הרצאה 8: משפט הגבול המרכזי והחוק החלש של המספרים הגדולים

בהרצאה זו אנחנו דנים בנושאי מעבר מתורת ההסתברות לתורת הסטטיסטיקה.

8.1. קירוב נורמלי למשתנה בינומי: משפט דה מואבר--לפלס

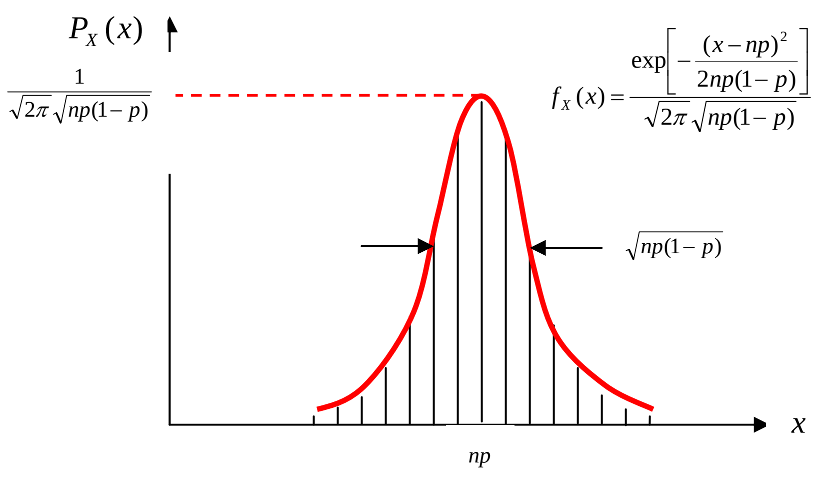
8.1.1. המקרה $n \to \infty$

נתבונן במשתנה בינומי $X \sim B(n,p)$ המתואר על ידי פונקצית ההסתברות

$$P(X=k) = C_n^k p^k (1-p)^{n-k}$$

עם $k = 0,1,\ldots,n$.

עבור $X \sim B(n,p)$ ידועות תוחלתו $E[X] = np$ ושונותו $Var[X] = \sigma^2 = np(1-p)$.



מספר מקלות בגרף עבור $P_X(x)$ הוא $n+1$ וכאשר $n >> 1$ המספר הזה הוא גדול מאוד. במקרה זה קיימת אפשרות לקרב את סדרת המקלות הארוכה על ידי פונקציה רציפה. מהי הפונקציה?

כדי למצוא אותה, נשתמש בעובדה כי המקסימום של $P_X(x)$ נמצא בנקודה $x_{\max} \approx np$. גובה של המקף עבור $x = x_{\max} \approx np$ שווה ל- $P_X(x=np) = C_n^{np} p^{np}(1-p)^{n(1-p)}$. נוסחת סטירלינג

$$n! \approx \sqrt{2\pi}\, n^{n+1/2} e^{-n} \qquad (n >> 1)$$

מביאה שבקירוב $P_X(x=np) \approx \dfrac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} = \dfrac{1}{\sigma\sqrt{2\pi}}$.

<u>שאלה 2:</u> הוכח את הנוסחה הנ"ל.

בנוסף לזה, אנחנו יודעים כי רוחב של פ. הסתברות ניתנת על ידי סטיית התקן $\sigma = \sqrt{np(1-p)}$. כאשר $n >> 1$, הרוחה הזה צר מאוד יחסית למיקום המקסימום $x_{\max} \approx np$:
$\dfrac{\sigma}{x_{\max}} = \dfrac{\sqrt{np(1-p)}}{n} \sim \dfrac{1}{\sqrt{n}} \xrightarrow{n\to\infty} 0$. אזי יש לצפות כי לפונקציה המקרבת ישנה <u>צורה של פעמון</u>.

אמנם, אם נחשב למה שואפת ההסתברות

$$P_X(x = np + \tau\sqrt{np(1-p)}) = C_n^{np+\tau\sqrt{np(1-p)}} p^{np+\tau\sqrt{np(1-p)}} (1-p)^{n(1-p)-\tau\sqrt{np(1-p)}}$$

כאשר $n \to \infty$, נקבל (בעזרת נוסחת סטירלינג) ש-

$$P_X(x = np + \tau\sqrt{np(1-p)}) \approx \frac{1}{\sqrt{2\pi}} e^{-\tau^2/2}.$$

הנוסחה בצד ימין של המשוואה היא <u>פונקצית צפיפות עבור משתנה מקרי רציף נורמלי סטנדרטי</u>: $\tau \sim N(0,1)$. מפה גם נובע כי $X = np + \tau\sqrt{np(1-p)}$ עצמו מתפלג נורמלית עם פרמטרים $\mu = np$ ו- $\sigma^2 = np(1-p)$: $X \sim N(np, np(1-p))$. זו הייתה הדגמה של <u>קירוב נורמלי למשתנה בינומי</u>.

<u>משפט דה מואבר—לפלס.</u> עבור משתנה בינומי $X \sim B(n,p)$ מתקיים:

$$\lim_{n\to\infty} P\left(\frac{X-np}{\sqrt{np(1-p)}} \le t\right) = \Phi(t)$$

8.1.2. המקרה $n >> 1$ (גדול אך סופי): תיקון רציפות

בנוסחה הנ"ל ישנה סתירה: $X \sim B(n,p)$ הוא <u>משתנה בדיד</u> ופונקצית-Φ קשורה <u>למשתנה רציף</u>! במקרה של $n >> 1$ (גדול אך <u>סופי</u>) הסתירה יכולה להוביל לטעויות חישוב כמותיות.

ניתן להקטין טעויות חישוב אם נבדיל בין משתנה בדיד למשתנה רציף. למשל, כדי לחשב את ההסתברות כי משתנה בינומי $X_B = k$, אנחנו נחשב אינטגרל מפונקצית צפיפות בגבולות $(k - \tfrac{1}{2}, k + \tfrac{1}{2})$:

$$P(X_B = k) = P(k - \tfrac{1}{2} < X_N < k + \tfrac{1}{2}) = \Phi\left(\frac{k + \tfrac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \tfrac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

החצי הנוסף הוא **תיקון רציפות** עבור קירוב נורמלי למשתנה בינומי.

מאותה הסיבה מתקיים:

$$P(X_B \le k) = P(X_N \le k + \tfrac{1}{2}) = \Phi\left(\frac{k + \tfrac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

שאלה 3: יהי $X \sim B(16,1/2)$. חשב $P(X \le 4), P(X < 7), P(X \ge 7)$ באופן מדויק ובקירוב נורמלי.

פתרון:

א. חישוב מדויק: $P(X \le 4) = \sum_{k=0}^{4} C_{16}^{k}\left(\frac{1}{2}\right)^{16} = 0.038$;

קירוב נורמלי: $P(X \le 4) = \Phi\left(\frac{4.5-8}{2}\right) = \Phi(-1.75) = 0.040$.

ב. חישוב מדויק: $P(X < 7) = P(X \le 6) = \sum_{k=0}^{6} C_{16}^{k}\left(\frac{1}{2}\right)^{16} = 0.227$;

קירוב נורמלי: $P(X < 7) = P(X \le 6) = \Phi\left(\frac{6.5-8}{2}\right) = \Phi(-0.75) = 0.227$.

ג. חישוב מדויק: $P(X \ge 7) = 1 - P(X < 7) = 1 - 0.227 = 0.773$;

קירוב נורמלי: $P(X \ge 7) = 1 - P(X < 7) = 1 - 0.227 = 0.773$.

שאלה 4: יש לקבוע בקירוב מהי ההסתברות לכך שבלפחות 28 מתוך 72 הטלות קובייה מאוזנת הופיעה התוצאה "1" או "6".

פתרון:

מספר המקרים X בהם הופיעה התוצאה "1" או "6" מתפלג בינומית עם פרמטרים $n = 72$ ו- $p = \tfrac{2}{6} = \tfrac{1}{3}$: $X \sim B(72,1/3)$. יש לחשב

$$P(X \ge 28) = 1 - P(X \le 27) = 1 - \Phi\left(\frac{27 + \tfrac{1}{2} - 72 \cdot \tfrac{1}{3}}{\sqrt{72 \cdot \tfrac{1}{3}(1 - \tfrac{1}{3})}}\right) \approx 1 - \Phi(0.875) \approx 0.19.$$

8.2. משפט הגבול המרכזי

מה הסיבה להופעת ההתפלגות הנורמלית בסעיף הקודם? האם היה חשוב שהיה מדובר על מ. בינומי? מתברר שלא! משהיה **כן חשוב** זה שמ. בינומי עם הפרמטר $n >> 1$ הוא **סכום של מספר רב של משתנים מקריים בלתי תלויים** (במקרה ההוא – משתני ברנולי).

משפט הגבול המרכזי: יהיו $X_1, X_2, ..., X_n$ משתנים מקריים בלתי תלויים בעלי התפלגות זהה שתוחלתה μ ושונותה σ^2. אזי

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le t\right) = \Phi(t).$$

במילים אחרות, עבור $n \to \infty$ מתקיים: $\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$ או $X = \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$.

כאשר $n >> 1$ (גדול אך **סופי**) ו-$X_1, X_2, ..., X_n$ הם משתנים בודדים יש לקחת בחשבון **תיקון רציפות:**

$$P(X \le k) \cong \Phi\left(\frac{k + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right),$$

$$P(X = k) \cong \Phi\left(\frac{k + \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{k - \frac{1}{2} - n\mu}{\sigma\sqrt{n}}\right).$$

שאלה 5: בישוב 10,000 משפחות, שמהן 500 הן בעלות 3 מכוניות, 2,000 בעלות 2 מכוניות, 4,500 בעלות מכונית אחת ול-3,000 מהן אין כלל מכונית. מהי ההסתברות שמגרש שבו 26 מקומות חניה יספיק לכל דיירי בית בן 24 דירות?

פתרון:
יהי X מספר מכוניות בבעלותה של משפחה שנבחרה מקרית. פונקצית ההסתברות של X ניתנת על ידי הטבלה

k	0	1	2	3
$P(X=k)$	0.30	0.45	0.20	0.05

כך שתוחלת $E[X]=1$ ושונות $Var[X] = \sigma^2 = 0.7$.

ההסתברות שמגרש יספיק היא $P(\sum_{i=1}^{24} X_i \le 26) = \Phi\left(\frac{26 + \frac{1}{2} - 24\cdot 1}{\sqrt{0.7\cdot 24}}\right) = \Phi(0.61) = 0.73$.

8.3. החוק החלש של המספרים הגדולים

נתחיל ממשפט עזר:

משפט: יהיו $X_1, X_2, ..., X_n$ תצפיות בלתי תלויות מתוך התפלגות בעלת תוחלת μ ושונות σ^2. עבור **ממוצע התצפיות** $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ מתקיים: $E[\overline{X}_n] = \mu$ ו-$Var[\overline{X}_n] = \frac{\sigma^2}{n}$.

שאלה 6: הוכח!

משפט (החוק החלש של המספרים הגדולים): נתבונן בסדרה של משתנים מקריים בלתי תלויים ושווי התפלגות $X_1, X_2, \dots, X_n, \dots$, בעלי תוחלת μ ושונות σ^2. אזי לכל מספר $\varepsilon > 0$ מתקיים:

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = 1$$

כלומר, ממוצע התצפיות $\overline{X}_n$ שואף חלש לתוחלת ההתפלגות μ כאשר $n \to \infty$.

הערה: המשפט הזה בונה בסיס לתורת הסטטיסטיקה ומסביר למה מותר להעריך תוחלת של משתנה מקרי באמצעות הממוצע!

הוכחה: על פי משפט הגבול המרכזי

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le t\right) = \Phi(t)$$

כך ש-

$$\lim_{n\to\infty} P\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\frac{\sigma\sqrt{n}}{n}} \le t\right) = \lim_{n\to\infty} P\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le t\right) = \Phi(t).$$

במילים אחרות $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ עבור $n \to \infty$. אזי,

$$\lim_{n\to\infty} P\left(\left|\overline{X}_n - \mu\right| < \varepsilon\right) = \lim_{n\to\infty} P\left(\mu - \varepsilon < \overline{X}_n < \mu + \varepsilon\right) = \lim_{n\to\infty}\left[\Phi\left(\frac{\mu + \varepsilon - \mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{\mu - \varepsilon - \mu}{\sigma/\sqrt{n}}\right)\right]$$

$$= \lim_{n\to\infty}\left[\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{\varepsilon\sqrt{n}}{\sigma}\right)\right] = 2\Phi(+\infty) - 1 = 1$$

סוף הוכחה.

שאלה 7: תשתמש במשפטים של הסעיף הזה ובתכונות של התפלגות נורמלית כדי לנמק למה ומתאי מותר להעריך תוחלת של משתנה מקרי באמצעות הממוצע!

הערה: פונקצית הצפיפות של ממוצע $\overline{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ הינה

$$f_{\overline{X}_n}(x) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \exp\left[-\frac{n}{2\sigma^2}(x - \mu)^2\right].$$