הרצאה 2

1 אי-שוויונות ריכוז ומשפטי גבול

משפט 1.1: אי שוויונות צ'רנוף-הופדינג:

יהיו $X_1 \dots X_n$ מ"מ בת"ל כך ש: $\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = 1/2$ לכל $1 \leq i \leq n$. ויהי $X = \sum_{i=1}^{n} X_i$. אזי, לכל $t > 0$ מתקיים:

$$\mathbb{P}(X \geq t) \leq e^{-t^2/(2n)}$$

$$\mathbb{P}(X \leq -t) \leq e^{-t^2/(2n)}$$

(התוחלת היא 0).

החסם פחות טוב מאשר אי"ש צ'רנוף א, (פי 4 במעריך) אבל ההוכחה הרבה יותר פשוטה.

הרעיון המרכזי של ההוכחה: אנחנו רוצים לחסום את ההסתברות של המאורע $\mathbb{P}(X \geq t)$ ע"י אי"ש מרקוב. אבל X הוא לא אי"ש. אז עוברים למשתנה מקרי אחר, $e^{\lambda X}$ שהוא תמיד חיובי. אז נבצע: $\mathbb{P}(X \geq t) = \mathbb{P}(e^{\lambda X} \geq e^{\lambda t})$ וזה יאפשר לנו להשתמש במרקוב ולחסום את ההסתברות עם התוחלת (של המ"מ החדש) ולחסום עם חסם עליון על התוחלת: $\mathbb{P}\left(e^{\lambda X} \geq e^{\lambda t}\right) \leq \frac{\mathbb{E}(e^{\lambda X})}{e^{\lambda t}} \leq$ $e^{\frac{\lambda^2 n}{2} - \lambda t}$, ומשם יהיה נוח להגיע לחסם של המשפט.

ההוכחה משתמשת בטענה הבאה:

למה 1.2: לכל מספר ממשי $\lambda > 0$ מתקיים:

$$\frac{e^{\lambda} + e^{-\lambda}}{2} \leq e^{\lambda^2/2}$$

ההוכחה בעזרת טורי טיילור של שלושת הפונקציות:

$$e^{\lambda} + e^{-\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} + \sum_{n=0}^{\infty} \frac{\overbrace{(-1)^n \lambda^n}^{(-\lambda)^n}}{n!} =^{ב} 2 \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{(2n)!} \leq^{א} 2 \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{n! \cdot 2^n} =^{ג} 2 \sum_{n=0}^{\infty} \frac{\left(\frac{\lambda^2}{2}\right)^n}{n!} = 2e^{\lambda^2/2}$$

א – כי $(2n)! \geq n! \cdot 2^n$ מתקיים לכל n שלם טבעי[1].
ב – כל האיברים עם n אי זוגי מצטמצמים (חיובי בראשון, שלילי בשני). אז זה בעצם כל ה-n הזוגיים, פעמיים.
ג – הביטוי השמאלי זה פשוט טור טיילור של הפונקציה השלישית.

הוכחת משפט 1.1: נוכיח ש: $\mathbb{P}(X \geq t) \leq e^{-t^2/2n}$. לכל שלם ממשי $\lambda > 0$, $1 \leq i \leq n$ מתקיים:

$$\mathbb{E}\left(e^{\lambda X_i}\right) = \frac{e^{\lambda} + e^{-\lambda}}{2} \leq e^{\lambda^2/2}$$

ע"פ הגדרת התוחלת, וע"פ טענת העזר. מכיוון שכל ה- X_i בת"ל ע"פ ההנחה, גם המ"מ $e^{\lambda X_i}$ הם בת"ל. אז:

$$\mathbb{E}\left(e^{\lambda X}\right) = \mathbb{E}\left(e^{\lambda \sum_{i=1}^{n} X_i}\right) =^{א} \mathbb{E}\left(\prod_{i=1}^{n} e^{\lambda X_i}\right) =^{ב} \prod_{i=1}^{n} \mathbb{E}\left(e^{\lambda X_i}\right) \leq^{ג} e^{\lambda^2 n/2}$$

א – כי $e^{a+b} = e^a \cdot e^b$.
ב – כי כל ה- $e^{\lambda X_i}$ בת"ל.
ג – ע"פ טענת העזר.

ולכן: (1)

---

[1] אפשר להוכיח אלגברית, אבל גם: $(2n)!/n!\,2^n$ זה מספר הדרכים לחלק 2n אנשים ל-n זוגות. אז זה מספר שלם חיובי.

$$\mathbb{P}(X \geq t) =^{א} \mathbb{P}\left(e^{\lambda X} \geq e^{\lambda t}\right) \leq^{ב} \frac{\mathbb{E}\left(e^{\lambda X}\right)}{e^{\lambda t}} \leq e^{\frac{\lambda^2 n}{2} - \lambda t}$$

א – כי $f(x) = e^{\lambda x}$ היא פונקציה מונוטונית עולה. אז אם $X \geq t$, גם $e^{\lambda X} \geq e^{\lambda t}$.
ב – אי"ש מרקוב, מכיוון שהמ"מ החדש הוא אי-שלילי.

אנחנו רוצים למצוא את החסם הטוב ביותר. מכיוון שאי השוויון מתקיים לכל למדא חיובי, אנחנו רוצים להקטין את:

$$g(\lambda) = e^{\frac{\lambda^2 n}{2} - \lambda t}$$

נגזור לפי למדא ונקבל:

$$g'(\lambda) = (\lambda n - t) e^{\frac{\lambda^2 n}{2} - \lambda t}$$

מכיוון שהחלק של e תמיד חיובי, 0 מתקבל רק כאשר $\lambda = t/n$. וניתן לראות שזו נקודת מינימום כי לכל $\varepsilon > 0$ מתקיים:
$g'(t/n + \varepsilon) > 0$, $g'(t/n - \varepsilon) < 0$. כלומר הפונקציה יורדת לפני ועולה אחרי. נציב $\lambda = t/n$ ב-(1):

$$\mathbb{P}(X \geq t) \leq e^{\frac{t^2}{2n} - \frac{t^2}{n}} = e^{-\frac{t^2}{2n}}$$

כנדרש.

1.1 משפטי גבולות:

משפט 1.3: החוק החלש של המספרים הגדולים.

יהי $\{X_i\}_{i=1}^{\infty}$ סדרה של מ"מ בת"ל, לכולם אותה התפלגות ותוחלת סופית μ. אזי, לכל $\varepsilon > 0$ מתקיים:

$$\lim_{n \to \infty} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

באופן אינטואיטיבי, ה"ממוצע" של המ"מ שואף לתוחלת. יותר פורמלית: ככל שסכום יותר מ"מ, ההסתברות שההפרש בין הממוצע שלהם לתוחלת גדול מאפסילון, שואפת לאפס.

ההבדל בין זה לצ'בישב: צ'בישב הוא תוצאה כמותית - חוסם במספר קבוע עבור מספר מסוים של מ"מ. החוק הזה הוא תוצאה איכותית – ככל שיהיו יותר מ"מ, נתכנס למשהו מסוים.

נוכיח מקרה פרטי של המשפט, שבו יש לכל המ"מ שונות סופית σ^2:

לפי לינאריות התוחלת, מתקיים לכל n טבעי: (2)

$$\mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n)}{n} = \mu$$

ומכיוון שכולם בת"ל, מתקיים: (3)

$$Var\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\overbrace{Var(X_1)}^{\sigma^2} + \cdots + Var(X_n)}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

ולכן, לכל $\varepsilon > 0$ מתקיים ע"פ אי"ש צ'בישב:

$$\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) =^{2} \mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right)\right| \geq \varepsilon\right)$$
$$\leq^{א} \frac{Var\left(\frac{X_1 + \cdots + X_n}{n}\right)}{\varepsilon^2} =^{3} \frac{\sigma^2}{\varepsilon^2 n}$$

א – אי"ש צ'בישב.

ובסה"כ:

$$0 \leq \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{\varepsilon^2 n} \overset{\text{א}}{=} 0$$

א – כי σ סופי. (דרשנו שונות סופית).

ולכן:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-\mu\right| \geq \varepsilon\right) = 0$$

לסיכום, המשפט חלש יותר מצ'בישב במובן מסוים, כי הוא לא נותן חסם ספציפי. הוא רק אומר שאם נלך מספיק רחוק אז זה יסתדר.

משפט 1.4: החוק החזק של המספרים הגדולים.

יהי $\{X_i\}_{i=1}^{\infty}$ סדרה של מ"מ בת"ל, לכולם אותה התפלגות, תוחלת סופית μ, ושונות סופית. אזי, לכל $\varepsilon > 0$ מתקיים:

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1+\cdots+X_n}{n} = \mu\right) = 1$$

כלומר, ההסתברות שהממוצע שואף לתוחלת כאשר n שואף לאינסוף, היא 1. לכאורה כל מה שעשינו זה להחליף בין ההסתברות לגבול, אבל זאת החלפה לא טריוויאלית והיא משנה את המשמעות. בשני החוקים נטען שהסדרה של המ"מ מתכנסת ל-μ. ההבדל הוא בצורת ההתכנסות. כדי להסביר את זה, נגדיר שתי צורות התכנסות:

הגדרה 1.5: נאמר שסדרה $\{X_n\}_{n=1}^{\infty}$ של מ"מ **מתכנסת בהסתברות** למ"מ X כאשר n שואף לאינסוף, אם לכל $\varepsilon > 0$ מתקיים:

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) = 0$$

את החלק השמאלי נכתוב בדרך כלל ככה: $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon)$

כלומר, ההסתברות שההפרש בין X_n ל- X גדול מאפסילון, שואפת לאפס כאשר n שואף לאינסוף. נסמן:

$$X_n \xrightarrow{p} X$$

הגדרה 1.6: נאמר שסדרה $\{X_n\}_{n=1}^{\infty}$ של מ"מ **מתכנסת כמעט בוודאות** למ"מ X כאשר n שואף לאינסוף, אם מתקיים:

$$\mathbb{P}\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

אם מרחב המדגם סופי, זה אומר שלכל מאורע, כאשר n שואף לאינסוף, $X_n(\omega) = X(\omega)$. אם מרחב המדגם אינסופי אז זה קורה ב"כמעט כולם".

כלומר, ההסתברות שנקבל מאורע שעבורו הסדרה שואפת לX, היא 1. נסמן:

$$X_n \xrightarrow{a.s} X$$

באופן כללי, התכנסות "כמעט בוודאות" גוררת התכנסות בהסתברות. בכיוון ההפוך זה לא תמיד נכון, לדוגמה:

יהי $X \equiv 0$ מ"מ ותהי $\{X_n\}_{n=1}^{\infty}$ סדרה בת"ל של מ"מ כך ש $\mathbb{P}(X_n = 1) = 1/n$, $\mathbb{P}(X_n = 0) = 1 - 1/n$ לכל n טבעי. יהי $\varepsilon > 0$ כלשהו, אז מתקיים:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \lim_{n\to\infty} \mathbb{P}(X_n = 1) = \lim_{n\to\infty} 1/n = 0$$

כלומר, $X_n \xrightarrow{p} X$.

מצד שני, נב"ש ש $X_n \xrightarrow{a.s} X$. ע"פ הגדרה 1.6 זה אומר שבהסתברות 1, לכל אפסילון חיובי מתקיים $|X_n - X| \leq \varepsilon$ לכל n החל ממקום מסוים. כלומר לפי ההתפלגות של ה-X_n, בהסתברות 1 מתקיים $X_n = 0$ לכל n החל ממקום מסוים. אבל, לכל m טבעי מתקיים שההסתברות ש $X_n = 0$ לכל $n > m$ היא:

$$\mathbb{P}(\forall n > m, X_n = 0) =^{א} \prod_{n=m}^{\infty}(1 - \frac{1}{n}) \leq^{ב} \prod_{n=m}^{\infty} e^{-1/n} = e^{\overbrace{-\sum_{n=m}^{\infty} 1/n}^{-\infty}} = 0$$

א - מכיוון שהמ"מ כולם בת"ל.
ב – כי $1 - x \leq e^{-x}$.

בסתירה לכך שזה קורה בהסתברות 1.

דוגמה 1: נטיל מטבע n פעמים באופן בת"ל. בהרצאה הקודמת ראינו כלים למציאת על ההסתברות שנקבל עץ הרבה יותר או פחות מחצי מהפעמים. הכלים האלה נתנו תוצאה כמותית – לכל n מסוים, נתנו חסם מסוים. משפטים 1.3, 1.4 נותנים תוצאה איכותית – כלומר, לא נותנים חסם מסוים עבור אף n אלא מבטיחים תוצאה מסוימת עבור n מספיק גדול.
החוק החלש אומר שעבור כל אפסילון ודלתא חיוביים, אם נטיל את המטבע מספיק פעמים אז בהסתברות לפחות $1 - \delta$ נקבל עץ לפחות $n(1/2 - \varepsilon)$ ולכל היותר $n(1/2 + \varepsilon)$ פעמים.

יש עוד דוגמאות בקובץ המקורי.