

הסתברות למתמטיקאים – הרצאה 7

21.03.11

תוספת קטנה על המשפט. איפה השתמשנו באי תלות? כשאמרנו ש- $Y_1 - \mathbb{E}Y_1 \perp Y_2 - \mathbb{E}Y_2$ מדוע? X_1, X_2 ב"ת לכן Y_1, Y_2 ב"ת לכן $Y_1 - \mathbb{E}Y_1, Y_2 - \mathbb{E}Y_2$ לכן

$$\mathbb{E}((Y_1 - \mathbb{E}Y_1)(Y_2 - \mathbb{E}Y_2)) = (\mathbb{E}(Y_1 - \mathbb{E}Y_1))(\mathbb{E}(Y_2 - \mathbb{E}Y_2)) = 0 \cdot 0 = 0$$

לכן למעשה מספיק ב"ת בזוגות.

תוספת שנייה:

הוכחנו התכנסות ב- L_2 :

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n}\right| < \varepsilon\right) \xrightarrow{n \rightarrow \infty} 1$$

אבל $\mathbb{P}(Y_n \geq \varepsilon) \geq \frac{\mathbb{E}Y_n}{\varepsilon}$ לכן $\mathbb{E}Y_n \geq \varepsilon \mathbb{P}(Y_n \geq \varepsilon)$: לכן גם $\mathbb{E}Y_n \geq \varepsilon$

משפט.

$$\forall \varepsilon > 0 \quad \mathbb{P}(\sup |F_n() - F()| \leq \varepsilon) \xrightarrow{n \rightarrow \infty} 1$$

תזכורת.

$$F_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(-\infty, x]}(X_k)$$

$X_1, \dots, X_n$ מ"מ ב"ת מתפלגים F .

למה. לכל מידת הסתברו μ ב- $\mathbb{R}$ ולכל $\varepsilon > 0$ קיימים n ו- $x_1, \dots, x_n$ כך ש- $-\infty < x_1 < x_2 < \dots < x_n < \infty$ ו- $\mu((-\infty, x_1)) \leq \varepsilon, \mu((x_1, x_2)) \leq \varepsilon, \dots, \mu((x_{n-1}, x_n)) \leq \varepsilon, \mu((x_n, \infty)) \leq \varepsilon$

הלמה עוזרת לעקוף בעיה של אטומים כבדים.

הוכחה. ניקח $x_1 = \sup\{x \in \mathbb{R} : \mu((-\infty, x)) < \varepsilon\}$. זאת קבוצה לא ריקה. מצד שני זאת קבוצה חסומה אז $x_1 \in \mathbb{R}$.

$$\mu((-\infty, x_1)) = \lim_{k \rightarrow \infty} \underbrace{\mu((-\infty, x_1 - \frac{1}{k}))}_{< \varepsilon} \leq \varepsilon$$

מצד שני:

$$\mu((-\infty, x_1]) = \lim_{k \rightarrow \infty} \underbrace{\mu((-\infty, x_1 + \frac{1}{k}))}_{\geq \varepsilon} \geq \varepsilon$$

יכול להיות ש- x_1 הוא אטום אבל אנחנו נזהרים.
אם $\mu((x_1, \infty)) \leq \varepsilon$ אז $n = 1$ וסגרנו עניין. אם $\mu((x_1, \infty)) > \varepsilon$ נמשיך בתהליך:

$$x_1 = \sup\{x \in (x_1, \infty) : \mu((x_1, x)) < \varepsilon\}$$

מדוע יש כאלה? כי אם שואפים מימין אז ההסתברות שואפת לאפס לכן מתישהו תהיה קטנה מאפסילון. וזה חסום.
שוב,

$$\mu((x_1, x_2)) \leq \varepsilon$$

-1

$$\mu((x_1, x_2]) \geq \varepsilon$$

כעת שואלים $\mu((x_2, \infty)) \leq \varepsilon$ אם כן $n = 2$ אחרת נמשיך.
התהליך סופי כי כל פעם מורידים ε .

□

כעת בעזרת הלמה נוכיח את המשפט:

הוכחה. יהי $\varepsilon > 0$. לפי הלמה קיימים $x_1, \dots, x_n$ כנ"ל.

$$\mathbb{P}(|\mu_n((-\infty, x_k)) - \mu((-\infty, x))| \leq \varepsilon) \xrightarrow{n \rightarrow \infty} 1$$

אפשר גם לכתוב במונחי פונקציות:

$$\mathbb{P}(|F_n(x_k-) - F(x-)| \leq \varepsilon) \xrightarrow{n \rightarrow \infty} 1$$

$$F(x) = \mathbb{P}(X_1 \leq x) \quad F(x-) = \lim_{\varepsilon \rightarrow 0+} F(x - \varepsilon) = \mathbb{P}(X < x)$$

וההבדל הוא אם יש אטום בנקודה.
השכיחות שואפת להסתברות. לפי החוק החזק של המספרים הגדולים. המאורעות $A_i = \{X_i \in (-\infty, x_k)\}$ כעת,

$$\begin{aligned} \mathbb{P}(\forall k = 1, \dots, n |\mu_n((-\infty, x_k)) - \mu((-\infty, x))| \leq \\ \leq \varepsilon, |\mu_n((-\infty, x_k]) - \mu((-\infty, x])| \leq \varepsilon) \xrightarrow{n \rightarrow \infty} 1 \end{aligned}$$

כי מדובר בהסתברות של חיתוך מאורעות.
נותר להוכיח שבכל נקודה שזה מתקיים זה גורר:

$$\sup_{x \in \mathbb{R}} |F_n() - F()| \leq 3\varepsilon$$

וזה יסיים.

אם x הוא אחד מהבחירה שלנו אז ברור וטוב.
נניח $x \in (x_k, x_{k+1})$. זה לא פייר כי יכול להיות ש- $x \in (-\infty, x_1)$ או $x \in (x_n, \infty)$. נטפל אח"כ.

$$\mu((-\infty, x]) \in [\mu((-\infty, x_k]), \mu((-\infty, x_{k+1}))]$$

זה טריוויאלי ממונוטוניות המידה.
זה נכון גם:

$$\begin{aligned} \mu_n((-\infty, k]) &\in [\mu_n((-\infty, x_k]), \mu_n((-\infty, x_{k+1}))] \subset \\ &\subset [\mu((-\infty, x_k)) - \varepsilon, \underbrace{\mu((-\infty, x_{k+1})) + \varepsilon}_{\mu((-\infty, x_k)) + \varepsilon}] \subset \mu((-\infty, x_k)) + [-\varepsilon, 2\varepsilon] \end{aligned}$$

לכן $\mu_n((-\infty, k])$ בתוך קטע באורך 3ε . באופן דומה,

$$\mu((-\infty, x]) \in [\mu((-\infty, x_k]), \mu((-\infty, x_{k+1}))] \subset \mu((-\infty, x_k)) + [0, \varepsilon]$$

הקטע הזה מוכל בקטע למעלה לכן המרחק הוא לכל היותר 3ε (ואפילו 2ε אם מתאמצים קצת). לכן,

$$|\mu_n((-\infty, x]) - \mu((-\infty, x])| \leq 3\varepsilon$$

בפרט,

$$\sup |\mu_n((-\infty, x]) - \mu((-\infty, x])| \leq 3\varepsilon$$

□

משפט הגבול המרכזי

מדובר על מ"מ ב"ת ש"ה כך ש- $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$. ורוצים להראות:

$$\mathbb{P}(a\sqrt{n} \leq X_1 + \dots + X_n \leq b\sqrt{n}) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du$$

אפשר לשאול גם שאלות קשות יותר, לגבי מה קורה אם המשתנים תלויים או אם לא ש"ה אבל אנחנו נתעניין במקרה הנ"ל בלבד. יש המון הוכחות למשפט הזה. יש משפט שאומר

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{x_1 + \dots + x_n}{\sqrt{n}} \right)^k = \frac{1}{\sqrt{2\pi}} \int u^k e^{-u^2/2} du$$

זה יהיה החלק הקל. החלק הקשה יהיה להראות שאפשר לקרב פונקציה אופיינית ע"י פולינום. אבל כשהפולינום ישתחרר הוא יעוף לאינסוף מאוד מהר. אפשר גם בגישה אחרת: נסתמך

$$\lim_{n \rightarrow \infty} \mathbb{E} e^{i\lambda \frac{X_1 + \dots + X_n}{\sqrt{n}}} = e^{-\lambda^2/2}$$

$\lambda \in \mathbb{R}$.

ואז אפשר לקרב עם פולינום טריגונומטרי והוא לא יעוף כשישתחרר אלא יישאר חסום. בכל זאת אפשר להשתמש בהוכחה עם מומנטים אם אין אי תלות אלא יש איזשהו שקר בסגנון אי-תלות מקורבת.

נשים לב שבהמון ספרים מספיקים להוכיח רק חלק ראשון בלי לקרב את הפונקציה האופיינית. אבל זה בספרים של לא מתמטיקאים. יש הוכחה שהופיע בתחילת המאה ה-20 שמתחילה בסגנון:

$$\lim \mathbb{E} f\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(u) e^{u^2/2} dt$$

כאשר f, f', f'', f''' קיימות ורציפות ו- f''' בכלל חסומה. זה קצת כבר להוכיח, אבל עכשיו ממש קל לקרב את הפונקציה האופיינית שלנו ע"י פונקציה חלקה כזאת. וריאציה של זה נוכיח:

$$|\mathbb{E} f\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right) - \mathbb{E} f\left(\frac{\tilde{X}_1 + \dots + \tilde{X}_n}{\sqrt{n}}\right)| \xrightarrow{n \rightarrow \infty} 0$$

כלומר נוכיח שבכלל אין תלות בהתפלגות ההתחלתית. ואז מה? ואז פשוט ניקח התפלגות נורמלית מראש. מה היתרון של זה? אפשר להוכיח את זה לא רק ל- $\mathbb{R}$ אלא גם לדברים כללים יותר כמו חבורת הסיבובים או חבורות ליי. נתחיל להתקדם לכיוון ההוכחה. נתחיל ממושג הקיפול (או בשפת בני-תמותה קונבולוציה).

הגדרה. נניח $f: \mathbb{R} \rightarrow \mathbb{R}$ רציפה וחסומה ו- μ התפלגות. נגדיר קיפול $\mu * f$ באופן הבא:

$$(\mu * f)(a) = \int f(a+x) \mu(dx)$$

מה זה אומר? נסתכל על מקרה פשוט, מידה אטומית.

$$\mu = \delta_b$$

$$(\delta_b * f)(a) = \int f(a+x) \delta_b(dx) = f(a+b)$$

זה פשוט הזה.

אם המידה היא קומבינציה של אטומים אז נקבל שקיפול הוא קומבינציה לינארית של הזות של f .

אם μ קוצבניציה אינטגרלית אז קיפול הוא קומבינציה אינטגרלית של הזות. מה זה אומר בשפה ההסתברותית?

$$P_x * f(a) = \mathbb{E} f(a+X)$$

מדוע? מפני ש- $\mathbb{E} f(a+X)$ תוחלת של פונקציה של מ"מ:

$$\mathbb{E} f(a+X) = \int_{\mathbb{R}} f(a+x) P_x(dx)$$

וזה אנחנו יודעים מתורת המידה.

עכשיו מעניין לשאול עבור X, Y מ"מ ב"ת:

$$p_{X+Y} * f = P_Y * P_X * f$$

נוכיח את זה:

$$\begin{aligned}(p_{X+Y} * f)(a) &= \mathbb{E}f(a + X + Y) = \int \int_{\mathbb{R}} f(a + x + y) P_X(dx) P_Y(dy) = \\ &= \int \underbrace{\left(\int f(a + y + x) P_x(dx) \right)}_{(P_X * f)(a+y)} P_Y(dy) = P_Y * (P_X * f)(a)\end{aligned}$$

מה לא בסדר? הגדרנו קיפול בין פונקציה רציפה וחסומה לפונקציה הסתברות. מה מבטיח שאחרי הקיפול יש פונקציה רציפה וחסומה? אנחנו נוכיח את זה בהמשך. מה שכן היה אפשר להגדיר קיפול לכל זוג פונקציות אבל זה סתם להיות לא עקביים עכשיו. הנחנו

$$p_{X+Y} * f = P_Y * (P_X * f)$$

אולי

$$p_{X+Y} * f = (P_Y * P_X) * f$$

חשוב להגדיר נכון: נגדיר: לכל קבוצת בורל X, Y, B ב"ת:

$$(\mu * \nu)(B) = (\mu \times \nu)(\{(x, y) \in \mathbb{R}^2 : x + y \in B\})$$

וזה לפי האינטואיציה ההסתברותית: $(P_X * P_Y) = \mathbb{P}(X + Y \in B)$

עבור X, Y ב"ת $X + Y \sim P_X * P_Y$

כלומר $P_{X+Y} = P_X * P_Y$.

מקרה בדיד:

$$\mathbb{P}(X + Y = z) = \sum_{(x,y): x+y=z} P(X=x)P(Y=y) = \sum_x P_X(\{x\})P_Y(\{z-x\})$$

ואז,

$$P_{X+Y}(z) = \int_{\mathbb{R}} P_X(x)P_Y(z-x)dx$$

קיפול של התפלגות בינומטית

$$Binom(n, p) * Binom(m, p) = Binom(m + n, p)$$

ממש מהגדרת התפלגות בינומית שהיא סכום אינדיקטורית. אבל גם,

$$N(a_1, \sigma_1^2) * N(a_2, \sigma_2^2) = N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$$

מה לגבי להוכיח את זה? קל לראות שמספיק להוכיח עבור תוחלות שוות 0:

$$N(0, \sigma_1^2) + N(0, \sigma_2^2) = N(0, \sigma_1^2 + \sigma_2^2)$$

כי מדובר בהזזה.

אפשר לחשב ישירות עם אינטגרלים. לא קשה אבל לא נעים. דרך אלגנטית:
 X, Y ב"ת מתפלגים $N(0, 1)$. צפיפות במישור:

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

וזה נס כי ההתפלגות אינווריאנטית ביחס לסיבוב. ואז $X \cos \alpha + Y \sin \alpha \sim N(0, 1)$ וזה לכל α . זה כמעט מה שרצינו:

$$N(0, \cos^2 \alpha) * N(0, \sin^2 \alpha) = N(0, 1)$$

תמיד נוכל למצוא α כך ש- $\sigma_1 = r \cos \alpha, \sigma_2 = r \sin \alpha$. כמובן שנצטרך גם להכפיל ב- r אבל זה מתיחה ולכן לא משפיעה.
 נחזור לתיאוריה כללית של קיפול:

למה. אם f רציפה וחסומה אז גם $\mu * f$ רציפה וחסומה ואפילו $\|\mu * f\| \leq \|f\|$ כאשר מדובר בנומרת סופרמום:

$$\|f\| = \sup \|f(\cdot)\|$$

הוכחה. אם חסימות אין בעיה: $X \sim \mu$:

$$|(\mu * f)(a)| = |\mathbb{E}f(a + X)| \leq \sup |f(\cdot)|$$

מה עם רציפות?

יהי $a_n \rightarrow a$ אז $\mathbb{E}(a_n + X) \xrightarrow{n \rightarrow \infty} \mathbb{E}f(a + X)$ ממשפט ההתכנסות החסומה יש התכנסות תוחלות. $\square$

למה. במקרה ש- f, f' רציפות וחסומות אז קיימת $(\mu * f)'$ רציפה וחסומה.