

האוניברסיטה העברית  
המחלקה לסטטיסטיקה

מבחן סיום בקורס: **(52609) שיטות חישוביות בסטטיסטיקה**

תאריך הבחינה: 17.1.11, שעה 9:00 (מועד א)

משך המבחן: שעתיים

חומר מותר בשימוש: 3 דפים כתובים בכתב ידו של הסטודנט.

הוראות: יש לענות על שתי השאלות ולנמק (בקיצור) כל אמירה. בהצלחה!

שאלה 1

פרמטר המאפשר להשוות בין שני משתנים בינומיים הוא יחס הסיכויים. פרמטר זה מוגדר כ-  
 $\theta = [p_1(1-p_2)]/[p_2(1-p_1)]$ . אומד נראות מרבית לפרמטר זה מתקבל על ידי אמידת  
הסתברויות ההצלחה בעזרת פרופורציות במדגם והצבה בנוסחה. אומד מקובל לשונות של  
לוג אומד הנראות המרבית של הפרמטר הנ"ל מתקבל על ידי שימוש בנוסחה המקורבת  
לשונות:  $Var[\log(\hat{\theta})] \approx 1/(n_1 p_1) + 1/(n_1(1-p_1)) + 1/(n_2 p_2) + 1/(n_2(1-p_2))$  בקוד  
שלהלן מוצעים שני אומדים שונים לשונות לוג האומד של יחס הסיכויים:

```
> p1 <- 25/100
> p2 <- 37/140
> or <- (p1*(1-p2))/(p2*(1-p1))
> sqrt(1/25 + 1/75 + 1/37 + 1/103)
[1] 0.3001151 (אומד 1)
> OR <- rep(0,200)
> for(i in 1:200)
+ {
+ P1 <- rbinom(1,100,p1)/100
+ P2 <- rbinom(1,140,p2)/140
+ OR[i] <- (P1*(1-P2))/(P2*(1-P1))
+ }
> sd(log(OR))
[1] 0.3403561 (אומד 2)
```

1. מבין האפשרויות הבאות, מי המתאימה ביותר לאפיון האלגוריתם בו משתמשים לחישוב אומד 2?

- i. Non-parametric Bootstrap
- ii. Parametric Bootstrap
- iii. EM
- iv. Gibbs

נמקו!

2. מה ערכו במדגם של לוג אומד הנראות המרבית של יחס הסיכויים? נמקו.
3. נטען כי אם במקום המספר 200 שמופיע בקוד היה רשום המספר 2,000 (בשני המקומות בהם המספר רשום) אזי שונות לוג אומד הנראות המרבית של יחס הסיכויים הייתה קטנה יותר. האם הנכם מסכימים עם טענה זו אם לאו?
4. רשמו פונקציה המקבלת כקלט את מספר ההצלחות ומספר הכישלונות בשני המשתנים הבינומיים ומוציאה כפלט אומד לסטיית התקן של אומד הנראות ליחס הסיכויים. (שימו לב: סטיית התקן של האומד עצמו, לא של לוג האומד.)

## שאלה 2

נתונה תצפית  $Y$ , שעשויה לקבל את הערך 0 או את הערך 1. בהינתן משתנה  $X$  מתקבל כי ההסתברות לערך  $Y=1$  הוא  $P(Y=1) = \exp(\beta X) / [1 + \exp(\beta X)]$ . ההתפלגות של  $X$  הנה אחידה (בדידה) על הערכים 0, 1, ו-2. בעבור סדרת תצפיות  $Y_1, Y_2, \dots, Y_n$  בלתי תלויות מעוניינים באומד נראות מרבית לפרמטר  $\beta$ . שימו לב כי המשתנים המתאימים  $X_1, X_2, \dots, X_n$  אינם נצפים.

1. רשמו את לוג פונקציית הנראות של הפרמטר בעבור סידרת התצפיות. מדוע יש קושי לחשב את אומד הנראות המרבית ישירות מתוך לוג פונקציית הנראות?
2. רשמו אל לוג פונקציית הנראות בעבור המקרה בו צופים בסדרת הזוגות הבלתי תלויים  $(Y_1, X_1), (Y_2, X_2), \dots, (Y_n, X_n)$ .
3. חשבו, בעבור ערך ידוע של  $\beta$ , את ההתפלגות המותנית של  $X_i$  בהנתן  $Y_i$ .
4. רשמו את הפונקציה  $Q(\beta' | \beta)$  שנדרשת לצורך הפעלת אלגוריתם EM-בבעיה זו. שימו לב כי הפונקציה המבוקשת היא התוחלת של הפונקציה שחושבה בסעיף 2 (עבור הערך  $\beta'$ ), כאשר התוחלת הנדרשת היא התוחלת המותנית בתצפיות (עבור הערך  $\beta$ ). כיצד ניתן להשתמש בפונקציה  $Q(\beta' | \beta)$  כדי למצוא את אומד הנראות המרבית לפרמטר?