פתרון תרגיל בית 3 בהסתברות וסטטיסטיקה מתמטית
88-373 סמסטר ב' תשפ"א

תוחלת ושונות

תרגיל 1. יהי X משתנה מקרי עם פונקציית התפלגות מצטברת F_X. כזכור, F_X רציפה מימין בכל נקודה. הוכיחו: $P(X=a)=0$ אם ורק אם F_X רציפה ב-a.

הוכחה. אנחנו יודעים ש-F_X רציפה מימין ב-a. כדי להראות רציפות משמאל, ניעזר ברציפות ההסתברות לסדרת מאורעות עולה:

$$P(X=a)=P(X\le a)-P(X<a)=F_X(a)-P\left(\bigcup_{n=1}^{\infty}\left\{X\le a-\frac{1}{n}\right\}\right)=$$
$$=F_X(a)-\lim_{n\to\infty}P\left(X\le a-\frac{1}{n}\right)=F_X(a)-\lim_{x\to a^-}F_X(x).$$

לכן $P(X=a)$ אם ורק אם $F_X(a)=\lim_{x\to a^-}F_X(x)$, אם ורק אם F_X רציפה משמאל ב-a. $\square$

לפני שתענו על השאלה הבאה, קראו על מידת סטילטיס בקישורים שבעמוד הקורס.

תרגיל 2. יהי X משתנה מקרי עם פונקציית התפלגות מצטברת

$$F_X(t)=\begin{cases}\frac{1}{4}e^t, & t<0\\ \frac{1}{2}t^2+\frac{1}{4}, & 0\le t<1\\ 1, & t\ge 1\end{cases}.$$

חשבו את ההסתברויות ואת התוחלות הבאות:

א. $P(X\ge 1)$.

ב. $P(0\le X\le 1)$.

ג. $\mathbb{E}[X]$.

ד. $\mathbb{E}[X\mid X\le 0]$.

פתרון. נסמן על ידי μ_{F_X} את מידת סטילטיס המתאימה.

א. לפי רציפות ההסתברות

$$P(X\geq 1)=1-P(X<1)=1-P\left(\bigcup_{n=1}^{\infty}\left(X\leq 1-\frac{1}{n}\right)\right)=1-\lim_{n\to\infty}P\left(X\leq 1-\frac{1}{n}\right)=$$
$$=1-\lim_{n\to\infty}\left(\frac{1}{2}\left(1-\frac{1}{n}\right)^2+\frac{1}{4}\right)=1-\left(\frac{1}{2}+\frac{1}{4}\right)=\frac{1}{4}.$$

ב. נשים לב ש-F_X רציפה ב-0 ולכן $P(X=0)=0$. מכאן

$P(0\leq X\leq 1)=P(0<X\leq 1)=\mu_{F_X}((0,1])=F_X(1)-F_X(0)=1-\frac{1}{4}=\frac{3}{4}$.

ג. נחלק את $\mathbb{R}$ לקטעים: $A_1=(-\infty,0)$, $A_2=[0,1)$, $A_3=\{1\}$, $A_4=(1,\infty)$. נחשב את האינטגרל על כל קטע בנפרד. על A_4, F_X קבועה, בפרט רציפה בהחלט, ולכן

$$\int_{A_4}t\,\mathrm{d}F_X(t)=\int_{A_4}t\cdot 0\,\mathrm{d}F_X(t)=0.$$

על A_3,

$$\int_{A_3}t\,\mathrm{d}F_X(t)=P(X=1)=\frac{1}{4}.$$

על A_2, $F_X|_{A_2}(t)=\frac{1}{2}t^2+\frac{1}{4}$ רציפה בהחלט ואינטגרבילית רימן, ולכן

$$\int_{A_2}t\,\mathrm{d}F_X(t)=\int_0^1 t\cdot t\,\mathrm{d}t=\left(\frac{t^3}{3}\right)_0^1=\frac{1}{3}.$$

על A_1, $F_X|_{A_1}(t)=\frac{1}{4}e^t$ רציפה בהחלט ואינטגרבילית רימן, ולכן

$$\int_{A_1}t\,\mathrm{d}F_X(t)=\int_{-\infty}^0\frac{te^t}{4}\,\mathrm{d}t=\left(\frac{te^t}{4}\right)_{-\infty}^0-\int_{-\infty}^0\frac{e^t}{4}\,\mathrm{d}t=-\left(\frac{e^t}{4}\right)_{-\infty}^0=-\frac{1}{4}.$$

בסך הכל התוחלת היא $\frac{1}{4}+\frac{1}{3}-\frac{1}{4}=\frac{1}{3}$.

ד. נזכור כי

$$\mathbb{E}[X\mid X\leq 0]=\frac{\mathbb{E}\left[X\cdot 1_{\{X\leq 0\}}\right]}{P(X\leq 0)}.$$

F_X רציפה ב-0, לכן $P(X\leq 0)=\int_{-\infty}^0\frac{e^t}{4}\,\mathrm{d}t=\frac{1}{4}$. מצד שני,

$$\mathbb{E}\left[X\cdot 1_{\{X\leq 0\}}\right]=\int_{(-\infty,0]}t\,\mathrm{d}F_X(t)=-\frac{1}{4}$$

ולכן $\mathbb{E}[X\mid X\leq 0]=-1$.

תזכורת. פונקציה $f:\mathbb{R}\to\mathbb{R}$ נקראת **קמורה** (convex), אם לכל $x,y\in\mathbb{R}$ ו-$0\leq\lambda\leq1$ מתקיים

$$f\left(\left(1-\lambda\right)x+\lambda y\right)\leq\left(1-\lambda\right)f\left(x\right)+\lambda f\left(y\right).$$

מבחינה גרפית: המשיק לפונקציה תמיד נמצא מתחת לפונקציה. אם f גזירה פעמיים ו-$f''\left(x\right)\geq0$, אז f היא קמורה.

תרגיל 3. יהי $\left(\Omega,\mathcal{F},P\right)$ מרחב הסתברות, יהי $X:\Omega\to\mathbb{R}$ משתנה מקרי, ותהי $f:\mathbb{R}\to\mathbb{R}$ פונקציה קמורה. בשאלה זו נוכיח את אי-שוויון ינסן.

א. הוכיחו שלכל $x\in\mathbb{R}$ קיימות הנגזרות החד-צדדיות $f'_{+}\left(x\right)=\lim_{h\to0^{+}}\frac{f\left(x+h\right)-f\left(x\right)}{h}$ ו-$f'_{-}\left(x\right)=\lim_{h\to0^{-}}\frac{f\left(x+h\right)-f\left(x\right)}{h}$.

ב. הוכיחו שלכל $x\in\mathbb{R}$ מתקיים $f'_{-}\left(x\right)\leq f'_{+}\left(x\right)$.

ג. יהי $t\in\mathbb{R}$. הראו שקיימים $a,b\in\mathbb{R}$ שעבורם $f\left(x\right)\geq ax+b$ לכל $x\in\mathbb{R}$ וגם $f\left(t\right)=at+b$.

ד. הסיקו את אי-שוויון ינסן: $f\left(\mathbb{E}\left[X\right]\right)\leq\mathbb{E}\left[f\left(X\right)\right]$.

הוכחה.

א. יהיו $x_1<x_2<x_3$. אפשר לכתוב $x_2=\frac{x_3-x_2}{x_3-x_1}x_1+\frac{x_2-x_1}{x_3-x_1}x_3$, ולפי הקמירות לקבל

$$f\left(x_2\right)\leq\frac{x_3-x_2}{x_3-x_1}f\left(x_1\right)+\frac{x_2-x_1}{x_3-x_1}f\left(x_3\right).$$

אם מסדרים את הכל מחדש, מקבלים

$$\frac{f\left(x_2\right)-f\left(x_1\right)}{x_2-x_1}\leq\frac{f\left(x_3\right)-f\left(x_2\right)}{x_3-x_2}$$

לכל $x_1<x_2<x_3$.

כעת נוכיח את קיום הנגזרות. יהי $x\in\mathbb{R}$, ונגדיר $g:\mathbb{R}\setminus\{0\}\to\mathbb{R}$ לפי $g\left(h\right)=\frac{f\left(x+h\right)-f\left(x\right)}{h}$. אי-השוויון שהוכחנו מראה ש-g עולה בקטעים $\left(0,\infty\right)$ ו-$\left(-\infty,0\right)$ בנפרד. בנוסף, לכל $h,h'>0$ מתקיים

$$\frac{f\left(x\right)-f\left(x-h'\right)}{h'}=g\left(-h'\right)\leq g\left(h\right)=\frac{f\left(x+h\right)-f\left(x\right)}{h}.$$

כיוון ש-g עולה ב-$\left(-\infty,0\right)$ וחסומה בו (למשל על ידי $g\left(1\right)$), קיים הגבול $\lim_{h'\to0^{-}}g\left(h'\right)=f'_{-}\left(x\right)$. בדומה קיים הגבול $\lim_{h\to0^{+}}g\left(h\right)=f'_{+}\left(x\right)$.

ב. נובע ישירות מאי-השוויון שהוכחנו.

ג. יהי $a=f'_{-}\left(t\right)$, ויהי $b=f\left(t\right)-f'_{-}\left(t\right)\cdot t$. נרצה להוכיח כי

$$f\left(x\right)\geq f'_{-}\left(t\right)\left(x-t\right)+f\left(t\right)=at+b$$

לכל $x\in\mathbb{R}$ (את השוויון ב-t מקבלים אוטומטית). אם $x>t$, זה נכון כי

$$f'_{-}\left(t\right)\leq f'_{+}\left(t\right)\leq\frac{f\left(x\right)-f\left(t\right)}{x-t}$$

ואם $x < t$ זה נכון כי

$$.f'_{-}(t) \geq \frac{f(t) - f(x)}{t - x}$$

ד. נסמן $t = \mathbb{E}[X]$, וניקח $a, b \in \mathbb{R}$ שמתאימים לתנאי הסעיף הקודם. אז

$$.\mathbb{E}[f(X)] \geq \mathbb{E}[aX + b] = a\mathbb{E}[X] + b = at + b = f(\mathbb{E}[X])$$

$\square$

ריכוז מידה

תרגיל 4. הראו כי אי-שוויון מרקוב ואי-שוויון צ'בישב הם הדוקים. כלומר, לכל $a \in \mathbb{R}$:

א. מצאו משתנה מקרי אי-שלילי X כך ש-$P(X \geq a) = \frac{\mathbb{E}[X]}{a}$.

ב. מצאו משתנה מקרי X עם שונות סופית כך ש-$P(|X - \mathbb{E}[X]| \geq a) = \frac{\mathrm{Var}(X)}{a^2}$.

(אם המשתנים שלכם מקבלים יותר משלושה ערכים, אפשר למצוא דוגמה יותר פשוטה.)

הוכחה. עבור אי-שוויון מרקוב ניקח משתנה מהצורה $X = \begin{cases} a, & p \\ 0, & 1-p \end{cases}$. במקרה הזה $\mathbb{E}[X] = ap$ וכן $P(X \geq a) = p = \frac{\mathbb{E}[X]}{a}$.

עבור אי-שוויון צ'בישב ניקח $X = \begin{cases} \mu - a, & p \\ \mu, & 1 - 2p \\ \mu + a, & p \end{cases}$. במקרה הזה $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = 2pa^2$, וכן

$$.P(|X - \mu| \geq a) = 2p = \frac{\mathrm{Var}(X)}{a^2}$$

$\square$

תרגיל 5. הוכיחו את אי-שוויון קנטלי: אם X משתנה מקרי בעל שונות סופית, אז לכל $\lambda > 0$,

$$.P(X - \mathbb{E}[X] \geq \lambda) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X) + \lambda^2}$$

(רמז: $X - \mathbb{E}[X] \geq \lambda$ אם ורק אם $X - \mathbb{E}[X] + t \geq \lambda + t$ לכל $t \in \mathbb{R}$.)

הוכחה. נסמן $Y = X - \mathbb{E}[X]$. כעת

$$P(X - \mathbb{E}[X] \geq \lambda) = P(Y \geq \lambda) = P(Y + u \geq \lambda + u) \leq P\left((Y + u)^2 \geq (\lambda + u)^2\right) \leq$$
$$\leq \frac{\mathbb{E}\left[(Y + u)^2\right]}{(\lambda + u)^2} = \frac{\sigma^2 + u^2}{(\lambda + u)^2}.$$

נחפש את המינימום של הביטוי שקיבלנו כתלות ב-u. אם נגזור נקבל

$$\frac{2u\left(\lambda+u\right)^2-2\left(\lambda+u\right)\left(\sigma^2+u^2\right)}{\left(\lambda+u\right)^4}=\frac{2u\left(\lambda+u\right)-2\left(\sigma^2+u^2\right)}{\left(\lambda+u\right)^3}=\frac{2\lambda u-2\sigma^2}{\left(\lambda+u\right)^3}.$$

לכן יש קיצון ב-$u=\frac{\sigma^2}{\lambda}$. אפשר לברר שזהו מינימום, ולכן

$$P\left(X-\mathbb{E}\left[X\right]\geq\lambda\right)\leq\frac{\sigma^2+\frac{\sigma^4}{\lambda^2}}{\left(\lambda+\frac{\sigma^2}{\lambda}\right)^2}=\frac{\sigma^2\left(\sigma^2+\lambda^2\right)}{\left(\sigma^2+\lambda^2\right)^2}=\frac{\sigma^2}{\sigma^2+\lambda^2}.$$

□

תרגיל 6 (בעיית אוסף הקופונים)**.** בכל יום גדי מגריל ספר מהספרייה של המחלקה, שכידוע מכילה בדיוק n ספרים. כיוון שהוא כל כך אוהב את כל תחומי המתמטיקה ועוד לא בחר תחום התמחות, הוא בוחר בכל יום ספר באופן אחיד וקורא אותו. לפעמים יוצא שהוא בוחר ספר שהוא כבר קרא, אבל הוא קורא אותו שוב, למה לא.

א. הראו שתוחלת כמות הימים עד שגדי יקרא את כל הספרים בספרייה היא בקירוב טוב $n\log n$.

ב. הראו שאם X היא כמות הימים עד שהוא יקרא את כל הספרים, מתקיים

$$P\left(\left|X-\mathbb{E}\left[X\right]\right|\geq\alpha n\right)\leq\frac{c}{\alpha^2}.$$

הוכחה. נסמן על ידי X את כמות הימים עד שגדי יקרא את כל הספרים.

א. נסמן על ידי X_i את כמות הימים עד שגדי יקרא i ספרים ($X_0=0$). אפשר לראות כי $X_i-X_{i-1}\sim\mathrm{Geo}\left(\frac{n-(i-1)}{n}\right)$ (אם הוא קרא $i-1$ ספרים, יש $n-(i-1)$ ספרים שהוא עוד לא קרא, ובכל יום הוא בוחר באופן אחיד אחד מהם). לכן

$$\mathbb{E}\left[X\right]=\mathbb{E}\left[X_n\right]=\mathbb{E}\left[\sum_{i=1}^{n}\left(X_i-X_{i-1}\right)\right]=\sum_{i=1}^{n}\frac{n}{n-(i-1)}=n\sum_{i=1}^{n}\frac{1}{i}.$$

אפשר לקרב את הסכום ההרמוני $\sum_{i=1}^{n}\frac{1}{i}\approx\log n+\gamma$ כאשר γ קבוע (קבוע אוילר-מסקרוני). לכן

$$\mathbb{E}\left[X\right]\approx n\log n+\gamma n.$$

ב. נחשב את השונות של X. המשתנים $X_1,X_2-X_1,\ldots,X_n-X_{n-1}$ הם בלתי-תלויים; לכן

$$\mathrm{Var}\left(X\right)=\mathrm{Var}\left(\sum_{i=1}^{n}\left(X_i-X_{i-1}\right)\right)=\sum_{i=1}^{n}\mathrm{Var}\left(X_i-X_{i-1}\right)=\sum_{i=1}^{n}\frac{1-\frac{n-(i-1)}{n}}{\left(\frac{n-(i-1)}{n}\right)^2}=$$

$$=\sum_{i=1}^{n}\frac{n^2}{\left(n-(i-1)\right)^2}-\sum_{i=1}^{n}\frac{n}{n-(i-1)}=n^2\sum_{i=1}^{n}\frac{1}{i^2}-n\sum_{i=1}^{n}\frac{1}{i}\leq\frac{\pi^2n^2}{6}$$

ולפי אי-שוויון צ׳בישב

$$.P\left(\left|X-\mathbb{E}\left[X\right]\right|\geq\alpha n\right)\leq\frac{\frac{\pi^{2}n^{2}}{6}}{\alpha^{2}n^{2}}=\frac{\pi^{2}}{6\alpha^{2}}$$

$\square$

בהצלחה וחג שמח!