תרגול 13 – אדם צ'פמן

משפט הגבול המרכזי

אם נתונים משתנים מקריים בלתי-תלויים רבים $Y_1, \dots, Y_n$ (יכולים להיות בדידים או רציפים) עם תוחלות $E_1, \dots, E_n$ ושונויות $V_1, \dots, V_n$, אז הסכום שלהם $X = Y_1 + \dots + Y_n$ מתפלג בקירוב נורמלית $X \sim N(\mu, \sigma^2)$ עם $\mu = E_1 + \dots + E_n$ ו$\sigma^2 = V_1 + \dots + V_n$.

דוגמא:

זכרו שמשתנה ברנולי $X \sim B(n,p)$ הוא סכום של משתני ברנולי בלתי-תלויים $Y_1, \dots, Y_n$ שהסיכוי של כל אחד מהם להיות אחד הוא p ולהיות 0 הוא $1-p$. לפי משפט הגבול המרכזי, כאשר n גדול, אז אפשר לומר שמתקיים בערך $X \sim N(\mu, \sigma^2)$ כאשר $\mu = np$ ו$\sigma^2 = np(1-p)$.

שאלה:

ישנה אוכלוסייה של 10,000 גורי תנים כאשר הסיכוי של כל אחד לשרוד הוא 0.8. מה הסיכוי שישרדו יותר מ8,100 גורים?

פתרון:

מספר השורדים מתפלג $X \sim B(10{,}000, 0.8)$ שזה בקירוב $X \sim N(8000, 1600)$. סטיית התקן היא 40. 8100 הוא אם כך $\frac{100}{40} = 2.5$ סטיות תקן מעל הממוצע. לפי הטבלה של ההתפלגות הנורמלית, יש סיכוי של 0.9938 למשתנה נומרלי לצאת עד 2.5 סטיות תקן מעל הממוצע. לכן הסיכוי שיהיו יותר מ8,100 גורים שורדים הוא $1 - 0.9938 = 0.0062$.

אי-שוויון מרקוב

אם X הוא משתנה חיובי אז $P(X \geq a) \leq \frac{E(X)}{a}$ לכל a חיובי

דוגמא:

אם $X \sim Exp(\frac{1}{2})$. אז $P(X \geq 2) = \int_2^\infty \frac{1}{2} e^{-\frac{1}{2}x} dx = -e^{-\frac{1}{2}x}|_2^\infty = e^{-1} = 0.367879$, $\frac{E(X)}{2} = \frac{2}{2} = 1$.

שאלה:

נתונים משתנים מקריים חיוביים X_1, X_2, X_3, X_4 עם תוחלות 1,2,3,4 בהתאמה. מצאו בעזרת משפט מרקוב חסם לסיכוי של $X_1 + X_2 + X_3 + X_4 > 100$.

פתרון:

התוחלת של הסכום היא $1 + 2 + 3 + 4 = 10$. לכן לפי משפט מרקוב הסיכוי שהסכום יהיה גדול מ100 הוא לכל היותר $\frac{1}{10} = \frac{10}{100}$.

פונקציית צפיפות של סכום משתנים

בהינתן X וY משתנים מקריים רציפים עם התפלגות משותפת $f_{X,Y}(x,y)$, את הצפיפות של $Z = X + Y$ מחשבים:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx$$

בפרט, אם הם בלתי-תלויים אז $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ ואז

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)\,dx$$

שאלה:

נגיד ש$X \sim U(0,1)$ ו$f_Y(y) = 2y$ עבור $0 < y < 1$ ו0 אחרת. נניח שהמשתנים בלתי-תלויים. מצאו את פונקציית הצפיפות של $Z = X + Y$.

פתרון:

הערכים של Z נעים בין 0 ל2. לכל $0 \leq z \leq 1$ צריך לחשב

$$f_Z(z) = \int_0^z 1 \cdot 2(z-x)dx = (2zx - x^2)|_0^z = z^2$$

לכל $1 < z \leq 2$ צריך לחשב

$$f_Z(z) = \int_{z-1}^{1} 1 \cdot 2(z-x)dx = (2zx - x^2)|_{z-1}^{1} = 2z - 1 - 2z(z-1) + (z-1)^2$$
$$= 2z - 1 - 2z^2 + 2z + z^2 - 2z + 1 = 2z - z^2$$

שאלה:

נתון $f_X(x) = 2xe^{-x^2}$ עבור $x > 0$ ו0 אחרת. מצאו את פונקציית הצפיפות של $Y = X^2$.

פתרון:

$$F_X(x) = P(X < x) = \int_0^x 2te^{-t^2}dt = -e^{-t^2}|_0^x = 1 - e^{-x^2}$$

נשים לב שמדובר במשתנים חיוביים, ואז לכל ערך חיובי y מתקיים

$$F_Y(y) = P(Y < y) = P(X^2 < y) = P\left(X < \sqrt{y}\right) = F_X\left(\sqrt{y}\right) = 1 - e^{-y}$$

נגזור ונקבל

$$f_Y(y) = -e^{-y}$$

עבור ערכי y חיוביים, ו0 אחרת.

שאלה:

אורך חיים של סוללה מתפלג מעריכית עם ממוצע של 1000 דקות. מנשה קונה 10 סוללות. מה הסיכוי ש7 סוללות יחיו יותר מהממוצע ו3 יחיו פחות?

פתרון:

נסמן את אורך החיים של סוללה i בX_i וידוע כי $X_i \sim Exp(\frac{1}{1000})$. נסמן משתנה ברנולי ("אינדיקטור") Y_i שמחזיר 1 אם $X_i > 1000$ ו0 אחרת. מעוניינים לדעת ?= $P(Y_1 + \cdots + Y_{10} = 7)$.

הסיכוי של X_i להיות גדול מ1000 הוא

$$P(Y_i = 1) = P(X_i > 1000) = \int_{1000}^{\infty} \frac{1}{1000} e^{-\frac{1}{1000}x} dx = -e^{-\frac{1}{1000}x}|_{1000}^{\infty} = \frac{1}{e} = 0.367879$$

אז $Y_1 + \cdots + Y_{10} \sim B(10, \frac{1}{e})$ ולכן

$$P(Y_1 + \cdots + Y_{10} = 7) = \binom{10}{7}\left(\frac{1}{e}\right)^7 \left(1 - \frac{1}{e}\right)^3 = 0.028$$