נספח א8

תוצאות מבחן הראנדומיזציה

בנספח זה נדון בתוצאות מבחן הראנדומיזציה ביתר פירוט וביתר עומק. בסעיף א, נציג את התוצאות שנתקבלו בניסוי, ובצדן התוצאות שנתקבלו עבור טקסטים נוספים לביקורת. בסעיף ב, נסביר כיצד חושבה המובהקות.

א. תוצאות מבחן הראנדומיזציה

תיאור מבחן הראנדומיזציה נמצא בפרק יט ובנספח א7. כאן נביא את התוצאות. בניסוי הרצנו ארבע תחרויות:

- בתחרות הראשונה, התחרתה מידת "הנטייה הכוללת לקרבה" P_1 עם 999,999 המידות P_1^π של המדגמים משובשים.
- בתחרות השניה, התחרתה מידת "הנטייה הכוללת לקרבה" P_2 עם 999,999 המידות P_2^π של *אותם* מדגמים משובשים.
- בצורה דומה נעשה הדבר בשתי התחרויות הנוספות עבור P_3 ו-P_4. (הגדרתן של המידות P_1-P_4 נמצאת בנספח א3).

הדירוג של מידות אלו בתחרויות הוא המעניין אותנו, והרי הוא לפנינו:

הדירוג של P_i מתוך 1,000,000 P_i^π

הטקסט	P_1	P_2	P_3	P_4
G	453	5	570	4
R	619,140	681,451	364,859	573,861
T	748,183	363,481	580,307	277,103
I	899,830	932,868	929,840	946,261
W	883,770	516,098	900,642	630,269
U	321,071	275,741	488,949	491,116
V	211,777	519,115	410,746	591,503

ספר בראשית מסומן בטבלה ב- G, והמספרים עבורו, הניתנים באותה שורה, הם הדירוגים של ה-P_i מתוך 1,000,000 ה-P_i^π המתאימים. למשל, הדירוג 4 עבור P_4 - פירושו, שבדיוק ב-3 מתוך 999,999 פרמוטציות (צימודים) אקראיות π, היתה המידה P_4^π קטנה מ-P_4 (אף אחת לא היתה שווה לה).

בטבלה זו, שהופיעה במאמר הסופי שלנו ב- *Statistical Science*, הובאו תוצאות לגבי טקסטים אחרים. החישובים נעשו באמצעות אותן 999,999 פרמוטציות אקראיות. הטקסט הראשון ששימש לביקורת היה R, שהתקבל מ-G על ידי ערבוב אותיותיו לפי פרמוטציה אקראית (הפרטים בנספח א7). אחד השופטים הציע להשתמש לביקורת בתרגום העברי של "מלחמה ושלום" של טולסטוי. השתמשנו בטקסט T, שהוא הקטע הראשון, באורך ספר בראשית, מן התרגום העברי של "מלחמה ושלום"[1]. שופט אחר ביקש, נחזור ונבצע ניסוי ביקורת על "טקסט עברי עתיק". נוסף לכך, הוא הציע לערבב את המלים ב- G בשתי צורות: האחת בכל הספר, והשניה בכל פסוק ופסוק. על כן בדקנו גם את הטקסטים I, U, ו- W. טקסט I הוא ספר ישעיהו[2]. הטקסט W התקבל מערבוב המלים ב- G לפי פרמוטציה אקראית, והטקסט U נוצר מ- G על ידי ערבוב המלים בתוך כל פסוק לפי פרמוטציה אקראית. נוסף לכך, יצרנו את הטקסט V על ידי ערבוב הפסוקים של G בצורה אקראית. (הפרטים של יצירת U, V ו- W נמצאים בנספח א7).

די במבט בטבלה כדי להבהיר את ההבדל התהומי בין הדירוגים עבור ספר בראשית (G), לבין הדירוגים עבור שאר הטקסטים. כל מלה נוספת – מיותרת.

ב. חישוב המובהקות

1. ארבעת הדירוגים עבור ארבע המידות בספר בראשית, קובעים 4 הסתברויות (על ידי חלוקת הדירוג במספר המתחרים. ראו נספח א7): נסמן אותן ב- r_1, r_2, r_3 ו- r_4. נסמן את הערך הנמוך ביותר מבין 4 ההסתברויות ב- $\min r_i$.

נשאלת השאלה: "בהשערת האפס, מהי ההסתברות שלפחות אחד מארבעת ה- r_i קטן מן הערך $\min r_i$ או שווה לו?"

באופן כללי, בהינתן מספר $0 \le x \le 1$, הסיכוי שלפחות אחד מארבעת המספרים r_i קטן מ-x או שווה לו, הוא לכל היותר $4x$. (אי-שוויון זה ידוע כאי-השוויון של בונְפֶרוֹני[3]). ציור של המקרים הקיצוניים יבהיר את הרעיון מאחורי אי-שוויון מתמטי זה.

<u>מקרה א</u>: נניח שארבעת המספרים r_i מייצגים ארבע הסתברויות של מאורעות זרים ובלתי תלויים זה בזה. אזי הסיכוי שלפחות אחד המספרים יהיה קטן מ-x או שווה לו, הוא בדיוק $4x$: כי עמדו לרשותנו ארבעה נִסיונות, או ארבע הזדמנויות בלתי תלויות כדי להגיע לערך x. זהו המספר המכסימלי של הזדמנויות עבור ארבעה מאורעות, ולכן הסיכוי מקבל את הערך המכסימלי - $4x$.

<u>מקרה ב</u>: נניח שארבעת המספרים r_i מייצגים ארבע הסתברויות של מאורעות זהים. אזי הסיכוי שלפחות אחד המספרים יהיה קטן מ-x או שווה לו, הוא בדיוק x: כי עמד לרשותנו ניסיון אחד

---

[1] ל"נ טולסטוי, מלחמה ושלום, תרגום לעברית ע"י ל' גולדברג, ספרית פועלים, מרחביה 1953.
[2] השתמשנו בקובץ מחשב מצוי של ספר זה:
Facility for Computer Analysis of Texts (FCAT) and Tools for Septuagint Studies (CATSS). *The Book of Isaiah*, file ISAIAH.MT. University of Pennsylvania, Philadelphia. April 1986.
[3] Bonferroni.

בלבד כדי להגיע לערך x. זהו המספר המינימלי של הזדמנויות עבור ארבעה מאורעות, ולכן הסיכוי מקבל את הערך המינימלי: x.

לעומת מקרים קיצוניים אלה, המקרה הרגיל מצוי אי שם "באמצע". ארבעת המספרים r_i מייצגים ארבע הסתברויות של מאורעות שאינם זרים ואשר יש תלות ביניהם. כאן אין עומדים לרשותנו ארבעה נִסיונות "שלמים", או ארבע הזדמנויות "במלואן" – ולכן הסיכוי להגיע בכל זאת למצב בו לפחות אחד המספרים r_i יהיה קטן מ-x או שווה לו, יהיה קטן מ- $4x$. מצד שני, עמדו לרשותנו יותר מניסיון בודד (כי המאורעות אינם חופפים לחלוטין), ולכן הסיכוי הכולל גדול מ-x. כך, שבסך הכל, ההסתברות היא לכל היותר $4x$.

פרופסור דיאקוניס ושאר השופטים דרשו לחשב את המובהקות בדרך זו. ולכן, החליט פרופסור אומן כי המובהקות המבוקשת r_0 תיקבע על פי הנוסחה:

$$r_0 = 4 \min r_i$$

זו היא רמת המובהקות הכוללת של הניסוי.

לעניות דעתנו, קביעה זו קיפחה אותנו. מי שיתבונן בהגדרות של המידות P_1-P_4 (הן נמצאות בנספח א3), יראה כי המקרה שלפנינו קרוב מאד למקרה ב: קיימת תלות חזקה, ולמעשה כמעט חפיפה, בין P_1 ל-P_3, ובין P_2 ל-P_4, וכן קיימת תלות ברורה בין P_1 ל-P_2. ולכן, הגורם 4, בו הוכפלה התוצאה – מוגזם בהחלט. הוכחנו זאת להלן, בסעיף ג.

2. מן הטבלה לעיל, אנו מקבלים עבור ספר בראשית כי $\min r_i = 0.000004$, ולכן:

$$r_0 = 4 \min r_i = 0.000016$$

מאותה טבלה עולה כי עבור הטקסט I, ערכו של $\min r_i$ הוא כ- 0.900; עבור R ערכו 0.365; עבור T ערכו 0.277; עבור U ערכו 0.276; ערכו עבור V הוא 0.212; וערכו עבור W הוא 0.515.

לכן, המובהקות הכוללת, שהיא $r_0 = 4 \min r_i$, אינה משמעותית כלל וכלל לגבי שום טקסט מהנ"ל.

ג. עוד על תיקון בונפרוני

כמבואר לעיל, בעבודתנו נעשה שימוש בפרוצדורה של בונפרוני, וזאת לפי דרישת השופטים. לפי בונפרוני, אם α הוא הסף המבוקש ויש m סטטיסטים בניסוי, הרי צריכה להיות תוצאה של (לפחות) אחד הסטטיסטים $p \le \alpha / m$.

השימוש בפרוצדורה של בונפרוני נחשב כשמרני עד שמרני מאד (חיפוש במרשתת מגלה כי צמד המלים בונפרוני ו"שמרני" שכיח. יש המשתמשים אפילו בתואר "דרקוני"). במיוחד התיקון נחשב שמרני כאשר קיימת תלות בין הסטטיסטים.

למעשה, כל אי שוויון כללי המשמש בהערכות הסתברותיות נותן הערכות שמרניות; ככל שהוא כוללני יותר – הוא פחות מדויק ופחות מחודד עבור מקרים פרטיים. אבל, לנו אין צורך

להזדקק לאי שוויונים כאלה: ביכולתנו למדוד ישירות את המובהקות הכוללת באמצעות אותו מבחן הפרמוטציות עצמו שבוצע בניסוי[1]!

1. בעבודתנו היו 4 סטטיסטים. תוצאותיהם המדורגות במבחן הפרמוטציות:

4/1,000,000   5/1,000,000   453/1,000,000   570/1,000,000

לפי תיקון בונפרוני: $p = 4\times 4/1{,}000{,}000 = 16/1{,}000{,}000$.

למעשה, הסטטיסטים הללו תלויים זה בזה תלות חיובית חזקה. נטפל בכך תוך ניצול התוצאות של מבחן פרמוטציות שבידינו.

2. נדגים זאת:

הסטטיסטים P_2 ו- P_4 תלויים חזק זה בזה (P_4 הוא בדיוק P_2 המיושם לגבי תת-רשימה של נתונים כפי שמבואר בנספח א3).

(א) הדירוג של P_2 היה חמישי מתוך מיליון פרמוטציות. ארבע פרמוטציות הקדימו את המדגם המקורי:

#300087, #620851, #788884, #808836.

(ב) הדירוג של P_4 היה רביעי מתוך מיליון פרמוטציות. שלוש פרמוטציות הקדימו את המדגם המקורי:

#777442, #788884, #808836.

(ג) אם נסכם שני סטטיסטים אלה לפי בונפרוני נקבל: $p = 2\times 4/1{,}000{,}000 = 8/1{,}000{,}000$.

אבל אם נשאל: כמה פרמוטציות שונות הגיעו לדירוג 1-4 מתוך 1,000,000 פרמוטציות, לפחות באחד משני "המירוצים"? – התשובה לכך: 6 (כולל המדגם המקורי), והמובהקות היא לא יותר מ-6/1,000,000!

3. אם נסכם באופן דומה לגבי כל ארבעת הסטטיסטים, נקבל כי המובהקות היא 12 למיליון ולא 16 למיליון (כפי שחושב בעבודתנו לפי דרישת השופטים).

---

[1] חישובי מובהקות כוללת באמצעות פרמוטציות או ראנדומיזציה אחרת כבר נעשו ע"י P. H. Westfall ואחרים.