

מפגשים בין שמות אישי ספר בראשית לתאריכי הלידה שלהם: מדידה חדשה של המובהקות באמצעות רנדומיזציה של הטקסט

מאת דורון ויצטום

תקציר

מבחן פרמוטציות שימש למדידת המובהקות בניסוי $WRR1$ [1]. לאחר מכן, נעשה שימוש במבחן זה בניסויים המודדים מפגשים בין מד"שים לרצפי אותיות בספר בראשית, ביניהם: ניסוי $WRR2$ [2] והניסוי על תאריכי הלידה של אישי ספר בראשית [3].

כאן מוצגת המובהקות המתקבלת כאשר משתמשים במבחן טקסטים. במבחן טקסטים מדידת המובהקות נעשית באמצעות השוואה לטקסטים רבים "דומים". בכל המדידות שהן בתחום הרזולוציה של הניסוי נתקבלה מובהקות חזקה מן המובהקות שנמדדה באמצעות מבחן הפרמוטציות.

מבוא

א. בעבודת $WRR1$ [1] בוצעו בספר בראשית מדידות של רשימות הנתונים המופיעים בו כמד"שים (מד"ש=מלה בדילוג שווה של אותיות). האפקט הכולל נמדד באמצעות סטטיסטים P_i , המקבלים ערך הקטן ככל שהמפגשים בין המד"שים יותר "מוצלחים". כדי למדוד את מובהקות התוצאות היה צריך לחזור על אותן מדידות בטקסטים רבים "דומים". דבר זה לא היה בהישג ידם של WRR לפני 25 שנים (מבחינת כמות החישובים). הוצע להם, לערוך מבחן פרמוטציות, אשר פרטיו סוכמו בין פרסי דיאקוניס וישראל אומן, ואושרו לפני עריכת הניסוי על ידי ארבעה סטטיסטיקאים נודעים נוספים. מבחן זה בוצע ב- [1]. מבחן הפרמוטציות מעריך את מובהקות הסטטיסטים P_i שנתקבלו בספר בראשית עבור רשימת הנתונים המקורית (מפגשים של שמות אישים עם תאריכי הלידה ו/או הפטירה שלהם), באמצעות השוואה לערכי הסטטיסטים P'_i המתקבלים בספר בראשית עבור הרבה רשימות נתונים "מעורבבים" (מפגשים של אותם שמות אישים עם תאריכי הלידה ו/או הפטירה של אישים אחרים מן הרשימה המקורית).

ב. כחמש וחצי שנים לאחר ביצוע מבחן הפרמוטציות המתואר ב[1], סבר אליהו ריפס שהתקדמות המהירה של אמצעי החישוב מאפשרת לשוב לרעיון הפשוט ביותר של מדידת המובהקות: מדידה באמצעות השוואה לטקסטים רבים "דומים". במכתב [4] לדוד קשדן הציע אליהו ריפס את העקרונות הבאים לגבי מדידת מובהקות של רשימה בספר בראשית:

- להשתמש ב- 1000 (או יותר) טקסטים T_j , כל T_j נוצר ע"י ערבוב אקראי של מלים בתוך פסוקים (הדבר נעשה בדיוק כפי שנעשה ב[1] לגבי טקסט U).
 - להעריך את מובהקות הסטטיסטים P_i בספר בראשית, באמצעות השוואתם לערכי P'_i המתקבלים בטקסטים T_j .
- מבחן מסוג זה יכול להיות "מבחן טקסטים".

ג. השוואה בין מבחן הפרמוטציות למבחן הטקסטים:

כל אחד משני המבחנים מודד דבר שונה:

- מבחן הפרמוטציות מודד עד כמה "מוצלחים" המפגשים של שמות האישים עם תאריכי הלידה ו/או הפטירה שלהם יותר מן המפגשים של אותם שמות אישים עם תאריכים אחרים (שהם תאריכי לידה/פטירה של אישים אחרים מאותה רשימת נתונים).
- מבחן הטקסטים מודד מה הסיכוי לקבל במקרה P_i כה קטנים. דהיינו: מה הסיכוי שהצטברו במקרה כל כך הרבה מפגשים "מוצלחים".

במלים אחרות: מבחן הטקסטים מודד את מובהקות הכמות המצטברת (אקומולציה) של מפגשים "מוצלחים", ואילו מבחן הפרמוטציות מודד את המובהקות של "ספציפיות הצופן".

לכן אנו מצפים אפריורי להבדל בתוצאות שני המבחנים עבור אותה רשימת נתונים, שהרי הם מודדים דברים שונים.

ד. במחקר המדעי של הצפנים בתורה נעשה מעקב אחר שתי תכונות:

תכונה א: נטייה למפגש בין מד"שים המייצגים "ביטוי א", לבין מד"שים המייצגים "ביטוי ב", כאשר קיים קשר מושגי מובהק בין "ביטוי א" ל"ביטוי ב".

תכונה ב: נטייה למפגש בין מד"שים המייצגים "ביטוי א", לבין "ביטוי ב" הנמצא כרצף אותיות בטקסט עצמו, כאשר קיים קשר מושגי מובהק בין "ביטוי א" ל"ביטוי ב".

במאמרו "מדידת מובהקות במחקר הצפנים בתורה: תוצאות מדידות במבחן הטקסטים" [5], הבאנו תוצאות מדידות של עבודות המודדות את תכונה א, והסקנו "כי יש הבדל מהותי ודרמטי בין התוצאות של מבחן הפרמוטציות לאלה של מבחן הטקסטים".

במאמר הנוכחי ידווחו תוצאות בדיקות של עבודות המודדות את תכונה ב.

ה. המבחן המתאים למצבו העכשווי של מחקר הצפנים בתורה:

המחקר המדעי של הצפנים בתורה, עד כמה שבא לידי ביטוי במדידת רשימות גדולות של נתונים (כגון אלה שבעבודות [3]-[1]), עסק עד כה בהוכחת קיומו של צופן מסויים. הוא לא עסק באיפיון מדויק או בפיענוח של הצפנים. הדבר הוצהר במפורש, בהקדמת *WRR* [1] (בעמוד הראשון של המאמר. בעניין זה בהרחבה ראה [6]).

לפי הניתוח בסעיף ג ברור כי, לעבודות המצויינות כאן, מתאים דווקא מבחן הטקסטים, ולא מבחן פרמוטציות המודד את המובהקות של "ספציפיות הצופן". (*WRR* השתמשו במבחן פרמוטציות רק באין ברירה, כמבואר בסעיף א לעיל).

ו. מלבד השיקולים הכלליים המורים על הבדל הצפוי אפריורית בין המובהקויות המתקבלות בשתי השיטות, ייתכנו גם סיבות ספציפיות להבדל, עבור סוגים שונים של רשימות נתונים. ניתוח הסיבות הספציפיות, יעשה במאמר נפרד.

מדידת העבודה "מפגשים בין שמות אישי ספר בראשית לתאריכי הלידה שלהם"

שתי הערות מקדימות

1. בעבודה המתוארת כאן, נמדדה הקירבה בין מד"שים לבין מלים המופיעות כרצף אותיות בטקסט של ספר בראשית. יש שרצו להגדיר ביטויים המופיעים כרצף אותיות, כביטויים המופיעים "בדילוג" של ± 1 . אולם, אין הדבר כן. המושג "מלה בדילוג אותיות" והמושג "מלה כרצף אותיות" אינם מושגים טכניים בלבד. אלה מושגים תורניים מתחום הרמז. כל אחד מהם שייך לשיטת רמז שונה. הראשון לדילוג אותיות, והשני לצירוף אותיות (זו הצורה הפשוטה ביותר של צירוף אותיות) [7].

2. בכל הבדיקות בהן השתמשתי במבחן טקסטים, נעשה שימוש מירבי ב- 50,000 טקסטים (לפי יכולות החישוב העומדות לרשותי, אני זקוק למספר חודשי עבודת מחשב, על מנת לערוך השוואה ב-50,000 טקסטים, עבור רשימת נתונים מסדר גודל של אלו המדווחות כאן).

חלק א: המדגמים

א. העבודה "מפגשים בין שמות אישי ספר בראשית לתאריכי הלידה שלהם" [3] נועדה למדוד את הקירבה בין שמות אישי ספר בראשית הנמצאים בטקסט של ספר בראשית לבין תאריכי הלידה ו/או הפטירה שלהם, המופיעים כמד"שים בספר בראשית. רשימת אישי ספר בראשית להם קיימים תאריכי

לידה או פטירה מפורשים במסורת חז"ל אותרה תוך הסתייעות במאגר ממוחשב [8]. למעשה, נמצאו ארבעה עשר אישים כאלה ולהם תאריכי לידה בלבד.

ב. על בסיס רשימה זו הוכנו שני מדגמים:

1. מדגם א: הוגדרו זוגות ביטויים (שם, תאריך לידה), כאשר המלה "שם" הוא שם האיש או כינויו, והביטוי "תאריך" הוא תאריך לידתו הנתון באחת משלוש תצורות, כפי שנעשה ב[1]. לדוגמא: עבור יוסף, שנולד בא תמוז, יהיו שלושה זוגות ביטויים:

(יוסף, א תמוז), (יוסף, בא תמוז), (יוסף, א בתמוז).

לכל אדם ברשימת האישים נלקחו בחשבון כל שמותיו או צורות שמו או כינויו הנזכרים בספר בראשית. מדגם א הוא קבוצת הזוגות השונים, אשר בהם התאריך הוא בן 5-8 אותיות (זה אילוץ טכני במדידה, שנקבע כבר לפני WRI).

2. מדגם ב – מדגם הלידה: בשים לב לעובדה כי נמצאו תאריכי לידה בלבד, עלה הרעיון להצמיד לכל אדם ברשימה אך ורק את צורת השם בו נקרא בסיפור לידתו. (רעיון דומה נמצא במקורות, בהקשר לשמות בני יעקב שהיו חקוקים באבני החושן שעל כתפי הכהן הגדול [9]). לכן מדגם הלידה הוא מדגם חלקי של מדגם א.

3. לאחר שפורסמה העבודה [3], טען אחד הקוראים, כי נשמט מן הרשימה הכינוי "אברך" עבור יוסף (לפי רש"י ורשב"ם לבראשית פרק מא פסוק מג). לכן, נבדוק כאן לראשונה גם אפשרות זו. נגדיר מדגם שלישי – מדגם א1: כמו מדגם א, בתוספת הכינוי "אברך" עבור יוסף.

חלק ב: מדידת המדגמים

תהליך המדידה דומה עבור שלושת המדגמים. נשים לב, כי שלב א וכן שלב ב1 דלקמן, כבר בוצעו ב[3] עבור מדגם א ומדגם הלידה, בהבדל אחד: שם נעשו מיליון פרמוטציות. כאן נסתפק במספר פרמוטציות השווה למספר הטקסטים "הדומים" במבחן הטקסטים.

א. בשלב ראשון נמדדה קרבת המפגשים בין התאריך כמד"שים לבין השם המופיע כרצף אותיות בטקסט של ספר בראשית, עבור כל זוגות המדגם, בדיוק כפי שנעשה בעבודות קודמות של WRR [2][10]. האפקט הכולל נמדד באמצעות הסטטיסטים P_i (שהוזכרו במבוא, סעיף א) כמו בכל העבודות הקודמות של WRR.

ב. בשלב שני נערכו עבור כל אחד משני הסטטיסטים:

1. מבחן פרמוטציות בו נמדדו ערכי P_i ב-50,000 מדגמים המתקבלים מפרמוטציות אקראיות על קבוצת 14 האישים. כלומר: במקום מדגם של זוגות (שם איש פלוני, תאריך לידה של פלוני) נמדדים הזוגות מסוג (שם איש פלוני, תאריך לידה של אלמוני), כאשר פלוני ואלמוני הם אישים בקבוצת 14 האישים.

2. מבחן טקסטים בו נמדדו ערכי P_i ב-50,000 טקסטים (המעורבים באופן אקראי, כנזכר במבוא סעיף ב).

ג. התוצאות מובאות בטבלה 1:

טבלה 1

מספר המתחרים n שלהם ערך $P_i \leq P_i$ בתחרות ובה 50,000 מתחרים.

מדגם א1		מדגם הלידה		מדגם א		המדגם
P_1	P_2	P_1	P_2	P_1	P_2	סטטיסטי
0	88	0	22	3	126	מבחן טקסטים
7	102	0	75	18	238	מבחן פרמוטציות

כמו ב[3], הסטטיסטי בעל המובהקות החזקה יותר הוא P_I . נראה, כי עבור מדגם הלידה ומדגם א1 המובהקות של P_I היא מעבר לרזולוציה העומדת לרשותנו (ראה לעיל הערה מקדימה 2).

חלק ג: מדידות נוספות הקשורות למדגמים אלה.

א. במדגמים המתוארים בחלק א נלקחו תצורות התאריך בדיוק כמו בעבודת WRI [1]. נסמן את תצורות התאריך כך: A - "א' תמוז", B - "בא' תמוז", C - "א' בתמוז". ברירת המחזל היתה לקחת את התצורות ABC. כך נעשה ב[3].

בעקבות רעיון של $MBBK$ (ראה על כך ב[11]) נערוך שתי מדידות נוספות לכל מדגם:

1. כאשר נשתמש גם בתצורה רביעית: D – "בא' בתמוז". כלומר, ניקח את התצורות ABCD.
2. כאשר נשתמש רק בשתי התצורות AB.

ב. תהליך המדידה הוא בדיוק כמו בחלק ב סעיפים א-ב.

ג. התוצאות עבור שלושת המדגמים מוצגות בטבלאות 2-4. מתברר, כי בכל התוצאות הסטטיסטי בעל המובהקות החזקה (הרבה) יותר הוא P_I ולכן נתמקד בו. כדי להקל על הקורא שבנו והצגנו בתוכן גם את התוצאות מחלק ב של המדידות.

1. התוצאות למדגם א:

טבלה 2

מספר המתחרים n שלהם ערך $P'_I \leq P_I$ בתחרות ובה 50,000 מתחרים.

המדגם	ABCD	ABC	AB
ערכי P_I	5.9178E-5	1.6101E-4	3.4952E-4
מבחן טקסטים	4	3	7
מבחן פרמוטציות	46	18	56

2. התוצאות למדגם הלידה:

טבלה 3

מספר המתחרים n שלהם ערך $P'_I \leq P_I$ בתחרות ובה 50,000 מתחרים.

המדגם	ABCD	ABC	AB
ערכי P_I	3.2832E-6	1.3359E-5	1.6542E-5
מבחן טקסטים	0	0	0
מבחן פרמוטציות	0	0	9

3. התוצאות למדגם א1:

טבלה 4

מספר המתחרים n שלהם ערך $P'_I \leq P_I$ בתחרות ובה 50,000 מתחרים.

המדגם	ABCD	ABC	AB
ערכי P_I	2.3983E-5	4.4572E-5	5.5656E-5
מבחן טקסטים	1	0	1
מבחן פרמוטציות	13	7	14

מסקנות:

- א. מובהקות P_I היא שקובעת את מובהקות הניסוי בכל אחד מתשעה הניסויים (שאינם בלתי תלויים).
- ב. בכל המקרים שבדקנו כאן, שהם בתחום הרזולוציה, מובהקות P_I הנמדדת באמצעות מבחן הטקסטים חזקה מזו הנמדדת באמצעות מבחן הפרמוטציות בכדי סדר גודל אחד.

ג. עבור מובהקות המדגמים שלפנינו אין הבדל מהותי אם היינו משתמשים בתצורות ABC או $ABCD$ או AB . בכל מקרה היינו מגלים, כי קיימת קורלציה מובהקת בין אישי ספר בראשית לתאריכי הלידה שלהם.

מקורות והערות

1. עבודת WRR:

Witztum, D., Rips, E. and Rosenberg, Y. (1994). *Equidistant letter sequences in the Book of Genesis*. Statist. Sci. 9 No. 3 429-438.

Available at: http://www.torahcode.co.il/pdf_files/pub/wrr.pdf.

2. המאמר:

Witztum, D., Rips, E. and Rosenberg, Y. (1995). *Equidistant Letter Sequences in the Book of Genesis: II. The Relation to the Text*, a preprint.

3. הניסוי על תאריכי הלידה של אישי ספר בראשית:

Witztum, D. (2006). *The Hidden Birth Dates of Personalities of Genesis*, Proceedings of the 18th International Conference on Pattern Recognition, August 2006.

Available at: http://www.torahcode.co.il/pdf_files/pub/leida1.pdf.

4. בדואר אלקטרוני מיום 15 במאי שעה 16:40:13, שנת 1997 (למניינם).

5. ד. ויצטום (התשע"ז). *מדידת מובהקות במחקר הצפנים בתורה: תוצאות מדידות במבחן הטקסטים*. נמצא באתר צופן בראשית בכתובת: http://www.torahcode.co.il/pdf_files/research/texts1.pdf

6. ד. ויצטום (התשס"ד). *צופן בראשית, פרק ג. ראה גם: תיאור המחקר המדעי של הצופן בתורה*, נמצא באתר צופן בראשית בכתובת: http://www.torahcode.co.il/description_heb.htm

7. עיין בספר "פרדס רמונים" לרבי משה קורדוברו, שער ל (שער הצירוף).
8. ד. ויצטום (התשס"ד). *צופן בראשית, פרק יח עמוד 202, ובנספח 15*. נמצא גם (באנגלית) באתר צופן בראשית בכתובת: http://www.torahcode.co.il/pdf_files/pub/personaldata.pdf

9. על שמות אלה נאמר בתורה כי צריכים הם להיכתב "כתולדתם" (שמות כח, י), ומבואר בתלמוד (סוטה לו ע"ב) כי "בנימין" צריך להיכתב מלא, כשם שנכתב בספור הולדתו. וכן כתב בעל "תרומת הדשן" (סימן רלג): "פירוש 'כתולדתם' משמע שיכתבו כמו שנכתבו בתחילת תולדותם."

10. ד. ויצטום, א. ריפס, י. רוזנברג (התשנ"ו). *צופן חבוי בדילוג שווה בספר בראשית: מובהקות סטטיסטית של התופעה*. מאמר נלווה להרצאה לפני האקדמיה הלאומית הישראלית למדעים, ביום כ"ח באדר התשנ"ו (19 במרס 1996 למניינם).

11. ד. ויצטום (התשע"ו). *סיגנאל של מערכת צפני ELS: ניסויים חדשים ב"תחום המורחב"*. נמצא באתר צופן בראשית בכתובת:

http://www.torahcode.co.il/pdf_files/research/murchav1.pdf