הסתברות מתמטית
מערכי תרגול קורס 88-373

יוני 2024

אוניברסיטת בר־אילן

סמסטר ב' תשפ"ד

תוכן העניינים

מבוא **4**

1 תרגול ראשון **5**
1.1 אלגברות, σ-אלגברות ומידות הסתברות 5
1.2 משתנים מקריים 7

2 תרגול שני **8**
2.1 מידת סטילטיס 8
2.2 התוחלת והשונות 9
2.3 התרחשות אסימפטוטית 10

3 תרגול שלישי **11**
3.1 בורל-קנטלי 11
3.2 חוק ה-$0-1$ של קולמוגורוב 12

4 תרגול רביעי **14**
4.1 אי־שוויוני מרקוב וצ'בישב 14
4.2 מודל ארדש-רייני 14
4.3 אי־שוויון הופדינג 16

5 תרגול חמישי **17**
5.1 מומנטים ואי־שוויון צ'רנוף 17
5.2 סוגי התכנסויות 19

6 תרגול שישי **20**
6.1 החוק החלש והחוק החזק 20
6.2 התכנסות חלשה 22

7 תרגול שביעי **23**
7.1 משפט הגבול המרכזי 23
7.2 הפונקציה האופיינית 24

8 תרגול שמיני **26**
8.1 תוחלת מותנית 27

9 תרגול תשיעי **29**
9.1 מרטינגלים 30

10 תרגול עשירי **32**
10.1 זמני עצירה 32

11 תרגול אחד עשר **34**
11.1 מרטינגלים וריכוז מידה 34
11.2 התכנסות מרטינגלים 36

12 תרגול שנים עשר **39**
12.1 צימודים 39
12.2 שליטה סטוכסטית 42

13 תרגול שלושה עשר **43**
13.1 תהליכי פואסון 43

14 תרגול ארבעה עשר **45**
14.1 אנטרופיה 45

מבוא

נתחיל עם כמה הערות:

- דף הקורס נמצא באתר למדה.
- שאלות בנוגע לחומר הלימודי מומלץ לשאול בפורום הנמצא במודל.
- יפורסמו תרגילי בית כל שבוע, ויהיו תרגילים להגשה.
- החומר בקובץ זה נאסף מכמה מקורות, ומבוסס בעיקרו על מערכי תרגול קודמים בקורס הסתברות מתמטית באוניברסיטת בר-אילן.
- נשמח לכל הערה על מסמך זה.

מחבר בשנת הלימודים תשפ"א: גיא בלשר
עדכונים בשנת הלימודים תשפ"ב, תשפ"ג ותשפ"ד: גיא בלשר

1 תרגול ראשון

1.1 אלגברות, σ-אלגברות ומידות הסתברות

הגדרה 1.1. בהינתן קבוצה Ω, משפחה של תת־קבוצות $\mathcal{F} \subseteq P(\Omega)$ נקראת **אלגברה**, אם:

א. $\varnothing \in \mathcal{F}$.

ב. סגירות למשלים: אם $A \in \mathcal{F}$ אז $A^c \in \mathcal{F}$.

ג. סגירות לאיחודים סופיים: אם $A, B \in \mathcal{F}$ אז $A \cup B \in \mathcal{F}$.

$\mathcal{F}$ נקראת **σ-אלגברה** אם היא סגורה לאיחודים בני-מנייה. במקרה זה, המרחב $(\Omega, \mathcal{F})$ נקרא **מרחב מדיד**.

הערה. משילוב התכונות נקבל שאלגברה סגורה גם לחיתוכים סופיים, ו-σ-אלגברה סגורה גם לחיתוכים בני-מנייה.

תרגיל 1.2 (בהרצאה)**.** תהי Ω קבוצה לא ריקה, ותהי $\{\mathcal{F}_i\}_{i \in I}$ משפחה של σ-אלגברות. הוכיחו כי $\mathcal{F} = \bigcap_{i \in I} \mathcal{F}_i$ היא σ-אלגברה על Ω.

הוכחה.

א. כיוון ש-$\varnothing \in \mathcal{F}_i$ לכל $i \in I$, $\varnothing \in \mathcal{F}$.

ב. נניח כי $A \in \mathcal{F}$. אז לכל $i \in I$ מתקיים $A \in \mathcal{F}_i$. לכן $A^c \in \mathcal{F}_i$ לכל $i \in I$, ומכאן $A^c \in \mathcal{F}$.

ג. תהי $A_1, A_2, \cdots \in \mathcal{F}$. לכן לכל $i \in I$ מתקיים $A_1, A_2, \cdots \in \mathcal{F}_i$, מכאן $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}_i$, ולכן $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

□

הגדרה 1.3. תהי Ω קבוצה ו-$\mathcal{A} \subseteq P(\Omega)$ אוסף של תת־קבוצות שלה. **ה-σ-אלגברה הנוצרת על ידי $\mathcal{A}$**, שתסומן $\sigma(\mathcal{A})$, היא ה-σ-אלגברה המינימלית המכילה את $\mathcal{A}$. באופן שקול,

$$\sigma(\mathcal{A}) = \bigcap_{\mathcal{F}: \mathcal{A} \subseteq \mathcal{F}} \mathcal{F}$$

כאשר החיתוך רץ על פני כל ה-σ-אלגברות $\mathcal{F}$ המכילות את $\mathcal{A}$.

הגדרה 1.4. **σ-אלגברת בורל** על מרחב טופולוגי (X, τ), שתסומן $\mathbb{B}(X)$, היא ה-σ-אלגברה הנוצרת על ידי הקבוצות הפתוחות של X.

דוגמה 1.5. $\mathbb{B}(\mathbb{R})$ היא ה-σ-אלגברה הנוצרת על ידי כל הקטעים הפתוחים, או על ידי כל הקטעים $[a, b)$, או על ידי כל הקטעים $[a, \infty)$...

הגדרה 1.6. יהי $(\Omega, \mathcal{F})$ מרחב מדיד. פונקציה $\mathrm{P}: \mathcal{F} \to [0,1]$ נקראת **מידת הסתברות** על $(\Omega, \mathcal{F})$, אם היא מקיימת את התנאים הבאים:

א. $P(\Omega) = 1$.

ב. לכל $A_1, A_2, \cdots \in \mathcal{F}$ סדרת קבוצות זרות בזוגות, $\mathrm{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathrm{P}(A_n)$.

המרחב $(\Omega, \mathcal{F}, \mathrm{P})$ נקרא **מרחב הסתברות**.

תרגיל 1.7 (בעיית ברטרנד)**.** מגרילים מיתר במעגל לפי אחת משלוש הדרכים הבאות. מה הסיכוי שהוא יהיה ארוך יותר מאורך צלע המשולש שווה הצלעות החסום במעגל?

א. בוחרים את המיתר דרך נקודות קצה אקראיות;

ב. מקצים את המיתר במאונך לנקודה אקראית על רדיוס אקראי;

ג. בוחרים את נקודת האמצע של המיתר באקראי.

פתרון. נניח בה"כ שרדיוס המעגל הוא 1; לכן אורך הצלע a של המשולש שווה הצלעות החסום בו מקיימת $a^2 = \left(\frac{1}{2}a\right)^2 + \left(\frac{3}{2}\right)^2$, כלומר $\frac{3}{4}a^2 = \frac{9}{4}$, ולכן $a = \sqrt{3}$.

א. בגלל הסימטריה של מעגל, אפשר להניח שהנקודה הראשונה היא $(1,0)$, ומגרילים באופן אחיד את הנקודה השנייה. כדי שאורך המיתר ביניהן יהיה יותר מאורך הצלע של המשולש שווה הצלעות החסום במעגל, צריך שהזווית המרכזית הנשענת על המיתר תהיה בין $\frac{2\pi}{3}$ ל-π . לכן

$$\mathrm{P}\left(\text{המיתר אורך} > \sqrt{3}\right) = \mathrm{P}\left(\theta \in \left(\frac{2\pi}{3}, \pi\right)\right) = \frac{1}{3}.$$

ב. גם פה בגלל הסימטריה הגרלת הרדיוס עצמו לא משנה. נחשב את אורך הצלע כביטוי של המרחק על הרדיוס: אם אנחנו בוחרים נקודה במרחק r ממרכז המעגל, האורך של המיתר המתאים (על פי פיתגורס) הוא $2\sqrt{1-r^2}$. לכן

$$2\sqrt{1-r^2} > \sqrt{3} \iff 4\left(1-r^2\right) > 3 \iff r^2 < \frac{1}{4} \iff r < \frac{1}{2}.$$

לכן הסיכוי הוא $\frac{1}{2}$.

ג. נניח שמגרילים את נקודת האמצע (x,y). המרחק שלה ממרכז המעגל הוא $\sqrt{x^2+y^2}$, ולכן אורך המיתר המתאים הוא $2\sqrt{1-x^2-y^2}$. בדומה לחישוב הקודם,

$$2\sqrt{1-x^2-y^2} > \sqrt{3} \iff x^2+y^2 < \frac{1}{4}.$$

לכן השאלה היא מה הסיכוי שנקודה אקראית במעגל היחידה תיפול בתוך מעגל מרדיוס $\frac{1}{2}$, והסיכוי הזה הוא יחס השטחים ששווה ל-$\frac{1}{4}$.

1.2 משתנים מקריים

תזכורת. יהיו $(\Omega_1, \mathcal{F}_1)$ ו-$(\Omega_2, \mathcal{F}_2)$ מרחבים מדידים. אומרים שפונקציה $f : \Omega_1 \to \Omega_2$ היא **מדידה**, אם לכל $B \in \mathcal{F}_2$ מתקיים $f^{-1}(B) \in \mathcal{F}_1$. לצורך הזהירות, נגיד ש-$f$ היא $(\mathcal{F}_1, \mathcal{F}_2)$-מדידה.

הגדרה 1.8. יהי $(\Omega, \mathcal{F})$ מרחב מדיד. **משתנה מקרי** על Ω הוא פונקציה מדידה $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathbb{B}(\mathbb{R}))$.

הגדרה 1.9. יהי $X : \Omega \to \mathbb{R}$ משתנה מקרי. **ההתפלגות** (או **החוק**) של X היא

$$\mathcal{L}_X(A) = \mathrm{P}(X(\omega) \in A)$$

(במילים אחרות, $\mathcal{L}_X = \mathrm{P} \circ X^{-1}$). **פונקציית ההתפלגות המצטברת** של X הינה

$$F_X(x) = \mathrm{P}(X \leq x).$$

טענה 1.10 (תכונות של פונקציית ההתפלגות המצטברת). יהי X משתנה מקרי.

א. F_X מונוטונית עולה.

ב. $\lim_{x\to-\infty} F_X(x) = 0$, $\lim_{x\to\infty} F_X(x) = 1$.

ג. F_X רציפה מימין, כלומר לכל $x \in \mathbb{R}$ מתקיים $\lim_{\varepsilon\to 0} F_X(x+\varepsilon) = F_X(x)$.

תרגיל 1.11. תהי F פונקציה המקיימת את שלוש התכונות הנ"ל. אזי קיים משתנה מקרי $X : ([0,1], \mathbb{B}([0,1])) \to \mathbb{R}$ שעבורו $F_X = F$.

הוכחה. הרעיון של ההוכחה הוא שאם X משתנה מקרי שעבורו F_X רציפה והפיכה, אז $F_X(X) \sim U[0,1]$. אכן, לכל $a \in [0,1]$ מתקיים

$$\mathrm{P}(F_X(X) \leq a) = \mathrm{P}\left(X \leq F_X^{-1}(a)\right) = F_X\left(F_X^{-1}(a)\right) = a.$$

נוכיח למקרה הכללי. נגדיר

$$F^{-1}(a) = \inf\{x \in \mathbb{R} \mid F(x) \geq a\}.$$

נשים לב כי:

- F^{-1} מונוטונית עולה.
- אם $t = F^{-1}(a)$, מהרציפות מימין נקבל $F(F^{-1}(a)) = F(t) \geq a$.
- מההגדרה נובע כי $F^{-1}(F(x)) \leq x$.

נגדיר את המשתנה המקרי $X = F^{-1}(U)$ עבור $U \sim U[0,1]$; כעת

$$\{U < F(x)\} \subseteq \left\{X = F^{-1}(U) \leq F^{-1}(F(x)) \leq x\right\} \subseteq \left\{U \leq F\left(F^{-1}(x)\right) \leq F(x)\right\}.$$

אך כיוון ש-U רציף, ההסתברות של שני המאורעות באגף ימין ובאגף שמאל שווה. לכן

$$\mathrm{P}(X \leq x) = \mathrm{P}(U \leq F(x)) = F(x)$$

כנדרש. □

2 תרגול שני

2.1 מידת סטילטיס

הגדרה 2.1. תהי $F:\mathbb{R}\to\mathbb{R}$ פונקציית הצטברות, כלומר פונקציה המקיימת את תנאי טענה 1.10. נגדיר את **מידת סטילטיס המתאימה ל-**F להיות המידה μ_F על $\mathbb{R}$ הנתונה על ידי $\mu_F((a,b])=F(b)-F(a)$ (ומורחבת ל-$\mathbb{B}(\mathbb{R})$ באופן היחיד האפשרי).

דוגמה 2.2.

א. אם ניקח $F(x)=x$, נקבל את מידת לבג הרגילה.

ב. אם $F=F_X$ פונקציית ההצטברות המצטברת של איזשהו משתנה מקרי X, אז

$$\mu_{F_X}((a,b])=F_X(b)-F_X(a)=P(X\le b)-P(X\le a)=P(a<X\le b).$$

כלומר μ_{F_X} היא בדיוק החוק של X.

צורת רישום 2.3. נרשום $\int_A g(t)\,\mathrm{d}\mu_F=\int_A g(t)\,\mathrm{d}F$. אם F רציפה בהחלט ב-$\mathbb{R}$, אז F גזירה כב"ם, $\mu_F\ll m$, ומתקיים $\int_A g(t)\,\mathrm{d}F(t)=\int_A g(t)F'(t)\,\mathrm{d}m$.

תרגיל 2.4. נניח ש-$F(t)=\begin{cases}0, & t<0\\ t^2, & 0\le t<\frac{1}{2}\\ 1 & t\ge\frac{1}{2}\end{cases}$. חשבו את $\int_{-\infty}^{\infty}t\,\mathrm{d}F(t)$.

פתרון. נחלק את $\mathbb{R}$ לארבעה קטעים: $A_1=(-\infty,0]$, $A_2=(0,\frac{1}{2})$, $A_3=\left\{\frac{1}{2}\right\}$, $A_4=(\frac{1}{2},\infty)$. מאדיטיביות האינטגרל, מספיק לחשב את האינטגרל על כל אחד מהקטעים האלו ולסכום את התוצאות.

- על הקטעים A_1 ו-A_4 הפונקציה F קבועה, לכן $\mu_F(A_1)=\mu_F(A_4)=0$, והאינטגרלים המתאימים מתאפסים.
- על A_2, הפונקציה F רציפה וגזירה, ולכן

$$\int_{A_2}t\,\mathrm{d}F(t)=\int_0^{\frac{1}{2}}2t\cdot t\,\mathrm{d}t=\left(\frac{2}{3}t^3\right)_0^{\frac{1}{2}}=\frac{1}{12}.$$

- על A_3 מתקיים $\mu_F(A_3)=1-\frac{1}{4}=\frac{3}{4}$, ולכן

$$\int_{A_3}t\,\mathrm{d}F(t)=\frac{1}{2}\cdot\frac{3}{4}=\frac{3}{8}.$$

בסך הכל $\int_{-\infty}^{\infty}t\,\mathrm{d}F(t)=0+\frac{1}{12}+\frac{3}{8}+0=\frac{11}{24}$.

2.2 התוחלת והשונות

הגדרה 2.5. יהי $X:\Omega\to\mathbb{R}$ משתנה מקרי. **התוחלת** של X הינה

$$\mathbb{E}[X]=\int_\Omega X\,\mathrm{d}F_X.$$

השונות של X הינה

$$\mathrm{Var}(X)=\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]=\mathbb{E}\left[X^2\right]-\mathbb{E}[X]^2.$$

תרגיל 2.6. יהי X משתנה מקרי אינטגרבילי המקיים $\mathrm{P}(X>0)=1$. אזי

$$\mathbb{E}[X]=\int_0^\infty \mathrm{P}(X>t)\,\mathrm{d}t.$$

בפרט, אם X בדיד ומקבל ערכים טבעיים, אז

$$\mathbb{E}[X]=\sum_{n=1}^\infty \mathrm{P}(X\geq n).$$

הוכחה. נחשב ישירות:

$$\mathbb{E}[X]=\int_\Omega X(\omega)\,\mathrm{dP}(\omega)=\int_{-\infty}^\infty t\,\mathrm{d}F_X(t)=\int_0^\infty t\,\mathrm{d}F_X(t)=\int_0^\infty\left(\int_0^t \mathrm{d}s\right)\mathrm{d}F_X(t)=$$
$$=\iint_{\{(s,t)|0\leq s\leq t\}}\mathrm{d}s\,\mathrm{d}F_X(t)=\int_0^\infty\left(\int_s^\infty dF_X(t)\right)\mathrm{d}s=\int_0^\infty \mathrm{P}(X>s)\,\mathrm{d}s.$$

□

תרגיל 2.7. לכל $f:\mathbb{R}\to\mathbb{R}$ מדידה בורל מתקיים $\mathbb{E}[f(X)]=\int_{-\infty}^\infty f(t)\,\mathrm{d}F_X(t)$.

הוכחה.

א. עבור $f=a\cdot 1_A$ כאשר $A\in\mathbb{B}(\mathbb{R})$,

$$\mathbb{E}[a\cdot 1_A(X)]=a\cdot\mathrm{P}(X\in A)=a\cdot\int_A \mathrm{d}F_X=\int_\Omega a\cdot 1_A(t)\,\mathrm{d}F_X(t).$$

ב. מלינאריות נקבל שהטענה נכונה לכל פונקציה פשוטה $f=\sum a_i 1_{A_i}$.

ג. ממשפט ההתכנסות המונוטונית, הטענה נכונה לכל פונקציה מדידה חיובית.

ד. עבור f מדידה כללית אפשר לכתוב $f=f^+-f^-$, ומפה נובעת הטענה הכללית. ($f^+=\max\{f,0\}$, $f^-=\max\{-f,0\}$).

□

תזכורת. יהי $I\subseteq\mathbb{R}$ קטע (פתוח או סגור). אומרים שפונקציה $f:I\to\mathbb{R}$ היא **קמורה**, אם לכל $x,y\in I$ ולכל $\lambda\in[0,1]$ מתקיים $f((1-\lambda)x+\lambda y)\leq(1-\lambda)f(x)+\lambda f(y)$.
אם f גזירה פעמיים, אז f קמורה אם ורק אם $f''\geq 0$.

משפט 2.8 (אי־שוויון ינסן)**.** אם $f:\mathbb{R}\to\mathbb{R}$ קמורה, אז $f(\mathbb{E}[X])\leq\mathbb{E}[f(X)]$.

2.3 התרחשות אסימפטוטית

מוטיבציה. בחלק זה, השאלה שנרצה לשאול את עצמנו היא: בהינתן סדרת מאורעות $A_1, A_2, \ldots$, כמה מהם קורים בו-זמנית?

הגדרה 2.9. יהי $(\Omega, \mathcal{F}, P)$ מרחב הסתברות, ותהי $\{A_n\}_{n=1}^{\infty}$ סדרת מאורעות.

א. ה-$\limsup$ של הסדרה הינו אוסף האיברים $\omega \in \Omega$ הנמצאים באינסוף מה-A_n-ים. בנוסחה:

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m .$$

ב. ה-$\liminf$ של הסדרה הינו אוסף האיברים $\omega \in \Omega$ הנמצאים בכמעט כל A_n (כלומר, לא נמצאים רק במספר סופי של A_n-ים). בנוסחה:

$$\liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m .$$

ג. אומרים ש-$\{A_n\}$ מתרחשת **אינסוף פעמים** (**often infinitely**), או בסימון המקוצר i.o., אם $\mathrm{P}(\limsup A_n) = 1$. אחרת, אומרים שהסדרה $\{A_n\}$ מתרחשת **מספר סופי של פעמים** (**often finitely**).

משפט 2.10 (למת פאטו)**.** תהי $\{A_n\}$ סדרת מאורעות. אזי

$$\mathrm{P}(\{A_n \text{ a.e.}\}) = \mathrm{P}(\liminf A_n) \leq \liminf P(A_n),$$
$$\mathrm{P}(\{A_n \text{ i.o.}\}) = \mathrm{P}(\limsup A_n) \geq \limsup P(A_n).$$

תרגיל 2.11. תנו דוגמה שבה מקבלים אי־שוויונות חזקים

$$\mathrm{P}(\liminf A_n) < \liminf \mathrm{P}(A_n) < \limsup \mathrm{P}(A_n) < \mathrm{P}(\limsup A_n).$$

פתרון. במרחב $\Omega = \{0,1\}$ עם $\mathcal{F} = P(\Omega)$ והמידה $\mathrm{P}(\{0\}) = \frac{1}{3}$ ו-$\mathrm{P}(\{1\}) = \frac{2}{3}$, נגדיר סדרת מאורעות $\{A_n\}$ לפי

$$A_{2n} = \{0\}, A_{2n+1} = \{1\}.$$

במקרה זה מקבלים

$$\mathrm{P}(\liminf A_n) = 0$$
$$\liminf \mathrm{P}(A_n) = \frac{1}{3}$$
$$\limsup \mathrm{P}(A_n) = \frac{2}{3}$$
$$\mathrm{P}(\limsup A_n) = 1.$$

3 תרגול שלישי

3.1 בורל-קנטלי

הגדרה 3.1. אומרים שמאורע A קורה **כמעט תמיד**, אם $\mathrm{P}(A)=1$.

הגדרה 3.2. יהי $(\Omega,\mathcal{F},\mathrm{P})$ מרחב הסתברות.

א. אומרים שמאורעות $A_1,\ldots,A_n\in\mathcal{F}$ **בלתי־תלויים**, אם לכל $I\subseteq\{1,\ldots,n\}$ מתקיים

$$\mathrm{P}\left(\bigcap_{i\in I}A_i\right)=\prod_{i\in I}\mathrm{P}(A_i).$$

ב. אומרים ש-σ-אלגברות $\mathcal{F}_1,\ldots,\mathcal{F}_n$ הן **בלתי־תלויות**, אם לכל $A_i\in\mathcal{F}_i$ מתקיים ש-$A_1,\ldots,A_n$ בלתי־תלויים.

טענה 3.3. מספיק לבדוק את התנאים האלו על אלגברות שיוצרות את ה-σ-אלגברות.

משפט 3.4 (למות בורל קנטלי)**.** תהי $\{A_n\}$ סדרת מאורעות.

א. אם $\sum_{n=1}^{\infty}\mathrm{P}(A_n)<\infty$, אז $\mathrm{P}(\limsup A_n)=0$.

ב. אם $\sum_{n=1}^{\infty}\mathrm{P}(A_n)=\infty$ והמאורעות A_n בלתי־תלויים, אז $\mathrm{P}(\limsup A_n)=1$.

דוגמה 3.5. דרישת האי־תלות הכרחית בלמה השנייה של בורל-קנטלי. אכן, ניתן לקחת $\Omega=[0,1]$ עם σ-אלגברת בורל ומידת לבג, ואת המאורעות $A_n=\left[0,\frac{1}{n}\right]$. במקרה זה $\sum_{n=1}^{\infty}\mathrm{P}(A_n)=\sum_{n=1}^{\infty}\frac{1}{n}=\infty$ הטור, אבל $\mathrm{P}(\limsup A_n)=0$.

תרגיל 3.6. יהיו $X_1,X_2,\ldots$ משתנים מקריים בלתי־תלויים עם ההתפלגות

$$\mathrm{P}(X_n=-n)=1-\frac{1}{n^2},\qquad \mathrm{P}\left(X_n=n^3-n\right)=\frac{1}{n^2}.$$

הראו כי כמעט תמיד $X_n=-n$ פרט אולי למספר סופי של n-ים.

הוכחה. נשים לב כי

$$\sum_{n=1}^{\infty}\mathrm{P}\left(X_n=n^3-n\right)=\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty,$$

לפי הלמה הראשונה של בורל-קנטלי, $X_n=n^3-n$ מתרחש מספר סופי של פעמים כמעט תמיד, כלומר $X_n=-n$ מתרחש כמעט תמיד. □

תרגיל 3.7. יהיו $X_1,X_2,\ldots$ משתנים מקריים בלתי־תלויים המתפלגים $\mathrm{N}(0,1)$. נגדיר את המאורע

$$A_n=\{\max\{X_{n^2+1},\ldots,X_{n^2+2n}\}>5\}.$$

הראו כי $\mathrm{P}(\{A_n\text{ a.e.}\})=1$.

הוכחה. נעבור בשנית למשלים, שהרי $\mathrm{P}\left(\{A_n \text{ a.e.}\}\right) = 1 - \mathrm{P}\left(\{A_n^c \text{ i.o.}\}\right)$. נשים לב כי

$$\mathrm{P}\left(A_n^c\right) = \mathrm{P}\left(\max\left\{X_{n^2+1}, \ldots, X_{n^2+2n}\right\} \leq 5\right) = \prod_{i=1}^{2n} \mathrm{P}\left(X_{n^2+i} \leq 5\right).$$

המשתנים $\max\left\{X_{n^2+1}, \ldots, X_{n^2+2n}\right\}$ בלתי־תלויים (כי אין חפיפה בין ה-X-ים) ושווי התפלגות; אם נסמן $p = \mathrm{P}\left(X_{n^2+i} \leq 5\right)$, נקבל $\mathrm{P}\left(A_n^c\right) = p^{2n}$. לכן

$$\sum_{n=1}^{\infty} P\left(A_n^c\right) = \sum_{n=1}^{\infty} p^{2n} < \infty$$

(כי $p < 1$), וזה מראה שהמאורעות A_n^c קורים מספר סופי של פעמים. □

תרגיל 3.8. יהי $X \sim \mathrm{U}(0,1)$. הראו כי בהסתברות 1, כל רצף סופי של ספרות יופיע בייצוג העשרוני של X אינסוף פעמים.

הוכחה. ראשית, נוכיח את הטענה עבור רצף מסוים. הספרה ה-n אחרי הנקודה שלו הינה $X_n = \lfloor X \cdot 10^n \rfloor \pmod{10}$. ניתן לוודא ש-$X_1, X_2, \ldots$ הם משתנים מקריים בלתי־תלויים עם התפלגות אחידה על $\{0, \ldots, 9\}$. אם $d_1, \ldots, d_k$ הוא רצף ספרות כלשהו, נגדיר $A_n = \{(X_{n+1}, \ldots, X_{n+k}) = (d_1, \ldots, d_k)\}$. הבעיה: המאורעות A_n הם לא בלתי־תלויים, אפילו לא בזוגות (אלא אם $k = 1$). אז נסתכל במקום זאת על המאורעות $A_k, A_{2k}, A_{3k}, \ldots$. אלו מאורעות בלתי־תלויים, ומתקיים $P\left(A_{nk}\right) = 10^{-k}$. הטור מתבדר, ולכן בהסתברות 1 המאורעות A_{nk} מתרחשים אינסוף פעמים.

כעת, יהי $\mathcal{S}$ אוסף כל הרצפים הסופיים של ספרות, ולכל $s \in \mathcal{S}$ יהי B_s המאורע לפיו הרצף s מופיע אינסוף פעמים בייצוג העשרוני של X. הראינו שלכל $s \in \mathcal{S}$ מתקיים $\mathrm{P}(B_s) = 1$; כיוון ש-$\mathcal{S}$ בת-מנייה, נקבל $P\left(\bigcap_{s \in \mathcal{S}} B_s\right) = 1$, כנדרש. □

3.2 חוק ה-$0-1$ של קולמוגורוב

הגדרה 3.9. יהי $(\Omega, \mathcal{F})$ מרחב מדיד, ותהי $\{\mathcal{F}_n\}_{n=1}^{\infty}$ סדרת σ-אלגברות המוכלות ב-$\mathcal{F}$. נגדיר $\mathcal{T}_n = \sigma\left(\bigcup_{k=n}^{\infty} \mathcal{F}_k\right)$. ל-$\sigma$-אלגברה $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$ קוראים σ-**אלגברת הזנב** של $\{\mathcal{F}_n\}$, ולכל מאורע בה קוראים **מאורע זנב**.

משפט 3.10 (חוק ה-$0-1$ של קולמוגורוב ל-σ-אלגברות)**.** יהי $(\Omega, \mathcal{F}, \mathrm{P})$ מרחב הסתברות, ותהי $\{\mathcal{F}_n\}$ סדרת σ-אלגברות בלתי־תלויות המוכלות ב-$\mathcal{F}$. אזי לכל מאורע זנב A מתקיים $\mathrm{P}\left(A\right) \in \{0, 1\}$.

משפט 3.11 (חוק ה-$0-1$ של קולמוגורוב למשתנים מקריים)**.** יהיו $Y, X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים כך ש-Y הוא $\sigma\left(\{X_n\}_{n=1}^{\infty}\right)$-מדיד. אזי Y קבוע כמעט תמיד, כלומר קיים $c \in \mathbb{R}$ כך ש-$\mathrm{P}\left(\{\omega \in \Omega \mid Y\left(\omega\right) = c\}\right) = 1$.

הערה 3.12. אם Y הוא $\sigma\left(X_n\right)$-מדיד לכל n, אז הוא בפרט $\sigma\left(\{X_n\}_{n=1}^{\infty}\right)$-מדיד.

תרגיל 3.13. יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים, ותהי $S_n = \sum_{i=1}^{n} X_n$ סדרת הסכומים החלקיים שלהם. תהי $\{b_n\}$ סדרה כך ש־$b_n \to \infty$. הראו כי קיים $-\infty \leq c \leq \infty$ שעבורו $\mathrm{P}\left(\limsup \frac{S_n}{b_n} = c\right) = 1$.

הוכחה. נסתכל על המאורע $A_c = \left\{\limsup \frac{S_n}{b_n} \leq c\right\}$, ונראה שהוא מאורע זנב. אכן, מתקיים

$$\frac{S_{k+n}}{b_{k+n}} = \frac{S_{k-1}}{b_{k+n}} + \frac{X_k + \cdots + X_{k+n}}{b_{k+n}}.$$

לכן שינוי של k הערכים הראשונים משנה רק את המחובר הראשון. מאחר ש־$b_{k+n} \to \infty$, המחובר הזה בכל מקרה לא ישפיע על ה־$\limsup$.

מחוק ה־$0-1$ של קולמוגורוב, לכל $-\infty \leq c \leq \infty$ מתקיים $\mathrm{P}(A_c) \in \{0, 1\}$. נגדיר את הפונקציה $F : \mathbb{R} \cup \{\pm\infty\} \to \mathbb{R}$ לפי $F(c) = P(A_c)$; זו פונקציה מונוטונית עולה, רציפה מימין, וכן $F(\infty) = 1$. אם $F(-\infty) = 1$, אז ניקח $c = -\infty$ וסיימנו. אחרת, ניקח $c_0 = \sup\{c \in \mathbb{R} \mid F(c) = 0\}$. בגלל הרציפות מימין, $\mathrm{P}\left(\limsup \frac{S_n}{b_n} = c\right) = 1$. □

דוגמה 3.14. תהי $X_0, X_1, \ldots$ סדרת משתנים מקריים בלתי־תלויים. נסתכל על טור החזקות האקראי

$$f(z, \omega) = \sum_{n=0}^{\infty} X_n(\omega) z^n.$$

כזכור מאינפי 2, לכל טור חזקות יש רדיוס התכנסות שהוא

$$R(\omega) = \frac{1}{\limsup \sqrt[n]{|X_n(\omega)|}}.$$

אבל זהו משתנה מקרי שהוא $\mathcal{T}$־מדיד, ולכן קבוע כמעט תמיד.

תרגיל 3.15. אם בנוסף $X_0, X_1, \ldots$ הם שווי התפלגות ו־$0 \neq \mathbb{E}[|X_0|] < \infty$, אז $R(\omega) = 1$ כמעט תמיד.

הוכחה. נראה שהטור מתכנס לכל z עם $|z| < 1$ ומתבדר לכל z עם $|z| > 1$. אם $|z| < 1$, אז

$$\mathbb{E}\left[\sum_{n=0}^{N} |X_n(\omega) z^n|\right] = \mathbb{E}[|X_0|] \sum_{n=0}^{N} |z|^n \leq \mathbb{E}[|X_0|] \sum_{n=0}^{\infty} |z|^n.$$

ממשפט ההתכנסות המונוטונית, הטור $\sum_{n=0}^{\infty} |X_n(\omega) z^n|$ הוא בעל תוחלת סופית, ובפרט מתכנס כמעט תמיד. לכן גם הטור $\sum_{n=0}^{\infty} X_n(\omega) z^n$ מתכנס כמעט תמיד לכל $|z| < 1$.

נניח כי $r = |z| > 1$. נניח בשלילה כי $\sum_{n=0}^{\infty} X_n(\omega) z^n$ מתכנסת על קבוצה מהסתברות חיובית. מחוק ה־$0-1$ של קולמוגורוב, נקבל שהטור מתכנס כמעט תמיד.

המשתנים $X_0, X_1, \ldots$ שווי־התפלגות, ולכן קיים $c > 0$ כך ש-$\mathrm{P}(|X_n| \geq c) > 0$ לכל n. המאורעות האלו בלתי־תלויים, ולכן לפי הלמה של בורל-קנטלי נקבל ש-$|X_n z^n| \geq cr^n$ לאינסוף ערכי n כמעט תמיד, ולכן $X_n(\omega)z^n$ אינו מתכנס ל-0 כמעט תמיד, בסתירה. □

4 תרגול רביעי

4.1 אי־שוויוני מרקוב וצ'בישב

המטרה. להשיג הערכה כמה שיותר טובה למשתנה מקרי X על סמך כמה שפחות מידע.

טענה 4.1 (אי־שוויון מרקוב). לכל משתנה מקרי אי־שלילי X בעל תוחלת,

$$\mathrm{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}.$$

טענה 4.2 (אי־שוויון צ'בישב). לכל משתנה מקרי X בעל שונות,

$$\mathrm{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

תרגיל 4.3. נניח $X \sim \mathrm{Bin}(n,p)$. מצאו הערכה להסתברות $\mathrm{P}(X \geq \alpha n)$ עבור $p < \alpha < 1$ באמצעות אי־שוויון מרקוב ובאמצעות אי־שוויון צ'בישב.

פתרון. נזכור כי $\mathbb{E}[X] = np$ ו-$\mathrm{Var}(X) = np(1-p)$. מאי־שוויון מרקוב,

$$\mathrm{P}(X \geq \alpha n) \leq \frac{\mathbb{E}[X]}{\alpha n} = \frac{p}{\alpha}.$$

אי־שוויון צ'בישב ייתן

$$\mathrm{P}(X \geq \alpha n) = \mathrm{P}(X - np \geq (\alpha - p)n) \leq \mathrm{P}(|X - np| \geq (\alpha - p)n) \leq$$
$$\leq \frac{np(1-p)}{(\alpha - p)^2 n^2} = \frac{p(1-p)}{(\alpha - p)^2 n}.$$

4.2 מודל ארדש-רייני

הגדרה 4.4. גרף הוא זוג (V, E), כאשר V היא קבוצה של קודקודים ו-$E \subseteq V \times V$ היא קבוצה של קשתות. **גרף לא מכוון** הוא גרף שבו E סימטרי, כלומר $(u,v) \in E$ אם ורק אם $(v,u) \in E$. **גרף פשוט** הוא גרף שאין בו לולאות, כלומר $(v,v) \notin E$ לכל $v \in V$.

גרפים הם אובייקט מרכזי במתמטיקה, ומודל ארדש-ריינ מציע דרך לדגום גרף אקראי על n קודקודים.

הגדרה 4.5. מודל ארדש-רייני לגרפים מקריים, המסומן $G(n,p)$, הוא ההתפלגות על כל הגרפים בעלי n קודקודים, כך שכל קשת מופיעה בסיכוי p באופן בלתי תלוי בקשתות האחרות. זה אחד המודלים השימושיים ביותר לגרפים מקריים. לצורך הנוחות, נסמן את קבוצת הקודקודים $[n] = \{1, \ldots, n\}$.

תרגיל 4.6. נתבונן במודל ארדש-רייני $G(n,p_n)$ כאשר n שואף ל-∞. הוכיחו:

א. הם $p_n = o\left(\frac{1}{n}\right)$ (כלומר $np_n \to 0$), אז הסיכוי שב-$G(n,p_n)$ יש משולש שואף ל-0.

ב. אם $p_n = \omega\left(\frac{1}{n}\right)$ (כלומר $np_n \to \infty$), אז הסיכוי שב-$G(n,p_n)$ יש משולש שואף ל-1.

פתרון. נסמן על ידי $T_{n,p}$ את מספר המשולשים ב-$G(n,p)$ (זהו משתנה מקרי).

א. נתחיל מלחשב את תוחלת כמות המשולשים ב-$G(n,p)$, כלומר את $\mathbb{E}[T_{n,p}]$. נשים לב כי $T_{n,p} = \sum_{A\subseteq[n],|A|=3} I_A$, כאשר I_A הוא האינדיקטור של האם הקבוצה A יוצרת משולש. מתקיים $\mathbb{E}[I_A] = p^3$, ומלינאריות התוחלת נקבל $\mathbb{E}[T_{n,p}] = \sum_A \mathbb{E}[I_A] = \binom{n}{3}p^3$.
כעת ניעזר באי־שוויון מרקוב בתנאי השאלה, ונקבל

$$\mathrm{P}(T_{n,p_n} > 0) = \mathrm{P}(T_{n,p_n} \geq 1) \leq \mathbb{E}[T_{n,p_n}] = \binom{n}{3}p_n^3 \leq \frac{(np_n)^3}{6} \xrightarrow[n\to\infty]{} 0$$

כנדרש.

ב. העובדה שהתוחלת שואפת ל-∞ לא אומרת שבהכרח יהיה משולש בהסתברות שואפת ל-1. לכן נצטרך מידע קצת יותר על הריכוז של $T_{n,p}$, ולצורך כך נעריך את השונות שלו. נשים לב כי

$$\mathbb{E}\left[T_{n,p}^2\right] = \mathbb{E}\left[\sum_{A,B} I_A I_B\right] = \sum_{A,B} \mathbb{E}[I_A I_B]$$

(הסכומים הם על תת־קבוצות $A, B \subseteq [n]$ כך ש-$|A| = |B| = 3$). נחלק למקרים:

- אם $A \cap B = \varnothing$, אז $\mathbb{E}[I_A I_B] = p^6$. כמות הזוגות של תת־קבוצות זרות כאלו היא $\binom{n}{3}\binom{n-3}{3} \sim \frac{n^6}{36}$.
- אם $A \cap B = \{v\}$, אז $\mathbb{E}[I_A I_B] = p^6$. כמות הזוגות של תת־קבוצות כאלו היא $3\binom{n}{3}\binom{n-3}{2} \sim \frac{n^5}{4}$.
- אם $A \cap B = \{v_1, v_2\}$, אז $\mathbb{E}[I_A I_B] = p^5$. כמות הזוגות של תת־קבוצות כאלו היא $3\binom{n}{3}\binom{n-3}{1} \sim \frac{n^4}{2}$.
- אם $A = B$, אז $I_A^2 = I_A$, ונקבל $\mathbb{E}[I_A^2] = p^3$.

בסך הכל

$$\mathbb{E}\left[T_{n,p}^2\right] \sim \frac{n^6p^6}{36} + \frac{n^5p^6}{4} + \frac{n^4p^5}{2} + \frac{n^3p^3}{6}$$

ומכאן

$$\mathrm{Var}(T_{n,p}) \sim \frac{n^5p^6}{4} + \frac{n^4p^5}{2} + \frac{n^3p^3}{6}.$$

מנתוני השאלה $np_n \to \infty$, ולכן $\mathrm{Var}(T_{n,p_n}) \sim \frac{n^5p^6}{4}$. ניעזר באי־שוויון צ'בישב:

$$\begin{aligned}\mathrm{P}\left(T_{n,p_n}=0\right) &= \mathrm{P}\left(T_{n,p_n}\le 0\right) = \mathrm{P}\left(T_{n,p_n} - \mathbb{E}\left[T_{n,p_n}\right] \le -\mathbb{E}\left[T_{n,p_n}\right]\right) \le \\ &\le \mathrm{P}\left(\left|T_{n,p_n} - \mathbb{E}\left[T_{n,p_n}\right]\right| \le \mathbb{E}\left[T_{n,p_n}\right]^2\right) \overset{\text{צ'בישב}}{\le} \frac{\mathrm{Var}\left(T_{n,p_n}\right)}{\mathbb{E}\left[T_{n,p_n}\right]^2} \sim \\ &\sim \frac{\frac{n^5p^6}{4}}{\frac{n^6p^6}{6}} = \frac{3}{2n} \xrightarrow[n\to\infty]{} 0\end{aligned}$$

4.3 אי־שוויון הופדינג

אי־שוויוני מרקוב וצ'בישב נותנים מידע רב כאשר כל הנתונים הם התוחלת והשונות, ושניהם הדוקים. עם זאת, במקרים רבים אפשר לשפר את אי־השוויונים. נתמקד כעת באי־שוויון הופדינג, ובתרגול הבא באי־שוויון צ'רנוף, שנותנים חסם על ההתנהגות של סכום משתנים מקריים בלתי־תלויים ושווי התפלגות.

משפט 4.7 (אי־שוויון הופדינג)**.** יהיו $X_1, \ldots, X_n$ משתנים מקריים בלתי־תלויים שעבורם מתקיים $|X_i| \le 1$ וגם $\mathbb{E}\left[X_i\right] = 0$. אז לכל $\lambda \ge 0$ מתקיים

$$\mathrm{P}\left(\sum_{i=1}^n X_i \ge \lambda\right) \le e^{-\frac{\lambda^2}{2n}}.$$

תרגיל 4.8. חסמו באמצעות אי־שוויון הופדינג את $\mathrm{P}\left(\left|\mathrm{Bin}\left(n,p\right) - np\right| \ge \lambda\right)$.

הוכחה. נזכור כי אם $X \sim \mathrm{Bin}\left(n,p\right)$ אז $X = \sum_{i=1}^n X_i$ כשכל $X_i \sim \mathrm{Ber}\left(p\right)$. כדי לנרמל את התוחלות נגדיר $Y_i = X_i - p$. אז

$$|X - np| = \left|\sum_{i=1}^n X_i - np\right| = \left|\sum_{i=1}^n Y_i\right|.$$

כעת

$$\mathrm{P}\left(|X-np| \ge \lambda\right) = \mathrm{P}\left(\left|\sum_{i=1}^n Y_i\right| \ge \lambda\right) = \mathrm{P}\left(\sum_{i=1}^n Y_i \ge \lambda\right) + \mathrm{P}\left(\sum_{i=1}^n Y_i \le -\lambda\right).$$

לפי אי־שוויון הופדינג,

$$\mathrm{P}\left(|X-np| \ge \lambda\right) \le 2e^{-\frac{\lambda^2}{2n}}.$$

□

תרגיל 4.9. יהי S_n הילוך מקרי על $\mathbb{Z}$. כלומר: $S_n = \sum_{i=1}^n X_i$ כשכל $X_i = \begin{cases} 1, & \frac{1}{2} \\ -1, & \frac{1}{2} \end{cases}$.

נניח ש-$S_0 = 0$. הראו ש-$S_n > 2\sqrt{n \log n}$ מספר סופי של פעמים.

הוכחה. ניעזר באי־שוויון הופדינג על S_n. אז מתקיים

$$\mathrm{P}\left(S_n > 2\sqrt{n \log n}\right) \leq e^{-\frac{4n \log n}{2n}} = e^{-2 \log n} = \frac{1}{n^2}.$$

לפי למת בורל-קנטלי הראשונה, $S_n > 2\sqrt{n \log n}$ אינסוף פעמים. □

משפט 4.10 (חוק הלוגריתם החוזר)**.** עבור הילוך מקרי S_n על $\mathbb{Z}$,

$$\mathrm{P}\left(\limsup_{n \to \infty} \frac{S_n}{\sqrt{2n \log \log n}} = 1\right) = 1.$$

5 תרגול חמישי

5.1 מומנטים ואי־שוויון צ'רנוף

ראשית, נזכיר הגדרה מקורס המבוא:

הגדרה 5.1. יהי X משתנה מקרי. **המומנט ה-n-י של** X הינו $\mathbb{E}[X^n]$, בהנחה שהאינטגרל מתכנס בהחלט.

שאלה 5.2. האם המומנטים קובעים את ההתפלגות?

פתרון. לא. נציג דוגמה של Heyde מ-1963 להתפלגויות שונות בעלות אותן מומנטים. נתבונן בהתפלגות הלוג־נורמלית

$$f_0(x) = \frac{1}{\sqrt{2\pi}x} e^{-\frac{(\log x)^2}{2}}$$

ובהזזות שלה

$$f_a(x) = f_0(x) \cdot (1 + a \sin(2\pi \log x))$$

לכל $-1 \leq a \leq 1$. כדי לוודא שזו אכן פונקציית צפיפות ועם אותם מומנטים, נראה שלכל $r = 0, 1, 2, \ldots$ מתקיים

$$\int_0^\infty x^r f_0(x) \sin(2\pi \log x)\, \mathrm{d}x = (\star) = 0.$$

אכן, נבצע החלפת משתנים $x = \exp(u + r)$, כלומר $u = \log x - r$, ואז $\mathrm{d}u = \frac{\mathrm{d}x}{x}$; אז האינטגרל נהיה

$$(\star) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(ru + r^2\right) \exp\left(-\frac{(u+r)^2}{2}\right) \sin(2\pi(u+r))\, \mathrm{d}u =$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{r^2 - u^2}{2}\right) \sin(2\pi u)\, \mathrm{d}u = 0$$

(כי זה אינטגרל על פונקציה אי־זוגית). לכן לכל ההתפלגויות מהצורה f_a עבור $-1 \leq a \leq 1$ יש אותם מומנטים. המומנטים של ההתפלגויות האלו הם $\mathbb{E}[X^n] = \exp\left(\frac{n^2}{2}\right)$.

הגדרה 5.3. הפונקציה יוצרת המומנטים של משתנה מקרי X הינה $M_X(t) = \mathbb{E}\left[e^{tX}\right]$ (בכל מקום שבו היא מוגדרת).

דוגמה 5.4. יהי $X \sim \mathrm{Poi}(\lambda)$. אז

$$M_X(t) = \mathbb{E}\left[e^{tX}\right] = \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!}e^{tk} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} = e^{\lambda\left(e^t-1\right)}.$$

משפט 5.5. אם הפונקציה יוצרת המומנטים קיימת בקטע פתוח סביב 0, אז היא קובעת את ההתפלגות.

אי שוויוני צ'רנוף. הרעיון: אם X משתנה מקרי כלשהו עם פונקציה יוצרת מומנטים, אפשר לכתוב לכל $t > 0$

$$\mathrm{P}(X \geq a) = \mathrm{P}\left(e^{tX} \geq e^{ta}\right) \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{ta}} = \frac{M_X(t)}{e^{ta}}$$

ולבחור t אופטימלי.

דוגמה 5.6. ניקח את הדוגמה שפתרנו: $X \sim \mathrm{Poi}(\lambda)$. אז $\mathrm{P}(X \geq a) \leq \frac{e^{\lambda\left(e^t-1\right)}}{e^{ta}} = e^{\lambda\left(e^t-1\right)-ta}$. נחפש מינימום: אם נגדיר $f(t) = \lambda\left(e^t-1\right) - ta$, אז הנגזרת היא $f'(t) = \lambda e^t - a$. לכן נקבל מינימום עבור $t = \log\frac{a}{\lambda}$. לכן החסם שיתקבל הוא

$$\mathrm{P}(X \geq a) \leq e^{\lambda\left(\frac{a}{\lambda}-1\right)-a\log\frac{a}{\lambda}} = e^{a-\lambda-a\log\frac{a}{\lambda}}.$$

אם $X = \sum_{i=1}^n X_i$ סכום של משתנים מקריים בלתי־תלויים,

$$\mathrm{P}(X \geq a) \leq e^{-ta}\mathbb{E}\left[e^{\sum_{i=1}^n tX_i}\right] = e^{-ta}\mathbb{E}\left[\prod_{i=1}^n e^{-tX_i}\right] = e^{-ta}\prod_{i=1}^n M_{X_i}(t)$$

לכל $t > 0$, ולכן

$$\mathrm{P}(X \geq a) \leq \min_{t>0} e^{-ta}\prod_{i=1}^n M_{X_i}(t).$$

משפט 5.7 (אי־שוויון צ'רנוף)**.** יהיו $X_1, \ldots, X_n$ משתנים מקריים בלתי־תלויים כך ש־$|X_i| \leq 1$, $\mathbb{E}[X_i] = 0$ ו־$\mathrm{Var}(X_i) = V_i < \infty$. נסמן $V = \sum_{i=1}^n V_i$. אז לכל $\lambda \geq 0$ מתקיים

$$\mathrm{P}\left(\sum_{i=1}^n X_i \geq \lambda\right) \leq \max\left\{e^{-\frac{\lambda^2}{4V}}, e^{-\frac{\lambda}{2}}\right\}.$$

מקבלים את החסם השמאלי אם $\lambda < 2V$, ואחרת את החסם הימני.

דוגמה 5.8. איך נראה אי־שוויון צ'רנוף במקרה של התפלגות בינומית $X \sim \mathrm{Bin}(n,p)$?

$$\mathrm{P}\left(\sum_{i=1}^{n} X_i - np \geq \lambda\right) \leq \max\left\{e^{-\frac{\lambda^2}{4np(1-p)}}, e^{-\frac{\lambda}{2}}\right\}.$$

אם $\lambda < 2np(1-p)$, מקבלים את החסם השמאלי. האם הוא יותר טוב מאשר הופדינג?

$$e^{-\frac{\lambda^2}{4np(1-p)}} \leq e^{-\frac{\lambda^2}{2n}} \iff -\frac{\lambda^2}{4np(1-p)} \leq -\frac{\lambda^2}{2n} \iff 2p(1-p) \leq 1 \iff 2p^2-2p+1 \leq 0.$$

וזה נכון לכל p.

5.2 סוגי התכנסויות

הגדרה 5.9. יהיו $X, X_1, X_2, \ldots$ משתנים מקריים.

א. אומרים ש-X_n **מתכנסת בהתפלגות** ל-X, ומסמנים $X_n \xrightarrow{d} X$, אם $F_{X_n}(t) \to F_X(t)$ בכל נקודת רציפות של F_X. (המשתנים יכולים להיות ממרחבי מדגם שונים!)

ב. אומרים ש-X_n **מתכנסת בהסתברות** ל-X, ומסמנים $X_n \xrightarrow{\mathrm{P}} X$, אם לכל $\varepsilon > 0$, $\lim_{n\to\infty} \mathrm{P}(|X_n - X| \geq \varepsilon) = 0$. (פה המשתנים הם מאותו מרחב מדגם)

ג. אומרים ש-X_n **מתכנסת כמעט תמיד** ל-X, ומסמנים $X_n \xrightarrow{\text{a.s.}} X$, אם מתקיים $\mathrm{P}(\{\omega \mid X_n(\omega) \to X(\omega)\}) = 1$. (פה המשתנים הם מאותו מרחב מדגם)

טענה 5.10. התכנסות כמעט תמיד $\Longrightarrow$ התכנסות בהסתברות $\Longrightarrow$ התכנסות בהתפלגות.

טענה 5.11. אם $X_n \xrightarrow{\mathrm{P}} X$, אז קיימת תת־סדרה $X_{n_k} \xrightarrow{\text{a.s.}} X$.

דוגמה 5.12. ניתן דוגמאות שההגדרות הנ"ל אינן שקולות.

א. X_n שמתכנסת בהסתברות ולא כמעט תמיד – נגדיר משתנים מקריים על $[0,1]$ באופן הבא: לכל $n \in \mathbb{N}$ ולכל $0 \leq k \leq 2^n - 1$ נגדיר $X_{2^n+k} = 1_{[k\cdot 2^{-n},(k+1)\cdot 2^{-n}]}$. זו סדרה המתכנסת בהסתברות ל-0, כי לכל $\varepsilon < 1$ מתקיים $\mathrm{P}(|X_{2^n+k} - 0| \geq \varepsilon) = 2^{-n}$. מצד שני, היא לא מתכנסת כמעט תמיד כי כל נקודה מופיעה באינסוף קטעים מהצורה $[k \cdot 2^{-n}, (k+1) \cdot 2^{-n}]$.

ב. X_n שמתכנסת בהתפלגות ולא בהסתברות – נגדיר משתנים מקריים על $[0,1]$ באופן הבא: $X_{2n-1}(\omega) = \omega$, $X_{2n}(\omega) = 1 - \omega$. ההתפלגות של כל X_n היא אחידה, לכן יש התכנסות בהתפלגות למשתנה מקרי אחיד. מצד שני, אין התכנסות בהסתברות כי $|X_{2n}(\omega) - \omega| = |1 - 2\omega|$, ולכן

$$\mathrm{P}(|X_{2n}(\omega) - \omega| \geq \varepsilon) = \mathrm{P}(|1 - 2\omega| \geq \varepsilon) = 1 - \mathrm{P}(-\varepsilon < 1 - 2\omega < \varepsilon) \geq 1 - 2\varepsilon.$$

תרגיל 5.13. יהיו $X, X_1, X_2, \ldots, Y, Y_1, Y_2, \ldots$ משתנים מקריים על Ω וננית ש-$X_n \xrightarrow{d} X$ ו-$Y_n \xrightarrow{d} Y$. האם בהכרח $X_n + Y_n \xrightarrow{d} X + Y$?

פתרון. לא. למשל, ניקח $X_n \sim U[0,1]$ ו-$Y_n = -X_n$. אז $X_n \xrightarrow{d} U[0,1]$, $Y_n \xrightarrow{d} U[-1,0]$, אבל $X_n + Y_n = 0$ לא שואף בהתפלגות להתפלגות של $X + Y$ כאשר $X \sim U[0,1]$ ו-$Y \sim U[-1,0]$ בלתי־תלויים.

משפט 5.14 (משפט סלוצקי (1925), מיוחס גם לקרמר)**.** אם $X_n \xrightarrow{d} X$ ו-$Y_n \xrightarrow{d} c$, אז:

א. $X_n + Y_n \xrightarrow{d} X + c$.

ב. $X_n Y_n \xrightarrow{d} cX$.

ג. $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$, כל עוד c הפיך.

6 תרגול שישי

6.1 החוק החלש והחוק החזק

משפט 6.1 (החוק החלש של המספרים הגדולים)**.** יהיו X_n משתנים מקריים בלתי־תלויים בזוגות ושווי התפלגות עם תוחלת $\mathbb{E}[X_i] = \mu$. אזי $\frac{X_1+\cdots+X_n}{n} \xrightarrow{\mathrm{P}} \mu$.

משפט 6.2 (החוק החזק של המספרים הגדולים)**.** יהיו X_n משתנים מקריים בלתי־תלויים בזוגות ושווי התפלגות עם תוחלת $\mathbb{E}[X_i] = \mu$. אזי $\frac{X_1+\cdots+X_n}{n} \xrightarrow{\text{a.s.}} \mu$.

תרגיל 6.3. יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים בעלי התפלגות אחידה בקטע $[1,2]$. מגדירים $Y_n = \sqrt[n]{X_1 \cdot \cdots \cdot X_n}$. הוכיחו כי הסדרה Y_n מתכנסת כמעט תמיד, וחשבו את הגבול שלה.

פתרון. נשים לב כי

$$\log Y_n = \frac{\log X_1 + \cdots + \log X_n}{n}.$$

המשתנים $\log X_1, \log X_2, \ldots$ הם בלתי־תלויים ושווי התפלגות. התוחלת שלהם היא

$$\mathbb{E}[\log X_1] = \int_1^2 \log x \,\mathrm{d}x = (x \log x - x)_1^2 = 2\log 2 - 2 + 1 = 2\log 2 - 1.$$

מהחוק החזק של המספרים הגדולים,

$$\log Y_n \xrightarrow{\text{a.s.}} 2\log 2 - 1,$$

ולכן

$$Y_n \xrightarrow{\text{a.s.}} e^{2\log 2 - 1} = \frac{4}{e}.$$

הגדרה 6.4. יהי $2 \leq b \in \mathbb{N}$, ויהי $x \in \mathbb{R}$. עבור רצף ספרות s בבסיס b, נסמן ב-$N_n(s,x)$ את כמות ההופעות של s ב-n הספרות הראשונות של x. אומרים ש-x הוא **נורמלי בבסיס** b, אם לכל s מתקיים $\frac{N_n(s,x)}{n} \xrightarrow[n\to\infty]{} b^{-|s|}$, כאשר $|s|$ הוא האורך של s. אומרים ש-x הוא **נורמלי**, אם הוא נורמלי ביחס לכל בסיס b.

משפט 6.5 (בורל, 1909)**.** כמעט כל מספר ממשי הוא נורמלי.

הוכחה. מספיק להוכיח שלכל $b \geq 2$, כמעט כל מספר ממשי הוא נורמלי בבסיס b. אכן,

$$\{\text{מספרים נורמליים}\} = \bigcap_{b=2}^{\infty} \{\text{מספרים נורמליים בבסיס } b\},$$

ואם כל המאורעות באגף ימין הם מהסתברות 1 אז גם אגף שמאל. מנימוק דומה, מספיק לבדוק שלכל רצף ספרות s מתקיים שכמעט לכל x,

$$\frac{N_n(s,x)}{n} \xrightarrow[n\to\infty]{} b^{-|s|}.$$

כיוון שאנחנו לא אוהבים להגריל מספרים אחיד בכל $\mathbb{R}$, נראה זאת על מספרים שמוגרלים אחיד ב-$[0,1]$. יהי $X \sim \mathrm{U}(0,1)$, ויהי X_k המאורע שבו רצף הספרות של X בבסיס b החל מהמקום ה-k הוא s. זהו משתנה אינדיקטור המקיים $\mathrm{P}(X_k = 1) = b^{-|s|}$. כעת,

$$\frac{N_n(s,x)}{n} = \frac{1}{n}\sum_{k=1}^{n-|s|} X_k.$$

לכל $1 \leq m \leq |s|$, המשתנים $X_m, X_{m+|s|}, X_{m+2|s|}, \ldots$ הם בלתי־תלויים, ובעלי תוחלת $b^{-|s|}$. לפי החוק החזק של המספרים הגדולים,

$$\frac{1}{n}\sum_{k=1}^{n-|s|} X_{m+|s|k} \xrightarrow{\text{a.s.}} \frac{1}{|s|} b^{-|s|}.$$

כעת כל מה שנותר הוא לסכום את ההזחות השונות, ולקבל

$$\frac{N_n(s,x)}{n} \xrightarrow{\text{a.s.}} |s| \cdot \frac{1}{|s|} b^{-|s|} = b^{-|s|}$$

כנדרש. □

מסקנה 6.6. לא ניתן לייצר מכונה עם זיכרון פנימי סופי שעובדת בזמן סופי שתפלוט מספרים אקראיים באמת.

הערה. עד כמה שהם נורמליים, מאוד קשה למצוא מספרים נורמליים.

תרגיל 6.7. תנו דוגמה לסדרת משתנים מקריים בלתי־תלויים בעלי אותה התוחלת μ שעבורם $\frac{X_1+\cdots+X_n}{n} \xrightarrow{\mathrm{P}} \mu$ אבל $\frac{X_1+\cdots+X_n}{n} \not\xrightarrow{\text{a.s.}} \mu$.

פתרון. נגדיר סדרת משתנים X_n לפי

$$\mathrm{P}\left(X_n = n+1\right) = \mathrm{P}\left(X_n = -\left(n+1\right)\right) = \frac{1}{2\left(n+1\right)\log\left(n+1\right)}.$$
$$\mathrm{P}\left(X_n = 0\right) = 1 - \frac{1}{2\left(n+1\right)\log\left(n+1\right)}$$

כדי להראות התכנסות בהסתברות ל-0, נשים לב כי $\mathbb{E}\left[X_n\right] = 0$ ו-$\mathrm{Var}\left(X_n\right) = \mathbb{E}\left[X_n^2\right] = \frac{n+1}{\log(n+1)}$. לכן, לפי צ'בישב,

$$\mathrm{P}\left(\left|\frac{1}{n}S_n\right| \geq \varepsilon\right) \leq \frac{\mathrm{Var}\left(S_n\right)}{\varepsilon^2} \leq \frac{n \cdot \frac{n+1}{\log(n+1)}}{\varepsilon^2 n^2} = \frac{n+1}{\varepsilon^2 n \log\left(n+1\right)} \xrightarrow[n\to\infty]{} 0$$

לעומת זאת, הסדרה אינה מתכנסת כמעט בכל מקום. נזכור כי $\sum_{n=1}^{\infty} \frac{1}{(n+1)\log(n+1)} = \infty$; לכן, לפי בורל-קנטלי, $X_n = n+1$ אינסוף פעמים, ומתקיים

$$\left\{X_n = n+1\right\} \subseteq \left\{|S_n| \geq \frac{n}{2}\right\} \cup \left\{|S_{n-1}| \geq \frac{n}{2}\right\}.$$

לכן לא יכול להיות שתהיה התכנסות כמעט בכל מקום ל-0.

6.2 התכנסות חלשה

הגדרה 6.8. אומרים שסדרת מידות μ_n על מרחב Ω **מתכנסת חלש** למידה μ על Ω, ונסמן $\mu_n \xrightarrow{w} \mu$, אם לכל $h\colon \mathbb{R} \to \mathbb{R}$ רציפה וחסומה מתקיים

$$\mathbb{E}\left[h\left(X_n\right)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h\left(X\right)\right]$$

כאשר $X_n \sim \mu_n$ ו-$X \sim \mu$. נסמן גם $X_n \xrightarrow{w} X$.

משפט 6.9 (למת *Portmanteau*)**.** יהיו $X, X_1, X_2, \ldots$ משתנים מקריים. אזי התנאים הבאים שקולים (בוויקיפדיה תוכלו למצוא עוד תנאים שקולים):

א. $X_n \xrightarrow{d} X$.

ב. $X_n \xrightarrow{w} X$.

ג. לכל פונקציה חסומה וליפשיץ h, $\mathbb{E}\left[h\left(X_n\right)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h\left(X\right)\right]$.

ד. לכל פונקציה רציפה ואי־שלילית h, $\liminf \mathbb{E}\left[h\left(X_n\right)\right] \geq \mathbb{E}\left[h\left(X\right)\right]$.

הוכחת ג $\Leftrightarrow$ ד. $\boxed{\text{ב} \Leftarrow \text{ד}}$ תהי h רציפה ואי־שלילית. לכל N, $h_N\left(x\right) = \min\left\{h\left(x\right), N\right\}$ היא פונקציה רציפה וחסומה. מההנחה, $\mathbb{E}\left[h_N\left(X_n\right)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[h_N\left(X\right)\right]$, $\mathbb{E}\left[h\left(X_n\right)\right] \geq \mathbb{E}\left[h_N\left(X_n\right)\right]$,

ובפרט $\liminf \mathbb{E}\left[h\left(X_n\right)\right] \geq \mathbb{E}\left[h_N\left(X\right)\right]$. כעת נוכל להפעיל את משפט ההתכנסות המונוטונית $\mathbb{E}\left[h_N\left(X\right)\right] \xrightarrow[N\to\infty]{} \mathbb{E}\left[h\left(X\right)\right]$, ונקבל את אי־השוויון הדרוש.

ב $\Rightarrow$ ד תהי h רציפה וחסומה. אם $|h\left(x\right)| \leq M$, אז $g\left(x\right) = h\left(x\right) + M$ היא פונקציה רציפה ואי־שלילית, ולכן

$$\liminf \mathbb{E}\left[h\left(X_n\right)\right] + M = \liminf \mathbb{E}\left[g\left(X_n\right)\right] \geq \mathbb{E}\left[g\left(X\right)\right] = \mathbb{E}\left[h\left(X\right)\right] + M$$

כלומר $\liminf \mathbb{E}\left[h\left(X_n\right)\right] \geq \mathbb{E}\left[h\left(X\right)\right]$. אם נפעיל את אותו הטריק על $-h$, נקבל $\limsup \mathbb{E}\left[h\left(X_n\right)\right] \leq \mathbb{E}\left[h\left(X\right)\right]$. בסך הכל נקבל שוויון $\lim \mathbb{E}\left[h\left(X_n\right)\right] = \mathbb{E}\left[h\left(X\right)\right]$.
□

7 תרגול שביעי

7.1 משפט הגבול המרכזי

משפט 7.1 (משפט הגבול המרכזי)**.** יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים ושווי התפלגות עם תוחלת μ ושונות σ^2, אזי

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathrm{N}\left(0,1\right).$$

תרגיל 7.2. חשבו את הגבול $\lim_{n\to\infty} e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!}$.

פתרון. נשים לב כי

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = \mathrm{P}\left(\mathrm{Poi}\left(n\right) \leq n\right).$$

נזכור כי אם $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים עם התפלגות $\mathrm{Poi}\left(1\right)$, אז $X_1 + \cdots + X_n \sim \mathrm{Poi}\left(n\right)$, ולכן

$$\mathrm{P}\left(\mathrm{Poi}\left(n\right) \leq n\right) = \mathrm{P}\left(\sum_{i=1}^{n} X_i \leq n\right) = \mathrm{P}\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - 1}{\sqrt{n}} \leq 0\right) \xrightarrow[n\to\infty]{\text{CLT}} \mathrm{P}\left(\mathrm{N}\left(0,1\right) \leq 0\right) = \frac{1}{2}.$$

תרגיל 7.3. חשבו את

$$\lim_{n\to\infty}\frac{1}{2^n}\int_{-1}^{1}\cdots\int_{-1}^{1}\cos\left(\frac{x_1^{-1/3}+\cdots+x_n^{-1/3}}{\sqrt{n}}\right)\mathrm{d}x_1\ldots\mathrm{d}x_n.$$

פתרון. נגדיר את מרחב ההסתברות $\Omega = [-1,1]$ עם מידת לבג מנורמלת. יהיו $X_1, X_2, \cdots \sim \mathrm{U}\left[-1,1\right]$ משתנים מקריים בלתי־תלויים, ויהיו $Y_i = X_i^{-1/3}$. נגדיר $S_n = Y_1 + \cdots + Y_n$. אזי

$$\frac{1}{2^n}\int_{-1}^{1}\cdots\int_{-1}^{1}\cos\left(\frac{x_1^{-1/3}+\cdots+x_n^{-1/3}}{\sqrt{n}}\right)\mathrm{d}x_1\ldots\mathrm{d}x_n = \mathbb{E}\left[\cos\left(\frac{S_n}{\sqrt{n}}\right)\right].$$

כדי להיעזר במשפט הגבול המרכזי, נחשב את התוחלת והשונות של Y_i.

$$\mathbb{E}[Y_i] = \frac{1}{2}\int_{-1}^{1} x^{-1/3}\,\mathrm{d}x = \frac{1}{2}\cdot\frac{3}{2}\left(x^{2/3}\right)_{-1}^{1} = 0$$
$$\mathbb{E}\left[Y_i^2\right] = \frac{1}{2}\int_{-1}^{1} x^{-2/3}\,\mathrm{d}x = \frac{1}{2}\cdot 3\cdot\left(x^{1/3}\right)_{-1}^{1} = 3$$
$$\mathrm{Var}(Y_i) = \mathbb{E}\left[Y_i^2\right] - \mathbb{E}[Y_i]^2 = 3$$

נזכור כי $Y_1, Y_2, \ldots$ בלתי־תלויים ושווי התפלגות, ולכן לפי משפט הגבול המרכזי $\frac{S_n}{\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,\sigma^2)$. כיוון ש-cos היא פונקציה רציפה וחסומה, נקבל

$$\mathbb{E}\left[\cos\left(\frac{S_n}{\sqrt{n}}\right)\right] \xrightarrow[n\to\infty]{} \mathbb{E}[\cos(\mathrm{N}(0,3))] = \ell.$$

כדי לחשב את הגבול, ניקח $Z \sim \mathrm{N}(0,3)$, ונכתוב

$$\ell = \mathbb{E}[\cos Z] = \mathbb{E}\left[\frac{e^{iZ}+e^{-iZ}}{2}\right].$$

אפשר לחשב שאם $X \sim \mathrm{N}(\mu,\sigma^2)$, אז לכל $t\in\mathbb{R}$ מתקיים

$$\mathbb{E}\left[e^{itX}\right] = e^{it\mu-\frac{1}{2}\sigma^2t^2},$$

ולכן

$$\ell = \frac{1}{2}\left(e^{-\frac{1}{2}\cdot\left(\sqrt{3}\right)^2} + e^{-\frac{1}{2}\cdot\left(-\sqrt{3}\right)^2}\right) = e^{-\frac{3}{2}}.$$

7.2 הפונקציה האופיינית

נזכיר את הפונקציה יוצרת המומנטים של משתנה מקרי X: $M_X(t) = \mathbb{E}\left[e^{tX}\right]$.

הגדרה 7.4. יהי X משתנה מקרי. **הפונקציה האופיינית** של X היא $\varphi_X(t) = \mathbb{E}\left[e^{itX}\right]$.

הערה. איך מחשבים את הפונקציה האופיינית?

$$\varphi_X(t) = \mathbb{E}\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itX}\,\mathrm{d}F_X(t).$$

אם ל-X יש פונקציית צפיפות f_X, אז

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itX} f_X(t)\,\mathrm{d}t.$$

טענה 7.5 (תכונות של הפונקציה האופיינית).

א. $|\varphi_X(t)| \leq 1$ לכל $t \in \mathbb{R}$.

ב. אם X, Y בלתי־תלויים, אז $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$.

ג. פונקציה לינארית: לכל $a, b \in \mathbb{R}$ מתקיים $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$, ובפרט $\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)}$.

ד. φ_X רציפה במ"ש בכל $\mathbb{R}$.

ה. φ_X היא פונקציה ממשית (כלומר מקבלת רק ערכים ב-$\mathbb{R}$) אם ורק אם X הוא משתנה מקרי סימטרי, כלומר $\mathrm{P}(X > x) = \mathrm{P}(X < -x)$ לכל $x \in \mathbb{R}$.

תרגיל 7.6. חשבו את הפונקציה האופיינית עבור:

א. התפלגות אחידה על $[-1, 1]$.

ב. התפלגות גיאומטרית עם פרמטר p.

פתרון.

א. יהי $X \sim U[-1, 1]$. לפי ההגדרה,

$$\varphi_X(t) = \frac{1}{2}\int_{-1}^{1} e^{itx}\,\mathrm{d}x = \frac{1}{2it}\left(e^{itx}\right)_{-1}^{1} = \frac{e^{it} - e^{-it}}{2it} = \frac{\sin t}{t}.$$

ב. יהי $Y \sim \mathrm{Geo}(p)$. לפי ההגדרה,

$$\varphi_Y(t) = \sum_{n=1}^{\infty}(1-p)^{n-1}pe^{int} = \frac{p}{1-p}\cdot\sum_{n=1}^{\infty}\left((1-p)e^{it}\right)^n =$$
$$= \frac{p}{1-p}\cdot\frac{(1-p)e^{it}}{1-(1-p)e^{it}} = \frac{pe^{it}}{1-(1-p)e^{it}}$$

תרגיל 7.7. הראו שלא קיימים ממבתש"ה X, Y כך ש-$X - Y \sim U[-1, 1]$.

הוכחה. נניח בשלילה שקיימים. אז

$$|\varphi_X(t)|^2 = \varphi_X(t)\,\varphi_{-Y}(t) = \varphi_{X-Y}(t) = \varphi_{U[-1,1]}(t) = \frac{\sin t}{t},$$

אבל זה לא יכול להיות כי אגף ימין מקבל ערכים שליליים עבור $t < 0$. □

8 תרגול שמיני

משפט 8.1 (משפט ההיפוך)**.** אם X הוא משתנה מקרי בעל התפלגות μ, אז לכל $a < b$ מתקיים

$$\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita}-e^{itb}}{it}\varphi_X(t)\,\mathrm{d}t=\mu((a,b))+\frac{1}{2}\mu(\{a,b\}).$$

בפרט, אם $\varphi_X(t)$ אינטגרבילית על $\mathbb{R}$, אז μ רציפה בהחלט, והצפיפות של μ נתונה על ידי הנוסחה

$$f_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi_X(t)\,\mathrm{d}t.$$

תרגיל 8.2. יהיו $X_1,\ldots,X_n$ משתנים מקריים בלתי־תלויים בעלי התפלגות קושי סטנדרטית, כלומר

$$f_{X_i}(x)=\frac{1}{\pi\left(1+x^2\right)}.$$

הוכיחו כי עבור $S_n=X_1+\cdots+X_n$, למשתנה $\frac{S_n}{n}$ גם יש התפלגות קושי סטנדרטית.

הוכחה. במקום להוכיח ישירות, נחשב את $\varphi_{X_i}(t)$. ובמקום לחשב את זה ישירות, ניזכר במשפט שהתמרת פורייה הופכית לעצמה. כלומר, אם ל־X יש פונקציית צפיפות $f_X(x)$, אזי

$$f_X(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi_X(t)\,\mathrm{d}t.$$

נסתכל על המשתנה המקרי בעל פונקציית הצפיפות $g(x)=\frac{1}{2}e^{-|x|}$. הפונקציה האופיינית שלו היא

$$\begin{aligned}\varphi(t)&=\int_{-\infty}^{\infty}\frac{1}{2}e^{-|x|}e^{itx}\,\mathrm{d}x=\frac{1}{2}\int_{0}^{\infty}e^{(-1+it)x}\,\mathrm{d}x+\frac{1}{2}\int_{-\infty}^{0}e^{(1+it)x}\,\mathrm{d}x=\\&=\frac{1}{2}\cdot\left(\frac{e^{(-1+it)x}}{-1+it}\right)_0^{\infty}+\frac{1}{2}\cdot\left(\frac{e^{(1+it)x}}{1+it}\right)_{-\infty}^{0}=\frac{1}{2}\cdot\frac{1}{1-it}+\frac{1}{2}\cdot\frac{1}{1+it}=\frac{1}{1+t^2}\end{aligned}$$

ולכן לפי משפט ההיפוך

$$\frac{1}{2}e^{-|x|}=g(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\varphi(t)\,\mathrm{d}t=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-itx}\frac{1}{1+t^2}\,\mathrm{d}t=\frac{1}{2}\varphi_X(-x).$$

לכן נקבל $\varphi_X(t)=e^{-|t|}$. כעת נוכל לחשב את הפונקציה האופיינית של $\frac{S_n}{n}$ בקלות:

$$\varphi_{\frac{S_n}{n}}(t)=\varphi_{S_n}\left(\frac{t}{n}\right)=\left(\varphi_{X_1}\left(\frac{t}{n}\right)\right)^n=\left(e^{-\left|\frac{t}{n}\right|}\right)^n=e^{-|t|}$$

כלומר $\frac{S_n}{n}$ מתפלג קושי סטנדרטי. □

תרגיל 8.3. יהי X משתנה מקרי. אם $\mathbb{E}\left[|X|^n\right] < \infty$, אז $\varphi_X^{(n)}(t)$ קיימת לכל t ומתקיים

$$\varphi_X^{(n)}(t) = \mathbb{E}\left[e^{itX}(iX)^n\right].$$

בפרט,

$$\mathbb{E}\left[X^n\right] = (-i)^n \varphi_X^{(n)}(0).$$

הוכחה. נוכיח עבור $n = 1$. לפי הגדרת הנגזרת,

$$\varphi_X'(t) = \lim_{h\to 0}\frac{\varphi_X(t+h) - \varphi_X(t)}{h} = \lim_{h\to 0}\mathbb{E}\left[\frac{e^{i(t+h)X} - e^{itX}}{h}\right] = \lim_{h\to 0}\mathbb{E}\left[e^{itX}\cdot\frac{e^{ihX}-1}{h}\right].$$

לפי משפט ההתכנסות הנשלטת, כיוון ש-$\left|\frac{e^{ihX}-1}{h}\right| \le |X|$ ו-$\mathbb{E}\left[|X|\right] < \infty$, אפשר להכניס את הגבול לתוחלת ולקבל

$$\varphi_X'(t) = \mathbb{E}\left[e^{itX}\lim_{h\to 0}\frac{e^{ihX}-1}{h}\right] = \mathbb{E}\left[e^{itX}(iX)\right]$$

כנדרש. □

הערה. עבור n-ים זוגיים, הקיום של $\varphi_X^{(n)}(0)$ שקול לקיום המומנט ה-n-י. עבור n-ים אי־זוגיים, זה לא בהכרח נכון.

8.1 תוחלת מותנית

הגדרה 8.4. יהי $(\Omega, \mathcal{F}, \mathrm{P})$ מרחב הסתברות, יהי X משתנה מקרי, ויהי $A \in \mathcal{F}$ מאורע מהסתברות חיובית. **התוחלת המותנית של X בהינתן** A היא התוחלת של X ביחס למידה $Q(B) = \mathrm{P}(A \mid B)$, ובנוסחה

$$\mathbb{E}\left[X \mid B\right] = \frac{\mathbb{E}\left[X\cdot 1_B\right]}{\mathrm{P}(B)} = \frac{1}{\mathrm{P}(B)}\int_B X\,\mathrm{dP}.$$

דוגמה 8.5. למשל אם $X = 1_A$, אז

$$\mathbb{E}\left[1_A \mid B\right] = \frac{\mathbb{E}\left[1_A\cdot 1_B\right]}{\mathrm{P}(B)} = \frac{\mathbb{E}\left[1_{A\cap B}\right]}{\mathrm{P}(B)} = \frac{\mathrm{P}(A\cap B)}{\mathrm{P}(B)}.$$

הגדרה 8.6. יהי $Y : \Omega \to \mathbb{R}$ משתנה מקרי בדיד. נגדיר את המשתנה $Z = \mathbb{E}\left[X \mid Y\right]$ להיות

$$Z(\omega) = \mathbb{E}\left[X \mid Y = Y(\omega)\right].$$

דוגמה 8.7. $\Omega=[0,1]$ עם מידת בורל, $X(\omega)=2\omega^2$, $Y(\omega)=\begin{cases}1, & 0\leq\omega\leq\frac{1}{3}\\ 2, & \frac{1}{3}<\omega\leq\frac{2}{3}\\ 0, & \frac{2}{3}<\omega\leq 1\end{cases}$.

בניסוח שקול, מגרילים $U\sim\mathrm{U}[0,1]$, $X=2U^2$, Y באופן דומה. נמצא את $Z=\mathbb{E}[X\mid Y]$.

$$Z(\omega)=\begin{cases}\mathbb{E}[X\mid Y=1], & 0\leq\omega\leq\frac{1}{3}\\ \mathbb{E}[X\mid Y=2], & \frac{1}{3}<\omega\leq\frac{2}{3}\\ \mathbb{E}[X\mid Y=0], & \frac{2}{3}<\omega\leq 1\end{cases}.$$

נחשב את התוחלת הראשונה כדוגמה:

$$\mathbb{E}[X\mid Y=1]=\frac{\mathbb{E}\left[X\cdot 1_{\{Y=1\}}\right]}{\mathrm{P}(Y=1)}=\frac{\mathbb{E}\left[X\cdot 1_{\left[0,\frac{1}{3}\right]}\right]}{\frac{1}{3}}=3\int_0^{\frac{1}{3}}2\omega^2\,\mathrm{d}\omega=3\cdot\left(\frac{2}{3}\omega^3\right)_0^{\frac{1}{3}}=\frac{2}{27}.$$

מטרה. להכליל להתנייה במשתנה מקרי כללי, ולמעשה ב-σ-אלגברה כללית.

הגדרה 8.8. יהי $(\Omega,\mathcal{F},\mathrm{P})$ מרחב הסתברות, תהי $\mathcal{G}\subseteq\mathcal{F}$ תת-σ-אלגברה, ויהי $X\in L^1(\Omega,\mathcal{F},\mathrm{P})$. **התוחלת המותנית של X בהינתן** $\mathcal{G}$ היא משתנה מקרי $Y=\mathbb{E}[X\mid\mathcal{G}]$ המקיים:

א. $Y\in L^1(\Omega,\mathcal{G},\mathrm{P})$. בפרט, Y הוא $\mathcal{G}$-מדיד.

ב. לכל $A\in\mathcal{G}$ מתקיים $\int_A X\,\mathrm{dP}=\mathbb{E}[X\cdot 1_A]=\mathbb{E}[Y\cdot 1_A]=\int_A Y\,\mathrm{dP}$.

דוגמה 8.9.

א. אם $\mathcal{G}=\{\varnothing,\Omega\}$, אז $\mathbb{E}[X\mid\mathcal{G}]=\mathbb{E}[X]$.

ב. אם $\mathcal{G}=\sigma(Y)$ לאיזשהו משתנה מקרי Y, אז מסמנים גם $\mathbb{E}[X\mid\mathcal{G}]=\mathbb{E}[X\mid Y]$.

יהיו $\Omega=[0,1]$, P מידת בורל. ניקח $X(\omega)=2\omega^2$, $Y(\omega)=\begin{cases}2, & 0\leq\omega<\frac{1}{2}\\ \omega, & \frac{1}{2}\leq\omega\leq 1\end{cases}$.

נמצא את $Z=\mathbb{E}[X\mid Y]$.

ראשית, נמצא את $\sigma(Y)=\{Y^{-1}(B)\mid B\in\mathbb{B}(\mathbb{R})\}$. תהי $B\in\mathbb{B}(\mathbb{R})$.

- אם $2\notin B$, אז $Y^{-1}(B)=Y^{-1}\left(B\cap\left[\frac{1}{2},1\right]\right)=B\cap\left[\frac{1}{2},1\right]$.
- אם $2\in B$, אז $Y^{-1}(B)=\left(B\cap\left[\frac{1}{2},1\right]\right)\cup\left[0,\frac{1}{2}\right)$.

בסך הכל,

$$\sigma(Y)=\left\{B\mid B\in\mathbb{B}(\mathbb{R}),B\subseteq\left[\frac{1}{2},1\right]\right\}\cup\left\{B\cup\left[0,\frac{1}{2}\right)\mid B\in\mathbb{B}(\mathbb{R}),B\subseteq\left[\frac{1}{2},1\right]\right\}.$$

נקבל בהכרח ש-Z קבוע על $\left[0,\frac{1}{2}\right)$. נמצא את ערכו c על הקטע הזה:

$$\mathbb{E}\left[Z\cdot 1_{\{Y=2\}}\right]=\mathbb{E}\left[X\cdot 1_{\{Y=2\}}\right]\Longrightarrow\frac{c}{2}=\int_0^{\frac{1}{2}}c\,\mathrm{d}\omega=\int_0^{\frac{1}{2}}2\omega^2\,\mathrm{d}\omega=\frac{1}{12}\Longrightarrow c=\frac{1}{6}.$$

לכן $Z\left(\omega\right)=\frac{1}{6}$ לכל $\omega\in\left[0,\frac{1}{2}\right)$.
לכל $\frac{1}{2}\leq\omega\leq 1$, נגדיר $Z\left(\omega\right)=X\left(\omega\right)$. נוכיח ש-$Z$ מקיים את ההגדרה של תוחלת מותנית. המדידות תנבע מכך ש-$\sigma\left(Z\right)=\sigma\left(Y\right)$. לגבי החלק השני, כל $A\in\sigma\left(Y\right)$ נפריד לשני חלקים: $A=\left[0,\frac{1}{2}\right)\uplus A'$, ואז

$$\mathbb{E}\left[Z\cdot 1_A\right]=\mathbb{E}\left[Z\cdot 1_{\left[0,\frac{1}{2}\right)}\right]+\mathbb{E}\left[Z\cdot 1_{A'}\right]=\mathbb{E}\left[X\cdot 1_{\left[0,\frac{1}{2}\right)}\right]+\mathbb{E}\left[X\cdot 1_{A'}\right]=\mathbb{E}\left[X\cdot 1_A\right].$$

9 תרגול תשיעי

טענה 9.1. יהי $(\Omega,\mathcal{F},\mathrm{P})$ מרחב הסתברות, תהי $\mathcal{G}\subseteq\mathcal{F}$ תת־σ-אלגברה, ויהיו $X,Y\in L^1\left(\Omega,\mathcal{F},\mathrm{P}\right)$.

א. $\mathbb{E}\left[\mathbb{E}\left[X\mid\mathcal{G}\right]\right]=\mathbb{E}\left[X\right]$.

ב. אם X הוא $\mathcal{G}$-מדיד, אז $\mathbb{E}\left[X\mid\mathcal{G}\right]=X$.

ג. לינאריות: $\mathbb{E}\left[aX+bY\mid\mathcal{G}\right]=a\mathbb{E}\left[X\mid\mathcal{G}\right]+b\mathbb{E}\left[Y\mid\mathcal{G}\right]$.

ד. חיוביות: אם $X\geq 0$, אז $\mathbb{E}\left[X\mid\mathcal{G}\right]\geq 0$.

ה. אי־שוויון ינסן: אם $\varphi\colon\mathbb{R}\to\mathbb{R}$ קמורה, אז $\varphi\left(\mathbb{E}\left[X\mid\mathcal{G}\right]\right)\leq\mathbb{E}\left[\varphi\left(X\right)\mid\mathcal{G}\right]$ כמעט תמיד.

ו. תכונת ההרכבה: אם $\mathcal{H}\subseteq\mathcal{G}\subseteq\mathcal{F}$, אז $\mathbb{E}\left[\mathbb{E}\left[X\mid\mathcal{G}\right]\mid\mathcal{H}\right]=\mathbb{E}\left[X\mid\mathcal{H}\right]$.

ז. תכונות TOWIK: אם Z הוא $\mathcal{G}$-מדיד, אז $\mathbb{E}\left[Z\cdot X\mid\mathcal{G}\right]=Z\cdot\mathbb{E}\left[X\mid\mathcal{G}\right]$.

כמו כן, יש משפטי התכנסות כמו שאנחנו מכירים.

תרגיל 9.2. יהיו X,Y משתנים מקריים אינטגרביליים, ונניח כי $\mathbb{E}\left[X\mid Y\right]=Y$ כמעט תמיד, וגם $\mathbb{E}\left[Y\mid X\right]=X$ כמעט תמיד. הוכיחו כי $X=Y$ כמעט תמיד.

הוכחה. מהנתון נקבל $\mathbb{E}\left[X-Y\mid X\right]=0$ כמעט תמיד. לכן לכל $t\in\mathbb{R}$ מתקיים

$$\mathbb{E}\left[\left(X-Y\right)1_{\{X\leq t\}}\right]=\mathbb{E}\left[\mathbb{E}\left[X-Y\mid X\right]1_{\{X\leq t\}}\right]=0$$

מפה נקבל

$$\mathbb{E}\left[\left(X-Y\right)1_{\{X\leq t,Y\leq t\}}\right]=\mathbb{E}\left[\left(X-Y\right)1_{\{X\leq t\}}\right]-\mathbb{E}\left[\left(X-Y\right)1_{\{X\leq t,Y>t\}}\right]=\\=-\mathbb{E}\left[\left(X-Y\right)1_{\{X\leq t,Y>t\}}\right].$$

מצד שני, באופן סימטרי לחלוטין נקבל

$$\mathbb{E}\left[(X-Y)\,1_{\{X\leq t,Y\leq t\}}\right]=-\mathbb{E}\left[(X-Y)\,1_{\{X>t,Y\leq t\}}\right].$$

מהשוואת הביטויים נקבל

$$\underbrace{\mathbb{E}\left[(X-Y)\,1_{\{X\leq t,Y>t\}}\right]}_{\leq 0}=\underbrace{\mathbb{E}\left[(X-Y)\,1_{\{X>t,Y\leq t\}}\right]}_{\geq 0}.$$

לכן נובע כי

$$\mathbb{E}\left[(X-Y)\,1_{\{X\leq t,Y>t\}}\right]=0.$$

קיבלנו משתנה מקרי אי־חיובי שהתוחלת שלו היא 0, ולכן הוא שווה 0 כמעט תמיד. כלומר לכל $t\in\mathbb{R}$ מתקיים

$$\mathrm{P}\left(X\leq t<Y\right)=0.$$

אבל אז

$$\{X<Y\}=\bigcup_{t\in\mathbb{Q}}\{X\leq t<Y\},$$

ולכן $\mathrm{P}\left(X<Y\right)=0$. מהסימטריה נקבל $\mathrm{P}\left(X>Y\right)=0$, ולכן $X=Y$ כמעט תמיד. □

9.1 מרטינגלים

הגדרה 9.3.

א. תהי I קבוצת אינדקסים (בדידה או רציפה), ונניח שלכל $t\in I$ מוגדר משתנה מקרי $X_t\colon(\Omega,\mathcal{F},\mathrm{P})\to(\mathbb{R},\mathbb{B}(\mathbb{R}))$. אז למשפחה $\{X_t\}_{t\in I}$ קוראים **תהליך סטוכסטי**.

ב. **פילטרציה** (או **סינון**) היא משפחת σ-אלגברות $\{\mathcal{F}_t\}_{t\in I}$ כך שלכל $t,s\in I$ עם $s\leq t$ מתקיים $\mathcal{F}_s\subseteq\mathcal{F}_t\subseteq\mathcal{F}$.

ג. אומרים שתהליך סטוכסטי $\{X_t\}_{t\in I}$ **מכבד** את הפילטרציה $\{\mathcal{F}_t\}_{t\in I}$, אם לכל $t\in I$ המשתנה X_t הוא $\mathcal{F}_t$-מדיד.

דוגמה 9.4. כל תהליך סטוכסטי $\{X_t\}_{t\in I}$ מכבד את **הסינון הטבעי** $\mathcal{F}_t=\sigma\left(X_s\mid s\leq t\right)$.

הגדרה 9.5. יהי $(\Omega,\mathcal{F},\mathrm{P})$ מרחב הסתברות, תהי $I\subseteq\mathbb{R}$ קבוצת אינדקסים, תהי $\{\mathcal{F}_t\}_{t\in I}$ פילטרציה, ויהי $\{X_t\}_{t\in I}$ תהליך סטוכסטי המכבד את $\{\mathcal{F}_t\}_{t\in I}$ כך ש-$X_t\in L^1(\Omega,\mathcal{F}_t,\mathrm{P})$.

א. אומרים ש-$\{X_t\}$ הוא **מרטינגל**, אם לכל $s\leq t$ מתקיים $\mathbb{E}\left[X_t\mid\mathcal{F}_s\right]=X_s$.

ב. אומרים ש-$\{X_t\}$ הוא **על־מרטינגל**, אם לכל $s\leq t$ מתקיים $\mathbb{E}\left[X_t\mid\mathcal{F}_s\right]\leq X_s$.

ג. אומרים ש-$\{X_t\}$ הוא **תת־מרטינגל**, אם לכל $s \leq t$ מתקיים $\mathbb{E}[X_t \mid \mathcal{F}_s] \geq X_s$.

אנחנו נתמקד במקרה של מרטינגלים בזמן בדיד. אז $I = \mathbb{N}_0$, ומספיק לבדוק את התנאים עבור $t = n+1$ ו-$s = n$.

דוגמה 9.6 (בהרצאה)**.** יהיו $X_1, X_2, \ldots$ ממבתש"ה המקיימים $\mathbb{E}[X_i] = 0$. נגדיר את $S_n = \sum_{i=1}^{n} X_i$. אז S_n מרטינגל ביחס לסינון הטבעי $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. אכן,

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[S_n + X_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = S_n + \mathbb{E}[X_{n+1}] = S_n.$$

מה היה קורה אם $\mathbb{E}[X_i] = \mu \neq 0$? היינו מקבלים

$$\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + \mu$$

ואז היינו יכולים להגדיר מרטינגל לפי $M_n = S_n - n\mu$. אז אכן

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[S_{n+1} - (n+1)\mu \mid \mathcal{F}_n] = S_n + \mu - (n+1)\mu = M_n.$$

תרגיל 9.7. יהיו $X_1, X_2, \ldots$ ממבתש"ה עם $\mathbb{E}[X_i] = 1$. נגדיר $M_n = \prod_{i=1}^{n} X_i$. אז M_n מרטינגל ביחס לסינון הטבעי.

הוכחה. אכן

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[M_n \cdot X_{n+1} \mid \mathcal{F}_n] = M_n \cdot \mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = M_n.$$

□

דוגמה 9.8 (בהרצאה)**.** נחזור לדוגמה 9.6. נניח ש-$X_1, X_2, \ldots$ ממבתש"ה עם $\mathbb{E}[X_i] = 0$ ו-$\mathrm{Var}(X_i) = \sigma^2$. האם אפשר לייצר מרטינגל מ-S_n^2? נבדוק:

$$\mathbb{E}\left[S_{n+1}^2 \mid \mathcal{F}_n\right] = \mathbb{E}\left[S_n^2 + 2S_n X_{n+1} + X_{n+1}^2 \mid \mathcal{F}_n\right] = S_n^2 + \sigma^2.$$

לכן נקבל ש-$S_n^2 - n\sigma^2$ מרטינגל.

דוגמה 9.9 (הכד של פוליה)**.** בכד יש שני כדורים, שחור ולבן. בכל שלב מוציאים באקראי כדור באופן אחיד, ומחזירים אותו יחד עם עוד כדור באותו הצבע. נגדיר $B_0 = 1$, ונגדיר את B_n להיות כמות הכדורים השחורים בשלב ה-n-י. אז $M_n = \frac{B_n}{n+2}$ הוא מרטינגל.

אכן,

$$\begin{aligned}\mathbb{E}[B_{n+1} \mid B_1, \ldots, B_n] &= \frac{B_n}{n+2}(B_n + 1) + \frac{n+2-B_n}{n+2} \cdot B_n = \\ &= \frac{B_n}{n+2}(B_n + 1 + n + 2 - B_n) = \frac{n+3}{n+2} B_n\end{aligned}$$

דוגמה 9.10 (מרטינגל חשיפת מידע)**.** יהי $(\Omega, \mathcal{F}, \mathrm{P})$ מרחב הסתברות, יהי $X \in L^1(\Omega, \mathcal{F}, \mathrm{P})$, ותהי $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ פילטרציה. אז $X_n = \mathbb{E}[X \mid \mathcal{F}_n]$ מגדיר מרטינגל שנקרא **מרטינגל חשיפת מידע**, כי

$$\mathbb{E}[X_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = \mathbb{E}[X \mid \mathcal{F}_n] = X_n.$$

טענה 9.11. אם X_n מרטינגל ו-$\varphi : \mathbb{R} \to \mathbb{R}$ פונקציה קמורה, אז $\varphi(X_n)$ הוא תת־מרטינגל.

טענה 9.12. אם X_n מרטינגל, אז לכל $n \in \mathbb{N}$ מתקיים $\mathbb{E}[X_n] = \mathbb{E}[X_0]$.

10 תרגול עשירי

10.1 זמני עצירה

הגדרה 10.1. משתנה מקרי $\tau : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ נקרא **זמן עצירה** ביחס לפילטרציה $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, אם לכל $n \in \mathbb{N}$ מתקיים $\{\tau \leq n\} \in \mathcal{F}_n$.

דוגמה 10.2. אם X_n מרטינגל, אז $\tau = \min\{n \in \mathbb{N} \mid X_n = 5\}$ הוא זמן עצירה. אכן,

$$\{\tau \leq n\} = \{X_n = 5\} \in \mathcal{F}_n .$$

לעומת זאת, $\tau = \min\{n \in \mathbb{N} \mid X_{n+1} = X_n\}$ הוא לא זמן עצירה, כי הוא תלוי ב-X_{n+1} שלא בהכרח $\mathcal{F}_n$-מדיד.

תרגיל 10.3 (בהרצאה)**.** יהיו τ_1, τ_2 זמני עצירה. אז גם $\tau_1 \wedge \tau_2 = \min\{\tau_1, \tau_2\}$, $\tau_1 \vee \tau_2 = \max\{\tau_1, \tau_2\}$, $\tau_1 + \tau_2$ הם זמני עצירה.

הוכחה. אכן,

$$\begin{aligned}
\{\tau_1 \wedge \tau_2 \leq n\} &= \{\tau_1 \leq n\} \cup \{\tau_2 \leq n\} \in \mathcal{F}_n \\
\{\tau_1 \vee \tau_2 \leq n\} &= \{\tau_1 \leq n\} \cap \{\tau_2 \leq n\} \in \mathcal{F}_n \\
\{\tau_1 + \tau_2 \leq n\} &= \bigcup_{i=0}^{n} (\{\tau_1 \leq i\} \cap \{\tau_2 \leq n-i\}) \in \mathcal{F}_n
\end{aligned}$$

□

משפט 10.4 (משפט העצירה של דוב)**.** יהי X_n מרטינגל ביחס לפילטרציה $\{\mathcal{F}_n\}$, ויהי τ זמן עצירה. אז X_τ מוגדר כמעט תמיד ומתקיים $\mathbb{E}[X_\tau] = \mathbb{E}[X_0]$ אם (לפחות) אחד משלושת התנאים הבאים מתקיים:

א. קיים קבוע $C \in \mathbb{R}$ שעבורו $\tau \overset{\text{a.s.}}{\leq} C$.

ב. $\tau \overset{\text{a.s.}}{<} \infty$ בהסתברות 1 וגם קיים קבוע $C \in \mathbb{R}$ שעבורו $|M_n| \overset{\text{a.s.}}{\leq} C$ לכל $n \leq \tau$.

ג. $\mathbb{E}[\tau] < \infty$ וגם קיים קבוע $C \in \mathbb{R}$ שעבורו $|M_{n+1} - M_n| \overset{\text{a.s.}}{\leq} C$ לכל $n \leq \tau$.

תרגיל 10.5. יהיו $X_1, X_2, \ldots$ ממבתש"ה עם התפלגות $X_i = \begin{cases} 1, & p \\ -1, & 1-p \end{cases}$. נגדיר $S_n = \sum_{i=1}^{n} X_i$. הוכיחו כי לכל $p \neq \frac{1}{2}$ קיים $\alpha \neq 1$ שעבורו α^{S_n} מרטינגל, ומצאו אותו.

פתרון. נחשב:

$$\mathbb{E}\left[\alpha^{S_{n+1}} \mid \mathcal{F}_n\right] = \mathbb{E}\left[\alpha^{S_n + X_{n+1}} \mid \mathcal{F}_n\right] = \alpha^{S_n} \cdot \left(\alpha p + \alpha^{-1}(1-p)\right).$$

לכן רוצים

$$\alpha p + \alpha^{-1}(1-p) = 1 \Longrightarrow p\alpha^2 - \alpha + (1-p) = 0$$

למשוואה הזו יש שני פתרונות: $\alpha = 1$ ו-$\alpha = \frac{1-p}{p}$. לכן כל אחד מהם מגדיר מרטינגל. כמובן $\alpha = 1$ מגדיר מרטינגל קבוע, אבל $\alpha = \frac{1-p}{p}$ מגדיר את **המרטינגל האקספוננציאלי**.

תרגיל 10.6. יהי S_n הילוך מקרי מוטה על $\mathbb{Z}$, כלומר $S_n = \sum_{i=1}^{n} X_i$ עבור $X_i = \begin{cases} 1, & p \\ -1, & q = 1-p \end{cases}$, $p \neq \frac{1}{2}$. לכל $a \in \mathbb{Z}$ נגדיר

$$\tau_a = \min\{t \geq 0 \mid S_t = a\}.$$

עבור $a, b > 0$, מהי $\mathrm{P}(\tau_{-a} < \tau_b)$?

פתרון. עבור הילוך מקרי פשוט השתמשנו במרטינגל S_n. הבעיה פה היא שזה לא מרטינגל. אפשר לתקן אותו אם ניקח $S_n - (2p-1)\,n$, אבל אז ה-OST לא יהיה מספיק. לכן נשתמש במרטינגל האקספוננציאלי שראינו בתרגול הקודם

$$M_n = \left(\frac{1-p}{p}\right)^{S_n}.$$

נגדיר $\tau = \tau_{-a} \wedge \tau_b$. זהו זמן עצירה סופי כמעט תמיד, כי הוא חסום על ידי משתנה גיאומטרי, וכן $|M_n| \overset{\text{a.s.}}{\leq} \max\{-a, b\}$, לכן אפשר להשתמש ב-OST:

$$1 = \mathbb{E}[M_0] = \mathbb{E}[M_\tau] = \left(\frac{1-p}{p}\right)^{-a} \cdot \mathrm{P}(\tau_{-a} < \tau_b) + \left(\frac{1-p}{p}\right)^{b} \cdot \mathrm{P}(\tau_b < \tau_{-a}).$$

מפה אפשר לפתור את המשוואה ולקבל

$$\mathrm{P}(\tau_{-a} < \tau_b) = \frac{1 - \left(\frac{1-p}{p}\right)^{b}}{\left(\frac{1-p}{p}\right)^{-a} - \left(\frac{1-p}{p}\right)^{b}}.$$

תרגיל 10.7 (בעיית המצביעים)**.** במדינה עמומה יש בחירות עם שני מועמדים. המועמד הראשון קיבל a קולות, והמועמד השני קיבל b קולות. נניח ש-$a > b$, והקולות נספרים אחד-אחד בסדר אקראי. מה הסיכוי שלאורך כל הספירה המועמד הראשון הוביל על השני?

פתרון. נגדיר $n = a + b$, a_k להיות כמות הקולות שהמועמד הראשון קיבל עד לקול ה-k-י, b_k עבור המועמד השני, $S_k = a_k - b_k$ לכל $0 \leq k \leq n-1$. נרצה לייצר מרטינגל

מ-S_k. ננסה ללכת בכיוון ההפוך, ולחשב את $\mathbb{E}\left[S_{n-k} \mid S_n, \ldots, S_{n-k+1}\right]$. נשים לב כי

$$S_{n-k} = \begin{cases} S_{n-k+1}+1, & \text{הקול ה-}(n-k+1)\text{ הוא של המועמד השני} \\ S_{n-k+1}-1, & \text{הקול ה-}(n-k+1)\text{ הוא של המועמד הראשון} \end{cases} =$$
$$= \begin{cases} S_{n-k+1}+1, & \frac{n-k+1-a_{n-k+1}}{n-k+1} \\ S_{n-k+1}-1, & \frac{a_{n-k+1}}{n-k+1} \end{cases}$$

מצד שני, $2a_{n-k+1} - n - k + 1 = S_{n-k+1}$, ולכן

$$\mathbb{E}\left[S_{n-k} \mid S_n, \ldots, S_{n-k+1}\right] = (S_{n-k+1}+1) \cdot \frac{n-k+1-a_{n-k+1}}{n-k+1} + (S_{n-k+1}-1) \cdot \frac{a_{n-k+1}}{n-k+1} =$$
$$= S_{n-k+1} + \frac{n-k+1-2a_{n-k+1}}{n-k+1} =$$
$$= S_{n-k+1} - \frac{S_{n-k+1}}{n-k+1} = \frac{n-k}{n-k+1} S_{n-k+1}$$

מכאן ש-$X_k = \frac{S_{n-k}}{n-k}$ הוא מרטינגל. נגדיר

$$\tau = \begin{cases} \min\left\{k \mid X_k = 0\right\}, & \text{אם קיים } k \text{ כזה} \\ n-1, & \text{אחרת} \end{cases}.$$

אם המועמד הראשון מוביל לאורך כל הספירה, אז $X_k > 0$ לכל $0 \le k \le n-1$, ולכן $\tau = n-1$, $X_\tau = X_{n-1} = S_1$. אחרת, $X_k = 0$ לאיזשהו $0 \le k < n-1$, וכן $X_\tau = 0$. τ הוא זמן עצירה חסום, ולכן לפי ה-OST,

$$\mathbb{E}\left[X_\tau\right] = \mathbb{E}\left[X_0\right] = \mathbb{E}\left[\frac{S_n}{n}\right] = \frac{a-b}{a+b},$$

ומצד שני

$$\mathbb{E}\left[X_\tau\right] = 1 \cdot \mathrm{P}(\text{המועמד הראשון תמיד מוביל}) + 0 \cdot \mathrm{P}(\text{המועמד הראשון לא תמיד מוביל}).$$

לכן

$$\mathrm{P}(\text{המועמד הראשון תמיד מוביל}) = \frac{a-b}{a+b}.$$

11 תרגול אחד עשר

11.1 מרטינגלים וריכוז מידה

טענה 11.1 (אי־שוויון אזומה-הופדינג). יהי $\{X_n\}$ מרטינגל המקיים $|X_n - X_{n-1}| \overset{\text{a.s.}}{\le} c_n$. אזי לכל $\lambda > 0$,

$$\mathrm{P}\left(|X_n - \mathbb{E}\left[X_n\right]| \ge \lambda\right) \le 2e^{-\frac{\lambda^2}{2\sum_{i=1}^n c_i^2}}.$$

טענה 11.2 (אי־שוויון דוב Doob's maximal inequality). יהי $\{X_n\}$ תת־מרטינגל. אז לכל $\lambda > 0$,

$$\mathrm{P}\left(\max_{0\le k\le n} X_n \ge \lambda\right) \le \frac{\mathbb{E}\left[\max\left\{X_n, 0\right\}\right]}{\lambda}.$$

דוגמה 11.3. מחלקים m כדורים ל-n תאים באקראי. נסמן על ידי $Z_{n,m}$ את כמות התאים הריקים. קל לוודא כי

$$\mathbb{E}\left[Z_{n,m}\right] = n \cdot \left(1 - \frac{1}{n}\right)^m.$$

האם אפשר להגיד משהו על הריכוז של המשתנה הזה? היינו רוצים להגיד $Z_{n,m} = \sum_{i=1}^n B_i$ כאשר B_i שואל האם התא ה-i ריק, אבל אז המשתנים $B_1, \ldots, B_n$ תלויים כמובן, ונצטרך להסתבך בשונויות משותפות. במקום זאת, יהי X_j המשתנה שעבור הכדור ה-j-י מחזיק את מספר התא שהכדור נחת בו. תהי $f\left(X_1, \ldots, X_m\right)$ הפונקציה הסופרת את כמות התאים הריקים. אז $M_k = \mathbb{E}\left[f\left(X_1, \ldots, X_n\right) \mid X_1, \ldots, X_k\right]$ הוא מרטינגל, וכן $M_m = Z_{n,m}$. שימו לב שההפרשים $|M_k - M_{k-1}| \le 1$ לכל k, כי הוספה של עוד משתנה לא יכולה לשנות את כמות התאים הריקים ביותר מ-1. לכן, לפי אזומה-הופדינג,

$$\mathrm{P}\left(\left|Z_{n,m} - n\left(1 - \frac{1}{n}\right)^m\right| \ge \lambda\right) \le 2e^{-\frac{\lambda^2}{2m}}.$$

בפרט, אם ניקח $\lambda = b\sqrt{m}$, נקבל

$$\mathrm{P}\left(\left|Z_{n,m} - n\left(1 - \frac{1}{n}\right)^m\right| \ge b\sqrt{m}\right) \le 2e^{-\frac{b^2}{2}}.$$

טענה 11.4 (אי־שוויון קולמוגורוב Kolmogorov's maximal inequality). יהיו $X_1, \ldots, X_n$ משתנים מקריים בלתי־תלויים עם תוחלת 0 ושונות סופית. נסמן $S_n = \sum_{i=1}^n X_i$. אז לכל $\lambda > 0$,

$$\mathrm{P}\left(\max_{1\le k\le n} |S_n| \ge \lambda\right) \le \frac{1}{\lambda^2}\mathrm{Var}\left(S_n\right) = \frac{1}{\lambda^2}\sum_{k=1}^n \mathrm{Var}\left(X_k\right).$$

הוכחה. כידוע, הסדרה S_n היא מרטינגל (נגדיר כרגיל $S_0 = 0$). הפונקציה $f(x) = x^2$ היא קמורה, לכן S_n^2 היא תת־מרטינגל. מאי־שוויון דוב נקבל

$$\mathrm{P}\left(\max_{1\le k\le n} |S_n| \ge \lambda\right) = P\left(\max_{1\le k\le n} S_n^2 \ge \lambda^2\right) \le \frac{1}{\lambda^2}\mathbb{E}\left[S_n^2\right].$$

כנדרש. □

מרציפות ההסתברות, אם נשאיף $n \to \infty$ נקבל גרסה דומה למקרה של אינסוף משתנים מקריים.

מסקנה 11.5 (משפט שני הטורים של קולמוגורוב)**.** יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים עם תוחלות μ_n ושונויות σ_n^2 סופיות. אם $\sum_{n=1}^{\infty} \mu_n < \infty$ וגם $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$, אז $\sum_{n=1}^{\infty} X_n$ מתכנס כמעט תמיד.

הוכחה. אפשר להניח כי $\mu_n = 0$, אחרת נחליף את X_n ב־$X_n - \mu_n$ (פה חשוב טור התוחלות). כדי להוכיח ש־$\sum_{n=1}^{\infty} X_n$ מתכנס כמעט תמיד, מספיק להראות שסדרת הסכומים החלקיים $S_n = \sum_{k=1}^{n} X_k$ היא סדרת קושי כמעט תמיד. יהי $\varepsilon > 0$ ויהי $N \in \mathbb{N}$. נשים לב כי מאי־שוויון קולמוגורוב למקרה האינסופי,

$$\mathrm{P}\left(\sup_{N<n\leq k} |S_n - S_N| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{i=N+1}^{k} \mathrm{Var}\left(X_i^2\right) = \frac{1}{\varepsilon^2} \sum_{i=N+1}^{k} \sigma_i^2 \leq \frac{1}{\varepsilon^2} \sum_{i=N+1}^{\infty} \sigma_i^2$$

לכן

$$L_{N,\varepsilon} = \mathrm{P}\left(\sup_{n>N} |S_n - S_N| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} \sum_{i=N+1}^{\infty} \sigma_i^2 \xrightarrow[N\to\infty]{} 0.$$

מאי־שוויון המשולש, $\sup_{m,n>N} |S_n - S_m| \leq 2 \sup_{n>N} |S_n - S_N|$. לכן לכל $\varepsilon > 0$ נקבל

$$\mathrm{P}\left(\bigcap_{N=1}^{\infty} \bigcup_{m,n>N} \{|S_n - S_m| > \varepsilon\}\right) = \lim_{N\to\infty} \mathrm{P}\left(\sup_{m,n>N} |S_n - S_m| > \varepsilon\right) \leq \limsup_{N\to\infty} L_{N,\varepsilon/2} = 0.$$

אם ניקח את האיחוד של זה על פני $\varepsilon_n = \frac{1}{n}$, נקבל שההסתברות ש־$S_n$ אינה סדרת קושי היא 0. □

מסקנה 11.6. נניח ε_n סימנים אקראיים. אז $\sum_{n=1}^{\infty} \frac{\varepsilon_n}{n^\alpha}$ מתכנס כמעט תמיד לכל $\alpha > \frac{1}{2}$.

משפט 11.7 (משפט שלושת הטורים של קולמוגורוב)**.** יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים. אז $\sum_{n=1}^{\infty} X_n$ מתכנס כמעט תמיד אם ורק אם קיים $A > 0$ שעבורו התנאים הבאים מתקיימים (ואז הם מתקיימים לכל $A > 0$):

א. $\sum_{n=1}^{\infty} \mathrm{P}\left(|X_n| \geq A\right)$ מתכנס.

ב. נגדיר $Y_n = X_n \cdot 1_{\{|X_n|\leq A\}}$. אז $\sum_{n=1}^{\infty} \mathbb{E}\left[Y_n\right]$ מתכנס.

ג. $\sum_{n=1}^{\infty} \mathrm{Var}\left(Y_n\right)$ מתכנס.

11.2 התכנסות מרטינגלים

משפט 11.8. על־מרטינגל שחסום ב־L^1 (כלומר $\sup_{n\in\mathbb{N}} \mathbb{E}\left[|X_n|\right] < \infty$) מתכנס כמעט תמיד וב־$L_1$.

מסקנה 11.9. על־מרטינגל אי־שלילי (באופן כללי יותר: חסום מלמטה) מתכנס כמעט תמיד.

תרגיל 11.10. הראו שהילוך מקרי על $\mathbb{Z}$ הוא נשנה, כלומר חוזר לנקודת ההתחלה בסיכוי 1.

הוכחה. נסמן את ההילוך על ידי S_n. נניח כי $S_0 = 1$, ונראה כי S_n מגיע ל-0 בסיכוי 1. ראינו בעבר כי S_n הוא מרטינגל. נגדיר $M_n = S_{n\wedge\tau_0}$, כאשר $\tau_0 = \min\{n \in \mathbb{N} \mid S_n = 0\}$. גם זה מרטינגל, אבל הפעם הוא אי־שלילי. לכן $M_n \xrightarrow{\text{a.s.}} M_\infty$. אבל M_n הוא מרטינגל של מספרים שלמים, לכן אם הוא מתכנס הוא חייב להיות קבוע לבסוף, ומכאן שהוא חייב להיות 0 לבסוף. לכן $\mathrm{P}(\tau_0 < \infty) = 1$. □

הגדרה 11.11. אומרים ש-$X_n \xrightarrow{L^p} X$ אם $\|X_n - X\|_p = \mathbb{E}\left[|X_n - X|^p\right]^{1/p} \xrightarrow[n\to\infty]{} 0$.

טענה 11.12. נניח ש-$X_n \xrightarrow{L^p} X$.

א. $\mathbb{E}\left[|X_n|^p\right] \to \mathbb{E}\left[|X_n|^p\right]$.

ב. $X_n \xrightarrow{\mathrm{P}} X$.

ג. לכל $1 \le q < p$, $X_n \xrightarrow{L^q} X$.

ד. אם $X_n \xrightarrow{\text{a.s.}} Y$, אז $X = Y$.

תרגיל 11.13. הראו שהסדרה $X_n = \begin{cases} n^2, & \frac{1}{n} \\ 0, & 1-\frac{1}{n} \end{cases}$ מתכנסת בהסתברות ל-0 אבל לא מתכנסת ל-0 באף L^p.

פתרון. לטענת ההתכנסות בהסתברות, לכל $\varepsilon > 0$ מתקיים

$$\lim_{n\to\infty} P\left(|X_n - 0| \ge \varepsilon\right) = \lim_{n\to\infty} P\left(X_n = n^2\right) = 0 .$$

לגבי טענת ההתכנסות ב-L^p, לכל $p \ge 1$ מתקיים

$$\lim_{n\to\infty} \mathbb{E}\left[|X_n|^p\right] = \lim_{n\to\infty} n^{2p} \cdot \frac{1}{n} = \lim_{n\to\infty} n^{2p-1} = \infty .$$

משפט 11.14. מרטינגל שחסום באופן אחיד ב-L^p מתכנס בנורמת L^p.

הגדרה 11.15. תהליך גלטון-ווטסון הוא תהליך סטוכסטי $\{Z_n\}$ המוגדר כך: בכל דור, כל פריט מוליד צאצאים לפי התפלגות X באופן בלתי־תלוי בשאר הפריטים ובכל הדורות הקודמים. פורמלית, $Z_0 = 1$, ולכל $n \ge 0$ מתקיים

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_j^{(n)}$$

כאשר $X_j^{(n)}$ ממבתש"ה בעלי ערכים שלמים אי־שליליים (עם התפלגות כמו של X). **הסתברות ההכחדה** של $\{Z_n\}$ היא $\lim_{n\to\infty} \mathrm{P}(Z_n = 0)$.

משפט 11.16. הסתברות ההכחדה של $\{Z_n\}$ היא 1 אם ורק אם $\mathbb{E}[Z_n] < 1$ או $\mathbb{E}[Z_n] = 1$ וגם $X_n \neq 1$.

הוכחה. נסמן $\mu = \mathbb{E}\left[X_j^{(n)}\right]$. נרצה לבנות מרטינגל מ-$\{Z_n\}$. נשים לב כי

$$\mathbb{E}[Z_{n+1} \mid Z_1, \ldots, Z_n] = \mathbb{E}\left[\sum_{j=1}^{Z_n} X_j^{(n)} \mid Z_1, \ldots, Z_n\right] = \mu Z_n.$$

לכן $M_n = \frac{Z_n}{\mu^n}$ הוא מרטינגל.

אם $\mu < 1$, אז החישוב הנ"ל מראה כי $\mathbb{E}[Z_n] = \mu^n$. באמצעות אי־שוויון מרקוב אפשר לראות כי

$$\mathrm{P}(Z_n > 0) = \mathrm{P}(Z_n \geq 1) \leq \mathbb{E}[Z_n] = \mu^n \xrightarrow[n\to\infty]{} 0,$$

מה שמראה ש-$\lim_{n\to\infty} \mathrm{P}(Z_n = 0) = 1$.

אם $\mu = 1$, נצטרך להתאמץ קצת יותר. אז המרטינגל הוא $M_n = Z_n$, והוא מרטינגל אי־שלילי, לכן מתכנס. כלומר קיים Z_∞ כך ש-$Z_n \xrightarrow{\text{a.s.}} Z_\infty$. אבל Z_n הוא מרטינגל של מספרים שלמים, לכן הוא חייב להיות קבוע לבסוף. אבל כל עוד $\mathrm{P}(X_1 = 1) \neq 1$, בהכרח נקבל $Z_\infty = 0$, כנדרש.

נותר המקרה $\mu > 1$. אפשר להניח שהמשתנה X של התפלגות הצאצאים חסום, ובפרט $1 < \mu < \infty$, אם נחליף את X ב-$X \wedge M$ ל-M מספיק גדול. אז שוב $M_n = \frac{Z_n}{\mu^n}$ הוא מרטינגל אי־שלילי, ולכן מתכנס לאיזשהו משתנה M_∞. נשים לב שמתקיים

$$\{M_\infty > 0\} \subseteq \{\forall n : Z_n > 0\}.$$

נרצה להוכיח ש-$\sup_{n\in\mathbb{N}} \mathbb{E}[M_n^2] \leq C$ לאיזשהו חסם אחיד C. נסמן $\sigma^2 = \mathrm{Var}(X)$. נתחיל מההפרשים:

$$\mathbb{E}\left[(M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}\right] = \mathbb{E}\left[\left(\frac{Z_k}{\mu^k} - \frac{Z_{k-1}}{\mu^{k-1}}\right)^2 \mid \mathcal{F}_{k-1}\right] = \frac{1}{\mu^{2k}}\mathbb{E}\left[(Z_k - \mu Z_{k-1})^2 \mid \mathcal{F}_{k-1}\right] =$$

$$= \frac{1}{\mu^{2k}}\mathbb{E}\left[\left(\sum_{j=1}^{Z_{k-1}} X_j^{(k)} - \mu Z_{k-1}\right)^2 \mid \mathcal{F}_{k-1}\right] =$$

$$= \frac{1}{\mu^{2k}}\mathrm{Var}\left(\sum_{j=1}^{Z_{k-1}} X_j^{(k)}\right) = \frac{1}{\mu^{2k}} Z_{k-1}\sigma^2$$

ולכן

$$\mathbb{E}\left[(M_k - M_{k-1})^2\right] = \frac{1}{\mu^{2k}} \cdot \mu^{k-1} \cdot \sigma^2 = \frac{1}{\mu^{k+1}}\sigma^2.$$

מטענה בתרגילי הבית,

$$\mathbb{E}\left[M_n^2\right] = \mathbb{E}\left[M_0^2\right] + \sum_{k=1}^{n} \mathbb{E}\left[(M_k - M_{k-1})^2\right] = 1 + \sum_{k=1}^{n} \frac{1}{\mu^{k+1}}\sigma^2 \leq 1 + \sigma^2 \sum_{k=1}^{\infty} \frac{1}{\mu^{k+1}} < \infty.$$

לכן $M_n \xrightarrow{L^2} M_\infty$, ובפרט $M_n \xrightarrow{L^1} M_\infty$. בפרט $1 = \mathbb{E}[M_n] \xrightarrow[n\to\infty]{} \mathbb{E}[M_\infty]$, ולכן $\mathrm{P}(M_\infty > 0) > 0$. זה מראה שיש הסתברות חיובית לשרוד. □

12 תרגול שנים עשר

12.1 צימודים

הגדרה 12.1.

א. **צימוד של שתי מידות** μ, ν על מרחב Ω הוא זוג משתנים מקריים (X, Y) כך ש-$X \sim \mu$ ו-$Y \sim \nu$.

ב. **צימוד של שני משתנים מקריים** W, Z הוא זוג משתנים מקריים X, Y כך ש-$X \sim W$ ו-$Y \sim Z$.

ג. **צימוד של שתי מידות הסתברות** $\mathrm{P}_1, \mathrm{P}_2$ על Ω הוא מידה $\hat{\mathrm{P}}$ על $\Omega \times \Omega$ כך ש-$\pi_i\left(\hat{\mathrm{P}}\right) = \mathrm{P}_i$.

דוגמה 12.2. אם יש לנו שני מטבעות, אחד עם סיכוי p ליפול על "עץ" והשני עם סיכוי $q > p$ ליפול על "עץ", יש שני צימודים טבעיים להגדיר עליהם:

א. אפשר לצמד ביניהם באופן בלתי־תלוי. אכן, ניקח $\Omega = \{0,1\}^2$ עם מידת ההסתברות $\mathrm{P}(a,b) = p^a q^b$, ואז הקואורדינטה הראשונה מתפלגת $\mathrm{Ber}(p)$ והשנייה $\mathrm{Ber}(q)$.

ב. אפשר לצמד ביניהם כך שאם המטבע הראשון נופל על "עץ", גם השני ייפול על "עץ". אכן, נסמן על ידי $X \sim \mathrm{Ber}(p)$ את תוצאת המטבע הראשון, ועל ידי $Y \sim \mathrm{Ber}(q)$ את תוצאת המטבע השני. נגריל מספר אחיד $U \sim \mathrm{U}(0,1)$, ונקבע

$$X = \begin{cases} 1, & U \leq p \\ 0, & U > p \end{cases}, \qquad Y = \begin{cases} 1, & U \leq q \\ 0, & U > q \end{cases}.$$

במקרה הזה ההתפלגות המשותפת היא

$$\begin{aligned} \mathrm{P}((X,Y) = (0,0)) &= 1 - q, & \mathrm{P}((X,Y) = (0,1)) &= q - p \\ \mathrm{P}((X,Y) = (1,0)) &= 0, & \mathrm{P}((X,Y) = (1,1)) &= p \end{aligned}$$

תרגיל 12.3. יהיו $X_1, X_2, \ldots$ סדרת משתנים מקריים בלתי־תלויים בעלי התפלגות

$$.X_n = \begin{cases} 1, & \frac{1}{2} - \frac{1}{2n^2} \\ -1, & \frac{1}{2} - \frac{1}{2n^2} \\ n^2, & \frac{1}{n^2} \end{cases}$$

שימו לב כי $\mathbb{E}[X_n] = 1$ לכל n. הוכיחו כי $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} 0$.

הוכחה. נגדיר $Y_n = \begin{cases} 1, & \frac{1}{2} \\ -1, & \frac{1}{2} \end{cases}$, ונצמד בין X_n ל-Y_n באופן הבא: ראשית נגריל את X_n. אם $X_n \neq n^2$, ניקח $Y_n = X_n$; אחרת, נגריל $Y_n = \begin{cases} 1, & \frac{1}{2} \\ -1, & \frac{1}{2} \end{cases}$. לכן $\mathrm{P}(X_n \neq Y_n) = \frac{1}{n^2}$. מבורל-קנטלי, $\mathrm{P}(X_n \neq Y_n \text{ i.o.}) = 0$.
הראינו שבהסתברות 1, $X_n = Y_n$ לכמעט כל n. נטען שאם הגבול $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i$ קיים, אז גם הגבול $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i$ קיים ושווה לו. אכן, לכל $\omega \in \Omega$ שעבורו $X_n = Y_n$ לכמעט כל n, ההבדל בין הסכומים הוא רק במספר סופי של מחוברים, וכשנשאיף $n \to \infty$ הגורם הזה ישאף ל-0.
מצד שני, לפי החוק החזק של המספרים הגדולים, $\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{\text{a.s.}} \mathbb{E}[Y_i] = 0$, כנדרש. □

הגדרה 12.4. יהיו μ, ν שתי מידות על $(\Omega, \mathcal{F})$. מגדירים

$$. \|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \nu(x)| = \sup_{A\in\mathcal{F}} |\mu(A) - \nu(A)|$$

טענה 12.5. לכל צימוד (X, Y),

$$. \|\mu - \nu\|_{\mathrm{TV}} \leq \mathrm{P}(X \neq Y)$$

בנוסף, קיים צימוד שמשיג את החסם הזה.

דוגמה 12.6. בדוגמה של צימוד $\mathrm{Ber}(p)$ ו-$\mathrm{Ber}(q)$ עם $q > p$,

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2}(|\mu(0) - \nu(0)| + |\mu(1) - \nu(1)|) = \frac{1}{2}(|(1-p) - (1-q)| + |p - q|) =$$
$$= \frac{1}{2} \cdot 2 \cdot |p - q| = q - p$$

ביחס לצימוד שבו הם בלתי־תלויים,

$$.\mathrm{P}(X \neq Y) = (1-p)\,q + p\,(1-q) = p + q - 2pq$$

ביחס לצימוד המונוטוני,

$$.\mathrm{P}(X \neq Y) = \mathrm{P}((X, Y) = (0, 1)) = q - p$$

זהו באמת הצימוד המקסימלי.

איך מוצאים את הצימוד המקסימלי? נגדיר $r = \min\{\mu, \nu\}$. נחשב

$$q = \int_{\mathbb{R}} r(x)\,\mathrm{d}x = 1 - \|\mu - \nu\|_{\mathrm{TV}}.$$

מגרילים $Z \sim \mathrm{Ber}(q)$. נסמן גם:

- V משתנה מקרי המתפלג לפי $\frac{r(x)}{q}$;
- W משתנה מקרי המתפלג לפי $\frac{\mu(x)-r(x)}{1-q}$;
- W' משתנה מקרי המתפלג לפי $\frac{\nu(x)-r(x)}{1-q}$.

אז הצימוד המקסימלי הוא

$$X = \begin{cases} V, & Z = 1 \\ W, & Z = 0 \end{cases}. \qquad Y = \begin{cases} V, & Z = 1 \\ W', & Z = 0 \end{cases}$$

תרגיל 12.7. מצאו צימוד מקסימלי של X, Y המוגדרים על $[0, \infty)$ עם $X \sim \mathrm{Exp}(1)$ ו-$Y \sim \mathrm{Exp}(2)$.

פתרון. כזכור, רוצים

$$f_X(t) = e^{-t}, \qquad f_Y(t) = 2e^{-2t}$$

נשים לב כי

$$f_X(t) > f_Y(t) \iff e^{-t} > 2e^{-2t} \iff e^t > 2 \iff t > \log 2.$$

לכן

$$r(x) = \min\{f_X(t), f_Y(t)\} = \begin{cases} e^{-t}, & t \le \log 2 \\ 2e^{-2t}, & t > \log 2 \end{cases}.$$

מכאן

$$q = \int_0^\infty r(x)\,\mathrm{d}x = \int_0^{\log 2} e^{-t}\,\mathrm{d}t + \int_{\log 2}^\infty 2e^{-2t}\,\mathrm{d}t = \left(-e^{-t}\right)_0^{\log 2} + \left(-e^{-2t}\right)_{\log 2}^\infty =$$
$$= -\frac{1}{2} + 1 + \left(-0 + \frac{1}{4}\right) = \frac{3}{4}.$$

בפרט מרחק ה-variation total הוא $\frac{1}{4}$.
ניקח משתנים מקריים:

- $Z \sim \mathrm{Ber}\left(\frac{3}{4}\right)$;
- יהי V משתנה מקרי בעל הצפיפות $\frac{4}{3}\min\{e^{-t}, 2e^{-2t}\}$ לכל $t \ge 0$;

- יהי W משתנה מקרי בעל הצפיפות $4\left(e^{-t}-2e^{-2t}\right)$ לכל $t\geq\log 2$;
- ויהי W' משתנה מקרי בעל הצפיפות $4\left(2e^{-2t}-e^{-t}\right)$ לכל $0\leq t\leq\log 2$.

אז הצימוד המקסימלי הוא

$$X=\begin{cases}V, & Z=1\\ W, & Z=0\end{cases},\qquad Y=\begin{cases}V, & Z=1\\ W', & Z=0\end{cases}.$$

נדגים למשל ש-X אקספוננציאלי:

$$F_X(a)=\mathrm{P}(X\leq a)=\mathrm{P}(X\leq a\mid Z=0)\cdot\mathrm{P}(Z=0)+\mathrm{P}(X\leq a\mid Z=1)\cdot\mathrm{P}(Z=1)=$$
$$=\mathrm{P}(W\leq a)\cdot\frac{1}{4}+P(V\leq a)\cdot\frac{3}{4}$$

אם $a<\log 2$ אז

$$F_X(a)=\frac{3}{4}\cdot\frac{4}{3}\cdot\int_0^a\min\left\{e^{-t},2e^{-2t}\right\}\,\mathrm{d}t=\int_0^a e^{-t}\,\mathrm{d}t=\left(-e^{-t}\right)_0^a=1-e^{-a}$$

ואם $a\geq\log 2$ אז

$$F_X(a)=\frac{1}{4}\cdot 4\cdot\int_{\log 2}^a\left(e^{-t}-2e^{-2t}\right)\,\mathrm{d}t+\frac{3}{4}\cdot\frac{4}{3}\cdot\int_0^a\min\left\{e^{-t},2e^{-2t}\right\}\,\mathrm{d}t=$$
$$=\int_{\log 2}^a e^{-t}\,\mathrm{d}t-\int_{\log 2}^a 2e^{-2t}\,\mathrm{d}t+\int_0^{\log 2}e^{-t}\,\mathrm{d}t+\int_{\log 2}^a 2e^{-2t}\,\mathrm{d}t=$$
$$=\int_0^a e^{-t}\,\mathrm{d}t=\left(-e^{-t}\right)_0^a=1-e^{-a}$$

זו בדיוק פונקציית ההתפלגות המצטברת של התפלגות מעריכית.

12.2 שליטה סטוכסטית

הגדרה 12.8. יהיו X,Y שני משתנים מקריים. אומרים ש-X **נשלט סטוכסטית על ידי** Y, אם לכל a מתקיים $P(X\geq a)\leq P(Y\geq a)$.

טענה 12.9. X נשלט סטוכסטית על ידי Y אם ורק אם קיים צימוד של (X,Y) שעבורו $P(X\leq Y)=1$.

תרגיל 12.10. יהיו שני משתנים מקריים $X=\begin{cases}0, & \frac{1}{2}\\ 1, & \frac{1}{2}\end{cases}$ ו-$Y=\begin{cases}0, & \frac{1}{3}\\ 1, & \frac{1}{3}\\ 2, & \frac{1}{3}\end{cases}$. הראו ש-$X$ נשלט סטוכסטית על ידי Y.

הוכחה. נגדיר צימוד באופן הבא: ראשית נגריל את Y, ואז נגריל את X באופן הבא: אם $Y=0$, ניקח $X=0$; אם $Y=2$, ניקח $X=1$; אם $Y=1$, נגריל $X=\begin{cases}0, & \frac{1}{2}\\ 1, & \frac{1}{2}\end{cases}$. קל לראות שבצימוד זה $P(X\le Y)=1$. □

תרגיל 12.11. נניח $X\sim \mathrm{Bin}(n,p)$, $Y\sim \mathrm{Bin}(m,p)$ עבור $m>n$. הראו ש-Y שולט סטוכסטית על X.

הוכחה. ניקח צימוד כך: ראשית נגריל את X, ואז נגריל $Z\sim \mathrm{Bin}(m-n,p)$ בלתי־תלוי ונגדיר $Y=X+Z$. אז $Y\sim \mathrm{Bin}(m,p)$, וכן $X\le Y$ תמיד. □

13 תרגול שלושה עשר

13.1 תהליכי פואסון

הגדרה 13.1. תהליך מנייה (Counting process) הוא תהליך הסופר את כמות המאורעות שמתרחשים לפי תנאים מסוימים. באופן פורמלי, זהו תהליך סטוכסטי $\{X_t\}_{t\ge 0}$ כך ש-X_t מייצג את כמות המאורעות שהתרחשו בין זמן 0 לזמן t. נדרוש את התנאים הבאים:

- $X_0=0$.
- לכל $t>0$, $X_t\in\mathbb{N}_0$.
- לכל $s<t$, X_t-X_s סופר את כמות המאורעות בין זמן s לזמן t.

נאמר של-$\{X_t\}$ יש **הפרשים בלתי־תלויים**, אם לכל $0\le t_1<t_2<\cdots<t_n$, המשתנים $X_{t_2}-X_{t_1},\ldots,X_{t_n}-X_{t_{n-1}}$ הם בלתי־תלויים.

הגדרה 13.2. נזכיר את התפלגות פואסון – $X\sim \mathrm{Poi}(\lambda)$ אם

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!},\qquad k=0,1,2,\ldots.$$

הגדרה 13.3. תהליך פואסון הוא תהליך מנייה הממדל התרחשות מאורעות בקצב מסוים (אם הוא קבוע בזמן – **תהליך פואסון הומוגני**). פורמלית, נתון $\lambda>0$. **תהליך פואסון עם קצב** λ הוא תהליך מנייה $\{X_t\}_{t\ge 0}$ המקיים:

א. $X_0=0$.

ב. X_t בעל הפרשים בלתי־תלויים.

ג. בכל קטע מאורך $\tau>0$, כמות המאורעות בקטע מתפלגת $\mathrm{Poi}(\tau\lambda)$. בפרט, כמות המאורעות אינה תלויה בקטע עצמו, אלא רק באורך שלו ("**הפרשים סטציונריים**").

תרגיל 13.4. מספר הלקוחות המגיעים למכולת הוא תהליך פואסון, כשבשעה נכנסים בממוצע 10 לקוחות.

א. מה ההסתברות שהגיעו שני לקוחות בין 10:00 ל-10:20?

ב. מה ההסתברות שהגיעו שני לקוחות בין 10:00 ל-10:20 וגם שמונה לקוחות בין 10:20 ל-11:00?

פתרון. נסמן את התהליך על ידי $\{X_t\}_{t\geq 0}$.

א. לפי הנתונים, $X_{10\frac{1}{3}} - X_{10} \sim \text{Poi}\left(\frac{10}{3}\right)$, לכן

$$P\left(X_{10\frac{1}{3}} - X_{10} = 2\right) = P\left(\text{Poi}\left(\frac{10}{3}\right) = 2\right) = e^{-\frac{10}{3}}\frac{\left(\frac{10}{3}\right)^2}{2!} \approx 0.198.$$

ב. כיוון שההפרשים בלתי־תלויים,

$$P\left(X_{10\frac{1}{3}} - X_{10} = 2, X_{11} - X_{10\frac{1}{3}} = 8\right) = P\left(X_{10\frac{1}{3}} - X_{10} = 2\right) \cdot P\left(X_{11} - X_{10\frac{1}{3}} = 8\right) =$$
$$= e^{-\frac{10}{3}}\frac{\left(\frac{10}{3}\right)^2}{2!} \cdot e^{-\frac{20}{3}}\frac{\left(\frac{20}{3}\right)^8}{8!} \approx 0.0244$$

תרגיל 13.5. יהי $\{X_t\}$ תהליך פואסון עם קצב λ. מהי הפונקציה $C(t_1, t_2) = \text{Cov}(X_{t_1}, X_{t_2})$?

פתרון. נניח בה"כ $t_1 \leq t_2$. לכן

$$C(t_1, t_2) = \text{Cov}(X_{t_1}, X_{t_2}) = \text{Cov}(X_{t_1}, X_{t_2} - X_{t_1}) + \text{Cov}(X_{t_1}, X_{t_1}).$$

כיוון שההפרשים בלתי־תלויים, הגורם הראשון הוא 0. נותרנו עם

$$C(t_1, t_2) = \text{Var}(X_{t_1}) = \lambda t_1.$$

לכן נקבל

$$C(t_1, t_2) = \lambda \min\{t_1, t_2\}.$$

טענה 13.6. יהי $\{X_t\}$ תהליך פואסון עם קצב λ. יהי T_i הזמן שבו קרה המאורע ה-i-י **(זמני ההגעה)**. נגדיר $W_n = T_n - T_{n-1}$ **(זמני ההמתנה)**. אזי

$$W_n \sim \text{Exp}(\lambda)$$

והמשתנים $W_1, W_2, \ldots$ בלתי־תלויים.

מסקנה 13.7. ההתפלגות של T_n היא $\Gamma(n, \lambda)$, ובפרט $\mathbb{E}[T_n] = \frac{n}{\lambda}$ ו-$\text{Var}(T_n) = \frac{n}{\lambda^2}$.

טענה 13.8 (איחוד, דילול ופיצול של תהליכי פואסון). יהיו $\{X_t\}_{t\geq 0}$ ו-$\{Y_t\}_{t\geq 0}$ תהליכי פואסון בלתי־תלויים עם קצבים λ ו-μ בהתאמה. אזי:

א. $\{X_t + Y_t\}$ הוא תהליך פואסון עם קצב $\lambda + \mu$.

ב. נניח שנתון $0 \leq p \leq 1$. נצבע את הנקודות של X_t בשני צבעים בכך שנטיל מטבע-p על כל נקודה כשהמטבעות בלתי־תלויים זה בזה; אם הוא יוצא 1 נצבע אותה באדום, ואם הוא יוצא 0 נצבע אותה בכחול. נסמן ב-R_t את תהליך הספירה של הנקודות האדומות וב-B_t את תהליך הספירה של הנקודות האדומות. אזי $\{R_t\}$ הוא תהליך פואסון עם קצב $p\lambda$, ו-$\{B_t\}$ הוא תהליך פואסון עם קצב $(1-p)\,\lambda$. בנוסף, R_t ו-B_t בלתי־תלויים.

תרגיל 13.9. יהיו $\{X_t\}_{t\geq 0}$ ו-$\{Y_t\}_{t\geq 0}$ תהליכי פואסון בלתי־תלויים עם קצבים $\lambda = 1$ ו-$\mu = 2$ בהתאמה. מצא את ההסתברות שזמן ההגעה השני ב-X_t קורה לפני זמן ההגעה השלישי ב-Y_t.

פתרון. ניקח תהליך פואסון $\{Z_t\}_{t\geq 0}$ עם קצב $\xi = 3$, ונפצל אותו לשני תהליכי פואסון $\{X_t\}_{t\geq 0}$ ו-$\{Y_t\}_{t\geq 0}$ באופן הבא: לכל מאורע שקורה נטיל מטבע-$\frac{1}{3}$; אם מקבלים 1 נוסיף אותו ל-X_t, אחרת ל-Y_t. הטלות המטבעות בלתי־תלויות ב-Z_t, ולכן X_t הוא תהליך פואסון עם קצב 1 ו-Y_t הוא תהליך פואסון עם קצב 2. כמו כן, מהטענה הם בלתי־תלויים. לכן ניתן לנסח מחדש את השאלה באופן הבא: אם מטילים מטבעות-$\frac{1}{3}$ בלתי־תלויים בזה אחר זה, מה הסיכוי שמתוך ארבע ההטלות הראשונות נראה לפחות פעמיים 1? התשובה היא

$$\sum_{k=2}^{4} \binom{4}{k} \left(\frac{1}{3}\right)^k \left(\frac{2}{3}\right)^{4-k} = \frac{11}{27}.$$

14 תרגול ארבעה עשר

14.1 אנטרופיה

הגדרה 14.1. יהי X משתנה מקרי בדיד המקבל את הערכים $x_1, \ldots, x_n$ בהסתברויות $p_1, \ldots, p_n$. **האנטרופיה** של X היא

$$\mathrm{H}(X) = -\mathbb{E}[\log_2 \mathrm{P}(X = k)] = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i.$$

אם מסכימים $0 \log_2 0 = 0$, נוכל לאפשר גם $p_i = 0$.

האנטרופיה של X מייצגת לנו את כמות ה"הפתעה" הממוצעת מכך שנדע את הערך של המשתנה X.

טענה 14.2.

א. לכל משתנה מקרי X, $\mathrm{H}(X) \geq 0$.

ב. אם X מקבל n ערכים, אז $\mathrm{H}(X) \leq \log_2 n$, ומתקיים שוויון אם ורק אם X אחיד על התומך שלו.

תרגיל 14.3. נניח $X \sim \mathrm{Ber}(p)$. חשבו את $\mathrm{H}(X)$. לאיזה ערך של p מתקבלת אנטרופיה מקסימלית?

פתרון. לפי ההגדרה,

$$H(X) = -p \log_2 p - (1-p) \log_2 (1-p).$$

המקסימום מתקבל כאשר $p = \frac{1}{2}$, שכן לכל p התומך של $\mathrm{Ber}(p)$ הוא $\{0,1\}$, וההתפלגות $\mathrm{Ber}(\frac{1}{2})$ היא התפלגות אחידה על התומך הזה.

תרגיל 14.4. לכל $0 \leq \alpha \leq 1$, העריכו את $\binom{n}{\alpha n}$ עבור n גדול.

פתרון. ניעזר בקירוב סטירלינג $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$:

$$\begin{aligned}
\binom{n}{\alpha n} = \frac{n!}{(\alpha n)!\,(1-\alpha n)!} &\approx \frac{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n}{\sqrt{2\pi\alpha n}\left(\frac{\alpha n}{e}\right)^{\alpha n}\sqrt{2\pi(1-\alpha)n}\left(\frac{(1-\alpha)n}{e}\right)^{(1-\alpha)n}} = \\
&= \frac{1}{\sqrt{2\pi n\alpha(1-\alpha)}} \cdot \left(\frac{1}{\alpha}\right)^{\alpha n} \cdot \left(\frac{1}{1-\alpha}\right)^{(1-\alpha)n} = \\
&= \frac{1}{\sqrt{2\pi n\alpha(1-\alpha)}} \cdot 2^{-n(\alpha \log_2 \alpha + (1-\alpha)\log_2(1-\alpha))} = \frac{2^{n \cdot \mathrm{H}(\alpha)}}{\sqrt{2\pi\alpha(1-\alpha)n}}
\end{aligned}$$

הגדרה 14.5. האנטרופיה המשותפת של זוג משתנים מקריים X, Y היא

$$\mathrm{H}(X,Y) = -\mathbb{E}[\log_2 \mathrm{P}(X=x, Y=y)] = -\sum_{x,y} \mathrm{P}(X=x, Y=y) \log_2 \mathrm{P}(X=x, Y=y).$$

האנטרופיה המותנית של X ב-Y היא

$$\mathrm{H}(X \mid Y) = -\mathbb{E}[\log_2 \mathrm{P}(X=x \mid Y=y)] = -\sum_{x,y} \mathrm{P}(X=x, Y=y) \log_2 \mathrm{P}(X=x \mid Y=y).$$

האינפורמציה ההדדית בין X ו-Y היא

$$\mathrm{I}(X;Y) = \mathrm{H}(X) - \mathrm{H}(X \mid Y).$$

טענה 14.6. לכל זוג משתנים X, Y מתקיים:

א. $\mathrm{H}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y \mid X)$.

ב. $\mathrm{H}(X \mid Y) \leq \mathrm{H}(X)$.

ג. $\mathrm{H}(X,Y) \leq \mathrm{H}(X) + \mathrm{H}(Y)$.

ד. אם X, Y בלתי־תלויים, אז $\mathrm{H}(X,Y) = \mathrm{H}(X) + \mathrm{H}(Y)$.

ה. $\mathrm{I}(X;Y) = \mathrm{H}(Y) - \mathrm{H}(Y \mid X) = \mathrm{H}(X) + \mathrm{H}(Y) - \mathrm{H}(X,Y)$.

ו. $\mathrm{I}(X,Y) \geq 0$.

תרגיל 14.7. נניח $0 \leq \alpha \leq \frac{1}{2}$. הראו כי $\sum_{i \leq \alpha n} \binom{n}{i} \leq 2^{n \cdot \mathrm{H}(\alpha)}$.

הוכחה. נסתכל על אוסף הקבוצות $\mathcal{C} = \{C \subseteq [n] \mid |C| \leq \alpha n\}$. אם X הוא משתנה מקרי אחיד על $\mathcal{C}$, אז

$$\mathrm{H}(X) = \log_2 |\mathcal{C}| = \log_2 \left(\sum_{i \leq \alpha n} \binom{n}{i} \right).$$

לכן מספיק להראות $\mathrm{H}(X) \leq \mathrm{H}(\alpha) \cdot n$. אכן, נחשוב על X בתור ווקטור $(X_1, \ldots, X_n)$, כאשר $X_i = 1_{i \in X}$. מתת־האדיטיביות של האנטרופיה, $\mathrm{H}(X) = \mathrm{H}(X_1, \ldots, X_n) \leq \sum_{i=1}^{n} \mathrm{H}(X_i)$. נחסום את $\mathrm{H}(X_i)$:

$$\mathrm{P}(i \in X) = \sum_{l=0}^{\alpha n} \mathrm{P}(i \in X \mid |X| = l) \cdot \mathrm{P}(|X| = l) = \sum_{l=0}^{\alpha n} \frac{l}{n} \cdot \mathrm{P}(|X| = l) \leq \sum_{l=0}^{\alpha n} \alpha \cdot \mathrm{P}(|X| = l) = \alpha.$$

לכן $\mathrm{H}(X_i) \leq \mathrm{H}(\alpha)$, כנדרש. □

הגדרה 14.8. יהיו P, Q שתי התפלגויות בדידות כך ש־P רציפה בהחלט ביחס ל־Q (P מייצגת את התפלגות המידע שנאסף בניסוי, Q את ההתפלגות התיאורטית). נגדיר את **סטיית Kullback-Leibler** (או **האנטרופיה היחסית**) ביניהן להיות

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_x P(x) \log_2 \frac{P(x)}{Q(x)} = \sum_x P(x) \log_2 P(x) - \sum_x P(x) \log_2 Q(x).$$

(שימו לב שזה לא סימטרי!)

תרגיל 14.9. יהיו P ו־Q התפלגויות בדידות. אז $D_{\mathrm{KL}}(P \parallel Q) \geq 0$, ויש שוויון אם ורק אם $P = Q$.

פתרון. נזכור כי לכל $a > 0$, $\log_2 a \leq \frac{1}{\log 2}(a-1)$, ויש שוויון אם ורק אם $a = 1$. לכן

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_x P(x) \log_2 \frac{P(x)}{Q(x)} = -\sum_x P(x) \log_2 \frac{Q(x)}{P(x)} \geq$$

$$\geq -\frac{1}{\log 2} \sum_x P(x) \cdot \left(\frac{Q(x)}{P(x)} - 1 \right) = \frac{1}{\log 2} \sum_x (P(x) - Q(x)) = 0$$

נקבל שיש שוויון אם ורק אם $\frac{Q(x)}{P(x)} = 1$ לכל x, כלומר $P = Q$.

תרגיל 14.10. הראו שמבין כל ההתפלגויות על הטבעיים עם תוחלת נתונה $\mu \geq 1$, ההתפלגות עם האנטרופיה המקסימלית היא גיאומטרית.

פתרון. יהי X משתנה מקרי עם תוחלת μ, ויהי $Y \sim \mathrm{Geo}(p)$ עבור $p = \frac{1}{\mu}$. לפי התרגיל הקודם, נשים לב כי

$$\begin{aligned}
\mathrm{H}(X) &= -\sum_{n=1}^{\infty} \mathrm{P}(X=n) \log_2 \mathrm{P}(X=n) \leq -\sum_{n=1}^{\infty} \mathrm{P}(X=n) \log_2 \left(p(1-p)^{n-1}\right) = \\
&= -\sum_{n=1}^{\infty} \mathrm{P}(X=n)(\log_2 p + (n-1)\log_2(1-p)) = \\
&= -\sum_{n=1}^{\infty} \mathrm{P}(X=n)(n\log_2(1-p) + \log_2 p - \log_2(1-p)) = \\
&= -\mu \log_2(1-p) - \log_2 p + \log_2(1-p) = -\log_2 p + (1-\frac{1}{p})\log_2(1-p)
\end{aligned}$$

אם היינו מתחילים מהתפלגות גיאומטרית היינו מקבלים שוויון בהתחלה, ולכן זו האנטרופיה של משתנה מקרי גיאומטרי.

כעת נראה הצצה לאחד היישומים המרכזיים של אנטרופיה – תורת הקידוד.

משפט 14.11 (משפט הקידוד של שאנון)**.** בהינתן n משתנים מקריים בלתי־תלויים ושווי התפלגות $X_1, \ldots, X_n$ עם אנטרופיה $\mathrm{H}(X)$, לכל $\varepsilon > 0$:

א. אפשר לכווץ אותם ל-$n(\mathrm{H}(X)+\varepsilon)$ ביטים כשהסיכוי לאבד מידע שואף ל-0 כאשר $n \to \infty$;

ב. לכל כיווץ שלהם ל-$n(\mathrm{H}(X)-\varepsilon)$ ביטים יש סיכוי חיובי לאבד מידע כאשר $n \to \infty$.

הוכחה. נקבע $\varepsilon > 0$, ונגדיר

$$p(x_1, \ldots, x_n) = \mathrm{P}(X_1 = x_1, \ldots, X_n = x_n).$$

נגדיר את קבוצת הסדרות ה-ε-טיפוסיות להיות

$$A_n^{\varepsilon} = \left\{ (x_1, \ldots, x_n) \mid \left| -\frac{1}{n} \log_2 p(x_1, \ldots, x_n) - \mathrm{H}(X) \right| < \varepsilon \right\}.$$

לפי החוק החלש של המספרים הגדולים, $\mathrm{P}(A_n^{\varepsilon}) \xrightarrow[n\to\infty]{} 1$. לתכונה הזו קוראים **Asymptotic Equipartition Property**, או בקיצור **AEP**. בפרט אפשר לבחור n מספיק גדול שעבורו $\mathrm{P}(A_n^{\varepsilon}) \geq 1-\varepsilon$.

מהגדרת A_n^{ε} אפשר לשים לב שלכל $(x_1, \ldots, x_n) \in A_n^{\varepsilon}$ מתקיים

$$2^{-n(\mathrm{H}(X)+\varepsilon)} \leq p(x_1, \ldots, x_n) \leq 2^{-n(\mathrm{H}(X)-\varepsilon)}.$$

בפרט, $(1-\varepsilon)2^{n(\mathrm{H}(X)-\varepsilon)} \leq |A_n^{\varepsilon}| \leq 2^{n(\mathrm{H}(X)+\varepsilon)}$. מפה נסיק את המשפט:

א. נניח שנתון $\varepsilon > 0$. אפשר להסתכל על A_n^ε, שהגודל שלה הוא לכל היותר $2^{n(\mathrm{H}(X)+\varepsilon)}$. בפרט אפשר לקודד כל איבר שם על ידי $n(\mathrm{H}(X)+\varepsilon)$ ביטים באופן חח"ע, ואת כל האיברים שאינם ב-A_n^ε אפשר לצורך העניין לקודד באפסים. אז הסיכוי לא להצליח לפענח סימבול מסוים הוא $1 - P(A_n^\varepsilon)$, וזה שואף ל-0 כאשר $n \to \infty$.

ב. כעת נניח שיש לנו קידוד שיודע לקודד כל תוצאה של X_i ל-$n(\mathrm{H}(X)-\varepsilon)$ ביטים, בפרט יש לו לכל היותר $2^{n(\mathrm{H}(X)-\varepsilon)}$ סימבולים אפשריים. נסתכל על $A_n^{\varepsilon/2}$. עבור n מספיק גדול, בקבוצה הזו יש לפחות $(1 - \varepsilon/2)2^{n(\mathrm{H}(X)-\varepsilon/2)}$ איברים. אבל כאשר $n \to \infty$ הגודל של $A_n^{\varepsilon/2}$ עולה מהר יותר מאשר הגודל של הסימבולים האפשריים, ולכן יהיו בקבוצה הזו הרבה התנגשויות.

□