פתרון מטלה 09 — תורת ההסתברות (1), 80420

10 בינואר 2025

שאלה 1

נחסום על־ידי אי־שוויון מרקוב ובעזרת אי־שוויון צ'בישב את ההסתברות שלתמורה אקראית על $[n]$ יש לפחות k נקודות שבת.

פתרון נגדיר X מספר נקודות השבת וכן X_i שבמספר ה־i יש נקודת שבת. מאי־שוויון מרקוב

$$\mathbb{P}(X \geq k) \leq \frac{\mathbb{E}(X)}{k}$$

ובתרגול 8 ראינו ש־$\mathbb{E}(X) = 1$ ולכן

$$\mathbb{P}(X \geq k) \leq \frac{1}{k}$$

ניזכר כי באותו תרגול ראינו גם ש־$\mathrm{Var}(X) = 1$, לפי אי־שוויון צ'בישב

$$\mathbb{P}(X \geq k) = \mathbb{P}(X - 1 \geq k - 1) = \mathbb{P}(|X - \mathbb{E}(X)| \geq k - 1) \leq \frac{\mathrm{Var}(X)}{(k-1)^2} = \frac{1}{(k-1)^2}$$

וכמובן אסימפטוטית זהו חסם טוב בהרבה.

שאלה 2

בכיתה n תלמידים. ימי ההולדת שלהם בלתי־תלויים וכל יום הולדת מתפלג אחיד על פני שנה כלשהי.

יהי X משתנה מקרי שסופר את זוגות התלמידים שחולקים יום הולדת.

סעיף א׳

נחשב את תוחלת X.

פתרון בשיעור 14 ביצענו את החישוב הדרוש, והתקבל שאם יש m ימים בשנה הנתונה, אז

$$\mathbb{E}(X) = \binom{n}{2}\frac{1}{m}$$

סעיף ב׳

יהיו $i, j, k, l \in [n]$ כך ש־$\{i, j\} \neq \{k, l\}$, נראה שהמאורע ש־i, j חולקים יום הולדת בלתי תלוי במאורע ש־k, l חולקים יום הולדת.

הוכחה. נניח ש־X_ι המשתנה המקרי שמייצג שלתלמיד ה־$\iota \in [n]$ יש יום הולדת ביום מסוים, אז $X_i \sim U([m])$.

נחלק למקרים על ארבעת התלמידים הנתונים.

במקרה הראשון $i \neq j \neq k \neq l$ ולכן חוסר התלות נובע מחוסר התלות של X_ι.

מקרה נוסף הוא המקרה ש־$i = j \neq k = l$, כלומר בחרנו שני תלמידים בלבד, אז ישירות מחוסר התלות הנתון של ימי ההולדת גם הפעם מאורעות אלה בלתי תלויים.

נשאר אפוא לבדוק את המקרה (ללא הגבלת הכלליות) ש־$i = k$ אך $j \neq l$,

$$\mathbb{P}(X_i = X_j, X_k = X_l) = \mathbb{P}(X_i = X_j, X_i = X_l) = \mathbb{P}(X_i = X_j = X_l = \alpha) = \mathbb{P}(X_i = \alpha)\mathbb{P}(X_j = \alpha)\mathbb{P}(X_l = \alpha)$$

כפי שרצינו להראות (כאשר $\mathbb{P}(X_\iota = X_\iota) = 1$). □

סעיף ג׳

נחשב את השונות של X.

פתרון לצורך חישוב זה נגדיר $X_{i,j}$ המשתנה המקרי ש־i ול־j יש יום הולדת משותף, לכן $X_{i,j} \sim Ber(\frac{1}{m})$.

נובע גם ש־$X = \sum_{i=1}^{n} \sum_{j=i+1}^{n} X_{i,j}$.

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Cov}(X, X) \\
&= \left(\sum_{1 \le i < j \le n} \left(\sum_{\substack{1 \le k < l \le n \\ \{i,j\} \neq \{k,l\}}} \mathrm{Cov}(X_{i,j}, X_{l,k})\right)\right) + \left(\sum_{1 \le i < j \le n} \left(\sum_{\substack{1 \le k < l \le n \\ \{i,j\} = \{k,l\}}} \mathrm{Cov}(X_{i,j}, X_{l,k})\right)\right) \\
&= \left(\sum_{1 \le i < j \le n} \left(\sum_{\substack{1 \le k < l \le n \\ \{i,j\} \neq \{k,l\}}} 0\right)\right) + \left(\sum_{1 \le i < j \le n} \mathrm{Cov}(X_{i,j}, X_{i,j})\right) \\
&= 0 + \left(\sum_{1 \le i < j \le n} \mathrm{Var}(X_{i,j})\right) \\
&= \sum_{1 \le i < j \le n} \frac{1}{m}(1 - \frac{1}{m}) \\
&= \frac{n(n-1)(m-1)}{2m^2}
\end{aligned}$$

סעיף ד׳

נחסום על־ידי אי־שוויון צ׳בישב את ההסתברות שבכיתה בת 30 תלמידים יש שלושה זוגות שחולקים יום הולדת.

פתרון מהנתונים $n = 30, m = 365, \lambda = 3$, ומאי־שוויון צ׳בישב

$$\begin{aligned}
\mathbb{P}(X \geq \lambda) &= \mathbb{P}(X - \mathbb{E}(X) \geq \lambda - \mathbb{E}(X)) \\
&= \mathbb{P}(\left| X - \binom{n}{2}\frac{1}{m} \right| \geq \lambda - \binom{n}{2}\frac{1}{m}) \\
&\leq \frac{\mathrm{Var}(X)}{\lambda - \binom{n}{2}\frac{1}{m}} \\
&= \frac{\frac{n(n-1)(m-1)}{2m^2}}{\lambda - \binom{n}{2}\frac{1}{m}} \\
&= \frac{n(n-1)(m-1)}{2m^2\lambda - 2m\binom{n}{2}}
\end{aligned}$$

נציב ונקבל $\mathbb{P}(X \geq 3) \lessapprox 0.657$, כלומר הסיכוי שיש לפחות שלושה זוגות חסום על־ידי כשני שלישים.

שאלה 3

אליס בוחרת באקראי מספר שלם ב־$[100]$ ובוב מנחש באקראי את המספר שהיא בחרה.
לאחר מכן בוב משלם לאליס את ריבוע ההפרש בין המספר שאליס בחרה לבין המספר שבוב ניחש.

סעיף א׳

נחשב את המספר שעל בוב לנחש כדי למזער את תוחלת התשלום שלו לאליס.
פתרון נגדיר b המספר שבחר בוב, וכן $X_a \sim U([100])$ה מספר שבחרה אליס, ולבסוף $Y = (X_a - b)^2$ הכסף שבוב ישלם לאליס בסוף.

$$\mathbb{E}(Y) = \mathbb{E}(X_a^2) - 2b\mathbb{E}(X_a) + b^2 = (\sum_{i\in[100]} \frac{i^2}{100}) - 2b\frac{100+1}{2} + b^2 = 3383.5 - 101b + b^2$$

ולכן נגדיר $f(b) = 3383.5 - 101b + b^2$ ונגזור,

$$f'(b) = 2b - 101, \qquad f''(b) = 2$$

ולכן בנקודה $b = 50.5$ מקבלת הפונקציה מינימום, וזהו גם המספר שעל בוב לבחור (כאשר אם אסור לו לבחור מספר לא שלם כי הוא לא מתחכם, אז הבחירה 50 או 51 יהיו טובות באותה המידה).

סעיף ב׳

נניח שבוב אכן בוחר במדיניות שממזערת את תוחלת התשלום שלו כמו בסעיף הקודם.
אליס משחקת בהוגנות ומציעה לבוב לשלם מראש סכום מסוים, נחשב את הסכום שעל אליס לשלם כדי שתוחלת ההפסד של בוב תהיה אפס.
פתרון בסעיף הקודם מצאנו ש־$Y = (X_a - 50.5)^2$ המשתנה המקרי שמייצג את ההפסד של בוב במקרה שבו הוא בוחר מספר שימזער את ההפסד שלו. מהחישוב שעשינו גם נובע

$$\mathbb{E}(Y) = 3383.5 - 101 \cdot 50.5 + 50.5^2 = 833.25$$

כלומר תוחלת ההפסד של בוב היא 833.25, ואם אליס תשלם לו מראש בדיוק את הסכום הזה לפני תחילת הסיבוב, אז בהתאם תוחלת ההפסד שלו תתאפס (ישירות מלינאריות התוחלת).

שאלה 4

יהי אספן פוגים שבכל יום קונה חטיף במבה בו יש פוג בהתפלגות אחידה מבין n סוגי פוגים.

סעיף א׳

נסמן ב־Y_k את המשתנה המקרי שמתאר את היום הראשון בו יש לאספן k פוגים שונים.
נראה שהמשתנים המקריים $\{Y_k - Y_{k-1}\}_{1<k\leq n}$ הם בלתי־תלויים ומתפלגים גאומטרית עם פרמטר $1 - \frac{k-1}{n}$.

הוכחה. מטעמי שימוש בנוסחת ההסתברות השלמה מספיק שנבחן את המקרה שבו כלל המשתנים המקריים מוגדרים, אז

$$\begin{aligned}
\mathbb{P}(Y_2 - Y_1 = s_1, \dots, Y_n - Y_{n-1} = s_{n-1}) &= \mathbb{P}(Y_1 = 1, Y_2 - Y_1 = s_1, \dots, Y_n - Y_{n-1} = s_{n-1}) \\
&= \mathbb{P}(Y_1 = 1, Y_2 = 1 + s_1, \dots, Y_n = 1 + \dots + s_{n-1}) \\
&= \mathbb{P}(Y_2 = 1 + s_1, \dots, Y_n = 1 + \dots + s_{n-1}) \\
&= \sum_{\{c_1,\dots,c_n\}\in S_n} \mathbb{P}(\forall 1 \leq i \leq n, X_1, \dots, X_{s_i} \in \{c_1, \dots, c_{i-1}\}, X_{1+s_i} = c_i) \\
&= \sum_{\{c_1,\dots,c_n\}\in S_n} \prod_{1<i\leq n} \mathbb{P}(X_1, \dots, X_{s_i} \in \{c_1, \dots, c_{i-1}\}, X_{1+s_i} = c_i) \\
&= n! \prod_{1<i\leq n} \mathbb{P}(X_1, \dots, X_{s_i} \in [i-1])\mathbb{P}(X_{s_i+1} = i) \\
&= n! \prod_{1<i\leq n} (\frac{i-1}{n})^{s_i} \frac{1}{n} \\
&= (n-1)! \prod_{1<i\leq n} (\frac{i-1}{n})^{s_i}
\end{aligned}$$

וכן

$$\begin{aligned}
\mathbb{P}(Y_k - Y_{k-1} = s_k) &= \mathbb{P}(X_{Y_{k-1}}, \dots, X_{Y_k - 1} \in [k-1], X_{Y_k} \in \{k, \dots, n\}) \\
&= \mathbb{P}(X_{Y_{k-1}} \in [k-1]) \cdots \mathbb{P}(X_{Y_k - 1} \in [k-1])\mathbb{P}(X_{Y_k} \in \{k, \dots, n\}) \\
&= (\frac{k-1}{n})^{Y_k - Y_{k-1} - 1} \cdot \frac{n-k+1}{n} \\
&= (\frac{k-1}{n})^{s_k - 1} \cdot \frac{n-k+1}{n}
\end{aligned}$$

כלומר $Y_k - Y_{k-1} \sim Geo(1 - \frac{k-1}{n})$ ומכפלת המשתנים הללו היא בדיוק התוצאה שחישבנו קודם לכן, לכן הם בלתי־תלויים ומתפלגים גאומטרית. □

סעיף ב׳

נחשב את תוחלת מספר הימים שיעבור עד שלאספן יהיו כלל הפוגים.

פתרון אנו רוצים לחשב את Y_n, אך זה בתורו מקיים $Y_n = Y_n - Y_{n-1} + Y_{n-1} - Y_{n-2} + \dots + Y_2 - Y_1 + Y_1$. לכן מלינאריות התוחלת, אי־התלות וההתפלגות

$$\mathbb{E}(Y_n) = \mathbb{E}(Y_n - Y_{n-1}) + \dots + \mathbb{E}(Y_2 - Y_1) + \mathbb{E}(Y_1) = 1 + \sum_{k=2}^{n} \frac{1}{1 - \frac{k-1}{n}} = 1 + n \sum_{l=0}^{n-2} \frac{1}{l+1} = 1 + n \sum_{l=1}^{n-1} \frac{1}{l}$$

נבחין כי מטור טיילור של $\log$ כאשר $n \to \infty$ גם $\mathbb{E}(Y_n) \to 1 + n\log(1+n)$ ולכן אסימפטוטית התוחלת מתנהגת כמו $n \log n$.

סעיף ג׳

נחשב את שונות מספר הימים שיעברו עד שלאספן יהיו כלל הפוגים.

פתרון תחילה נחשב את השונות של $Y_k - Y_{k-1}$ לכל $2 \leq k \leq n$, זוהי כמובן שונות של התפלגות גאומטרית ולכן

$$\mathrm{Var}(Y_k - Y_{k-1}) = \frac{1 - (1 - \frac{k-1}{n})}{(1 - \frac{k-1}{n})^2}$$

נשתמש בחוסר התלות ובתכונות השונות והשונות המשותפת,

$$\begin{aligned}
\mathrm{Var}(Y_n) &= \mathrm{Var}(Y_n - 1) \\
&= \mathrm{Cov}(Y_n - Y_{n-1} + \cdots Y_2 - Y_1, Y_n - Y_{n-1} + \cdots Y_2 - Y_1) \\
&= \sum_{k=1}^{n} \mathrm{Cov}(Y_k - Y_{k-1}, Y_k - Y_{k-1}) + 2 \sum_{1 \leq k < l \leq n} \mathrm{Cov}(Y_k - Y_{k-1}, Y_l - Y_{l-1}) \\
&= \sum_{k=1}^{n} \mathrm{Var}(Y_k - Y_{k-1}) + 2 \sum_{1 \leq k < l \leq n} 0 \\
&= \sum_{k=1}^{n} \mathrm{Var}(Y_k - Y_{k-1}) \\
&= \sum_{l=0}^{n} \frac{\frac{l}{n}}{\left(1 - \frac{l}{n}\right)^2} \\
&\xrightarrow[n \to \infty]{} \int_0^{1-\frac{1}{n}} \frac{x}{(1-x)^2}\, dx \\
&= n + \log \frac{1}{n} \\
&= o(n^2)
\end{aligned}$$

ולכן הפונקציה מתנהגת אסימפטוטית כמו n^2.

סעיף ד׳

נראה שאחרי $Cn \log n$ ימים ההסתברות שלאספן יש את כל הפוגים שואפת ל־1 אם $C > 1$ ושואפת ל־0 אם $C < 1$.

הוכחה.

$$\mathbb{P}(Y_n \geq Cn \log n) \leq \frac{\mathbb{E}(Y_n)}{Cn \log n} = \frac{o(n \log n)}{Cn \log n} = \frac{o(n \log n)}{n \log(n^C)}$$

ומכאן נוכל להסיק את האמור. □

שאלה 5

נחשב כמה קטנה יכולה להיות השונות המשותפת $\mathrm{Cov}(X_i, X_j)$ אם ידוע ש־$X_1, \ldots, X_n$ הם משתנים מקריים המתפלגים $Ber(\frac{1}{2})$, וכי השונות המשותפת של כולם שווה.

נראה דוגמה שבה המינימום מתקבל.

פתרון נגדיר $X = \sum_{i=1}^{n} X_i$ וכן נגדיר $\mathrm{Cov}(X_1, X_2) = K$, אז

$$\mathbb{E}(X) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{n}{2}$$

וכן מתכונות השונות,

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Cov}(X, X_i) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j) = (n^2 - n)K + \frac{n}{4} \cdot n$$

אבל שונות היא אי־שלילית ולכן

$$(n^2 - n)K + \frac{n}{4} \cdot n \geq 0 \iff K \geq \frac{1}{4(1-n)}$$

ומצאנו ערך מינימלי לשונות המשותפת.

נגדיר $n = 2$ וכן $X_1 \sim Ber(\frac{1}{2})$ ו־$X_2 = 1 - X_1$, כלל ההנחות מתקיימות באופן ריק וכן מתכונות השונות המשותפת

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_1, 1 - X_1) = \mathrm{Cov}(X_1, -X_1) = -\mathrm{Cov}(X_1, X_1) = -\mathrm{Var}(X_1) = -p(1-p) = -\frac{1}{4}$$

ומהצד השני

$$\frac{1}{4(1-n)} = -\frac{1}{4}$$

ומצאנו שהמינימום אכן מתקבל.