הסתברות 1 ־ תרגול 13
התכנסות בהתפלגות

9 בינואר 2019

נזכר ראשית בהגדרה:
נאמר שסדרת מ"מ (לא בהכרח באותו מרחב הסתברות) $\{X_n\}$ מתכנסת למ"מ X בהתפלגות ונסמן

$$X_n \xrightarrow{d} X$$

אם לכל a שהיא נקודת רציפות של F_X מתקיים

$$\lim_n F_{X_n}(a) = F_X(a)$$

כאשר F_X פונקצית ההתפלגות הצוברת

$$F_X(a) = P(X \leq a)$$

הערה 0.1 $X_n \xrightarrow{d} X$

1. מדוע דורשים התכנסות רק בנקודות הרציפות? התשובה היא שאנו רוצים לאפשר דוגמאות בהם ההתפלגות הגבולית איננה רציפה, מה שקורה למשל כהגבול הוא מ"מ קבוע.
2. ההתכנסות בהתפלגות היא חלשה יותר מהתכנסות בהסתברות (ראיתם שהתכנסות בהסתברות גוררת התכנסות בהתפלגות).
3. כל פונקצית התפלגות מצטברת היא
   (א) מונוטונית עולה חלש.
   (ב) רציפה מימין.
   (ג) $\lim_{t\to\infty} F_X(t) = 1$ ו $\lim_{t\to-\infty} F_X(t) = 0$.

למעשה, כל פונקציה המקיימצת את תכונות אלו היא פונקצית התפלגות של איזשהו מ"מ. (אנו יודעים איך לחלץ את ההתפלגות מההתפלגות המצטברת לפחות במקרה הדיסקרטי).

נפתח בדוגמא שאולי תבהיר את ההבדל שבין התכנסות בהתפלגות להתכנסויות הקודמות ובפרט תראה מדוע היא חלשה יותר.

דוגמא א: (דוגמא להתכנסות בהתפלגות אך לא בהסתברות)

יהיו $X \sim X_n$ מ"מ ב"ת כאשר X איננו קבוע. אזי, משאלה 5 בתרגיל 11 אם $X_n \xrightarrow{p} X$ אז X קבוע (מה שלא יתכן כמובן) ולכן אין התכנסות בהסתברות. מצד שני ההתכנסות בהתפלגות היא מיידית שכן לכל n ו $a \in \mathbb{R}$

$$F_{X_n}(a) = F_X(a)$$

שימו לב שאי התלות של X_n כלל לא רלוונטית בחישוב. אנו כבר לא בודקים אם המ"מ קרובים זה לזה (כאן האי־תלות רלוונטית מאוד) אלא רק עם ההתפלגויות הצוברות קרובות. לכן אין צורך לבדוק באילו מאורעות מתקבלים אלו ערכים וכולי אלא רק האם בסה"כ ההסתברויות מתכנסות.

נראה מתי בכל זאת התכנסות בהתפלגות גוררת בהסתברות:

טענה 0.2 תהי $\{X_n\}$ סדרת מ"מ המתכנסת בהתפלגות לקבוע a, אז X_n מתכנסות בהסתברות ל־a.

הוכחה: תהי

$$F_a(x) = \begin{cases} 1 & x \geq a \\ 0 & x < a \end{cases}$$

פונקצית ההתפלגות של המ"מ הקבוע a. כעת, יהי $\epsilon > 0$, נוכיח כי $P(|X_n - a| \leq \epsilon) \to 1$. אכן,

$$|X_n - a| \leq \epsilon \iff a - \epsilon \leq X_n \leq a + \epsilon$$

ומכאן ש

$$\begin{aligned} P(|X_n - a| \leq \epsilon) &= P(a - \epsilon \leq X_n \leq a + \epsilon) \\ &= P(X_n \leq a + \epsilon) - P(X_n < a - \epsilon) \\ &\leq F_{X_n}(a + \epsilon) - F_{X_n}(a - \epsilon) \end{aligned}$$

(שימו לב כי הא"ש האחרון הוא חזק אם"ם $P(X_n = a - \epsilon) > 0$).

מאחר ו $X_n \xrightarrow{d} a$ נקבל

$$F_{X_n}(a + \epsilon) \to F_a(a + \epsilon) = 1, \quad F_{X_n}(a - \epsilon) \to F_a(a - \epsilon) = 0$$

ולכן

$$\lim_n P(|X_n - a| \leq \epsilon) = 1$$

כנדרש. ■

דוגמא ב:

יהיו X_n מ"מ רציפים בהחלט עם התפלגות צוברת

$$F_{X_n}(t) = \begin{cases} \left(1 - \frac{1}{1+nt}\right)^n & t > 0 \\ 0 & t \leq 0 \end{cases}$$

(מדוע זו פונקצית התפלגות מצטברת לגיטימית)

לכל $t>0$

$$\lim_n \left(1-\frac{1}{1+nt}\right)^n = \lim_n \left(1-\frac{1}{nt}\right)^n = \lim_n \left(1-\frac{1/t}{n}\right)^n = e^{-1/t}$$

ולכן

$$F_X(t) = \begin{cases} e^{-1/t} & t>0 \\ 0 & t\le 0 \end{cases}$$

דוגמא ג:

נניח כי

$$F_{X_n}(t) = \frac{e^{nx}}{1+e^{nx}}$$

אזי

$$F_{X_n}(t) \to \begin{cases} 0 & t<0 \\ 1/2 & t=0 \\ 1 & t>0 \end{cases}$$

לכאורה יש כאן בעיה שכן הגבול איננו רציף מימין ב $t=0$ (ולכן איננו התפלגות מצטברת) אולם, $x=0$ איננה נקודת רציפות ולכן ניתן לומר כי

$$X_n \xrightarrow{d} X$$

כאשר $P(X=0)=1$. (לצייר ־קירוב של $1_{[0,\infty]}$)

משפט הגבול המרכזי

משפט 0.3 תהי $\{X_n\}$ סדרה של מ"מ ב"ת ושווי התפלגות בעלי תוחלת 0 ושונות 1. אזי

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \xrightarrow{d} Z$$

כאשר $Z=N(0,1)$. כלומר

$$P(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \le a) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-x^2/2}dx =: \Phi(a)$$

הערות:

- אם $E(X_i) = \mu$ ו $Var(X_i) = \sigma^2$ אזי, נוכל להגדיר את הסכום המנורמל (כך שנעמוד בתנאי המשפט)

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - n\mu}{\sigma}$$

ונקבל

$$\lim_n P(S_n \le a) = \Phi(a),\ a \in \mathbb{R}$$

- משפט הגבול המרכזי עוסק בממוצע הנכפל בפקטור $\sqrt{n}$.זה קצת מבלבל, מה זה בעצם אומר על הממוצע (היותר אינטואיטבי עבורנו)?
ניקח

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

אזי

$$P(S_n \le a) = P(Y_n \le \frac{a}{\sqrt{n}}) \to \Phi(a), \quad \frac{1}{\sqrt{n}} = \sigma(S_n) = \sqrt{Var(S_n)}$$

במילים אחרות, המשפט נותן הערכה של סטיות מהממוצע (המנורמל כאן לאפס) בסדר גודל של $\frac{1}{\sqrt{n}}$ (סטיית התקן).

דוגמא א (מתוך מבחן משנה שעברה):

יהי $X_n \sim 40Bin(n, 1/2)$ (בשאלה זה יוצא מה שאוהד משלם לאורי על פלאפל על n ימים כאשר ארוחה עולה 40 ש"ח והם מגרילים כל ארוחה מי ישלם)

מצאו סדרות מספרים a_n, b_n כך ש

$$a_n(X_n - b_n) \xrightarrow{d} Z,\ Z \sim N(0,1)$$

פתרון:

$$X_n = 40 \sum_{i=1}^{n} Y_i, \quad Y_i \sim Ber(1/2)$$

נשים לב כי $E(40Y_i) = \frac{40}{2} = 20$ וכן

$$Var(40Y_i) = 40^2 \frac{1}{4} = 400, \quad (\sigma = \sqrt{400} = 20)$$

ולכן מההערה לעיל

$$\frac{X_n - 20n}{20\sqrt{n}} \xrightarrow{d} Z$$

כלומר $a_n = \frac{1}{20\sqrt{n}}$ ו $b_n = 20n$.

23.8.1 תכונות של התפלגות פואסון

יהי $X \sim Po(\lambda)$

1. אם $Y \sim Po(\eta)$ ב"ת ב $X \sim Po(\lambda)$ אז $(X+Y) \sim Po(\eta+\lambda)$

דוגמא ב (מועד א' 2015)

יהי $X \sim Po(10000)$. תנו קירוב להסתברות $P(X \leq 9800)$. הקרוב יכול להיות בצורת אינטגרל.

פתרון:

ראשית, כפי שראיתם בכיתה, אם $X_i \sim Po(1)$ ב"ת, אזי הסכום מתפלג פואסוני עם פרמטר שהוא סכום הפרמטרים

$$X = \sum_{i=1}^{10000} X_i \sim Po(10000)$$

נזכר כי $E(X) = 1$ ו $Var(X) = 1$ (התוחלת והשנות שווים לפרמטר) ולכן אם ניקח $Y_i = X_i - 1$ נקבל כי $E(Y_i) = 0$ ו $Var(Y_i) = 1$ והם ב"ת כמובן. ולכן ל

#. טענה תכונות בשונות

יהא X מ"מ בעל שונות סופית ויהי $a \in \mathbb{R}$ אזי

1. $Var(X+a) = Var(X)$

$$Y := \frac{1}{\sqrt{10000}} \sum_{i=1}^{10000} (X_i - 1)$$

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - n\mu}{\sigma}$$

יש התפלגות מצטברת קרובה יחסית למשתנה נורמלי סטנדרטי. כעת

$$X \leq 9800 \iff X - 10000 \leq -200 \iff \frac{X - 10000}{100} = Y \leq -2$$

ולכן

$$P(X \leq 9800) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-2} e^{-t^2/2} dt$$