

## הרביעייה הטובה ביותר

כאן נתאר את המדידות, שצוינו בנספח ה' למאמר "מדגם העמים" (בסעיף ה'), ואשר תוכנו ובוצעו על ידי יוסף ברמז.

### א. כיצד נקבעת "הרביעייה הטובה ביותר" עבור טקסט $T$ כלשהו לפי $P_I$ :

1. מאמצים את רשימת הקידומות  $k_i$  של המבקרים,  $1 \leq i \leq 136$ .
2. כל קידומת  $k_i$  מגדירה קבוצה  $K_i$  של 68 זוגות ביטויים ( $X$ , קידומת  $X$ ), כאשר  $X$  הוא שם מ"לוח העמים" (בלוח העמים יש 68 שמות שונים. ראה טבלה 1 בנספח ג).
3. כאשר מחשבים את ערכי  $c(w, w')$  עבור הזוגות בקבוצה  $K_i$ , מתקבלת הקבוצה  $K_i(T)$  שבה  $n(K_i, T)$  ערכי  $c(w, w')$ , כאשר  $n(K_i, T)$  הוא מספר בין 0 ל-68.
4. מחשבים את הערך  $P_I$  עבור הקבוצה  $K_i(T)$ .
5. מדרגים את הקבוצות  $K_i(T)$  לפי ערך עולה של  $P_I$ , כך ש-  $K_1(T)$  היא בעלת ערך  $P_I$  הנמוך ביותר (את קבוצת זוגות הביטויים, ש-  $K_1(T)$  היא קבוצת ערכי  $c(w, w')$  שלה, נסמן ב-  $K_1$ ).
6. "הרביעייה הטובה ביותר",  $Q(T)$ , מוגדרת כאיחוד של ארבע הקבוצות:  $K_1(T) - K_4(T)$ .
7. כדי למדוד את המובהקות של הסטטיסטיקה  $P_I(T)$  עבור הרביעייה  $Q(T)$ , נפעיל את הראנדומיזציה  $PRWL$  על קבוצת הזוגות  $Q$ , שהיא איחוד ארבע הקבוצות  $K_1 - K_4$ .
  - בדרך זו נמצא כי בספר בראשית, ערכו של  $P_I$  עבור הרביעייה  $Q(G)$  הוא:  $P_I(G) = 1.65 \times 10^{-9}$ . ובאמצעות הראנדומיזציה  $PRWL$ , עם 10,000,000,000 פרמוטציות, התקבלה המובהקות:  $r_I(G) = 4.0 \times 10^{-10}$ .
  - ובטקסט  $WGP$ , היה ערכו של  $P_I$  עבור הרביעייה  $Q(WGP)$ :  $P_I(WGP) = 4.23 \times 10^{-5}$ . ובאמצעות הראנדומיזציה  $PRWL$ , עם 100,000,000 פרמוטציות, התקבלה המובהקות:  $r_I(WGP) = 6.16 \times 10^{-6}$ .

### ב. הערכת ההסתברות לקבלת הערך $r_I(T)$ (על ידי סימולציה):

בגלל הפער העצום בין הערכים  $r_I(G)$  ו-  $r_I(WGP)$ , מתעורר הצורך להעריך את ההסתברות, שערכו של  $r_I(T)$  נמוך כפי שהוא. לא ניתן לעשות זאת באופן פשוט, על ידי חישוב מספר הרביעיות שניתן לבחור מתוך 136 פריטים, בגלל שרביעיות כאלו אינן זרות, ויש תלות ביניהן. ולכן נשתמש בסימולציה באופן הבא:

1. במקום הקבוצה  $K_i(T)$  בשלב 3 דלעיל, מגרילים קבוצה  $K'_i(T)$  ובה  $n(K_i, T)$  מספרים אקראיים בקטע  $(0, 1]$  (זה טווח ערכי  $c(w, w')$ , המיוצרים על ידי פונקציה ראנדומלית במחשב.
2. מבצעים שלבים 4-6 דלעיל, כאשר משתמשים בקבוצות  $K'_i(T)$ , במקום  $K_i(T)$ .

3. כך מקבלים את הקבוצה  $Q'(T)$ , ומחשבים את ערך  $P'_I(T)$ . (שלב 7 דלעיל, שהיה נחוץ כדי למנוע השפעת תלויות אפשריות בערכי  $c(w, w')$ , הוא מיותר כאשר משתמשים במספרים אקראיים.)

4. חוזרים פעמים רבות על סדרת ההגרלות והחישובים שבשלים 1-3, ובוחנים מהו הדירוג של  $r_I(T)$  מתוך קבוצת ערכי  $P'_I(T)$ .

### ג. תוצאות ומסקנות:

- לגבי ספר בראשית נעשו 1,000,000 סדרות הגרלות, וחושבו 1,000,000 ערכי  $P'_I(G)$ . הדירוג של  $r_I(G)$  בקבוצת ערכים זו היה 420.

לכן, הסיכוי שערכו של  $r_I(G)$  כה קטן הוא  $p=0.00042$ .

- לגבי טקסט WGP נעשו 1,000,000 סדרות הגרלות, וחושבו 1,000,000 ערכי  $P'_I(WGP)$ . הדירוג של  $r_I(WGP)$  בקבוצת ערכים זו היה 353,949.

לכן, הסיכוי שערכו של  $r_I(WGP)$  כה קטן הוא  $p=0.354$ .

מסקנה: לפנינו הוכחה מתמטית מוצקה לכך, שבספר בראשית אכן יש צופן חבוי בדלוג שווה של אותיות.