הסתברות כללית למדעי המחשב, 89-262

סמסטר א', בוחן, תשס"ט

1. החברים בבית הנבחרים במדינה מסויימת משתייכים לשתי מפלגות: 40 שמרנים, ו-60 ליברלים. שמרנים מתנגדים לכל ההצעות, וליברלים מתנגדים להצעה בסיכוי $4/5$. העמדה בכל הצבעה בלתי תלויה בעמדה בהצבעות אחרות. בוחרים חבר בית הנבחרים באקראי. מה ההסתברות לכך שהוא יתנגד להצעה בפעם השניה, אם ידוע שהוא התנגד לה בפעם הראשונה?

2. בכל לידה נולדים בן או בת, בהסתברויות שוות ובאופן בלתי תלוי בלידות אחרות. בעיר מסויימת ל-10% מהמשפחות אין ילדים כלל, ל-30% יש ילד אחד, ל-40% יש שני ילדים ול-20% יש 3 ילדים. נסמן ב-B את מספר הבנים במשפחה, וב-G את מספר הבנות.

   (א) מצא את פונקציית ההסתברות המשותפת של B ו-G, כלומר את כל ההסתברויות $P(B=i,G=j)$.

   (ב) חשב את ההתפלגות של B.

   (ג) חשב את התוחלת והשונות של B בהנתן $G=2$.

   (ד) מצא את מקדם המתאם של B ו-G.

3. יעל ואלון מבצעים מדי יום שני ניסויי ברנולי, שהסתברות ההצלחה של כל אחד מהם היא p. הניסויים בלתי תלויים גם זה בזה, וגם בניסויים שנערכו בימים אחרים.

   (א) מהי ההסתברות לכך שההצלחה הראשונה של שניהם תהיה באותו יום?

   (ב) סדרת הניסויים נעצרת עם ההצלחה הראשונה של אלון. מהי ההסתברות לכך שעד אותו יום (ועד בכלל), יעל תצליח בדיוק פעמיים?

   (ג) מהי תוחלת מספר ההצלחות של יעל עד שסדרת הניסויים עצרה?

הסתברות כללית למדעי המחשב, 89-262

סמסטר א', בוחן, תשס"ט
—פתרון—

1. נסמן ב־A את המאורע 'החבר הוא שמרן', ב־B את המאורע 'החבר הצביע נגד בהצבעה הראשונה', וב־C את המאורע 'החבר הצביע נגד בהצבעה השניה'. לפי הנתונים, $P(A)=\frac{2}{5}$, $P(B|A)=P(C|A)=1$ ו־ $P(B|A^{\mathrm{c}})=P(C|A^{\mathrm{c}})=\frac{3}{5}$. כמו כן $P(C|B,A)=P(C|A)$ וגם $P(C|B,A^{\mathrm{c}})=P(C|A^{\mathrm{c}})$, משום שעבור מצביע נתון, שתי ההצבעות בלתי תלויות.

   אפשר לחשב ש־ $P(B)=P(B|A)\cdot P(A)+P(B|A^{\mathrm{c}})\cdot P(A^{\mathrm{c}})=1\cdot\frac{2}{5}+\frac{4}{5}\cdot\frac{3}{5}=\frac{22}{25}$. לכן $P(A|B)=\frac{P(A\cap B)}{P(B)}=P(B|A)\cdot\frac{P(A)}{P(B)}=\frac{2/5}{22/25}=\frac{5}{11}$.

   מבקשים לחשב את $P(C|B)$. לפי נוסחת ההסתברות השלמה, $P(C|B)=P(C|B,A)\cdot P(A|B)+P(C|B,A^{\mathrm{c}})\cdot P(A^{\mathrm{c}}|B)=P(C|A)\cdot P(A|B)+P(C|A^{\mathrm{c}})\cdot P(A^{\mathrm{c}}|B)=\frac{5}{11}+\frac{4}{5}\cdot\frac{6}{11}=\frac{49}{55}$.

2. נסמן $N=B+G$; כך $B\sim\mathrm{Bin}(N,\frac{1}{2})$ וגם $G\sim\mathrm{Bin}(N,\frac{1}{2})$, למרות שכמובן B,G תלויים. ההתפלגות של N היא $P(N=0)=\frac{1}{10}$, $P(N=1)=\frac{3}{10}$, $P(N=2)=\frac{4}{10}$, $P(N=3)=\frac{2}{10}$.

   (א) על ידי נוסחת ההסתברות השלמה (ועל־פי ההסתברויות המותנות $(B|N)$, אפשר לחשב ש:

$$\begin{aligned}
P(B=G=0) &= 0.1;\\
P(B=0,G=1)=P(B=1,G=0) &= 0.15;\\
P(B=2,G=0)=P(B=0,G=2) &= 0.1;\\
P(B=G=1) &= 0.2;\\
P(B=3,G=0)=P(B=0,G=3) &= 0.025;\\
P(B=2,G=1)=P(B=1,G=2) &= 0.075.
\end{aligned}$$

   (ב) סיכום כל הערכים האפשריים על G מביא ל־ $P(B=0)=\frac{15}{40}$, $P(B=1)=\frac{17}{40}$, $P(B=2)=\frac{7}{40}$, $P(B=3)=\frac{1}{40}$.

   (ג) בהנתן $G=2$, ההתפלגות של B היא $P(B=1|G=2)=\frac{P(B=1,G=2)}{P(G=2)}=\frac{3/40}{7/40}=\frac{3}{7}$, ו־ $P(B=0|G=2)=1-P(B=1|G=2)=\frac{4}{7}$, כי הרי $3\leq B+G\leq G$. לכן $B|G=2\sim b(\frac{3}{7})$, ואז $E(B|G=2)=\frac{3}{7}$ ו־ $V(B|G=2)=\frac{3}{7}\frac{4}{7}=\frac{12}{49}$.

   (ד) מקדם המתאם הוא $\rho=\frac{E(BG)-E(B)E(G)}{\sqrt{V(B)V(G)}}$. אצלנו,

$$\begin{aligned}
P(BG=2) &= 0.075+0.075=0.15,\\
P(BG=1) &= 0.2,\\
P(BG=0) &= 0.65;
\end{aligned}$$

לכן $E(BG)=0.5$. אפשר לחשב ש־ $E(G)=E(B)=\frac{0+17\cdot1+7\cdot2+1\cdot3}{40}=\frac{17}{20}$, וש־ $E(G^2)=E(B^2)=\frac{0+17\cdot1+7\cdot4+1\cdot9}{40}=\frac{27}{20}$. לכן

$$V(G)=V(B)=\frac{27}{20}-\frac{17^2}{20^2}=\frac{540-289}{400}=\frac{251}{400}$$

(בכל הנ"ל השתמשנו בעובדה ש־ B ו־G שווי התפלגות). לכן

$$\rho=\frac{\frac{1}{2}-\frac{17^2}{20^2}}{\frac{251}{400}}=\frac{200-289}{251}=-\frac{89}{251}.$$

3. (א) נסמן ב־ A את זמן ההמתנה להצלחה הראשונה של אלון, וב־ Y את זמן ההמתנה להצלחה הראשונה של יעל. כמובן $A,Y\sim G(p)$, והם בלתי תלויים. לכן $P(A=Y)=\sum_{n=1}^{\infty}P(A=Y|Y=n)P(Y=n)=\sum_{n=1}^{\infty}P(A=n|Y=n)P(Y=n)=\sum_{n=1}^{\infty}P(A=n)P(Y=n)=\sum_{n=1}^{\infty}p^2q^{2n-2}=\frac{p^2}{1-q^2}=\frac{p}{2-p}$.

(ב) נסמן ב־ X את מספר ההצלחות של יעל עד שסדרת הניסויים נעצרה. מכיוון שהניסויים בלתי תלויים, $X|A\sim\mathrm{Bin}(A,p)$, ולכן

$$P(X=2|A)=\frac{A(A-1)}{2}p^2q^{A-2}.$$

מכאן ש־

$$\begin{aligned}P(X=2) &= \sum_{n=1}^{\infty}P(X=2|A=n)\cdot P(A=n)\\ &= \sum_{n=1}^{\infty}\frac{n(n-1)}{2}p^2q^{n-2}pq^{n-1}\\ &= \sum_{n=1}^{\infty}\frac{n(n-1)}{2}p^3q^{2n-3}.\end{aligned}$$

אבל אם גוזרים פעמיים את השוויון $\sum_{i=1}^{\infty}x^i=\frac{1}{1-x}$, מקבלים $\frac{2}{(1-x)^3}=\sum_{i=1}^{\infty}i(i-1)x^{i-2}$, ולכן

$$\begin{aligned}P(X=2) &= \frac{p^3q}{2}\sum_{n=1}^{\infty}n(n-1)(q^2)^{n-2}\\ &= \frac{p^3q}{2}\frac{2}{(1-q^2)^3}\\ &= \frac{p^3q}{(2p-p^2)^3}=\frac{1-p}{(2-p)^3}.\end{aligned}$$

(ג) עם X כבסעיף הקודם, $E(X)=E(E(X|A))=E(Ap)=E(A)p=1$ כי $X|A\sim\mathrm{Bin}(A,p)$ ו־ $A\sim G(p)$.

ניתוח של שאלה 3: תשובות שגויות (אמיתיות) והסברים.

המבוא לשאלה: יעל ואלון מבצעים מדי יום שני ניסויי ברנולי, שהסתברות ההצלחה של כל אחד מהם היא p. הניסויים בלתי תלויים גם זה בזה, וגם בניסויים שנערכו בימים אחרים.

a. מהי ההסתברות לכך שההצלחה הראשונה של שניהם תהיה באותו יום?

- פתרון: נסמן ב־A את מספר הניסויים עד להצלחה הראשונה של אלון, וב־Y את מספר הניסויים עד להצלחה הראשונה של יעל. לפי הנתונים, A ו־Y מתפלגים גאומטרית $G(p)$, והם בלתי תלויים. לכן

$$P(A=Y)=\sum_{n=1}^{\infty}P(A=Y=n)=\sum_{n=1}^{\infty}p^2(1-p)^{2n-2}=\frac{p}{2-p}.$$

- התשובה $\frac{p^2}{1-(1-p)^2}$ נכונה, אבל יש להמשיך ולצמצם.

שימו לב ש־ $(1-p)^{2^k}\neq((1-p)^2)^k$, משום שלפי כללי החזקה, $a^{b^c}=a^{(b^c)}$.

"$\sum p^2(1-p)^{2(i-1)}=0$ כי האיבר הכללי שואף לאפס". לא כך מסכמים טורים.

התשובה היא $\sum(p(1-p)^{k-1})^2$, ולא $(\sum p(1-p)^{k-1})^2$.

שימו לב ש־ $(1-p^2)^{k-1}p^2$ היא ההסתברות לכך שההצלחה *המשותפת* הראשונה תהיה ביום k, ולא לכך שאותה הצלחה משותפת אכן תהיה ראשונה עבור כל אחד מהם בנפרד. לכן הפתרון $\sum_k p^2(1-p^2)^{k-1}$ שגוי - הוא מסכם את ההסתברויות לכך שביום k כלשהו תהיה, לראשונה, הצלחה בו־זמנית ראשונה (בניגוד ל"הצלחה ראשונה בו־זמנית"!). התשובה היא, כצפוי, 1.

נתקלתי בביטוי $P(A\cap Y)$. אין משמעות לחיתוך של משתנים מקריים, אלא רק של מאורעות. גם את המאורע $A=Y=k$ כותבים כך, או $\{A=k\}\cap\{Y=k\}$, ולא $A\cap Y=k$.

"נסמן את ההצלחה הראשונה של אלון ב־A", או לקצרנים, "A - ההצלחה הראשונה של אלון", "A - הצלחה ביום מסויים", "Y - הניסוי של יעל" וכן הלאה. משתנה מקרי חייב להיות *מספר*. איך מחשבים מתוך הגדרות תמוהות שכאלה את ערכו של המשתנה? אותה בעיה קיימת בהגדרה "X - משתנה מקרי של הצלחה בניסוי ביום מסויים".

"נסמן ב־ k את הסיכוי לכך שההצלחה הראשונה של שניהם תהיה ביום k". כאן העברית תקינה, אבל הלוגיקה מעגלית.

התשובה "np^2 אם נניח שמבצעים את הניסוי n ימים" אינה נכונה, וגם אינה קבילה. התשובה "$2t$, כאשר $t=\frac{1}{2}\int_0^1\log(x)dx$" אינה עונה לשאלה "מהו האינטגרל $\int_0^1\log(x)dx$", אפילו אם היא נכונה. כנ"ל - "בהנחה ש־k הוא מספר הניסויים עד לעצירת הסדרה": הייתי אמור לשאול - ומה התשובה ללא ההנחה הזו?

"לשניים אותה הסתברות להצליח ולכן התשובה היא 1". לא רק שהנימוק שגוי בתכלית (העובדה שלשני אירועים אותה הסתברות אינה מחזקת את הנטיה שלהם לקרות בו זמנית), אלא שהפתרון לחלוטין אינו סביר. איך יכולה ההסתברות להיות 1, אם יתכן שביום הראשון אלון יצליח ויעל תכשל? התשובה $\frac{1}{p}$ גרועה אף יותר: הרי זו אינה הסתברות (אלא אם $p=1$).

נתקלתי, לדוגמא, בחישוב $\frac{1}{1-p} = \sum_0^k \frac{1}{p}$. יש שגיאות, ויש ביטויים שמאותתים "אין צורך לבדוק מחברת זו".

"אנחנו מחפשים את $P(X=Y=k)$, כאשר k יכול לשאוף לאינסוף". ראשית, אנחנו מחפשים את $P(X=Y)=\sum_k P(X=Y=k)$, ולא כפי שנטען. שנית, k אינו "יכול לשאוף לאינסוף"; k הוא מספר טבעי כלשהו, אבל מספרים, מטבעם, אינם שואפים לכלום. הם יכולים להיות גדולים כרצוננו, אבל לא לשאוף לאינסוף.

האירוע המבוקש "שקול לכך ששניהם יצליחו בניסוי הראשון". למה?

b. סדרת הניסויים נעצרת עם ההצלחה הראשונה של אלון. מהי ההסתברות לכך שעד אותו יום (ועד בכלל), יעל תצליח בדיוק פעמיים?

- פתרון: נסמן ב־ X את מספר ההצלחות של יעל עד שסדרת הניסויים נעצרה. מכיוון שהניסויים בלתי תלויים, $X|A \sim \mathrm{Bin}(A,p)$, ולכן

$$P(X=2|A) = \frac{A(A-1)}{2}p^2q^{A-2}.$$

מכאן ש־

$$\begin{aligned} P(X=2) &= \sum_{n=1}^{\infty} P(X=2|A=n)\cdot P(A=n) \\ &= \sum_{n=1}^{\infty} \frac{n(n-1)}{2}p^3q^{2n-3} = \cdots = \frac{1-p}{(2-p)^3}. \end{aligned}$$

- פתרון פופולרי: $P(X=2)=\binom{k}{2}p^2(1-p)^{k-2}$. זוהי כמובן ההסתברות המותנית $P(X=2|A=k)$, ולא $P(X=2)$, אלא שלא זה ולא זה פותרים את השאלה. הרי k אינו מופיע בה.

"הסיכוי של יעל להצליח ביום מסויים הוא הסתברות משתנה מקרי גאומטרי" (?) ־ רוצה לומר, "ההסתברות של יעל להצליח ביום שמספרו k שווה לסיכוי של משתנה מקרי גאומטרי להיות שווה ל־ k", או בקיצור "מספר הימים שחלפו עד להצלחה הראשונה של יעל מתפלג גאומטרית". גם "הצלחה של כל אחד מהם היא p" אינו משפט בעברית.

גם כאן הופיעה התשובה $P(X=2)=\frac{1}{p}$. תשובה בלתי אפשרית בעליל גרועה מסתם תשובה שגויה.

בחישוב $\left(\sum_{k=1}^{\infty} p(1-p)^{k-1}\right)\binom{k}{2}p^2(1-p)^{k-2}$ יש כוונה טובה, אבל מיקום הסוגריים מותיר את ההמשך ביטוי חסר משמעות (מהו k?)

רבים הציעו ש־ $X \sim NB(2,p)$. ההתפלגות הבינומית השלילית סופרת את מספר הנסיונות (של יעל) עד להצלחה השניה, ואינה קשורה לכאן.

"$2 \sim \mathrm{Bin}(n,p)$" – ?

"$\sum \binom{i}{2} q^{2i}$ הוא טור גאומטרי".

c. מהי תוחלת מספר ההצלחות של יעל עד שסדרת הניסויים עצרה?

- פתרון: עם X כבסעיף הקודם, $E(X)=E(E(X|A))=E(Ap)=E(A)p=$ 1 כי $X|A\sim \mathrm{Bin}(A,p)$ ו־ $A\sim G(p)$.

  התאכזבתי מכך שהפתרון הקצר לעיל לא נמצא באף מחברת.

- "הוכחנו שההסתברות לשתי הצלחות היא אפס" - ואי־אפשר לדמיין סיטואציה שבה יעל מצליחה בדיוק פעמיים עד להצלחה הראשונה של אלון?

  אחת התשובות השכיחות לשאלה הזו היתה $E(X)=kp$. כשקוראים את השאלה בתשומת לב מגלים שהפרמטר k אינו מופיע בה. לא יתכן שהתשובה תהיה תלויה בפרמטר כזה. אפילו כאשר מגדירים את k בתור "הזמן עד להצלחה הראשונה של אלון", אין בזה תשובה ממשית לשאלה. המטרה היא הרי למצוא את התוחלת על־פי הנתונים, ולא על־פי מידע פנימי נוסף שאינו נתון. באותה מידה אפשר היה לענות שהתשובה היא p^2 בהנחה שאלון הצליח בניסוי השני, או לתת נוסחה מורכבת יותר לסיכוי בהנחה שאלון הצליח לראשונה בניסוי ה־k: השאלה היתה מה התוחלת, ולא מה התוחלת המותנית במידע נוסף.

  לביטוי $E(Y=k)$ אין משמעות.