מבוא להסתברות

ארזים ©

22 במאי 2016

ראינו בשיעור שעבר את משפט הגבול המרכזי. ניזכר בו.
ראשית, נגדיר את התקנון של משתנה מקרי Y בעל שונות סופית להיות

$$\hat{Y} = \frac{Y - \mathbb{E}(Y)}{\sqrt{\mathrm{Var}(Y)}}$$

נקבל כי תמיד

$$\begin{aligned} \mathbb{E}\left(\hat{Y}\right) &= 0 \\ \mathrm{Var}\left(\hat{Y}\right) &= 1 \end{aligned}$$

משפט 0.1 (משפט הגבול המרכזי) יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי תלויים שווי התפלגות ובעלי שונות סופית. נגדיר $S_n = X_1 + \cdots + X_n$. לכן

$$\hat{S}_n = \frac{S_n - n \cdot \mathbb{E}(X_1)}{\sqrt{n \cdot \mathrm{Var}(X_1)}}$$

כעת, לכל $-\infty \le a < b \le \infty$, מתקיים

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

משמעות ההסתברות הזו היא ההסתברות שהמשתנה S_n סוטה מן התוחלת שלו בין a לבין b סטיות תקן.

דוגמא מטילים קוביה $12,000$ פעם ומסמנים בתור S את כמות הפעמים שקיבלנו 6. מה ההסתברות, בקירוב, שנקבל כמות 6 בין 1800 לבין 2100?

פתרון נשים לב כי

$$\begin{aligned} S &\sim \mathrm{Bin}\left(12000, \frac{1}{6}\right) \\ \mathbb{E}(S) &= 2000 \\ \mathrm{Var}(S) &= 12000 \cdot \frac{1}{6} \cdot \frac{5}{6} = \frac{10000}{6} \\ \mathrm{std}(S) &= \sqrt{\mathrm{Var}(S)} = 40.82\ldots \end{aligned}$$

לפי אי־שוויון צ'בישב:

$$\begin{aligned}\mathbb{P}\left(1800 \le S \le 2100\right) &= 1 - \mathbb{P}\left(S > 2100\right) - \mathbb{P}\left(S < 1800\right) \\ \mathbb{P}\left(S > 2100\right) &\le \mathbb{P}\left(\left|S - 2000\right| > 100\right) \le \frac{\mathrm{Var}\left(S\right)}{100^2} = \frac{1}{6}\end{aligned}$$

לעומת זאת, לפי משפט הגבול המרכזי,

$$\mathbb{P}\left(1800 \le S \le 2100\right) = \mathbb{P}\left(-4.9 \le \hat{S} \le 2.45\right) \approx \frac{1}{\sqrt{2\pi}} \int\limits_{-4.9}^{2.45} e^{-\frac{x^2}{2}}\,\mathrm{d}x \approx 0.9929$$

מקרה פרטי - משפט דה־מואבר־לפלס

זהו מקרה פרטי של משפט הגבול המרכזי, כאשר $X_i \sim \mathrm{Ber}\left(p\right)$, ולכן $S_n \sim \mathrm{Bin}\left(n, p\right)$.
נדבר על ההוכחה:

$$\mathbb{P}\left(S_n = k\right) = \binom{n}{k} \cdot p^k \cdot \left(1-p\right)^{n-k}$$

ובמקרה זה

$$\hat{S}_n = \frac{S_n - np}{\sqrt{np\left(1-p\right)}}$$

סימונים: עבור k שלם, נגדיר

$$\hat{k} = \frac{k - np}{\sqrt{np\left(1-p\right)}}$$

כך שמתקיים

$$\mathbb{P}\left(S_n = k\right) = \mathbb{P}\left(\hat{S}_n = \hat{k}\right)$$

כמו כן, נגדיר

$$\begin{aligned}I_n &= \left\{k \in \mathbb{Z} \mid a \le \hat{k} \le b\right\} = \\ &= \left\{k \in \mathbb{Z} \mid np + a\sqrt{np\left(1-p\right)} \le k \le np + b\sqrt{np\left(1-p\right)}\right\}\end{aligned}$$

אנו רוצים להוכיח כי

$$\begin{aligned}\mathbb{P}\left(a \leq \hat{S}_n \leq b\right) &= \mathbb{P}\left(S_n \in I_n\right) = \sum_{k \in I_n} \mathbb{P}\left(S_n = k\right) = \\ &= \sum_{k \in I_n} \binom{n}{k} \cdot p^k \left(1-p\right)^k \xrightarrow[n\to\infty]{?} \int\limits_a^b \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \,\mathrm{d}x\end{aligned}$$

משפט עזר: עבור $k \in I_n$ (במידה שווה)

$$\mathbb{P}\left(S_n = k\right) = \binom{n}{k} \cdot p^k \cdot \left(1-p\right)^{n-k} \approx \frac{1}{\sqrt{2\pi np\left(1-p\right)}} e^{-\frac{\hat{k}}{2}}$$

קירוב סטירלינג:

$$\frac{m!}{\sqrt{2\pi m}\left(\frac{m}{e}\right)^m} \xrightarrow[m\to\infty]{} 1$$

כעת,

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!\left(n-k\right)!} \approx \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \sqrt{2\pi\left(n-k\right)}\left(\frac{n-k}{e}\right)^{n-k}} = \sqrt{\frac{n}{2\pi k\left(n-k\right)}} \frac{n^n}{k^k\left(n-k\right)^{n-k}} \approx \\ &\approx \sqrt{\frac{1}{2\pi np\left(1-p\right)}} \left(\frac{n}{k}\right)^k \left(\frac{n}{n-k}\right)^{n-k} = \sqrt{\frac{1}{2\pi np\left(1-p\right)}} e^{k\log\left(\frac{n}{k}\right)} e^{\left(n-k\right)\log\left(\frac{n}{n-k}\right)}\end{aligned}$$

רעיון המשך ההוכחה: נשתמש בקירוב טיילור של $\log$ כדי להעריך את מה שקיבלנו, ולבסוף נשתמש בקירוב רימן לאינטגרל כדי להראות את השאיפה שרצינו.