיבר הקטן ביותר של V, שאינו מופיע ב- s. כדי למצוא את $F(V, H(V, s))$ עלינו למצוא (לפי הגדרה 1) את n_0, a ו- n_1. n_0 הינו האיבר הקטן ביותר של V, המהווה עלה של $<V, H(V, s)>$. לפי למה 2, זהו בדיוק k_0. מהנוסחה עבור $H(V, s)$ למעלה נובע לכן, שבמקרה זה $a = \{k_0, \text{hd}(s)\}$, ולכן $n_1 = \text{hd}(s)$ ו- $H(V, s) - \{a\} = H(V - \{k_0\}, \text{tail}(s))$.

לכן, מהגדרה 1:

$$F(V, H(V, s)) = \text{hd}(s)^\wedge F(V - \{k_0\}, H(V - \{k_0\}, \text{tail}(s)))$$

אבל מהנחת האינדוקציה לגבי $\text{tail}(s)$ (שהיא רשימה באורך n) נובע:

$$F(V - \{k_0\}, H(V - \{k_0\}, \text{tail}(s))) = \text{tail}(s)$$

לכן:

$$F(V, H(V, s)) = \text{hd}(s)^\wedge\text{tail}(s) = s$$

מ.ש.ל.

אנו יכולים לפנות סוף-סוף אל

הוכחת משפט קיילי:

תהי V קבוצה חלקית סופית של **N** כך ש- $|V| \geq 2$.

נגדיר:

$$F_V = (Cu(F))(V) = \lambda E \in T(V).\ F(V, E)$$

$$H_V = (Cu(H))(V) = \lambda s \in V^{|V|-2}.\ H(V, s)$$

ממסקנה 1 ומלמה 2 ברור, ש- $F_V : T(V) \to V^{|V|-2}$ ו- $H_V : V^{|V|-2} \to T(V)$. נראה שהן הפוכות זו לזו. ואכן, לפי למה 3, אם $s \in V^{|V|-2}$, אז

$$(F_V \circ H_V)(s) = F_V(H_V(s)) \overset{\beta}{=} F_V(H(V, s)) \overset{\beta}{=} F(V, H(V, s)) = s$$

ולפי אותה למה, אם $E \in T(V)$, אז

$$(H_V \circ F_V)(E) = H_V(F_V(E)) = H_V(F(V, E)) = H(V, F(V, E)) = E$$

מכאן ש- $F_V \circ H_V = i_{V^{|V|-2}}$ ו- $H_V \circ F_V = i_{T(V)}$.

קיבלנו ש- F_V היא פונקציית שקילות מ- $T(V)$ על $V^{|V|-2}$. לכן

$$|T(V)| = |V^{|V|-2}| = |V|^{|V|-2}$$

מ.ש.ל.

חשוב לציין, שההוכחה של משפט קיילי מספקת אינפורמציה נוספת על עצים סופיים מעבר לנוסחה עצמה. אינפורמציה כזו מתבטאת למשל בלמות 1 ו-2 למעלה. הלמות יוצרות קשר בין עצים על קבוצה סופית V ובין רשימות באורך $|V| - 2$ של איברי V, כך שבקשר זה התכונה "להיות עלה של עץ $\langle V, E \rangle$" מוחלפת בתכונה "להיות איבר של V שאינו מופיע ברשימה s". זה מסייע לפתור בעיות קומבינטוריות שונות הקשורות בעצים. נביא עתה מספר דוגמאות.

בעיה 1:

תהי V קבוצה סופית כך ש- $|V| = n \geq 2$, ונניח $v_1, v_2 \in V$, ו- $v_1 \neq v_2$. בכמה עצים על V יהיו v_1 ו- v_2 מחוברים על-ידי קשת?

נביא עתה שני פתרונות שונים לבעיה זו. האחד מתבסס רק על נוסחת קיילי (יחד עם שיקולים קומבינטוריים נוספים). השני מתבסס גם על האינפורמציה שנותנת הוכחת

משפט קיילי (כמעט ללא שיקולים קומבינטוריים נוספים). עבור שני הפתרונות נוכל להניח, ללא הגבלת הכלליות, ש- $v_1 = 1$, $v_2 = 2$ ו- $V = \{1, 2, \ldots, n\}$, כך ש- $n \geq 2$.

פתרון א:

נסמן את המספר המבוקש ב- x. נגדיר עתה:

$$U = \{<E, e> \in T(V) \times P_2(V) \mid e \in E\}$$

נחשב את $|U|$ בשתי צורות:

מצד אחד:

$$U = \biguplus_{E \in T(V)} \{<E, e> \mid e \in E\} \qquad (1)$$

כיוון שלכל $E \in T(V)$ מתקיים ש- $|\{<E, e> \mid e \in E\}| = |E| = n - 1$, הרי נובע מ- (1) ש- $|U| = |T(V)| \cdot (n - 1)$, ולכן, לפי נוסחת קיילי: $|U| = n^{n-2}(n - 1)$.

מצד שני:

$$U = \biguplus_{e \in P_2(V)} \{<E, e> \mid E \in T(V) \wedge e \in E\} \qquad (2)$$

עתה ברור, שלכל e: $|\{<E, e> \mid E \in T(V) \wedge e \in E\}| = |\{E \in T(V) | e \in E\}| = x$.

לכן, מ- (2) נובע ש- $|U| = |P_2(V)| \cdot x = \binom{n}{2} x$.

משתי הנוסחאות שמצאנו עבור $|U|$ אנו מקבלים:

$$\binom{n}{2} \cdot x = n^{n-2} \cdot (n - 1)$$

ולכן

$$x = 2 \cdot n^{n-3}$$

פתרון ב:

הבה נבדוק, כיצד נראות רשימות שמצפינות עצים על $\{1, 2, \ldots, n\}$, שבהם $\{1, 2\}$ היא קשת. נעשה זאת על-ידי בחינת הפרוצדורה לבניית רשימות אלה, כפי שניתנה בהוכחת המשפט.

נניח ראשית ש- $n \geq 3$. האבחנה הראשונה שלנו היא, שלא ייתכן ש- $\{1, 2\}$ היא הקשת, שנותרת אחרונה בתהליך בניית הרשימה. הסיבה: אילו היה זה המצב, אז בשלב שלפני האחרון (שקיים, כי $n \geq 3$) 1 או 2 היה עלה, ולכן היינו בוחרים אותו יחד עם הקשת $\{1, 2\}$ כבר אז. מכאן ש- $\{1, 2\}$ נבחרת בשלב כלשהו, וייתכנו שני מקרים:

(I) הקשת $\{1,2\}$ נבחרת מיד בהתחלה. זה קורה אם 1 או 2 הוא עלה של העץ. לא ייתכן ששניהם עלים, כי יש בעץ קשת ביניהם ו- $n \geq 3$. מכאן שמקרה זה מתחלק לשני תת-מקרים זרים:

(I.1) 1 עלה של הגרף. במקרה זה הרשימה המצפינה תתחיל ב- 2, ו- 1 אינו מופיע בה כלל;

(I.2) 2 עלה של הגרף. במקרה זה הרשימה המצפינה תתחיל ב- 1, ו- 2 אינו מופיע בה כלל.

(II) הקשת $\{1,2\}$ נבחרת, אבל לא בהתחלה. גם מקרה זה מתחלק באופן טבעי לשני תתי-מקרים:

(II.1) בשלב הבחירה של $\{1,2\}$ 1 הינו עלה ו- 2 נרשם. אחרי 2 זה לא יופיע יותר שום 1. מצד שני, עד אותו שלב 1 לא היה עלה (אחרת היה נבחר עוד קודם). מכאן, שבדיוק לפני שלב זה נעשתה בחירה, שהפכה את 1 לעלה. כלומר, בדיוק לפניו נבחרה קשת המכילה את 1, הקצה השני שלה היה עלה, ו- 1 הוא זה שנרשם. היוצא מכך הוא, שבמקרה זה יש 1 ברשימה, ואחרי ה- 1 האחרון מופיע 2;

(II.2) בשלב הבחירה 2 הינו עלה ו- 1 נרשם. לפי שיקול דומה לזה של (II.1), לפני 1 זה מופיע 2, וזהו ה- 2 האחרון ברשימה.

קיבלנו שרשימות המצפינות עצים שבהם $\{1,2\}$ היא קשת, מתחלקות לארבעה סוגים, שהם בבירור זרים זה לזה:

(א) רשימות המתחילות ב- 1, ו- 2 אינו מופיע בהן.

(ב) רשימות המתחילות ב- 2, ו- 1 אינו מופיע בהן.

(ג) רשימות שגם 1 וגם 2 מופיעים בהן, ואחרי ה- 2 האחרון מופיע 1.

(ד) רשימות שגם 2 וגם 1 מופיעים בהן, ואחרי ה- 1 האחרון מופיע 2.

נבנה עתה פונקציה מקבוצת הרשימות הנ"ל אל קבוצת הרשימות באורך $n-3$ של איברי V באופן הבא:

במקרה (א) נשמיט את ה- 1 בהתחלה.

במקרה (ב) נשמיט את ה- 2 בהתחלה.

במקרה (ג) נשמיט את ה- 1, שמופיע אחרי ה- 2 האחרון.

במקרה (ד) נשמיט את ה- 2, שמופיע אחרי ה- 1 האחרון.

עתה, לכל רשימה ב- V^{n-3} יש בדיוק שני מקורות לפי פונקציה זו. האחד מתקבל על-ידי הוספת 1 בהתחלה (אם אין ברשימה 2) או אחרי ה- 2 האחרון ברשימה (אם יש

ברשימה 2). המקור השני מתקבל על-ידי הוספת 2 באופן דומה. מכאן, שמספר הרשימות מהסוג המבוקש הוא $2 \cdot n^{n-3}$, ולכן זה גם מספר העצים מהסוג המבוקש. השיקול עד כה היה בהנחה ש- $n \geq 3$. במקרה ש- $n = 2$, רואים ישירות, שמספר העצים בהם מדובר הוא 1. לכן הנוסחה $2 \cdot n^{n-3}$ נכונה גם במקרה זה.

בעיה 2:

בכמה עצים על $V = \{1, 2, \ldots, 9\}$ הקדקודים 6, 7, 8, 9 הם עלים?

פתרון:

עצים כאלה מוצפנים על-ידי רשימות באורך 7 של איברי V, שהמספרים 6, 7, 8, 9 אינם מופיעים בהם. כלומר: מדובר ברשימות באורך 7 של איברי $\{1, 2, 3, 4, 5\}$. מספר הרשימות הללו הוא 5^7.

בעיה 3:

בכמה עצים על $V = \{1, 2, \ldots, 9\}$ קבוצת העלים היא בדיוק $\{6, 7, 8, 9\}$?

פתרון:

ההבדל בין בעיה זו והבעיה הקודמת היא, שבבעיה הקודמת יכלו להיות עלים נוספים מלבד 6, 7, 8 ו- 9. הפעם הקדקודים האחרים אינם עלים, ולכן הם חייבים להופיע ברשימות המצפינות את העצים. הבעיה הופכת לכן להיות: כמה רשימות באורך 7 אפשר להרכיב מ- $\{1, 2, 3, 4, 5\}$, אם כל איברי קבוצה זו חייבים להופיע. בעיות מסוג זה פתרנו כבר בעבר, אם בעזרת עיקרון ההכלה וההפרדה, או בעזרת פונקציות יוצרות מעריכית (ראו דוגמה 2 בפרק ד.2, ואותה דוגמה מחדש בפרק ד.4). התשובה היא:

$$\sum_{k=0}^{5} (-1)^k \binom{5}{k} (5-k)^7$$

בעיה 4:

בכמה עצים על $V = \{1, 2, \ldots, 9\}$ יש בדיוק 4 עלים?

פתרון:
בבעיה הקודמת לא הסתמכנו בעצם כלל על הזהות של העלים $\{6, 7, 8, 9\}$. כל בחירה אחרת של ארבעת העלים היתה נותנת אותה תשובה בדיוק. כיוון שניתן לבחור את ארבעת העלים ב- $\binom{9}{4}$ צורות, התשובה (לאור הבעיה הקודמת) היא

$$\binom{9}{4}\sum_{k=0}^{5}(-1)^k\binom{5}{k}(5-k)^7$$

בעיה 5:
בכמה עצים על $V = \{1, 2, \ldots, 9\}$ יש בדיוק שני עלים?

פתרון:
ניתן לפתור בעיה זו כמו שפתרנו את הבעיה הקודמת. ברם, במקרה זה,אחרי שבחרנו שני עלים, נשאר להרכיב רשימה באורך 7 משבעת העלים הנותרים,כך שכל אחד מהם יופיע. זה אפשר לעשות, כמובן, ב- $7!$ צורות. לכן התשובה כאן היא:

$$\binom{9}{2}\cdot 7! = \frac{9!}{2}$$

לסיום נציין, שבדיקה לעומק של הוכחת משפט קיילי מראה, שאת למה 1 ניתן להכליל למשפט הבא (שאת הוכחתו נשאיר לקוראים):

משפט
אם $<V, E>$ עץ ו- $a \in V$, אז מספר הפעמים ש- a מופיע ברשימה $F(V, E)$ (המקודדת עץ זה) הוא $d(a) - 1$ (כאשר $d(a)$ היא, כזכור, הדרגה של a ב- $<V, E>$).

בעיה 6:
בכמה עצים על $V = \{1, 2, \ldots, 9\}$ דרגת 1 היא 2 ?

פתרון:
לאור המשפט האחרון, השאלה שקולה לבעיה: כמה רשימות, יש באורך 7 של מספרים מ- $\{1, 2, \ldots, 9\}$, בהן 1 מופיע פעם אחת בדיוק? התשובה, כמובן, היא: $7 \cdot 8^6$.