האוניברסיטה העברית בירושלים
החוג למתמטיקה

בחינה בקורס "מבוא להסתברות וסטטיסטיקה" (80430)

מועד א' תשפ"ב (8/02/2022)

שאלה 1

ראו תשובה בספר

שאלה 2

לאלי ולבלה סיכויים שווים ובלתי תלויים לנצח בכל סיבוב של משחק דוקים. הם משחקים שבעה משחקים והמנצח המוחלט הוא מי שזכה ביותר משחקים.

1. מה ההסתברות שתוצאת המשחק האחרון קבעה את זהות המנצח?

פתרון: נשים לב שזהות המנצח נקבעת כאשר אחד מהשחקנים מגיע ל4 נצחונות. לכן במשחק השישי חייב להיות תיקו כדי שההכרעה תקבע במשחק האחרון. מתקיים תיקו אם ורק אם בלה מנצחת 3 משחקים מבין 6 המשחקים הראשונים. נגדיר את X להיות משתנה מקרי שסופר את כמות הנצחונות של בלה בששת המשחקים הראשונים ונקבל

$$\mathbb{P}(X=3) = \begin{pmatrix} 6 \\ 3 \end{pmatrix} 0.5^3(1-0.5)^3 = \begin{pmatrix} 6 \\ 3 \end{pmatrix} \frac{1}{2^6}.$$

שזה הערך של פונקציית ההסתברות הנקודתית של משתנה מקרי בינומי $X \sim Bin(6, \frac{1}{2})$.

2. בהנתן שבלה היא המנצחת המוחלטת, מה ההסתברות שהיא ניצחה בחמישה משחקים או פחות?

פתרון: נשים לב שההסתברות שבלה תהיה המצחת המוחלטת הוא 0.5 כיוון שהבעיה סימטרית. בתרגול ראינו שפונקצית ההסתברות הנקודתית של משתנה מקרי בדיד מותנה, נתונה על ידי

$$p_{(X|A)}(x) = \mathbf{1}_A(x)\frac{p_X(x)}{\mathbb{P}(A)}.$$

מספר הנצחונות של בלה X, הוא משתנה מקרי שמתפלג $Bin(7, \frac{1}{2})$ ולכן נשאר לנו לחשב את ההסתברות שלה לנצח 4 או 5 פעמים.

$$p_X(4) = \begin{pmatrix} 7 \\ 4 \end{pmatrix} 0.5^7, \qquad p_X(5) = \begin{pmatrix} 7 \\ 5 \end{pmatrix} 0.5^7$$

נחשב את ההסתברות למאורע

$$\mathbb{P}(X\leq 5|X\geq 4)=\frac{0.5^7\left(\left(\begin{array}{c}7\\4\end{array}\right)+\left(\begin{array}{c}7\\5\end{array}\right)\right)}{0.5}=\frac{\left(\begin{array}{c}8\\5\end{array}\right)}{2^6}$$

כאשר הזהות האחרון היא מכלל נסיגה של משולש פסקל.

3. בהנתן שאלי הוא המנצח המוחלט, מהי תוחלת מספר הזכיות שלו ומהי תוחלת מספר הזכיות של בלה?

פתרון: יהי X מספר הנצחונות של בלה ו־Y מספר הנצחונות של אלי. נשתמש בנוסחת התוחלת למשתנה מקרי בדיד שנתמך על הטבעיים

$$\begin{aligned}\mathbb{E}\left(X|Y\geq 4\right)&=\sum_{i=1}^{\infty}\mathbb{P}\left(X\geq i|Y\geq 4\right)\\&=\mathbb{P}\left(X\geq 1|Y\geq 4\right)+\mathbb{P}\left(X\geq 2|Y\geq 4\right)+\mathbb{P}\left(X\geq 3|Y\geq 4\right)\\&=1-\mathbb{P}(Y=7|Y\geq 4)+\mathbb{P}(Y\leq 5|Y\geq 4)+\mathbb{P}\left(Y=4|Y\geq 4\right)\\&=1+\frac{-1+\left(\begin{array}{c}8\\5\end{array}\right)+\left(\begin{array}{c}7\\4\end{array}\right)}{2^6}\end{aligned}$$

כאשר היינו צרכים לחשב רק את $\mathbb{P}(Y=7|Y\geq 4)$ בסעיף הזה. בנוסף בשיוויון השלישי השתמשנו בכך שמתקיים $X+Y=7$. עוד פעם נשתמש בכך שמתקיים $X+Y=7$ ולכן

$$7=\mathbb{E}\left(X+Y|Y\geq 4\right)=\mathbb{E}\left(X|Y\geq 4\right)+\mathbb{E}\left(Y|Y\geq 4\right).$$

מכך נסיק שמתקיים

$$\mathbb{E}\left(Y|Y\geq 4\right)=7-\left(1+\frac{-1+\left(\begin{array}{c}8\\5\end{array}\right)+\left(\begin{array}{c}7\\4\end{array}\right)}{2^6}\right)$$

.

שאלה 3

יהי X,Y וקטור מקרי המתפלג אחיד במשולש התחום על ידי הישרים $x+y=2$, $y=0, x=0$. נסמן ב־A וב־B את המאורעות $A=\{X<1,Y<1\}$ ו־$B=\{X+Y<1\}$.

1. יש לחשב את צפיפותו המותנית של Y בהנתן $X=a$ לכל $a\in[0,2]$.

פתרון: נסמן את המשולש ב־Δ. מתקיים מהנתון ש

$$f_{X,Y}(x,y)=\frac{\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in\Delta\}}(x,y)}{Area(\Delta)}=\frac{\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in\Delta\}}(x,y)}{2}$$

הצפיפות השולית של הוקטור המקרי לפי X נתונה על ידי

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \begin{cases} 0, & x < 0 \\ \int_0^{2-x} \frac{1}{2}dy, & 0 \le x \le 2 \\ 0, & x > 2 \end{cases}$$

כלומר נקבל שהצפיפות המותנית מוגדרת רק עבור נקודות הרציפות של פונקצית הצפיפות $a \in (0,2)$. מהנוסחה נקבל שהצפיפות המותנית של Y בהנתן $X = a$ היא

$$f_{Y|X=a}(y) := \frac{f_{X,Y}(a,y)}{f_X(a)} = \frac{\frac{\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in\Delta\}}(a,y)}{2}}{\int_0^{2-a}\frac{1}{2}dy} = \frac{\mathbf{1}_{\{y\in\mathbb{R}:0\le y\le 2-a\}}(y)}{2-a}.$$

דרך נוספת (ניסוח פורמאלי של דרך גאומטרית): עבור כל $a \in (0,2)$ הצפיפות השולית חיובית כיוון שהיא אינטגרל של פונקציה בעלת נקודת רציפות חיובית.

$$f_{Y|X=a}(y) := \frac{f_{X,Y}(a,y)}{f_X(a)} = \frac{\frac{\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in\Delta\}}(a,y)}{2}}{f_X(a)} = \frac{\mathbf{1}_{[0,2-a]}(y)}{2f_X(a)}$$

כלומר פונקצית הצפיפות קבועה על קטע ואפס מחוץ לקטע. נסיק שמתקיים

$$Y|X = a \sim Unif[0, 2-a].$$

הצפיפות של המשתנה הזה היא כמובן מה שקיבלנו בדרך הראשונה.

2. יש לחשב את הסתברותם של המאורעות A ו־B.

פתרון: ראינו שעבור תחומים ההסתברות של המאורע היא האינטגרל על התחום של הצפיפות המושתפת. האינטגרל של פונקציה על תחום הוגדר להיות הנפח בין הגרף של פונקצית למישור הx,y בתחום. במקרה שלנו התומך של פונקציית הצפיפות הוא Δ. ולכן מתקיים

$$\iint_A f_{X,Y}(x,y)dxdy = \iint_{A\cap\Delta} f_{X,Y}(x,y)dxdy$$

מתקיים $A \cap \Delta = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 1, 0 \le y \le 1\}$. השטח מעל הריבוע הזה והגרף של הפונקציה הקבועה חצי הוא כמובן $\frac{1}{2} \cdot 1$ לפי הנוסחה של שטחים מהתיכון. מכאן נקבל

$$\mathbb{P}(A) = \frac{1}{2}$$

נשים לב ש$B \cap \Delta \subset A \cap \Delta$ הוא משולש עם צלעות משותפות עם הריבוע $A \cap \Delta$ ולכן שטח המשולש הוא חצי ממנו. כלומר נקבל

$$\mathbb{P}(B) = \frac{1}{4}.$$

3. האם X ו־Y בלתי תלויים? האם בהנתן A הם בלתי־תלויים?

פתרון: עבור משתנים מקריים בלתי תלויים מתקיים $f_X(x)\cdot f_Y(y)=f_{X,Y}(x,y)$.
חישבנו את f_X בסעיף א' ונשים לב שחישוב ההסתברות השולית של Y סימטרי לחלוטין. נבחר נקודה במישור שבה השיוויון לא מתקיים

$$f_X(1.9)\cdot f_Y(1.9)\neq 0=f_{X,Y}(1.9,1.9)$$

ולכן הם לא בלתי תלויים.
דרך נוספת: חישבנו את הצפיפות המותנית $f_{Y|X=a}$, אם הם בלתי תלויים היה מתקיים $f_{Y|X=a}(t)=f_Y(t)$ לכל a שבה הצפיפות השולית לא מתאפסת. מכך אפשר להסיק שעבור $0<a_1<a_2<2$ מתקיים

$$\forall t\in\mathbb{R}:\ f_{Y|X=a_1}(t)=f_{Y|X=a_2}(t).$$

מסעיף א' זה לא נכון, לכן הם תלויים.
מנוסחה שראינו בכיתה מתקיים

$$\begin{aligned}f_{X,Y|A}(x,y)&=\frac{f_{X,Y}(x,y)\cdot\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in A\}}(x,y)}{\mathbb{P}(A)}\\&=\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in\Delta\}}\cdot\mathbf{1}_{\{(x,y)\in\mathbb{R}^2:(x,y)\in A\}}(x,y)\end{aligned}$$

כלומר מתקיים ש$X,Y|A$ מתפלג אחיד על הריבוע $\Delta\cap A$.
דרך א': קל לחשב את ההתפלגות השולית ולראות שמתקיים

$$f_{Y|A}(x)=f_{X|A}(x)=\begin{cases}0, & x<0\\ \int_0^1 dy & 0\le x\le 1\\ 0 & x>1\end{cases}$$

כלומר הם מתפלגים אחיד על הקטע $[0,1]$. מתקיים שהם בלתי תלויים כי

$$f_{Y|A}(y)\cdot f_{X|A}(x)=f_{X,Y|A}(x,y)$$

עבור כל $(x,y)\in\mathbb{R}^2$.
דרך ב' : מתקיים עבור כל $a\in(0,1)$ הצפיפות השולית $f_{X|A}(a)$ חיובית כיוון שהיא אינטגרל של פונקציה בעלת נקודת רציפות חיובית. לכן מתקיים

$$f_{(Y|A)|(X|A)=a}(t)=\frac{f_{X,Y|A}(a,t)}{f_{X|A}(a)}=\frac{\mathbf{1}_{[0,1]}(t)}{f_{X|A}(a)}.$$

כלומר $(Y|A)|(X|A)=a\sim Unif[0,1]$. מתכונת הנרמול חייב להתקיים $f_{X|A}(a)=1$ עבור כל $a\in(0,1)$. מסימטריה נסיק שמתקיים לכל a שבה הצפיפות השולית לא מתאפס

$$f_{(Y|A)|(X|A)=a}(t)=f_{Y|A}(t).$$

לכן הם בלתי תלוים.

שאלה 4

בטבלה בגודל $(n+1)\times(n+1)$ נצבע כל ריבוע בשחור או בלבן בהסתברות שווה ובאופן בלתי תלוי. שני ריבועים נקראים מחוברים אם הם סמוכים (אחד לצד השני או אחד מעל השני) וצבועים באותו הצבע. נסמן ב־$A_{i,j}$ את המארע שהריבוע במיקום ה־i,jבטבלה מחובר לשכנו מימין, הריבוע במקום ה־$i+1,j$. נסמן ב־$B_{i,j}$ את המאורע שהריבוע במקום i,j בטבלה מחובר לשכנו מלמטה, הריבוע במקום $i,j+1$. נסמן ב Z את מספר הזוגות המחוברים בסך הכל.

1. האם אוסף המאורעות $\{A_{i,j},B_{i,j}\}_{i,j\in[n]}$ בלתי־תלויים? האם הם בלתי־תלויים בזוגות?

פתרון: המאורעות תלווים כיוון ש־

$$\mathbb{P}\left(A_{11},B_{11},B_{12},A_{21}\right)=\frac{1}{2}^{3}$$

כי A_{21}תמיד קורה בהנתן שלושת המאורעות האחרים. מצד שנים מתקיים

$$\mathbb{P}(A_{ij})=\mathbb{P}(B_{ij})=\frac{1}{2}$$

ולכן מתקיים

$$\mathbb{P}\left(A_{11},B_{11},B_{12},A_{21}\right)\neq\mathbb{P}(A_{11})\cdot\mathbb{P}\left(B_{11}\right)\mathbb{P}(B_{12})\mathbb{P}\left(A_{21}\right)$$

נראה שהמאורעות הם בלתי תלויים בזוגות. נסתכל רק על זוגות מאורעות שלהן יש ריבוע משותף. זוגות אחרים הם בוודאי בלתי תלויים. יהי $C,D\in\{A_{i,j},B_{i,j}\}$ זוג מאורעות שיש להם ריבוע משותף. מתקיים שההסתברות של המאורע $C\cap D$ הוא ההסתברות ששלושת הריבועים שחורים או ששלושת הריבועים הם לבנים כלומר

$$\mathbb{P}(C,D)=\left(\frac{1}{2}\right)^{3}+\left(\frac{1}{2}\right)^{3}=\frac{1}{4}=\mathbb{P}(C)\mathbb{P}(D)$$

2. יש לבטא באמצעות n את התוחלת והשונות של Z_n.

פתרון: מתקיים ש$\mathbf{1}_{A_{i,j}},\mathbf{1}_{B_{i,j}}\sim ber(\frac{1}{2})$ ונשים לב שמתקיים

$$Z_n=\sum_{\substack{1\leq i\leq n\\1\leq j\leq n+1}}\mathbf{1}_{A_{i,j}}+\sum_{\substack{1\leq j\leq n\\1\leq i\leq n+1}}\mathbf{1}_{B_{i,j}}$$

מלינאריות התוחלת נקבל שמתקיים

$$\mathbb{E}(Z_n)=n(n+1)\mathbb{E}(\mathbf{1}_{A_{1,1}})+n(n+1)\mathbb{E}(\mathbf{1}_{B_{1,1}})=2n(n+1)\frac{1}{2}=n(n+1).$$

נחשב את השונות באותה הדרך כיוון שהמאורעות הם בלתי תלויים בזוגות ולכן ראינו בתרגול שפונקציות המציינות של המאורעות גם הן יהיו בלתי תלויות בזוגות. לכן מתקיים

$$Var(Z_n)=n(n+1)Var(\mathbf{1}_{A_{1,1}})+n(n+1)Var(\mathbf{1}_{B_{1,1}})=\frac{n(n+1)}{2}.$$

3. יש להראות כי לכל $\epsilon > 0$ מתקיים

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{Z_n - \mathbb{E}(Z_n)}{2n(n+1)}\right| > \epsilon\right) = 0.$$

פתרון: נזכר באי שיוויון צבישב

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

שמתקיים אם השונות והתוחלת של משתנה מקרי X קיימים, וגם $C > 0$. כלומר עבור כל $n \in \mathbb{N}$ מתקיים

$$\mathbb{P}\left(\left|\frac{Z_n - \mathbb{E}(Z_n)}{2n(n+1)}\right| > \epsilon\right) \leq \mathbb{P}\left(\left|\frac{Z_n - \mathbb{E}(Z_n)}{2n(n+1)}\right| \geq \epsilon\right) = \mathbb{P}\left(|Z_n - \mathbb{E}(Z_n)| \geq \epsilon 2n(n+1)\right)$$
$$\leq \frac{\frac{n(n+1)}{2}}{(\epsilon 2n(n+1))^2} \overset{n\to\infty}{\longrightarrow} 0.$$

שאלה 5

נתונות שתי השערות בנוגע להתפלגותו של משתנה מקרי X: השערת האפס היא שהוא מתפלג אחיד ב־$[0,10]$ וההשערה החלופית ־ שהוא מתפלג מעריכית עם פרמטר 0.2.

1. יש למצא מבחן מיטבי עם מובהקות 0.05 (הלא היא α) להפרדה בין ההשערות. מה עוצמתו?

פתרון: מהלמה של ניימן פירסון מבחן רף הוא מיטבי. פונקציית יחס הנראות היא

$$r(x) = \frac{\mathbf{1}_{[0,10]}(x)\frac{1}{10}}{\mathbf{1}_{[0,\infty)}(x)\frac{1}{5}e^{-\frac{x}{5}}}.$$

נחפש רף η מתאים בכדי לקבל את המובהקות הרצויה.

$$r(x) < \eta_\alpha.$$

עבור x בתומך של ההתפלגות אחידה מתקיים

$$\begin{aligned} & r(x) && < \eta_\alpha \\ \iff & e^{\frac{x}{5}}/2 && < \eta_\alpha \\ \iff & \frac{x}{5} && < \ln(2\eta_\alpha) \\ \iff & x && < 5\ln(2\eta_\alpha) \end{aligned}$$

לכן נפתור את המשוואה האינטגרלית

$$\mathbb{P}_{H_0}(X < 5\ln(2\eta_\alpha)) = 0.05.$$

נחשב את האינטגרל

$$\mathbb{P}_{H_0}(X < 5\ln(2\eta_\alpha)) = \int_0^{5\ln(2\eta_\alpha)} \frac{1}{10} = \frac{1}{2}\ln(2\eta_\alpha)$$

נשווה ל-α ונקבל שמבחן רף מתאים יהיה

$$\eta = e^{0.1}/2.$$

נבדוק את העוצמה של המבחן. כלומר נחשב את

$$\begin{aligned}\mathbb{P}_{H_1}(r(X) < \eta_\alpha) &= \mathbb{P}_{H_1}(X < 5\ln(2\eta_\alpha) \cup (10,\infty)) \\ &= \mathbb{P}_{H_1}(X < 0.5) + \mathbb{P}_{H_1}(10 < X) \\ &= \int_0^{0.5} \frac{1}{5}e^{-\frac{x}{5}}dx + \int_{10}^{\infty} \frac{1}{5}e^{-\frac{x}{5}} \\ &= 1 - e^{-0.1} + e^{-2}.\end{aligned}$$

2. במקרה זה מבחן מיטבי עם מובהקות 0 אינו מקבל תמיד את השערת האפס. מהו מבחן זה.

פתרון: כדי לקבל מבחן עם מובהקות 0 אנו צרכים למצוא אזור דחייה $R \subset \mathbb{R}$ כך שמתקיים $\mathbb{P}_{H_0}(X \in R) = 0$. אם מתקיים $R \cap [0,10] = \emptyset$ נקבל מובהקות 0, כי התומך של פונקצית הצפיפות של התפלגות לפי H_0 הוא $[0,10]$. כיוון שבמודל רציף הלמה של ניימן פירסון מבטיחה שמבחן הוא מיטבי אם ורק אם הוא מבחן רף, אז נחפש מבחן רף מתאים כדי להגדיר את R. בסעיף הקודם ראינו שאם $x \in [0,10]$ אז עבור מבחן עם רף η מתקיים ש-$x \in R$ (כלומר בתחום הדחייה) אם

$$e^{\frac{x}{5}}/2 < \eta$$

נשים לב שאם נבחר שעבור $\eta < \frac{1}{2}$ האי שיוויון לא יתקיים עבור כל איבר בתומך של ההתפלגות האחידה על $[0,10]$, כלומר $R \cap [0,10] = \emptyset$. לכן מתקיים במבחן רף עם $\eta < \frac{1}{2}$ שההסתברות לקבל טעות מסוג ראשון היא אפס. נראה שאפשר לבחור $\eta < \frac{1}{2}$ כך שיש הסתברות חיביות לקבל את ההשערה החלופית במידה והיא נכונה. עבור $x > 10$ מתקיים

$$r(x) = 0.$$

מכך נסיק שאם נחבר את הרף η להיות מספר בקטע $(0, \frac{1}{2})$ אז אם ההשערה החלופית מתקיימת אז יש הסתברות חיובית

$$\mathbb{P}_{H_1}(X > 10) = \int_{10}^{\infty} \frac{1}{5}e^{-\frac{x}{5}} = e^{-2} > 0$$

לקבל את ההשערה החלופית.

3. הנסיין ביצע דגימות חוזרות, בלתי תלויות ושוות התפלגות של של המשתנה המקרי. נסמן דגימות אלה ב־$\{X_i\}_{i\in\mathbb{N}}$. הוא הסתכל על

$$\frac{\sum_{i=1}^{n} X_i - 5n}{\sqrt{n}}$$

לאיזו התפלגות תשאפנה הדגימות לפי השערת האפס ולאיזו התפלגות לפי ההשערה החליפית?

פתרון: נשים לב שהתוחלת של ההתפלגויות היא 5. מתקיים שהשונות לפי השערת האפס היא $\frac{100}{12}$ ולפי ההשערה החליפית היא 25. ממשפט הגבול המרכזי מתקיים

$$\frac{\sum_{i=1}^{n} X_i - 5n}{SD(X_1)\sqrt{n}} \xrightarrow{n\to\infty} Z$$

כאשר ההתכנסות היא בהתפלגות ו־Z הוא משתנה מקרי נורמלי סטנדרטי. ראינו בכיתה שמתקיים $aZ \sim N(a\cdot 0, a^2\cdot 1)$ ולכן נסיק שהמשתנה המקרי שנקבל לפי השערת האפס מתפלג $N(0,\frac{100}{12})$ ולפי ההשערה החלפית יתפלג $N(0,25)$.