תזכורות מתמטיות

אהוד לם

אוגוסט 2011

מבוא

רוב התהליכים הביולוגיים עליהם נחשוב הינם תהליכים שמתאים לייצגם מתמטית בעזרת הכלים של **תורת ההסתברות**: השכיחות של תכונה/גן באוכלוסיה, סיכויי ההשרדות של פרט בעל תכונות מסוימות והגעתו לבגרות, מספר הצאצאים הממוצע של הורה בעל תכונות מסוימות, הכשירות הממוצעת באוכלוסיה, הסיכוי שפרט אחד יבחר לעזור לפרט שני ועוד.

בהרבה מקרים נרצה לעקוב מתמטית אחר תהליך אבולוציוני רב דורי , ולהגיד משהו על תכונותיו, וזאת לפני ובמנותק מנתונים אמפיריים ספציפיים כלשהם. לשם כך נצטרך לתאר את התנהגות המערכת ברמה **סטטיסטית.** בפועל נציג את מצב המערכת הביולוגית על ידי כך שנתאר, למשל, את השכיחות של תכונות מסוימות בכל דור, ונבנה מודל מתמטי שמראה כיצד שכיחויות אלה משתנות בעקבות ברירה טבעית. כך נוכל לומר האם תכונה מתפשטת באוכלוסיה, או הולכת ומעלמת. נוכל לאפיין מתי האוכלוסיה נמצאת במצב יציב ומתי לא, וכמה זמן ייקח למערכת להתייצב.

הידע המתמטי הדרוש הוא בעיקרו אלגברה בסיסית (רמה של חטיבת ביניים), וידע בסיסי מאוד על הסתברות שעליו נחזור בכיתה במידת הצורך. להלן הגדרות של מושגים הסתברותיים בסיסיים שישמשו אותנו, ותכונות מתמטיות מרכזיות שיש להכיר אותן.

מושגים בסיסיים

- משתנה מקרי (random variable) הוא משתנה שערכו הוא תוצאה של תהליך הסתברותי כלשהו. לכל אחד מהערכים האפשריים של המשתנה ניתן לייחס הסתברות. למשל: גובה של אדם אקראי הוא משתנה מקרי, וניתן להציג פונקצייה הנותנת עבור כל גובה את ההסברות שזה יהיה גובהו של האדם הנבחר. לעתים קרובות נסמן את ערך המשתנה ב x ואת ההסתברות של ערך מסוים כ $p(x)$.
- תוחלת של משתנה מקרי (expectation) היא הערך הממוצע של המשתנה, במדגם גדול מספיק. עבור משתנים מקריים בדידים, ניתן לחשב את התוחלת בהתאם לנוסחא הבאה (זה כמו ממוצע משוקלל):

$E(x) = \sum_x xp(x)$

ערכו הממוצע של המשתנה x מסומן לרוב $\bar{x}$

- השונות של משתנה מקרי (variance) היא מדד של מידת הפיזור של ערכי המשתנה. פורמלית השונות מוגדרת כתוחלת של ריבוע המרחקים מהממוצע (סטיית תקן היא השורש של השונות):

$\mathrm{var(x)}=\mathrm{E}\left\{(x-\bar{x})^2\right\} = \mathrm{E}(x^2)\text{-}\mathrm{E(x)}^2$

- השונות המשותפת של שני משתנים מקריים x ו y (covariance) היא מדד של עוצמת הקשר בין שני משתנים מקריים. תוחלת המכפלה פחות מכפלת התוחלות:

$\mathrm{cov}(x,y) = E\left[(x-E(x))(y-E(y))\right]$

$\mathrm{cov(x,y)}=\mathrm{E(xy)\text{-}E(x)E(y)}$

(חשב את השונות המשותפת של x עם עצמו)

- מקדם הרגרסיה הוא השיפוע של הקו הישר הטוב ביותר המשמש לניבוי ערכו של y בהינתן x:

$\beta(y,x) = \frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)}$

(מה מקדם הגרסיה של x על x?)

- נוסחת קו הרגרסיה של y לפי x:

$y = E(y) + \beta(y,x)\,(x - E(x))$

תכונות

יהי ...a,b קבועים כלשהם, ו ...x,y משתנים מקריים.

- $E(a) = a$
- $E(ax) = aE(x)$
- $E(x+y) = E(x) + E(y)$
- $\mathrm{var}(x) \geq 0(\mathrm{var(c)}=0, \text{ when c is a constant})$
- $\mathrm{var}(ax+b) = a^2\mathrm{var}(x)$
- $\mathrm{var}(x+y) = \mathrm{var}(x) + 2\mathrm{cov}(x,y) + \mathrm{var}(y)$
- $\mathrm{cov}(x,y) = \mathrm{cov}(y,x)$

- $\mathrm{cov}(x+z,y) = \mathrm{cov}(x,y) + \mathrm{cov}(z,y)$
- $\mathrm{cov}(ax,by) = ab \times \mathrm{cov}(x,y)$
- אם x ו y הם בלתי תלויים השונות המשותפת שלהם היא 0. (שימו לב שההפך לא נכון: אם השונות המשותפת היא אפס לא נובע מכך שהמשתנים לא תלויים!)
- אם השונות המשותפת חיובית אז ככל ש x גדל y נוטה לגדול. אם השונות המשותפת שלילית אז ככל ש x גדל הערך של y נוטה לרדת.

תרגול

1. הראה ש $\mathrm{cov}(x,y) = E\left\{x\left(y - E(y)\right)\right\}$