פתרון מטלה 08 — תורת ההסתברות (1), 80420

30 בדצמבר 2024

שאלה 1

יהי X משתנה מקרי בעל תוחלת, נראה שלכל $c > 0$ מתקיים

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{2\mathbb{E}(|X|)}{c}$$

הוכחה. נבחין כי מתקיים

$$\mathbb{E}(|\mathbb{E}(X)|) = \sum_{s \in \text{Supp } |\mathbb{E}(X)|} s\mathbb{P}(|\mathbb{E}(X)| = s) = |\mathbb{E}(X)|\mathbb{P}(|\mathbb{E}(X)| = |\mathbb{E}(X)|) = |\mathbb{E}(X)|$$

ולכן מאי־שוויון מרקוב

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq c) \leq \frac{\mathbb{E}(|X - \mathbb{E}(X)|)}{c} \leq \frac{\mathbb{E}(|X| + |\mathbb{E}(X)|)}{c} = \frac{\mathbb{E}(|X|) + \mathbb{E}(|\mathbb{E}(X)|)}{c} = \frac{2\mathbb{E}(|X|)}{2}$$

□

שאלה 2

נבחן מודל לתיאור גודל אוכלוסיית חיידקים במושבה, בדור 0 יש חיידק בודד, בסוף כל דור על חיידק מתפצל למספר חיידקים המתפלג פואסון עם פרמטר c באופן בלתי־תלוי. נסמן לכל $0 \leq t \in \mathbb{Z}$ את Z_t מספר המחיידקים החיים באוכלוסיה בדור t.

סעיף א׳

יהי $t \in \mathbb{N}$ ו־$z \geq 0$, נחשב את ההתפלגות $Z_t \mid \{Z_{t-1} = z\}$ ונסיק $\mathbb{E}(Z_t \mid \{Z_{t-1} = z\}) = zc$.

הוכחה. למעשה, נתון כי ישנם z חיידקים, כל אחד מהם מתפלג פואסונית, וראינו בהרצאה שחיבור של משתנים מקריים בלתי־תלויים פואסון מתפלג פואסון בחיבור הפרמטרים שלהם, לכן נובע ישירות

$$Z_t \mid Z_{t-1} \sim Poi(zc)$$

נבחין כי אפשר להוכיח את הטענה פורמלית על־ידי יצירת וקטור מקרי של משתנים מתפלגים פואסונית ולהשתמש בחיבורם ובמסקנה מההרצאה באינדוקציה.

מעוד תוצאה מההרצאה נובע

$$\mathbb{E}(Z_t \mid Z_{t-1} = z) = zc$$

□

סעיף ב׳

נסיק $\mathbb{E}(Z_t) = c^t$.

הוכחה. נוכיח את הטענה באינדוקציה על t.
כבסיס האינדוקציה נתון כי $\mathbb{E}(Z_0) = c$ (מנוסחת תוחלת התפלגות פואסון והנתון).
נניח כי הטענה נכונה עבור $t-1$ ולכן $\mathbb{E}(Z_{t-1}) = c^{t-1}$, וכן מנוסחת התוחלת השלמה

$$\mathbb{E}(Z_t) = \sum_{z=1}^{\infty} \mathbb{E}(Z_t \mid Z_{t-1} = z)\mathbb{P}(Z_{t-1} = z) = c\sum_{z=1}^{\infty} z\mathbb{P}(Z_{t-1} = z) = c \cdot c^{t-1} = c^t$$

והשלמנו את מהלך האינדוקציה. □

סעיף ג׳

נאמר שהאוכלוסיה נכחדה אם קיים t עבורו $Z_t = 0$.
נניח כי $c < 1$ ונוכיח שההסתברות שאוכלוסיית החיידקים תיכחד היא 1.

הוכחה.

$$\mathbb{P}(Z_t = 0) = 1 - \mathbb{P}(Z_t \geq 1) \geq 1 - \frac{\mathbb{E}(Z_t)}{1} = 1 - c^t$$

ולכן $\mathbb{P}(Z_t = 0) \xrightarrow[t\to\infty]{} 1$.
אנו מחפשים את הסיכוי שעבור איזשהו t האוכלוסייה תיכחד ולכן עלינו להגדיר $Z = \min\{Z_i \mid i \in \mathbb{N} \cup \{0\}\}$, וכמובן מהערך שמצאנו בהכרח $\mathbb{P}(Z = 0) = 1$. □

שאלה 3

1000 אנשים מגיעים למעבדה להיבדק למחלה, ההסתברות שאדם חולה היא p באופן בלתי־תלוי.
מחלקים את האנשים לקבוצות של 20, $A_1, \dots, A_{50}$. בשלב הראשון בודקים בבדיקה יחידה אם לפחות אחד האנשים חולה בקבוצה, לכל קבוצה, ובשלב השני בודקים כל אדם בנפרד בקבוצות בהן היו חולים.

סעיף א׳

נבטא באמצעות p את תוחלת מספר הבדיקות שהמעבדה מבצעת. נמצא עבור אילו ערכי p תוחלת מספר הבדיקות גדולה מ־1000.

פתרון כל אדם חולה בהתפלגות ברנולי p, ולכן התפלגות החולים בקבוצה A_i היא $Bin(20, p)$.

נבחין כי הקבוצה תיבדק בשלב השני אם ורק אם יש לפחות אדם חולה אחד, כלומר $\mathbb{P}(X_i > 0) = 1 - \mathbb{P}(X_i = 0)$ כאשר $X_i \sim Bin(20, p)$,
ולכן הקבוצה תיבדק בבודדים בהתפלגות ברנולי $1 - (1-p)^{20}$.

יש 50 קבוצות בלתי־תלויות וכל אחת נבדקת ביחידים בהתפלגות ברנולי, ולכן שוב ההתפלגות המשותפת היא בינומית, מספר הבדיקות בשלב השני, ונוסיף את 50 הבדיקות הקבוצתיות מההתחלה, מתקבל

$$X = 50 + 20 \cdot Bin(50, 1 - (1-p)^{20})$$

כאשר X מייצג את מספר הבדיקות, וכאשר עשינו שימוש בסימון לא סטנדרטי כדי להבהיר את ההתבססות של X על התפלגות בינומית, זאת לצורך החישוב הבא.

כדי לחשב את התוחלת נשתמש בלינאריות התוחלת ותוחלת הערך הקבוע, יחד עם תוחלת בינומית,

$$\mathbb{E}(X) = 50 + 20 \cdot \mathbb{E}(Bin(50, 1 - (1-p)^{20})) = 50 + 20 \cdot 50 \cdot (1 - (1-p)^{20})$$

נבדוק מתי $\mathbb{E}(X) > 1000$,

$$\mathbb{E}(X) > 1000 \iff 1 - (1-p)^{20} > 0.95 \iff 0.05 > (1-p)^{20} \iff \sqrt[20]{0.05} > 1 - p \iff p > 1 - \sqrt[20]{0.05}$$

ולכן עבור בערך $p > 0.139$ התוחלת תהיה גדולה מ־1000, כלומר אם אחוז החולים באוכלוסייה גדול מכ־13.9, אז לא משתלם להשתמש בשיטת הבדיקה הזו, ועדיף לחלק לקבוצות בגדלים שונים או לעבור לאסטרטגיית בדיקה אחרת.

סעיף ב׳

המעבדה מבצעת את 50 הבדיקות הראשונות הסדר מקרי אחיד, ולאחר מכן מבצעת את הבדיקות הבודדות גם כן בסדר מקרי אחיד.
נסמן ב־Z את מספר הבדיקות שהמעבדה מבצעת עד שיש תוצאה לנבדק נתון, ונחשב את התוחלת של Z על־ידי p.

פתרון נחלק למקרים ואז נשתמש בנוסחת התוחלת השלמה.

נניח שהנבדק הנתון לא חולה, ובמקרה זה או שהקבוצה שלו חולה או שהקבוצה שלו לא חולה.

אם הקבוצה שלו לא חולה, אז ההתפלגות של הבדיקות שייעשו עד שהוא יקבל תשובה היא $U([50])$, זאת שכן הקבוצה שלו תיבדק בסדר מקרי אחיד ותקבל תשובה שלילית.

נניח שהקבוצה שלו חולה ושהיו k קבוצות חולות, אז ייעשו בכולל $50 + 20 \cdot k$ בדיקות, כאשר ההתפלגות היא אחידה על $20 \cdot k$ הבדיקות האחרונות.
נעבור למצב בו הנבדק חולה, במקרה זה די שנדע את מספר הקבוצות החולות הכולל k ולכן ההתפלגות תהיה תוספת של 50 להתפלגות אחידה על $20 \cdot k$.

נסמן A המאורע שהנבדק הנתון חולה ולכן

$$\begin{aligned}\mathbb{E}(Z) &= \mathbb{E}(Z \mid A)\mathbb{P}(A) + \mathbb{E}(Z \mid A^C)\mathbb{P}(A^C) \\ &= \mathbb{E}(Z \mid A)p + \mathbb{E}(Z \mid A^C)(1-p)\end{aligned}$$

אנו יודעים שיש התפלגות אחידה על מספר הבדיקה עבור $Z \mid A$, ולכן די שנחשב על תוחלת מספר הבדיקות, כלומר על $\mathbb{E}(X \mid A)$, מספר זה הוא $70 + 20 \cdot 49(1 - (1-p)^{19})$ ולכן

$$\mathbb{E}(Z \mid A) = 70 + 10 \cdot 49(1 - (1-p)^{19})$$

נעבור לחישוב $\mathbb{E}(Z \mid A^C)$, נסמן Y המאורע שהקבוצה של הנבדק מכילה חולים, ולכן

$$\mathbb{E}(Z \mid A^C) = \mathbb{E}(Z \mid A^C, Y)\mathbb{P}(Y) + \mathbb{E}(Z \mid A^C, Y^C)\mathbb{P}(Y^C) = \mathbb{E}(Z \mid A^C, Y)(1-(1-p)^{20}) + \mathbb{E}(Z \mid A^C, Y^C)(1-p)^{20}$$

עבור $\mathbb{E}(Z \mid A^C, Y)$ זהו המקרה שהקבוצה חולה ולכן $\mathbb{E}(Z \mid A^C, Y) = 70 + 10 \cdot 49(1-(1-p)^{19})$ שכן זהו המקרה שיש חולים בקבוצה, והיא תיבדק מחדש. עבור $\mathbb{E}(Z \mid A^C, Y^C) = \mathbb{E}(Z \mid Y^C)$ ולכן $\mathbb{E}(Z \mid Y^C) = \frac{21}{2}$ שכן זהו המקרה שבו הנבדק מקבל תשובה כבר בהתחלה. לכן

$$\mathbb{E}(Z \mid A^C) = (70 + 10 \cdot 49(1-(1-p)^{19}))(1-(1-p)^{20}) + \frac{21}{2}(1-p)^{20}$$

וכן

$$\begin{aligned}\mathbb{E}(Z) &= \mathbb{E}(Z \mid A)p + \mathbb{E}(Z \mid A^C)(1-p)\\ &= (70 + 10 \cdot 49(1-(1-p)^{19}))p + ((70 + 10 \cdot 49(1-(1-p)^{19}))(1-(1-p)^{20}) + \frac{21}{2}(1-p)^{20})(1-p)\end{aligned}$$

סעיף ג'

נקבע עבור $p = 0.01, 0.1, 0.5$ מה האיגום הטוב ביותר מבחינת מזעור כמות הבדיקות בחלוקה לקבוצות של $5, 10, 50$ אנשים.

פתרון אם x מייצג את כמות האנשים בקבוצה אז נגדיר X_x^p מספר הבדיקות הכללי שיש לבצע בחלוקה לקבוצות של x אנשים.

$$\mathbb{E}(X_x^p) = \frac{1000}{x} + 1000(1-(1-p)^x) = 1000(\frac{1}{x} + 1 - (1-p)^x)$$

ובהתאם

$$\mathbb{E}(X_5^{0.01}) \approx 249, \qquad \mathbb{E}(X_{10}^{0.01}) \approx 195.6, \qquad \mathbb{E}(X_{50}^{0.01}) \approx 414$$

והכי ישתלם לבחור קבוצות של 10 אנשים.

באופן דומה גם

$$\mathbb{E}(X_5^{0.1}) \approx 609.5, \qquad \mathbb{E}(X_{10}^{0.1}) \approx 751.3, \qquad \mathbb{E}(X_{50}^{0.1}) \approx 1014.8$$

ולכן נבחר לחלק את האנשים לקבוצות של 5.

לבסוף גם

$$\mathbb{E}(X_5^{0.5}) \approx 1168.7, \qquad \mathbb{E}(X_{10}^{0.5}) \approx 1099, \qquad \mathbb{E}(X_{50}^{0.5}) = 1020$$

ולכן נבחר קבוצות של 50 אנשים.

שאלה 4

יהיו X, Y משתנים מקריים בעלי שונות המקיימים $\mathrm{Var}(X) = \mathrm{Var}(Y)$ ונוכיח ש־$X - Y, X + Y$ הם משתנים מקריים בלתי־מתואמים.

הוכחה.

$$\begin{aligned}
\mathrm{Cov}(X - Y, X + Y) &= \mathrm{Cov}(X, X + Y) - \mathrm{Cov}(Y, X + Y) \\
&= \mathrm{Cov}(X, X) + \mathrm{Cov}(X, Y) - \mathrm{Cov}(Y, X) - \mathrm{Cov}(Y, Y) \\
&= \mathrm{Cov}(X, X) - \mathrm{Cov}(Y, Y) \\
&= \mathrm{Var}(X) - \mathrm{Var}(Y) \\
&= 0
\end{aligned}$$

כאשר המעברים נבעו מלינאריות, קומוטטיביות, הזהות $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$ והנתון. □

שאלה 5

יהי $X \sim U(\{-n, \ldots, n\})$ עבור $n \in \mathbb{N}$ ותהי $f : \mathbb{R} \to \mathbb{R}$ פונקציה זוגית.
נוכיח ש־$X, f(X)$ הם משתנים מקריים בלתי־מתואמים.

הוכחה. נתחיל בחישוב,

$$\mathbb{E}(X) = \sum_{m=-n}^{n} m\mathbb{P}(X = m) = 0 \cdot \mathbb{P}(X = 0) + \sum_{m=1}^{n} m\mathbb{P}(X = m) + \sum_{m=1}^{n} -m\mathbb{P}(X = -m) = \sum_{m=1}^{n} m\frac{1}{2n+1} - m\frac{1}{2n+1} = 0$$

ולכן

$$\begin{aligned}\mathrm{Cov}(X, f(X)) &= \mathbb{E}((X - \mathbb{E}(X))(f(X) - \mathbb{E}(f(X)))) \\ &= \mathbb{E}(Xf(X) - X\mathbb{E}(f(X))) = \mathbb{E}(Xf(X)) - \mathbb{E}(X)\mathbb{E}(f(X)) = \mathbb{E}(Xf(X))\end{aligned}$$

נבחין כי X מתפלגת סימטרית וגם $f(X)$ ולכן גם $Xf(X)$, זאת אומרת $\mathbb{P}(Xf(X) = n) = \mathbb{P}(Xf(X) = -n)$, ולכן

$$\mathbb{E}(Xf(X)) = \sum_{s \in \mathrm{Supp}\, Xf(X)} s\mathbb{P}(Xf(X) = s) = \sum_{s > 0 \in \mathrm{Supp}\, Xf(X)} s\mathbb{P}(Xf(X) = s) + \sum_{s > 0 \in \mathrm{Supp}\, Xf(X)} -s\mathbb{P}(Xf(X) = -s) = 0$$

ובהתאם $\mathrm{Cov}(X, f(X)) = 0$ כפי שרצינו להראות. □

שאלה 6

יהיו X, Y משתנים מקריים בלתי־תלויים כך שמתקיים

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \mathbb{E}(X) = \mathbb{E}(Y)$$

סעיף א׳

נחשב את התוחלת והשונות של $X + Y$.

פתרון מלינאריות התוחלת

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) = 1 + 1 = 2$$

וכן מתוצאה מההרצאה ומאי־התלות הגורר אי־תיאום נובע גם

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = 2$$

סעיף ב׳

נחשב את התוחלת והשונות של XY.

פתרון מאי־התלות של X, Y נסיק

$$\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y) = 1$$

נעבור לחישוב השונות. נבחין כי

$$1 = \mathrm{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}((X - 1)^2) = \mathbb{E}(X^2) - 2\mathbb{E}(X) + 1 = \mathbb{E}(X^2) - 1$$

ולכן $\mathbb{E}(X^2) = 2$, ובאופן דומה גם $\mathbb{E}(Y^2) = 2$. מההגדרה השקולה לשונות

$$\mathrm{Var}(XY) = \mathbb{E}((XY)^2) - (\mathbb{E}(XY))^2 = \mathbb{E}(X^2Y^2) - 1^2 = \mathbb{E}(X^2)\mathbb{E}(Y^2) - 1 = 3$$