

מבחן סיום בקורס: (52606) שיטות חישוביות בסטטיסטיקה גנטית

תאריך הבחינה: 19.4.07, שעה 16:30. (מועד ב)

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 התפלגות גמה מאופיינת עלידי שני פרמטרים: הסקלה (scale) והצורה (shape). להל"ן תוצאות חלקיות ממדגם של 100 תצפיות שנלקח מאוכלוסייה המפולגת גמה:

4.138, 7.560, 6.001, 7.457, 2.938, ..., 3.318, 4.406, 5.133, 6.056, 3.604, 1.916

התוצאות נשמרו ב-R באובייקט ששמו X. האומדים בשיטת הממומנטים לשני הפרמטרים חושבו באופן הבא:

```
> shape <- mean(X)^2/var(X)
> scale <- mean(X)/var(X)
> c(shape,scale)
[1] 7.243600 1.264871
```

מעונינים בחסם סמך עליון (L) לערכו של פרמטר הצורה. לצורך כך נכתב הקוד הבא:

```
> shape.bs <- vector(length=200)
> for (b in 1:200)
+ {
+   X.bs <- rgamma(100,shape,scale)
+   shape.bs[b] <- mean(X.bs)^2/var(X.bs)
+ }
> sd(shape.bs)
[1] 1.137762
> L <- shape + 1.645*sd(shape.bs)
```

נבחן את רמת הסמך של החסם הסמך העליון:

```
> mean(shape.bs <= L)
[1] 0.92
```

- א. הצדיקו את הטענה כי הצורה והסקלה מחושבים בשיטת המומנטים.
- ב. נטען כי השיטה לבנית החסם העליון מבוססת על שיטת ה-bootstrap. האם הינכם מסכימים עם טענה זו? אם כן, האם המדובר בגישה הפרמטרית או בגישה שאיננה פרמטרית?
- ג. מהי רמת הסמך הצפויה של L? הסבירו כיצד הגעתם להשערתכם.
- ד. האם הדרך שהוצעה לבחון את רמת הסמך של חסם הסמך העליון סבירה לדעתכם? האם ניתן לשפר את הדרך? אם כן, כיצד?

2 מערכת מכילה רכיבים זהים חיצונית. מחצית מן הרכיבים יוצרה על ידי יצרן אחד ומחצית על ידי יצרן שני. 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 . ברצוננו לאמוד את הפרמטרים λ_1 ו- λ_2 בעזרת מדגם של אורכי חיים.

א. יהי X משך החיים של רכיב אקראי. הראה כי ניתן לבטא את הנראות של X בצורה $0.5 \cdot \lambda_1 e^{-\lambda_1 X} + 0.5 \cdot \lambda_2 e^{-\lambda_2 X}$. בהנתן מדגם של n רכיבים כנ"ל רשום את הנראות המשותפת. מה נדרש לעשות בכדי לחשב את אומד הנראות המירבית של λ_1 ו- λ_2 ?

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:

$$-n \log 2 + \sum_{i=1}^n [Y_i (\log \lambda_1 - \lambda_1 X_i) + (1 - Y_i) (\log \lambda_2 - \lambda_2 X_i)]$$

ג. הראה כי התוחלת המותנית של Y_i , בהנתן X_i ובהנתן כי ערכי הפרמטרים הם λ_1^* ו- λ_2^* בהתאמה, שווה ל-

$$\frac{\lambda_1^* e^{-\lambda_1^* X_i}}{\lambda_1^* e^{-\lambda_1^* X_i} + \lambda_2^* e^{-\lambda_2^* X_i}}$$

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