מבוא להסתברות

© ארזים

20 במאי 2016

משפט 0.1 (משפט השונות השלמה)
יהיו X,Y משתנים מקריים בעלי תוחלת ושונות סופית. אזי

$$V\left(X\right)=\mathbb{E}\left(V\left(X\mid Y\right)\right)+V\left(\mathbb{E}\left(X\mid Y\right)\right)$$

הוכחה: זאת משום שמתקיים

$$\begin{aligned}V\left(X\right) &= \mathbb{E}\left(X^{2}\right)-\left(\mathbb{E}\left(X\right)\right)^{2}=\mathbb{E}\left(\mathbb{E}\left(X^{2}\mid Y\right)\right)-\left(\mathbb{E}\left(\mathbb{E}\left(X\mid Y\right)\right)\right)^{2}=\\ &= \mathbb{E}\left(V\left(X\mid Y\right)+\left(\mathbb{E}\left(X\mid Y\right)\right)^{2}\right)-\left(\mathbb{E}\left(\mathbb{E}\left(X\mid Y\right)\right)\right)^{2}=\\ &= \mathbb{E}\left(V\left(X\mid Y\right)\right)+\mathbb{E}\left(\left(\mathbb{E}\left(X\mid Y\right)\right)^{2}\right)-\left(\mathbb{E}\left(\mathbb{E}\left(X\mid Y\right)\right)\right)^{2}=\\ &= \mathbb{E}\left(V\left(X\mid Y\right)\right)+V\left(\mathbb{E}\left(X\mid Y\right)\right)\end{aligned}$$

■

תרגיל יהי $n\geq1$. נגדיר T_{n} - מספר ההטלות של מטבע הוגן עד לקבלת רצף של n עצים.

1. מה היא $\mathbb{E}\left(T_{n}\mid T_{n-1}\right)$ כפונקציה של T_{n-1}?
2. $\mathbb{E}\left(T_{n}\right)$.

פתרון

1. יש שתי אפשרויות: לאחר שמגיעים אל T_{n-1}, אם יוצא עץ, אז $T_{n}=T_{n-1}+1$, אחרת $T_{n}=T_{n-1}+1+T_{n}'$, כאשר $T_{n}'\sim T_{n}$. כעת,

$$T_{n}\mid T_{n-1}\sim\begin{cases}T_{n-1}+1\mid T_{n-1} & \frac{1}{2}\\ T_{n-1}+1+T_{n}'\mid T_{n-1} & \frac{1}{2}\end{cases}$$

לכן, נקבל

$$\mathbb{E}\left(T_{n}\mid T_{n-1}\right)=\frac{1}{2}\mathbb{E}\left(T_{n-1}+1\mid T_{n-1}\right)+\frac{1}{2}\mathbb{E}\left[T_{n-1}+1+T_{n}'\mid T_{n-1}\right]$$

מתקיים כי T_n' בלתי תלוי במשתנה T_{n-1}, ולכן השוויון הקודם מתפשט לכדי

$$=\frac{1}{2}\left(T_{n-1}+1\right)+\frac{1}{2}\left(T_{n-1}+1\right)+\frac{1}{2}\mathbb{E}\left(T_n\right)=T_{n-1}+1+\frac{1}{2}\mathbb{E}\left(T_{n-1}\right)$$

.2

$$\begin{aligned}
\mathbb{E}\left(T_n\right) &= \mathbb{E}\left(\mathbb{E}\left(T_n \mid T_{n-1}\right)\right)=\mathbb{E}\left(T_{n-1}+1+\frac{1}{2}\mathbb{E}\left(T_n\right)\right)\\
\frac{1}{2}\mathbb{E}\left(T_n\right) &= \mathbb{E}\left(T_{n-1}\right)+1\\
\mathbb{E}\left(T_n\right) &= 2\mathbb{E}\left(T_{n-1}\right)+2=2\left(2\mathbb{E}\left(T_{n-2}\right)+2\right)+2=\\
&= 4\mathbb{E}\left(T_{n-2}\right)+4+2=\cdots=2^{n-1}\mathbb{E}\left(T_1\right)+\sum_{i=1}^{n-1}2^i=\\
&= 2^n+\sum_{i=1}^{n-1}2^i=\sum_{i=1}^{n}2^i=2^{n+1}-2
\end{aligned}$$

1 אי שוויונות

1.1 מרקוב

יהי X משתנה מקרי אי־שלילי ויהי $\alpha>0$, אזי

$$\mathbb{P}\left(X\geq\alpha\right)\leq\frac{\mathbb{E}\left(X\right)}{\alpha}$$

1.2 צ'בישב

יהי X משתנה מקרי בעל שונות סופית ויהי $\alpha>0$, אזי

$$\mathbb{P}\left(\left|X-\mathbb{E}X\right|\geq\alpha\right)\leq\frac{V\left(X\right)}{\alpha^2}$$

1.2.1 חד צדדי

אי שוויונות אלה חשובים כאשר ההתפלגות מסובכת או כשמחפשים גבול.

תרגיל יהי X משתנה מקרי ונתון כי $\mathbb{P}\left(0\leq X\leq 1\right)=1$, $\mathbb{E}\left(X\right)=\frac{1}{2}$. נגדיר $p=\mathbb{P}\left(X>\frac{1}{4}\right)$. מה הערך המינימלי והמקסימלי של p?

1. מקסימום 1, מינימום 0

2. $p=\frac{3}{4}$
3. מקסימום 1,מינימום $\frac{1}{3}$
4. אף תשובה

פתרון אם $X\equiv\frac{1}{2}$, אזי $p=1$. לא ייתכן מצב בו $p=0$, כי אז $X\le\frac{1}{4}$ בהסתברות 1, ולכן $\mathbb{E}(X)\le\frac{1}{4}$, בסתירה. נניח כי

$$X\sim\begin{cases}\frac{1}{4} & \frac{2}{3}\\ 1 & \frac{1}{3}\end{cases}$$

במצב זה מתקבל $\mathbb{E}(X)=\frac{1}{4}\cdot\frac{2}{3}+1\cdot\frac{1}{3}=\frac{1}{2}$. נוכיח בעזרת מרקוב כי $p\ge\frac{1}{3}$ ונסיים:

$$\begin{aligned}p &= \mathbb{P}\left(X>\frac{1}{4}\right)=1-\mathbb{P}\left(X\le\frac{1}{4}\right)=1-\mathbb{P}\left(1-\frac{1}{4}\le 1-X\right)\ge 1-\frac{\mathbb{E}(1-X)}{\frac{3}{4}}=\\ &= 1-\frac{2}{3}=\frac{1}{3}\end{aligned}$$

תרגיל (מפברואר 2015) יהי $n\ge 2$. תהי σ פרמטוציה על $\{1,\ldots,n\}$ שנבחרה באופן אחיד מבין כל האפשרויות. נגדיר $I=|\{(i,j)\mid i<j,\sigma(i)>\sigma(j)\}|$, כלומר מספר ההחלפות בסדר שהפרמוטציה מייצרת.

1. חשבו את $\mathbb{E}(I)$
2. חשבו את $V(I)$
3. הוכיחו כי לכל $\varepsilon>0$, מתקיים

$$\lim_{n\to\infty}\mathbb{P}\left(\left|I-\frac{n^2}{4}\right|>n^{\frac{3}{2}+\varepsilon}\right)=0$$

פתרון

1. נגדיר $X_{i,j}$ לכל $i<j$ להיות האינדיקטור

$$X_{i,j}=\begin{cases}1 & \sigma(i)>\sigma(j)\\ 0 & \text{otherwise}\end{cases}$$

ההתפלגות שלהם היא

$$X_{i,j}\sim\text{Ber}\left(\frac{1}{2}\right)$$

כעת מתקיים

$$\mathbb{E}(I)=\sum_{i<j}\mathbb{E}(X_{i,j})=\frac{1}{2}\cdot\binom{n}{2}=\frac{n(n-1)}{4}=\frac{n^2-n}{4}$$

.2

$$\begin{aligned} V(I) &= V\left(\sum_{i<j} X_{i,j}\right) = \sum_{i<j} V(X_{i,j}) + \sum_{\substack{i<j \\ \neq \\ k<l}} \operatorname{cov}(X_{i,j}, X_{k,l}) = \\ &= \frac{1}{4}\cdot\binom{n}{2} + \sum_{i,j,k,l} \operatorname{cov}(X_{i,j}, X_{k,l}) \end{aligned}$$

אם האינדקסים שונים לחלוטין אזי $X_{i,j}, X_{k,l}$ בלתי תלויים, לכן $\operatorname{cov}(X_{i,j}, X_{k,l}) = 0$. לכן צריך לבחור רק מקרה שבו יש אינקס אחד משותף. במילים אחרות, נסתכל על שלשות של אינדקסים $i<j<k$, ונתבונן במצבים

$$\operatorname{cov}(X_{i,j}, X_{i,k}), \operatorname{cov}(X_{i,j}, X_{j,k}), \operatorname{cov}(X_{i,k}, X_{j,k})$$

נחשב

$$\begin{aligned} \operatorname{cov}(X_{i,j}, X_{i,k}) &= \mathbb{E}(X_{i,j}\cdot X_{i,k}) - \mathbb{E}(X_{i,j})\cdot\mathbb{E}(X_{i,k}) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \\ \operatorname{cov}(X_{i,j}, X_{j,k}) &= \frac{1}{6} - \frac{1}{4} = -\frac{1}{12} \\ \operatorname{cov}(X_{i,k}, X_{j,k}) &= \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \end{aligned}$$

ולכן בסך הכל

$$V(I) = \frac{1}{4}\cdot\binom{n}{2} + 2\cdot\binom{n}{3}\cdot\left(\frac{1}{12} + \frac{1}{12} - \frac{1}{12}\right) = \frac{n^2-n}{8} + \frac{1}{6}\cdot\frac{n(n-1)(n-2)}{3!}$$

.3

$$\mathbb{P}\left(|I - \mathbb{E}I| \geq n^{\frac{3}{2}+\varepsilon}\right) \leq \frac{V(I)}{\left(n^{\frac{3}{2}+\varepsilon}\right)^2} \approx \frac{n^3}{n^{3+2\varepsilon}} \xrightarrow[n\to\infty]{} 0$$

כעת, מתקיים

$$\begin{array}{ccc} \left|I - \dfrac{n^2-n}{4}\right| \geq n^{\frac{3}{2}+\varepsilon} & \Rightarrow & \begin{cases} I - \frac{n^2-n}{4} & \geq n^{\frac{3}{2}+\varepsilon} \\ I - \frac{n^2-n}{4} & \leq -n^{\frac{3}{2}+\varepsilon} \end{cases} \\ & & \Downarrow \\ \left|I - \dfrac{n^2}{4}\right| \geq n^{\frac{3}{2}+\frac{\varepsilon}{2}} & \Leftarrow & \begin{cases} I - \frac{n^2}{4} & \geq n^{\frac{3}{2}+\varepsilon} - \frac{n}{4} \geq n^{\frac{3}{2}+\frac{\varepsilon}{2}} \\ I - \frac{n^2}{4} & \leq -n^{\frac{3}{2}+\varepsilon} - \frac{n}{4} \leq -n^{\frac{3}{2}+\frac{\varepsilon}{2}} \end{cases} \end{array}$$

ולכן עבור n גדול מספיק מתקיים הגבול שרצינו להראות.