תרגול כיתה 9 – מבוא להסתברות וסטטיסטיקה

התפלגות נורמלית, קירובים, משפט הגבול המרכזי

מתרגלים: ליאור דקל ואדם צ'פמן

התפלגות נורמלית

המעבר מהתפלגות נורמלית כלשהי להתפלגות נורמלית סטנדרטית (תקנון)

$$X \sim N(\mu,\sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$$

כללים לשימוש בטבלת ההתפלגות הנורמלית סטנדרטית z

$(a \geq 0)$

$$P(Z \leq a) \equiv \phi(a); \quad P(Z \geq a) = 1-\phi(a); \quad P(Z \leq -a) = 1-\phi(a)$$

$$P(|Z| \leq a) = 2\phi(a)-1; \quad P(a \leq Z \leq b) = \phi(b)-\phi(a)$$

שאלה 1

בכיתה מסויימת ממוצע הקפיצה לגובה הוא $\bar{X} = 95$ (ס"מ) עם סטיית תקן $S_X = 5$ (ס"מ).
ממוצע הזמן של ריצת 60 מטר הוא $\bar{Y} = 10.5$ (שניות) עם סטיית תקן $S_Y = 0.3$ (שניות).
האם תלמיד הקופץ 100 ס"מ לגובה ורץ 60 מטר ב- 10.8שניות, מצטיין בשני ענפי הספורט באותה מידה?

פתרון:

כדי להשוות את הצטיינות היחסית של התלמיד נתקן את הישגיו.

קפיצה לרוחק: $Z_X = \frac{x-\bar{X}}{S_X} = \frac{100-95}{5} = 1$

ריצה: $Z_Y = \frac{y-\bar{Y}}{S_Y} = \frac{10.8-10.5}{0.3} = 1$

קיבלנו שבשני המקרים התלמיד נמצא יחידת תקן אחת (סטיית תקן) מעל לממוצע. אבל בריצה המטרה לרוץ כמה שפחות זמן בניגוד לקפיצה שהמטרה לקפוץ כמה שיותר גבוה.
לכן התלמיד טוב יותר בקפיצה לגובה. הערה: אילו היינו מקבלים שבריצה סטיית התקן שלו היא 1-, אזי ניתן היה לומר שהוא מצטיין אותה מידה בשני ענפי הספורט.

שאלה 2

המכירות של שפופרות משחת שיניים מסויימת מתפלג נורמלית עם תוחלת של 10,000 וסטיית תקן של 1,500 יחידות בשבוע.

1. מהי ההסתברות שיותר מ-12,000 שפופרות תמכרנה בשבוע?

2. כמה שפופרות לפחות צריך לייצר כדי להיות בטוחים ב- 95% שלא ייגמר המלאי?

פתרון:

א). נסמן ב- X את מספר השפופרות המיוצר בשבוע, אזי $X \sim N\left(10,000;1500^2\right)$.

פונקציית הצפיפות המתאימה $f_X(t)=\dfrac{1}{\sqrt{2\pi}\cdot 1500}e^{-\frac{(t-10000)^2}{4500000}}$.

צריך לחשב את ההסתברות שתמכרנה יותר מ-12000 שפופרות, דהיינו $P\left(X>12,000\right)$.

כדי להשתמש בטבלה נתקנן לנורמלית סטנדרטית $Z \sim N\left(0,1\right)$, מקבלים

$$P\left(X>12,000\right)=P\left(Z>\frac{12,000-10,000}{1,500}\right)=1-P\left(Z\le\frac{4}{3}\right)=1-\Phi\left(\frac{4}{3}\right)=0.091$$

ב). נדרשים למצוא את הקבוע c כך שמתקיים $P\left(X\ge c\right)=0.95$.

האחוזון ה- 95% בהתפלגות נורמלית סטנדרטית הוא $\Phi^{-1}(0.95)=1.6449$ סטיות תקן מעל הממוצע, דהיינו מספר השפופרות שצריך לייצר כדי להבטיח שהמלאי יתאים לביקוש ב- 95% מהפעמים הוא

$$\Phi\left(\frac{c-10,000}{1,500}\right)=0.95$$

$$c=\Phi^{-1}(0.95)\cdot 1500+10000=12467$$

שאלה 3

נתון שקבוצת תצפיות מתפלגת נורמלית עם ממוצע m וסטיית תקן s.

1. מהו אחוז התצפיות שמתחת לערך $(m-0.675s)$?
2. מצא קבוע a כך שאחוז התצפיות בתוך הטווח (m ± as) הוא 75% מהקבוצה.

פתרון:

א). צריך למצוא $P(X \le m - 0.675s)$
נתקנן:

$$P\left(\frac{x-m}{s} \le \frac{m-0.675s-m}{s}\right)$$
$$\Rightarrow P(Z \le -0.675) = \Phi(-0.6745) = 1-\Phi(0.6745) = 1-0.75 = 0.25$$

ב). צריך למצוא a כך שמתקיים $P(m-as \le X \le m+as) = 0.75$
נתקנן:

$$\Rightarrow P\left(\frac{m-as-m}{s} \le \frac{X-m}{s} \le \frac{m+as-m}{s}\right) = 0.75$$
$$\Rightarrow P(-a \le Z \le a) = 0.75$$
$$\Rightarrow \Phi(a) - (1-\Phi(a)) = 0.75$$
$$\Rightarrow 2\Phi(a) - 1 = 0.75$$
$$\Phi(a) = 0.875$$
$$a = \Phi^{-1}(0.875) = 1.15$$

<u>משפט הגבול המרכזי</u>

יהיו $X_i \sim N(\mu,\sigma^2)$ $(i=1,2,...,n)$ משתנים מקריים בלתי תלויים ושווי התפלגות. אזי:

$$(1) \qquad \sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$
$$(2) \qquad \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

<u>הקירוב הנורמלי להתפלגות בינומית</u>

אם- $X \sim Bin(n,p)$ ומתקיים $n \ge 30$, אזי- $X \sim N\left(np, \left(\sqrt{npq}\right)^2\right)$

<u>שאלה 4</u>

מטילים קובייה הוגנת [6..1] אלף פעמים.

1. חשב בקירוב את הסיכוי שיוטל "6" בין 150 ל-200 פעמים.
2. אם ידוע שהתוצאה "6" התקבלה בדיוק 200 פעמים, חשב בקירוב את הסיכוי ש-"5" הוטל פחות מ-150 פעמים.

פתרון:

א). המשתנה המחזיר 1 אם הוטל 6 בפעם ספציפית ואפס אחרת מתפלג ברנולי עם $p=1/6$.

מספר הפעמים שמוטל 6 באלף ההטלות X הוא סכום של אלף משתני ברנולי כאלה ומתפלג

$X \sim Bin(1000,\frac{1}{6})$. לפי משפט הגבול המרכזי, המשתנה הזה מתפלג בקירוב כמו

$N\left(np, np\left(1-p\right)\right) = N\left(\frac{1000}{6},\frac{5000}{36}\right) = N\left(166.667,138.889\right)$. לכן:

$$P(150 \le X \le 200) =$$
$$\Phi\left(\frac{200-166.667}{\sqrt{138.889}}\right)-\Phi\left(\frac{150-166.667}{\sqrt{138.889}}\right)$$
$$=\Phi\left(2.828\right)-\Phi\left(-1.414\right)=\Phi\left(2.828\right)-\left[1-\Phi\left(-1.414\right)\right]=0.92$$

ב). אם ידוע ש-6 הוטל 200 פעם, אזי המשתנה Y ששווה למספר ההטלות שבהן הוטל 5 מתפלג

$Bin(800,\frac{1}{6})$ שזה בקירוב $N\left(\frac{800}{6},\frac{4000}{36}\right)=N\left(133.334,111.111\right)$.

לכן, $P\left(Y<150\right)=P\left(Z<\frac{150-133.334}{\sqrt{111.111}}\right)=\Phi\left(1.58\right)=0.943$.

<u>שאלה 5</u>

משקלו של עלון פרסום הוא משתנה מקרי X המתפלג מעריכית עם ממוצע של 100 גרם.

מותר לשלוח בדואר חבילת עלונים במשקל מקסימלי של 3 ק"ג. העלונים ב"ת האחד במשנהו.

1.מהי בקירוב ההסתברות ש-36 עלוני פרסום שנבחרו באקראי ייכנסו בחבילה אחת?

2.מהי בקירוב ההסתברות שממוצע המשקל של 25 עלוני פרסום שנבחרו באקראי יעלה על 110 גרם?

3. מהי ההסתברות המדוייקת שלפחות 9 מתוך 36 עלוני פרסום שנבחרו באקראי ישקלו למעלה מ- 110 גרם? חשב בקירוב הסתברות זו.

4. מהי בקירוב ההסתברות שבדיוק 12 מתוך 36 עלוני פרסום שנבחרו באקראי ישקלו למעלה מ- 110 גרם?

פתרון:

$X \sim Exp(\lambda)\,; E(X) = 1/\lambda = 100 \Rightarrow \lambda = 1/100$ (בגרמים)

א). לפי משפט הגבול המרכזי $S_{36} = \sum_{i=1}^{36} X_i \sim N\left(36\cdot 100, 36\cdot 100^2\right)$,

לכן- $P(S_{36} \le 3000) = \Phi\left(\frac{3000-3600}{\sqrt{36\cdot 100^2}}\right) = \Phi(-1) = 0.1587$

ב). לפי משפט הגבול המרכזי $\bar{X}_{25} \sim N\left(100, \frac{100^2}{25}\right)$

לכן- $P(\bar{X}_{25} > 110) = 1-\Phi\left(\frac{110-100}{\sqrt{400}}\right) = 1-\Phi(0.5) = 1-0.69 = 0.31$

ג). נסמן Y – מס' העלונים ששוקלים יותר מ-110 גרם.

הסיכוי שכל עלון בנפרד ישקול יותר מ-110 גר' הוא $e^{-1.1}$,

זאת מהחישוב: $P(X>110) = 1-P(X\le 110) = 1-\left[1-e^{-\frac{110}{100}}\right] = e^{-\frac{110}{100}} = e^{-1.1}$.

ההתפלגות של Y היא $Y \sim Bin\left(36, e^{-1.1}\right)$. מכאן שההסתברות המדוייקת:

$$P(Y\ge 9) = 1-P(Y<9) = 1-\sum_{k=0}^{8}\binom{36}{k}\left(e^{-1.1}\right)^k\left(e^{-1.1}\right)^{36-k}$$

הקירוב הנורמלי :

מסעיף א' כאמור, Y = מתפלג בינומית. אזי: $E(Y) = 36\cdot e^{-1.1}$, $V(Y) = 36\cdot e^{-1.1}\left(1-e^{-1.1}\right)$.

נכניס את תיקון הרציפות (חיסור 0.5), ונקבל

$$P(Y\ge 9) = 1-P(Y<9) = 1-\Phi\left(\frac{9-0.5-36\cdot e^{-1.1}}{\sqrt{36\cdot e^{-1.1}\cdot\left(1-e^{-1.1}\right)}}\right) = 1-\Phi(-1.23) = 1-0.11 = 0.89$$

ד). נפעיל את משפט הגבול המרכזי ונשתמש בתיקון רציפות:

$$P(Y=12)\Rightarrow P(11.5\le Y\le 12.5)$$

$$=\Phi\left(\frac{12.5-36\cdot e^{-1.1}}{\sqrt{36\cdot e^{-1.1}\cdot\left(1-e^{-1.1}\right)}}\right)-\Phi\left(\frac{11.5-36\cdot e^{-1.1}}{\sqrt{36\cdot e^{-1.1}\cdot\left(1-e^{-1.1}\right)}}\right)=\Phi(0.183)-\Phi(-0.171)=0.140$$

שאלה 6

אסטרונום מעוניין למדוד את מרחקו של כוכב מסויים. כל מדידה היא משתנה מקרי שתוחלתו d שנות-אור כלשהן והשונות היא ארבע שנות-אור. כמה מדידות עליו לבצע כדי להגיע לכך שלממוצע המדידות יהיה סיכוי של 95% להיות בסביבת ± 0.5 שנות אור מהמרחק האמיתי?

פתרון:

נסמן את המדידות ב- X_i עבור $1\le i\le n$.

לפי משפט הגבול המרכזי, $\dfrac{\overline{X}-d}{\frac{2}{\sqrt{n}}}\approx Z:\ N(0,1)$.

לכן,

$$P(-0.5\le \overline{X}-d\le 0.5)\approx P(-0.5\frac{\sqrt{n}}{2}\le Z\le 0.5\frac{\sqrt{n}}{2})=1-2\Phi(-\frac{\sqrt{n}}{4})$$

דורשים כי $1-2\Phi(\frac{\sqrt{n}}{4})=0.95$ אזי $\Phi(\frac{\sqrt{n}}{4})=0.025$, ולכן $\frac{\sqrt{n}}{4}\approx -1.96$,

משמע $n\approx 61.4633$. אולם אין משמעות למספר לא שלם של דגימות ולכן $n\approx 62$.

[הערה חשובה לסטודנטים: במטלב, הפונקציה normcdf מחשבת את פונקציית ההצטברות של התפלגות נורמלית והפונקציה norminv מחשבת את הפונקצייה ההופכית]