

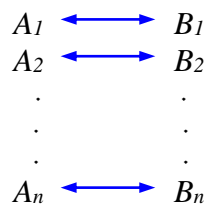
נספח א7

**ראנדומיזציות**

בנספח זה נטפל בכמה נושאים הנוגעים לראנדומיזציה. בסעיף א נגדיר את המקרה הכללי של מבחן הראנדומיזציה עבור מדגמים מן הצורה של המדגמים הגדולים – הראשון והשני. בסעיף ב נביא את הפרטים הנוגעים לראנדומיזציה שבוצעה עבור המדגם השני. בסעיף ג יוצגו הפרטים הנחוצים ליצירת הטקסטים  $U, V, W$  ששימשו כביקורת בניסויים.

**א. מבחן ראנדומיזציה למדגמים מטיפוס  $P$**

המדגמים של גדולי חכמי התורה (המוגדרים בנספחים 4 וא6), ששימשו לעריכת שני הניסויים הגדולים, ניתנים להצגה סכמתית בצורה הבאה:



כאשר  $A_i$  ו- $B_i$  הן קבוצות של ביטויים, כך שהביטויים ב- $A_i$  קשורים מושגית לאלה שבקבוצה  $B_i$ . המדגם כולל את כל זוגות הביטויים  $(w, w')$  כך ש- $w$  שייך ל- $A_i$  ו- $w'$  שייך ל- $B_i$ , לכל  $1 \leq i \leq n$ . למדגם כזה אנו קוראים מדגם מטיפוס  $P$ .

את  $n$  הקבוצות  $A_i$  אפשר להתאים ל- $n$  הקבוצות  $B_i$  ב- $n!$  אופנים שונים (הנקראים פרמוטציות או צימודים). לכל פרמוטציה  $\pi$  מ- $n!$  הפרמוטציות, אפשר להתאים מדגם, הכולל זוגות ביטויים  $(w, w')$  כך ש- $w$  שייך ל- $A_i$  ו- $w'$  שייך ל- $B_{\pi(i)}$ , לכל  $1 \leq i \leq n$ . עבור מדגם זה, מקבל הסטטיסטי  $j$  (שהוא מידת "הנטייה הכוללת לקרבה") את הערך  $P_j^\pi$ .

את  $n!$  ערכי  $P_j^\pi$  המתקבלים בפרמוטציות אלו, נסדר לפי הסדר הרגיל של המספרים הממשיים. אם התכונה שבמעקב אינה אלא אקראית, הסיכוי ש- $P_j$  (ערך הסטטיסטי  $j$  עבור המדגם המקורי) יאכלס כל אחד מ- $n!$  המקומות בסדור זה, הינו שווה. זו היא השערת האפס שלנו. כדי לחשב את רמת המובהקות הסטטיסטית, אנו מגרילים  $M$  פרמוטציות אקראיות  $\pi$  של  $n$  עצמים, בדרך שנתאר בסעיף ב. כל אחת מן הפרמוטציות  $\pi$  קובעת את הסטטיסטי  $P_j^\pi$ ; יחד עם  $P_j^\pi$  יש לנו  $M+1$  מספרים. נגדיר את הדירוג של  $P_j$  בתוך  $M+1$  המספרים הללו, כמספר ה- $P_j^\pi$  שאינם קטנים מ- $P_j$ ; אם  $P_j$  שווה ל- $P_j^\pi$  אחרים, חציים של אלה ייחשב כ"מקדים" את  $P_j$  בדירוג.

נסמן ב-  $r_j$  את הדירוג של  $P_j$ , מחולק ב-  $M+1$ . בהשערת האפס,  $r_j$  הוא ההסתברות ש- $P_j$  קרוב כל כך לראש הדירוג.

עד כאן התיאור הכללי של מבחן ראנדומיזציה עבור מדגם מטיפוס  $P$ . בספר, בפרק יט, ניתן תיאור מבחן הראנדומיזציה במקרה הפרטי של המדגם השני של גדולי חכמי התורה. בשער הרביעי, בנספח המטפל במדגמי "כותרת", יינתן התיאור הכללי של מבחן ראנדומיזציה עבור מדגם מטיפוס  $S$  (מדגם "כותרת").

ב. הראנדומיזציה בניסוי על המדגם השני של גדולי חכמי התורה  
כפי שכתבתי בגוף הספר, מבחן הראנדומיזציה הראשון בוצע לגבי המדגם השני. בחירת 999,999 הפרמוטציות האקראיות של 32 האישים הכלולים בו, נעשתה באמצעות "אלגוריתם  $P$ " של קנות' עמ' 125. המחולל הפסודו-רנדומלי ששימש כקלט לאלגוריתם זה, היה של Turbo-Pascal 5.0 המיוצר על ידי Borland Inter Inc. "הזרע הסטטיסטי" הנדרש להפעלת המחולל הוא מספר שלם בן 32 ספרות בינאריות (כלומר, מספר שלם בעל 32 ספרות כאשר הוא נכתב לפי בסיס 2). הוא נקבע על ידי פרופסור אומן בעזרת השופטים (כפי שמתואר בספר) להיות  
01001 10000 10011 11100 00101 00111 11  
(בהצגתו הרגילה, העשרונית, המספר הוא 1,277,674,143).

לפני עריכת הניסוי, דרש אחד השופטים לבצע אותו באמצעות 10 "זרעים" שונים, שכל אחד מהם ייצור סדרה של 100,000 פרמוטציות אקראיות. בכך הביע חשש מפני אפשרות של תלויות פנימיות במחולל הפסודו-רנדומלי. אולם הוחלט, שחשש זה רחוק והאפשרות הנ"ל היא בעלת סבירות נמוכה מאד. בכל זאת, לפני ביצוע הניסוי נדרשנו להסכים כי במקרה שנצליח בניסוי, והדרישה לערוך את הניסוי באמצעות 10 "זרעים" תועלה מחדש – נהיה חייבים לעשות זאת.  
ואכן, לאחר שהצלחנו בניסוי, שבוצע כפי שנקבע מראש עם סדרה של 999,999 פרמוטציות אקראיות, נדרשנו (על ידי שופט של ה- *Statistical Science*, כתב העת אליו שלחנו את המאמר לפרסום) לחזור על הניסוי ולבצעו באמצעות 10 "זרעים" כנ"ל. תוצאות הניסוי בגירסתו זו היו אף טובות יותר. התוכנה שביצעה את הראנדומיזציה המתוארת בסעיף זה הוכנה בידי יעקב רוזנברג.

### ג. יצירת טקסטים באמצעות ראנדומיזציה

יצרנו ארבעה טקסטים על ידי ערבוב אקראי.

1. טקסט  $R$ , ששימש לביקורת, נוצר מפרמוטציה אקראית יחידה של 78,064 אותיות ספר בראשית. הפרמוטציה נוצרה בתהליך שתואר בסעיף הקודם, עם "זרע סטטיסטי" שנבחר להיות המספר העשרוני 10 (כלומר, השלם הבינארי 1010). טקסט  $R$  עורבב, אם כן, ברמה של האותיות.
2. טקסט  $W$ , ששימש לביקורת, נוצר על ידי פרמוטציה של המלים בספר בראשית, שנעשתה באותה צורה ובאותו "זרע". טקסט  $W$  עורבב, אם כן, ברמה של המלים: האותיות בכל מלה נשארו כסדרן.

<sup>1</sup>D. E. Knuth, *The Art of Computer Programming*, Vol. 2, Addison-Wesley, Reading, MA. 1969.

3. טקסט V, ששימש לביקורת, נוצר מפרמוטציה אקראית יחידה של פסוקי ספר בראשית, שנעשתה באותה צורה ובאותו "זרע". טקסט V עורבב, אם כן, ברמה של הפסוקים: האותיות בכל פסוק נשארו כסדרן.

4. טקסט U, ששימש לביקורת, נוצר על ידי פרמוטציה של המלים בכל פסוק מספר בראשית, באותה צורה ובאותו "זרע", בעוד שהאותיות בכל מלה וכן הפסוקים - נשארו כסדרם. ביתר פירוט: "אלגוריתם P של קנות זקוק ל-  $n-1$  מספרים אקראיים כדי לייצר פרמוטציה אקראית של n פריטים. המחולל הפסודו-רנדומלי שהשתמשנו בו, מייצר לכל "זרע סטטיסטי", שרשרת ארוכה של מספרים אקראיים. על ידי "הזרע" 1010 יצרנו שרשרת כזאת. ששת המספרים הראשונים בשרשרת נוצלו כדי לייצר פרמוטציה אקראית של שבע המלים המרכיבות את הפסוק הראשון בספר בראשית. שלושה-עשר המספרים הבאים (במקומות 7-19 בשרשרת הנ"ל) נוצלו כדי לייצר פרמוטציה אקראית של ארבע עשרה המלים המרכיבות את הפסוק השני, וכן הלאה. הטקסטים הנ"ל נוצרו באמצעות תוכניות מחשב שהכין יעקב רוזנברג.