

מבחן סיום בקורס: (52606) שיטות חישוביות בסטטיסטיקה גנטית

תאריך הבחינה: 8.2.07, שעה 8:30. (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!

1 הממוצע הקטום (trimmed mean) משמש כאומד עמיד (robust) לתוחלת אוכלוסיה בעלת התפלגות סימטרית. באומד זה מחשבים את הממוצע של התצפיות המרכזיות בלבד. האופציה "trim" בפונקציה "mean" ב-R מאפשרת חישוב ממוצע קטום וקובעת את פרופורצית הקטימה בכל אחד מן הזנבות של המדגם. להל"ן תוצאות מדגם שנלקח מאוכלוסיה סימטרית:

12.98, 8.79, -3.85, 25.43, 11.19, 6.90, 13.31, 5.40, 7.54, 9.82, 7.50, 9.02, 21.14, 9.79
13.37, 9.31, 6.93, 10.90, 13.97, 9.79

התוצאות נשמרו ב-R באובייקט ששמו X. נחשב אומד לתוחלת האוכלוסיה ולסטיית התקן של האומד לתוחלת:

```
> mean(X,trim=0.05)
[1] 10.425
> n <- 20
> sd(X)/sqrt(n)
[1] 1.312
```

הוצע אומד אלטרנטיבי לסטיית התקן של האומד לתוחלת. האומד האלטרנטיבי יושם בקוד שלהל"ן:

```
> B <- 200
> Xbar.bs <- vector(length=B)
> for (b in 1:B) Xbar.bs[b] <- mean(sample(X,n,rep=TRUE),trim=0.05)
> sd(Xbar.bs)
[1] 1.298
```

- א. איזה משני האומדים לסטיית התקן של האומד לתוחלת תעדיפו? נמקו!
- ב. נטען כי האומד האלטרנטיבי מבוסס על שיטת ה-bootstrap. האם הינכם מסכימים עם טענה זו? אם כן, האם המדובר בגישה הפרמטרית או בגישה שאיננה פרמטרית?
- ג. כיצד, לדעתכם, יושפע הערך של האומד האלטרנטיבי לסטיית התקן אם ערכו של הפרמטר B יקבע להיות 800, במקום 200?
- ד. הציעו דרך לבנית רווח סמך לתוחלת האוכלוסיה. (ניתן ורצוי לכתוב קוד שיישם את הדרך שהצעתם.)

2 מערכת מכילה רכיבים זהים חיצונית, שיוצרו על-ידי שני יצרנים שונים. משך אורך החיים של רכיב שיוצר על-ידי יצרן 1 מפולג מעריכית עם פרמטר λ_1 ואילו משך אורך החיים של רכיב שיוצר בידי היצרן השני מפולג מעריכית עם פרמטר λ_2 . הנח כי הפרמטרים ידועים. ברצוננו לאמוד את פרופורצית הרכיבים שבמערכת שיוצרו על-ידי יצרן 1. נסמן פרופורציה זו ב- π .

א. יהי X משך החיים של רכיב אקראי. הראה כי ניתן לבטא את הנראות של X בצורה $\pi \lambda_1 e^{-\lambda_1 X} + (1 - \pi) \lambda_2 e^{-\lambda_2 X}$. בהנתן מדגם של n רכיבים כנ"ל רשום את הנראות המשותפת. מה נדרש לעשות בכדי לחשב את אומד הנראות המירבית של π ?

ב. נסמן ב- Y_i את המציינת של היות הרכיב ה- i מיוצר על-ידי היצרן 1. (כלומר, שווה לאחד אם הרכיב יוצר בידי יצרן 1 ולאפס אם הרכיב יוצר על-ידי יצרן 2). הראה כי ניתן להציג את לוג הנראות המשותפת של X_1, \dots, X_n ו- Y_1, \dots, Y_n בצורה:

$$\sum_{i=1}^n [Y_i \log \pi + (1 - Y_i) \log(1 - \pi)] + \sum_{i=1}^n [Y_i (\log \lambda_1 - \lambda_1 X_i) + (1 - Y_i) (\log \lambda_2 - \lambda_2 X_i)]$$

ג. הראה כי התוחלת המותנית של Y_i , בהנתן X_i ובהנתן כי התוחלת השולית של Y_i היא π^* , שווה ל-

$$\frac{(\pi^*) \lambda_1 e^{-\lambda_1 X_i}}{(\pi^*) \lambda_1 e^{-\lambda_1 X_i} + (1 - \pi^*) \lambda_2 e^{-\lambda_2 X_i}}$$

ד. הצע אלגוריתם לחישוב אומד הנראות המירבית ל- π המבוסס על גישת ה-EM.