

מבוא להסתברות

© ארזים

18 במאי 2016

1 אי שוויון מרקוב

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

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

הוכחה:

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

■

יש מקרים שבהם אי-השוויון הוא שוויון.

דוגמא

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

2 אי שוויון צ'בישב

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

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

הוכחה: נגדיר משתנה מקרי חדש

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

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

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

■

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

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

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,

$$\mathbb{E}N_k = \sum_{i=1}^{n-k+1} \mathbb{E}(I_i) = \frac{n-k+1}{3^k}$$

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

$$\text{Var}(N_k) = \sum_{i=1}^{n-k+1} \text{Var}(I_i) + 2 \sum_{1 \leq i < j \leq n-k+1} \text{Cov}(I_i, I_j)$$

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

$$\begin{aligned} \text{Cov}(I_i, I_j) &= \mathbb{E}(I_i \cdot I_j) - (\mathbb{E}I_i) \cdot (\mathbb{E}I_j) = 3^{-(k+m)} - 3^{-2k} = 3^{-k}(3^{-m} - 3^{-k}) = \\ &= \frac{1}{3^m} - \frac{1}{3^k} = \frac{3^k - 3^m}{3^{2k+m}} \end{aligned}$$

בסך הכל, נסיק כי

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

חסם עליון על השונות:

$$\text{Var}(N_k) \leq n \cdot 3^{-k} + 2kn \cdot 3^{-k} = (2k+1)n \cdot 3^{-k}$$

3. יהי $\varepsilon > 0$. הוכיחו כי

$$\mathbb{P}(M \geq (1+\varepsilon)\log_3 n) \xrightarrow{n \rightarrow \infty} 0$$

נשים לב כי

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

לכן, מאי שוויון מרקוב,

$$\begin{aligned} \mathbb{P}(M \geq (1+\varepsilon)\log_3 n) &= \mathbb{P}(N_{\lceil (1+\varepsilon)\log_3 n \rceil} \geq 1) \leq \mathbb{E}(N_{\lceil (1+\varepsilon)\log_3 n \rceil}) = \\ &= \frac{n - \lceil (1+\varepsilon)\log_3 n \rceil + 1}{3^{\lceil (1+\varepsilon)\log_3 n \rceil}} \leq \\ &\leq \frac{n}{3^{(1+\varepsilon)\log_3 n}} = \frac{n}{n^{1+\varepsilon}} = \frac{1}{n^\varepsilon} \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

4. יהי $\varepsilon > 0$. הוכיחו כי

$$\mathbb{P}(M \geq (1-\varepsilon)\log_3 n) \xrightarrow{n \rightarrow \infty} 1$$

כמו קודם,

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

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

$$\begin{aligned} \mathbb{P}(X \geq 1) &= 1 - \mathbb{P}(X \leq 0) = 1 - \mathbb{P}(X - \mathbb{E}X \leq -\mathbb{E}X) \geq 1 - \mathbb{P}(|X - \mathbb{E}X| \geq \mathbb{E}X) \geq \\ &\geq 1 - \frac{\text{Var}(X)}{(\mathbb{E}X)^2} \end{aligned}$$

במקרה שלנו, נקבל

$$\mathbb{P}(M \geq (1 - \varepsilon) \log_3 n) = \mathbb{P}(N_{\lceil (1 - \varepsilon) \log_3 n \rceil} \geq 1) \geq 1 - \frac{\text{Var}(N_{\lceil (1 - \varepsilon) \log_3 n \rceil})}{(\mathbb{E}(N_{\lceil (1 - \varepsilon) \log_3 n \rceil}))^2} \xrightarrow{n \rightarrow \infty} 1$$

צריך להוכיח כי

$$\frac{\text{Var}(N_{\lceil (1 - \varepsilon) \log_3 n \rceil})}{(\mathbb{E}(N_{\lceil (1 - \varepsilon) \log_3 n \rceil}))^2} \xrightarrow{n \rightarrow \infty} 0$$

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

$$\frac{\text{Var}(N_k)}{(\mathbb{E}N_k)^2} \leq \frac{(2k + 1)n \cdot 3^{-k}}{((n - k + 1)3^{-k})^2} \xrightarrow{n \rightarrow \infty} 0$$

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, \dots, X_n משתנים מקריים בלתי תלויים ושווי התפלגות, בעלי שונות סופית. אזי לכל $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}X_1\right| > \varepsilon\right) \leq \frac{\text{Var}(X_1)}{n\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$$

הוכחה:

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

ולכן, מאי-שוויון צ'בישב

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

■

3 משפט הגבול המרכזי

סכום של משתנים מקריים בלתי תלויים ושווי התפלגות, לא משנה אילו משתנים מקריים, מתפלג בקירוב בהתפלגות נורמלית או גאוסיינית - מסומן $N(0, 1)$, המרכז, 0 - 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

$$\mathbb{P}(a \leq X \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

בצורה פורמלית:

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

$$\mathbb{P}(a \leq \hat{S}_n \leq b) \xrightarrow{n \rightarrow \infty} \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

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

$$\mathbb{P}\left(\left|\hat{S}_n\right| \leq 2\right) \xrightarrow{n \rightarrow \infty} 0.9545 \dots$$