מבוא להסתברות

ⓒ ארזים

18 במאי 2016

1 אי שוויון מרקוב

טענה 1.1 אם X משתנה מקרי אי־שלילי אזי לכל $t > 0$,

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

הוכחה:

$$\mathbb{E}X \geq \mathbb{E}\left[X \cdot 1_{(X \geq t)}\right] \geq \mathbb{E}\left[t \cdot 1_{(X \geq t)}\right] = t \cdot \mathbb{E}\left[1_{(X \geq t)}\right] = t \cdot \mathbb{P}\left(X \geq t\right)$$

■

יש מקרים שבהם אי־השוויון הוא שוויון.

דוגמא

$$\begin{aligned} X &\sim \left\{\begin{cases} 0 & 0.99 \\ 100 & 0.01 \end{cases}\right\} \\ \mathbb{E}X &= 1 \\ \mathbb{P}\left(X \geq 100\right) &= \frac{1}{100} \end{aligned}$$

2 אי שוויון צ'בישב

טענה 2.1 יהי X משתנה מקרי בעלת תוחלת ושונות סופיות. אזי לכל $t > 0$,

$$\mathbb{P}\left(\left|X - \mathbb{E}X\right| \leq t\right) \leq \frac{\mathrm{Var}\left(X\right)}{t^2}$$

הוכחה: נגדיר משתנה מקרי חדש

$$Y = \left(X - \mathbb{E}X\right)^2$$

Y הוא אי־שלילי, ולכן נפעיל את אי־שוויון מרקוב עבור Y: לכל $t>0$,

$$\mathbb{P}\left(|X-\mathbb{E}X| \geq t\right)=\mathbb{P}\left(Y \geq t^{2}\right) \leq \frac{\mathbb{E}Y}{t^{2}}=\frac{\mathbb{E}\left((X-\mathbb{E}X)^{2}\right)}{t^{2}}=\frac{\operatorname{Var}(X)}{t^{2}}$$

■

יש אי־שוויונות נוספים שניתן להוכיח תחת הנחות מסויימות, למשל אם ידוע כי $\mathbb{E}\left(e^{X}\right) \leq \infty$, אז

$$\mathbb{P}(X \geq t)=\mathbb{P}\left(e^{X} \geq e^{t}\right) \leq \frac{\mathbb{E}\left(e^{X}\right)}{e^{t}}$$

דוגמא מטילים מטבע עם הסתברות $\frac{1}{3}$ לעץ n פעמים באופן בלתי־תלוי. נסמן בתור M את אורך הרצף הכי ארוך של עצים בסדרת ההטלות. מטרת השאלה: $M \approx \log_{3} n$ בסיכוי גבוה.

נגדיר משתנים מקריים N_k להיות כמות הרצפים באורך k של עצים בסדרת ההטלות, לכל $1 \leq k \leq n$, עם חפיפות. נבחין כי מתקיים:

1. חשבו את $\mathbb{E}N_k$.
נרשום

$$N_{k}=\sum_{i=1}^{n-k+1} I_{i}$$

כאשר I_i הוא האינדיקטור של המאורע בו יש רצף של k עצים שמתחיל במקום i. מלינאריות התוחלת,

$$\mathbb{E}N_{k}=\sum_{i=1}^{n-k+1} \mathbb{E}\left(I_{i}\right)=\frac{n-k+1}{3^{k}}$$

2. חשבו את $\operatorname{Var}(N_k)$.
ידוע לנו כי

$$\operatorname{Var}\left(N_{k}\right)=\sum_{i=1}^{n-k+1} \operatorname{Var}\left(I_{i}\right)+2 \sum_{1 \leq i<j \leq n-k+1} \operatorname{Cov}\left(I_{i}, I_{j}\right)$$

ידוע לנו, משונות של משתנה ברנולי, כי $\operatorname{Var}(I_i)=3^{-k}\left(1-3^{-k}\right)=\frac{3^k-1}{3^{2k}}$. כמו כן, אם $j-i \geq k$ אז I_i, I_j בלתי תלויים, ובפרט בלתי־מתואמים. כעת, נניח כי החפיפה בין הרצפים היא בגודל $k-m$, עבור $1 \leq m < k$. אזי=

$$\begin{aligned}
\operatorname{Cov}\left(I_{i}, I_{j}\right) &= \mathbb{E}\left(I_{i} \cdot I_{j}\right)-\left(\mathbb{E}I_{i}\right) \cdot\left(\mathbb{E}I_{j}\right)=3^{-(k+m)}-3^{-2k}=3^{-k}\left(3^{-m}-3^{-k}\right)= \\
&= \frac{\frac{1}{3^{m}}-\frac{1}{3^{k}}}{3^{k}}=\frac{3^{k}-3^{m}}{3^{2k+m}}
\end{aligned}$$

בסך הכל, נסיק כי

$$\mathrm{Var}\left(N_k\right) = \frac{\left(n-k+1\right)\left(3^k-1\right)}{3^{2k}} + 2\sum_{m=1}^{k-1}\frac{\left(n-k-m+1\right)\left(3^k-3^m\right)}{3^{2k+m}}$$

חסם עליון על השונות:

$$\mathrm{Var}\left(N_k\right) \le n\cdot 3^{-k} + 2kn\cdot 3^{-k} = \left(2k+1\right)n\cdot 3^{-k}$$

3. יהי $\varepsilon > 0$. הוכיחו כי

$$\mathbb{P}\left(M \ge \left(1+\varepsilon\right)\log_3 n\right) \xrightarrow[n\to\infty]{} 0$$

נשים לב כי

$$\left\{M \ge \left(1+\varepsilon\right)\log_3 n\right\} = \left\{M \ge \left\lceil\left(1+\varepsilon\right)\log_3 n\right\rceil\right\} = \left\{N_{\left\lceil\left(1+\varepsilon\right)\log_3 n\right\rceil} \ge 1\right\}$$

לכן, מאי שוויון מרקוב,

$$\begin{aligned}\mathbb{P}\left(M \ge \left(1+\varepsilon\right)\log_3 n\right) &= \mathbb{P}\left(N_{\left\lceil\left(1+\varepsilon\right)\log_3 n\right\rceil} \ge 1\right) \le \mathbb{E}\left(N_{\left\lceil\left(1+\varepsilon\right)\log_3 n\right\rceil}\right) = \\ &= \frac{n-\left\lceil\left(1+\varepsilon\right)\log_3 n\right\rceil+1}{3^{\left\lceil\left(1+\varepsilon\right)\log_3 n\right\rceil}} \le \\ &\le \frac{n}{3^{\left(1+\varepsilon\right)\log_3 n}} = \frac{n}{n^{1+\varepsilon}} = \frac{1}{n^{\varepsilon}} \xrightarrow[n\to\infty]{} 0\end{aligned}$$

4. יהי $\varepsilon > 0$. הוכיחו כי

$$\mathbb{P}\left(M \ge \left(1-\varepsilon\right)\log_3 n\right) \xrightarrow[n\to\infty]{} 1$$

כמו קודם,

$$\left\{M \ge \left(1-\varepsilon\right)\log_3 n\right\} = \left\{M \ge \left\lceil\left(1-\varepsilon\right)\log_3 n\right\rceil\right\} = \left\{N_{\left\lceil\left(1-\varepsilon\right)\log_3 n\right\rceil} \ge 1\right\}$$

נשים לב כי לכל משתנה מקרי X עם תוחלת ושונות סופיות שמקבל ערכים שלמים, נובע מאי שוויון צ'בישב

$$\begin{aligned}\mathbb{P}\left(X \ge 1\right) &= 1-\mathbb{P}\left(X \le 0\right) = 1-\mathbb{P}\left(X-\mathbb{E}X \le -\mathbb{E}X\right) \ge 1-\mathbb{P}\left(\left|X-\mathbb{E}X\right| \ge \mathbb{E}X\right) \ge \\ &\ge 1-\frac{\mathrm{Var}\left(X\right)}{\left(\mathbb{E}X\right)^2}\end{aligned}$$

במקרה שלנו, נקבל

$$\mathbb{P}\left(M \geq (1-\varepsilon)\log_3 n\right) = \mathbb{P}\left(N_{\lceil (1-\varepsilon)\log_3 n\rceil} \geq 1\right) \geq 1 - \frac{\operatorname{Var}\left(N_{\lceil (1-\varepsilon)\log_3 n\rceil}\right)}{\left(\mathbb{E}\left(N_{\lceil (1-\varepsilon)\log_3 n\rceil}\right)\right)^2} \xrightarrow[n\to\infty]{?} 1$$

צריך להוכיח כי

$$\frac{\operatorname{Var}\left(N_{\lceil (1-\varepsilon)\log_3 n\rceil}\right)}{\left(\mathbb{E}\left(N_{\lceil (1-\varepsilon)\log_3 n\rceil}\right)\right)^2} \xrightarrow[n\to\infty]{} 0$$

לפי החישובים שערכנו, כאשר נסמן $k = \lceil (1-\varepsilon)\log_3 n\rceil$,

$$\frac{\operatorname{Var}\left(N_k\right)}{\left(\mathbb{E}N_k\right)^2} \leq \frac{(2k+1)\, n\cdot 3^{-k}}{\left((n-k+1)\, 3^{-k}\right)^2} \xrightarrow[n\to\infty]{?} 0$$

כיוון שמתקיים $3^{-k} \approx n^{1-\varepsilon}$ נקבל את הגבול.

2.1 יישום - החוק החלש של המספרים הגדולים

משפט 2.2 יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים ושווי התפלגות, בעלי שונות סופית. אזי לכל $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mathbb{E}X_1\right| > \varepsilon\right) \leq \frac{\operatorname{Var}\left(X_1\right)}{n\varepsilon^2} \xrightarrow[n\to\infty]{} 0$$

הוכחה:

$$\mathbb{E}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{n\mathbb{E}\left(X_1\right)}{n} = \mathbb{E}X_1$$

ולכן, מאי־שוויון צ'בישב

$$\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n} - \mathbb{E}X_1\right| > \varepsilon\right) \leq \frac{\operatorname{Var}\left(\frac{X_1+\cdots+X_n}{n}\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}\cdot n\cdot \operatorname{Var}\left(X_1\right)}{\varepsilon^2} = \frac{\operatorname{Var}\left(X_1\right)}{n\varepsilon^2}$$

■

3 משפט הגבול המרכזי

סכום של משתנים מקריים בלתי תלויים ושווי התפלגות, לא משנה אילו משתנים מקריים, מתפלג בקירוב בהתפלגות נורמלית או גאוסיינית - מסומן $N(0,1)$, 0 - המרכוז, 1 - סטיית התקן. ישנה עקומה שמתארת את צפיפות ההסתברות -

$$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

ואם $X \sim N(0,1)$, אזי ההסתברות שהוא שווה למספר קונקרטי היא תמיד 0 (ההתפלגות רציפה), אבל

$$\mathbb{P}(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

בצורה פורמלית:

משפט 3.1 יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי תלויים, שווי התפלגות ובעלי שונות סופית. המשתנה המקרי $\hat{S}_n = \frac{X_1+\cdots+X_n-n\mathbb{E}X_1}{\sqrt{n\mathrm{Var}(X)}}$ (שמקיים $\mathbb{E}\hat{S}_n = 0, \mathrm{Var}\left(\hat{S}_n\right) = 1$) מתכנס בהתפלגות להתפלגות גאוסיינית - כלומר לכל $-\infty \le a < b \le \infty$ ממשיים,

$$\mathbb{P}\left(a \le \hat{S}_n \le b\right) \xrightarrow[n\to\infty]{} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

למשל, אם נרצה את ההסתברות ללכל היותר שתי סטיות תקן היא

$$\mathbb{P}\left(\left|\hat{S}_n\right| \le 2\right) \xrightarrow[n\to\infty]{} 0.9545\ldots$$