14.01.18

מבוא לסטטיסטיקה
הרצאה

פונ' הסתברות מצטברת $P_X(x)=P(X\le x)$

$$F_X(x)=\int_{-\infty}^{x} f(x)(t)\,dt$$

אם F רציפה, אז אומרים שהמשתנה X הוא רציף.

אם ~~~~קיימת פונקציה $f_X(x)$ שמקיימת $F_X(x)=\int_{-\infty}^{x} f_X(t)\,dt$ אז $f_X(t)$ נקראת פונקציית צפיפות של המשתנה.

באופן טבעי אם F גזירה אז F' מוגדרת להיות פונקציית צפיפות

אם קיימת פונקציית צפיפות אז אומרים שהמשתנה רציף בהחלט.

דוגמה: משתנה $X\sim exp(\lambda)$ משתנה מעריכי

$$F_X(x)=\begin{cases}0 & x<0\\ 1-e^{-\lambda x} & x\ge 0\end{cases}$$

יש למשתנה פונקציית צפיפות
$$f_X(x)=\begin{cases}0 & x<0\\ \lambda\cdot e^{-\lambda x} & x\ge 0\end{cases}$$

מהו $P(-1<X<2)$?

תשובה: זה $P(X\le 2)-P(X\le -1)=1-e^{-2\lambda}-0$

(מותר לכתוב $\le$ במקום $<$ כי פונ' ההסתברות רציפה) (הסתברות מצטברת ב-1-)

דרך אחרת: נשתמש באינטגרל של פונקציית הצפיפות בין 1- ל-2.

$$P(-1\le X\le 2)=\int_{-1}^{2} f_X(x)\,dx=\int_{-1}^{0} 0+\int_{0}^{2}\lambda\cdot e^{-\lambda x}\,dx=\left[-e^{-\lambda x}\right]_0^2=-e^{-2\lambda}-(-e^{-\lambda 0})=1-e^{-2\lambda}$$

<u>חישוב תוחלת של משתנה רציף</u>

לחישוב $E(X)$ כאשר X רציף בהחלט נשתמש בנוסחה

$$E(X)=\int_{-\infty}^{+\infty} f_X(x)\cdot x\,dx$$

ה-$f_X(x)$ אנלוגי להסתברות של המקרה הבדיד. במקום $\sum P(X=x)\cdot x$ עושים $\int f_X(x)\cdot x$

<u>דוגמה:</u> $X\sim U(a,b)$ אז $f_X(x)=\frac{1}{b-a}$ עבור $a\le x\le b$

ושווה לאפס אחרת.

$$\int_a^b \frac{1}{b-a}=1$$

$$E(X)=\int_a^b f_X(x)\,x\,dx=\int_a^b\frac{1}{b-a}\cdot x\,dx=\frac{1}{b-a}\int_a^b x\,dx=$$

$$=\frac{1}{b-a}\cdot\frac{1}{2}\cdot x^2\Big]_a^b=\frac{1}{b-a}\cdot\frac{1}{2}(b^2-a^2)=\frac{1}{2\cancel{(b-a)}}\cdot\cancel{(b-a)}\cdot(b+a)=\frac{a+b}{2}$$

שאלה: נתון משתנה מקרי X המקיים

$$f_X(x)=\begin{cases} cx & 0\le x\le 1 \\ 0 & \text{אחרת} \end{cases}$$

א) מצאו את c. ב) מצאו את $f_X(x)$ של X. ג) מצאו את $E(X)$

א) האינטגרל של הצפיפות צריך להיות שווה 1 (צריך להתקיים בכל המקרים)

$$\int_{-\infty}^{+\infty} f_X(x)\,dx=\int_0^1 cx\,dx=\left[\tfrac{1}{2}cx^2\right]_0^1=\tfrac{1}{2}c\cdot 1^2=\tfrac{1}{2}c$$

$$\tfrac{1}{2}c=1 \Rightarrow c=2$$

ב)

$$F_X(x)=\int_{-\infty}^{x} f_X(t)\,dt$$

עבור $x<0$: $F_X(x)=0$

עבור $0\le x\le 1$:

$$F_X(x)=\int_0^x f_X(t)\,dt=\int_0^x 2t\,dt=t^2\Big]_0^x=x^2$$

עבור $x\ge 1$: $F_X(x)=1$

בצורה אחרת עבור $x\ge 1$:

$$F_X(x)=\int_{-\infty}^{x} f_X(t)\,dt=\int_0^1 2t\,dt=1.$$

$$F_X(x)=\begin{cases} 0 & x<0 \\ x^2 & 1\ge x\ge 0 \\ 1 & 1<x \end{cases}$$

ג)

$$E(X)=\int_{-\infty}^{+\infty} f_X(x)\cdot x\,dx=\int_0^1 2x\cdot x\,dx=\int_0^1 2x^2\,dx=\tfrac{2}{3}x^3\Big]_0^1=\tfrac{2}{3}$$

<u>תוחלת של פונקציה של משתנה</u>:

$$E(g(x))=\int_{-\infty}^{+\infty} f_X(x)\cdot g(x)\,dx$$

דוגמא: נחשב את $E(X^3)$ עבור הדוגמא האחרונה שהייתה.

$$f_X(x)=\begin{cases} 2x & 1\ge x\ge 0 \\ 0 & \text{אחרת} \end{cases}$$

$$E(X^3)=\int_{-\infty}^{+\infty} f_X(x)\cdot x^3\,dx=\int_0^1 2x\cdot x^3\,dx=\int_0^1 2x^4\,dx=\tfrac{2}{5}\cdot x^5\Big]_0^1=\tfrac{2}{5}$$

חישוב שונות

נוסחה: $V(X) = E(X^2) - E^2(X)$

את $E(X)$ אנו כבר יודעים לחשב. את $E(X^2)$ נחשב לפי $\int_{-\infty}^{+\infty} f_X(x)\cdot x^2 dx$

דוגמה: חישוב שונות של משתנה $X \sim U(a,b)$

$$V(X) = E(X^2) - E^2(X)$$

$$E^2(X) = \left(\frac{a+b}{2}\right)^2$$

$$E(X^2) = \int_a^b f_X(x)\cdot x^2 dx = \int_a^b \frac{1}{b-a} x^2 dx = \frac{1}{3(b-a)} x^3 \Big]_a^b = \frac{b^3-a^3}{3(b-a)}$$

אם מחסירים $E(X^2) - E^2(X)$ מקבלים $V(X) = \frac{(b-a)^2}{12}$

אם $\underset{\text{בדיד}}{X \sim U[a,b]}$ אז $V(X) = \frac{(b-a+1)^2-1}{12}$

בחזרה למשתנה נורמלי, ונדבר על משתנה נורמלי סטנדרטי.

$X \sim N(0,1)$ אז $f_X(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$ לכל $+\infty > x > -\infty$

פונקציית ההסתברות המצטברת שלו היא $\Phi(x) = \int_{-\infty}^{x} f_X(t) dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} dt$

פונקציית ההסתברות המצטברת עבור ערכי $x \geq 0$ נתונה בטבלה.

(טבלת התפלגות משתנה נורמל סטנדרטי - ניתן לחפש באינטרנט).

לפי הטבלה מוצאים מהו $P(X \leq a)$ לכל a אי שלילי. מהו $\Phi(0)$?

מכיוון שהצפיפות סימטרית סביב 0, אז $\Phi(0) = 0.5$.

מה לגבי $\Phi(a)$ עבור $a<0$? $\Phi(-a) = 1 - \Phi(a)$

הערות: אפשר להוכיח שעבור $X \sim N(0,1)$ מתקיים $E(X) = 0$, $V(X) = 1$.

שאלה: $X \sim N(0,1)$. מהו $P(-1 \leq X < 2)$?

תשובה: $\Phi(2) - \Phi(-1) = \Phi(2) - [1 - \Phi(1)] = \Phi(2) + \Phi(1) - 1$.

משפט הגבול המרכזי:

תהי $\{X_k\}_{k=1}^{\infty}$ סדרת משתנים מקריים ב"ת שווי התפלגות ובעלי שונות סופית σ^2 ותוחלת μ. אז המשתנה $\frac{\frac{\sum_{k=1}^{n} X_k}{n} - \mu}{\sqrt{\frac{\sigma^2}{n}}}$ עבור ערכי n גדולים, והמשתנים האלה מתנהגים בקירוב כמו משתנה נורמלי סטנדרטי.

למשתנה הנ"ל יש פונקציית הסתברות מצטברת שקרובה לפונקציית ההסתברות המצטברת של משתנה $N(0,1)$.

אזי $X \sim N(0,1)$: המונח שהשתמשנו מתפלגת בקירוב כמו X.

שאלה: נניח סדרת משתנים ב"ת שווי התפלגות שמקיימים $P(X_k=1)=P(X_k=-1)=\frac{1}{2}$, מהי בקירוב ההסתברות שהממוצע של 100 משתנים ראשונים יהיה קטן מ- 0.2?

פתרון: מתקיים $\mu=E(X_k)=0$, $\sigma^2=1$

ההסתברות שווה בקירוב ל $\Phi\left(\frac{0.2-0}{\sqrt{\frac{1}{100}}}\right)=\Phi(2)$

שאלה נוספת: מהי בקירוב ההסתברות שהממוצע גדול מ- -0.1?

תשובה

$$1-\Phi\left(\frac{-0.1-0}{\sqrt{\frac{1}{100}}}\right)=1-\Phi(-1)=\Phi(1)$$

שאלת המשך: מה הסיכוי ש- $\sum_{k=1}^{100} X_k>30$?

זה שקול ל- $\frac{\sum_{k=1}^{100} X_k}{100}>\frac{30}{100}$

סכום גדול מ-30 שקול לממוצע גדול מ- 0.3, ועכשיו קל לקחת

$$1-\Phi\left(\frac{0.3-0}{\sqrt{\frac{1}{100}}}\right)=1-\Phi(3)$$

<u>תוחלת סופית ואין סופית</u>

$$E(X)=\sum P(X=k)\cdot k=\sum_{k<0} P(X=k)\cdot k+\sum_{k>0} P(X=k)\cdot k$$

אם שתי הסכומות סופיות, אז אומרים שיש תוחלת סופית.

אם השניה אינסופית (הראשונה סופית) אז אומרים שיש תוחלת $+\infty$.

אם הראשונה אינסופית והשניה סופית אז יש תוחלת $-\infty$

אם שתיהן אין סופיות אז התוחלת לא מוגדרת.

דוגמא: אם יש משתנה X שמקיים $P(X=k)=\frac{c}{k^2}$

עבור כל $k\geq 0$ כ-1 מתאים (כך ש: $\sum \frac{c}{k^2}=1$)

אז למשתנה יש תוחלת

$$E(X)=\sum P(X=k)\cdot k=\sum \frac{c}{k^2}\cdot k=\sum \frac{c}{k}=\infty$$

אז למשתנה יש תוחלת אינסופית

אם $E(X)$ סופי ו- $E(X^2)$ סופי, אז למשתנה יש שונות סופית.

החוק החלש

החוק החלש הוא תכונה של סדרת משתנים מקריים. אז החוק החלש אומר מתקיים עבור כל $0<\varepsilon$

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \frac{\sum_{i=1}^{n} E(X_i)}{n}\right| > \varepsilon\right) = 0$$

<u>הערה:</u> אם כל המשתנים יש להם אותה תוחלת μ, אז התנאי של החוק החלש הוא $\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) = 0$.

החוק החלש הוא תכונה שמקיימים חלק מסדרות משתנים, ולפעמים לא חלק מהם.

<u>משפט:</u> תהי סדרת משתנים בלתי תלויים שווי התפלגות בעלי שונות סופית. אז החוק החלש.

<u>הוכחה:</u> יהי μ התוחלת של כל אחד מהמשתנים.

יהי σ^2 השונות שלהם.

נתבונן על $0<\varepsilon$ קבוע. (אי-שוויון צ'בישב)

$$P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu\right| > \varepsilon\right) \le \frac{V\left(\frac{\sum X_i}{n}\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}\cdot V\left(\sum X_i\right)}{\varepsilon^2} = \frac{\frac{1}{n^2}\cdot n\cdot V(X_1)}{\varepsilon^2} = \frac{\sigma^2}{n\cdot\varepsilon^2}$$

לכן עבור כל $0<\varepsilon$, ההסתברות לסטייה יותר גדולה מ-ε מהממוצע נתון שואפת לאפס.

<u>דוגמא:</u> נתונה סדרה של משתנים מקריים ב"ת שווי התפלגות המקיימת

$$P(X_i=2) = P(X_i=0) = \frac{1}{2}$$

אז הסדרה את החוק החלש אם מתקיים

$$\lim_{n\to\infty} P\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - 1\right| > \varepsilon\right) = 0 \qquad \text{לכל } 0<\varepsilon$$

1 הוא התוחלת של כל אחד מהמשתנים. לכן ההסתברות שהממוצע יסטה מ-1 ביותר מ-$\varepsilon>0$, נתון שואפת לאפס כאשר $n\to\infty$.

<u>דוגמא:</u> יש 2 מטבעות. המטבע הראשון נופל על עץ בסיכוי $\frac{1}{4}$, והמטבע השני נופל על עץ בסיכוי $\frac{3}{4}$. אני בוחר באקראי אחד מהם בסיכוי שווה ומבצע בו סדרת הטלות בלתי תלויות בהינתן המטבע.

האם החוק החלש חל על הסדרה?

אם בחרנו במטבע הראשון אז $\lim P\left(\left|\frac{\sum X_i}{n} - \frac{1}{4}\right| > 0.01\right) = 0$

כי מדובר בסדרת הטלות ב"ת של מטבע שנופל על עץ בסיכוי $\frac{1}{4}$.

לכן, על פי המשפט שהוכחנו על התנאי המספיק, יש שאיפה ל- $\frac{1}{4}$

אם בחרנו במטבע השני, אז

$$\lim P\left(\left|\frac{\sum x_i}{n} - \frac{3}{4}\right| > 0.01\right) = 0$$

מתקיים מנוסח ;

$$E(X_i) = \frac{1}{2}\cdot\frac{3}{4} + \frac{1}{2}\cdot\frac{1}{4} = \frac{1}{2}$$

לכן מטבע יש תוחלת של $\frac{1}{2}$, אבל אנחנו נתקרב או ל-0.25 או ל-0.75 לא נתקרב ל-0.5.

<u>דוגמא נוספת:</u> נניח שיש 2 מטבעות. במטבע אחד יש 2 צדדים של 0 ו-4, במטבע השני יש שני צדדים של 1 ו-3. אני בוחר רנדומלי שווה אחד מהם ומטיל אותו סדרת הטלות.

ההטלות תלויות, כי אם פעם קיבלתי 3 אז לעולם לא אקבל 4.

אם בחרתי במטבע הראשון, אז מתקיים שהממוצע יתקרב ל- $\frac{0+4}{2}$, וזאת לפי החוק החלש, שכן יש לנו סדרת משתנים בלתי תלויים שווי התפלגות בעלי שונות סופית ותוחלת $\frac{0+4}{2} = 2$.

אם בחרתי במטבע השני, אז ההסתברות שהממוצע יהיה קרוב ל- $\frac{1+3}{2} = 2$ שואפת ל-1.

בכל מקרה, ההסתברות שהממוצע סוטה מהתוחלת ביותר מ- $\varepsilon > 0$ נתון שואפת לאפס.

דוגמא בה יש תוחלת סופית ושונות אינסופית:

$P(X=k) = \frac{c}{k^3}$ עבור $k \geq 1$ טבעי, רק ש $\sum \frac{c}{k^3} = 1$

$$E(X) = \sum \frac{c}{k^3}\cdot k = \sum \frac{c}{k^2} < \infty$$

$$E(X^2) = \sum \frac{c}{k^3}\cdot k^2 = \sum \frac{c}{k} = \infty$$

$$E(X^2) = \infty$$