

מעבדה לסטטיסטיקה: מודל פוואסון אחיד

נסמן ב Y_i את מספר הפרגמנטים המתחילים בבסיס $i=1, \dots, J$. בשיעור, ניסיתי לשכנע אתכם שהמודל הפשוט ביותר של התפלגות פרגמנטים יהיה להניח שכל Y_i מתפלג פוואסונית עם פרמטר זהה, ושה Y_i ים בלתי תלויים. מטרת התקציר הזה להבליט את ההנחות הדרושות לכל שלב בחישוב.

המודל הראשוני שלנו התנה על מספר הפרגמנטים הנצפים. נסמן מספר זה ב J . כעת, נגדיר סדרת משתנים $X_j, j = 1, \dots, J$ אשר מייצגים את הבסיס ההתחלתי של הפרגמנט. כלומר, אם הפרגמנט ה-10 מתחיל במקום ה-1000 אז $X_{10} = 1000$. אם כן, X_j הוא משתנה דיסקרטי המקבל ערכים בין 1 ל J . הערה: קל יותר אם נניח שהסדר שרירותי, לדוגמה כפי שהפרגמנטים מסודרים בקובץ המיפוי הראשון.

מהי התוחלת של Y_i ?

ניתן לכתוב את Y_i כסכום אינדיקטורים התלויים במשתנים X_1, \dots, X_J בצורה הבאה:

$$Y_i = \sum_{j=1}^J 1(X_j = i),$$

ולכן מלינראיות התוחלת:

$$E[Y_i] = \sum_{j=1}^J E[1(X_j = i)] = \sum_{j=1}^J P(X_j = i).$$

הנחה 1: כל המשתנים X_1, \dots, X_J הם משתנים שווי התפלגות.

על פי הנחה זו, הסיכוי של כל פרגמנט להיות במקום הראשון הוא סיכוי זהה, $P(X_1 = 1) = P(X_2 = 1) = \dots = P(X_J = 1)$ וניתן לסמן אותו בעזרת קבוע יחיד p_1 . באופן דומה נסמן גם קבועים p_2, p_3, \dots, p_J שמתייחסים לסיכוי של פרגמנט (כלשהו) להתחיל במקום ה-2 בכרומוזום, 3 בכרומוזום, וכן הלאה. מתוך ההנחה הזו מתקבל הביטוי המקוצר:

$$E[Y_i] = \sum_{j=1}^J p_i = J \cdot p_i.$$

כעת, נדון בשתי הנחות הנוספות המרכיבות את המודל הפשוט:

הנחה 2: הסיכוי של הפרגמנט הראשון להתחיל בכל תא הוא שווה.

בפרט, נובע מכך ש $p_1 = p_2 = \dots = p_I$ וכתוצאה מכך ש $p_i = 1/I$ לכל i .

הצבה של ההנחה הזו בנוסחא הקודמת לתוחלת תיתן לנו:

$$E[Y_i] = J \cdot (1/I) = J/I.$$

ועדיין, איננו יודעים מה ההתפלגות.

הנחה 3: המשתנים X_1, \dots, X_J בלתי תלויים.

אם אכן אין תלות בין המשתנים, הרי שגם אין תלות בין משתני האינדיקטור

$1(X_1 = i), \dots, 1(X_J = i)$. אם כן, Y_i הוא סכום של J משתני אינדיקטור בלתי תלויים ובעלי סיכוי

$1/I$. הצלחה כלומר, $Y_i \sim \text{Binomial}(J, 1/I)$. מכיוון ש J גדול ואילו $1/I$ קטן, הרי שניתן לקרב

בעזרת $Y_i \sim \text{Poisson}(J/I)$.

מחשבות להמשך הדרך:

1. חשבו על דוגמה שבה הנחות 1 ו-2 מתקיימות אך הנחה 3 איננה מתקיימת.

2. אם רק הנחות 1 ו-3 מתקיימות. מהי ההתפלגות של Y_i ?

3. אם J איננו קבוע אלא מתפלג פואסון, מהי ההתפלגות של Y_i ?

(רמז: אם J איננו קבוע, בעצם חישובנו ש $(Y_i | J = j) \sim \text{Binomial}(j, 1/I)$)