

פרק ששי

"מופעים נדירים"

המחקר במקורו נערך בספר בראשית, שהוא הספר הראשון מחמישה חלקי התורה. בספר בראשית מצויות 78,064 אותיות. הבה נבחן את אוכלוסיית המד"שים של מלה קצרה, בת שלוש אותיות שכיחות, כמו "איש". מעטים ינחשו, כי בספר בראשית – או בכל טקסט בעל אורך דומה – ימצאו לפחות מיליון מד"שים כאלה! אפשר לוודא זאת בחישוב מתימטי. כאשר נבדוק במחשב טקסט בעל תכולת אותיות כזו, אכן נמצא מספר דומה של מד"שים – ולא רק כאשר נציב את המלה "איש", אלא גם כשתהא זו המלה "אשה" או "משה", או כל מלה אחרת בעל שלוש אותיות מן השכיחות בטקסט. אכן, כמות עצומה של מד"שים! אבל, בסריקה אחר המד"שים למלה בת ארבע אותיות, למשל "האיש", הם ימצאו בכמויות פחותות: "רק" כמה עשרות אלפי פעמים. למלה "אנשים", שהיא בת חמש אותיות, ימצאו "רק" כמה אלפי מד"שים. את המלה "האנשים" בת שש האותיות יצגו רק כמה מאות מד"שים. ביטוי כמו "של אנשים", שהוא בן שבע אותיות ייצג רק על ידי כעשרה מד"שים, בעוד שלביטוי "של האנשים" (שמונה אותיות) צפוי להימצא אולי מד"ש בודד, אם ימצא בכלל. זהו ציור אופייני למצב המובן שבו אוכלוסיית המד"שים של ביטויים ארוכים יותר – קטנה יותר. בדוגמא דלעיל השתמשתי בביטויים הבנויים מאותיות שכיחות בטקסט הנבדק. מד"שים המייצגים ביטוי המורכב מאותיות פחות שכיחות או אפילו נדירות, יהיו נדירים יחסית. דוגמא קיצונית: כאשר מלה בת ארבע אותיות תורכב כולה מאותיות נדירות בטקסט, ירד מספר המד"שים שלה מעשרות אלפים לכמה בודדים!

מצב זה אינו מאפשר לבצע את המשימה העומדת לפנינו: לעקוב אחר שתי התכונות שהוגדרו בסוף הפרק הקודם. הכמויות העצומות של מד"שים המייצגים "ביטוי א" ושל מד"שים המייצגים "ביטוי ב", עשויות לגרום מפגשים קרובים גם באקראי. ולכן, אפילו קיים טקסט נסתר בספר בראשית – לא נוכל להבחין בכך. יש כאן בעיה עקרונית: כיצד ניתן להיפטר מ"רעש הרקע"? – שהרי כל "סיגנאל" של תכנון מועד להיבלע ב"רעש הרקע". הווה אומר, כי רק שימוש ב"מסנן" (פילטר) מתאים יוכל למנוע את ההפרעות החמורות הנגרמות בגלל "רעש הרקע", וגם לחסוך זמן חישוב רב ביותר. קשיים דומים עומדים בפני אנשי מחקר במגוון של תחומים: בפני החוקרים המנסים לפענח את הצופן הגנטי, או בפני חוקרים רבים אחרים המנסים "לעלות" על תשדורות של יצורים נבונים משוערים מכוכבים רחוקים, וכן בפני אנשי המודיעין הצבאי המנסים לנפות תשדורות סרק כדי להגיע לתשדורות הכוללות מידע.

"הסינון" המבוקש אמור להפחית את כמות המד"שים המייצגים "ביטוי א" והמד"שים המייצגים "ביטוי ב" באופן דרסטי. לאחר הסינון, נקבל מספר קטן של מד"שים המייצגים את הביטויים הנדרשים. נקרא למד"שים אלה "מופעים נדירים".

אפשרות אחת ל"סינון", היא הגבלת גודל הדילוג. אם גודל הדילוג נקבע מראש, יש בכך הגבלה חריפה על מספר המד"שים המייצגים ביטויים בני שלוש אותיות (לפחות אחת מהן נדירה), או ארבע אותיות שכיחות ומעלה. אלה יהיו "מופעים נדירים". הדוגמאות מעבודתו של הרב וייסמנדל מבוססות כולן על קביעת גודל מסוים לדילוג ולכן עוסקות במופעים נדירים. כפי שצויין לעיל, יכולתו של הרב וייסמנדל איפשרו לו להשתמש במסנן כזה. באשר לנו – נאלצנו להגדיר ולהשתמש במסנן אחר, ללא קביעה מראש של גודל הדילוג, כפי שיבואר מיד – בפרק הבא.