תורת ההסתברות 2: (או הסתברות ותהליכים סטוכסטים)

סוכם על ידי תום חן - tomhen@gmail.com

1 בדצמבר 2014

שימו לב - יתכנו שגיאות בטקסט עידכונים יתבצעו במהלך הסמסטר
נא לדווח שגיאות ל gidi.amir@gmail.com או לחלופין שלשמור כהערות בקובץ ולשלוח לי חזרה

תקציר

סיכום ההרצאות בקורס בתורת ההסתברות 2 - סמסטר ב' 2014 מהרצאותיו של אורי גוראל-גורביץ.

תוכן עניינים

ספרים:

- Durrett: Probability Theory & Examples.
- Varadhan: Porbability Theory.
- Williams - Probability with Martingales.

1 כמה דוגמאות ראשונות לשאלות הסתברותיות:

להלן כמה דוגמאות לשאלות מעניינות בהסתברות:

דוגמה 1.1 הכד של פוליה (Polya's Urn):: בכד בשלב ראשון יש שני כדורים, שחור ולבן. בכל שלב מוציאים כדור באקראי (באופן אחיד) ומחזירים אותו ועוד כדור באותו צבע. נסמן ב־ a_n את מספר הכדורים הלבנים בצעד ה־ n. נשאלת השאלה מה ההתפלגות של סדרת המשתנים המקריים $a_1, a_2, \ldots$. מתברר ש־ U $\frac{a_n}{n+2} \xrightarrow{\text{a.s}} U$ כאשר $U \sim \text{Uniform}\,(0,1)$.

תרגיל: עבור $N \in \mathbb{N}$ קבוע מתקיים $a_N \sim U\{1, \ldots, N+1\}$.

דוגמה 1.2 הילוך השיכור (Random Walk): נגדיר $X_0 = 0$ ובאינדוקציה:

$$X_{n+1} = \begin{cases} X_n + 1 & \mathbb{P} = \frac{1}{2} \\ X_n - 1 & \mathbb{P} = \frac{1}{2} \end{cases}$$

שאלות אפשריות שעולות:

- מה ההסתברות שקיים $N > 0$ כך ש־ $X_N = 0$ ־ התשובה היא 1 ומכך בהסתברות 1 קיימים אינסוף ערכי N כך ש־ $X_N = 0$.
- מהי ההתפלגות של זמן החזרה הראשונה ובפרט מהי התוחלת שלה ־ מתברר שהתוחלת היא ∞.

ניתן להכליל את הבעיה הזו למהלך על שריג דו־ממדי כאשר נגדיר $X_0 = (0,0)$ ובאינדוקציה:

$$X_n = \begin{cases} X_n + (0,1) & \mathbb{P} = \frac{1}{4} \\ X_n + (0,-1) & \mathbb{P} = \frac{1}{4} \\ X_n + (1,0) & \mathbb{P} = \frac{1}{4} \\ X_n + (-1,0) & \mathbb{P} = \frac{1}{4} \end{cases}$$

וכן הלאה ניתן להכליל את התהליך למימדים יותר גבוהים. בפרט מתברר שעבור $n = 1, 2$ ההילוך הוא **נשנה (Recurrent)** כלומר בהסתברות 1 קיים $N > 0$ כך ש־ $X_N = X_0$. מאידך עבור $n > 3$ הנ"ל כבר לא מתרחש בהסתברות 1. אפשר להראות שמבחינה אסימפטוטית הסיכוי לחזור לנקודת ההתחלה מתנהג כמו $\frac{1}{2n}$ עבור $n > 3$.

דוגמה 1.3 פרקולציה על שריג:

נתבונן בשריג הדו־ממדי $\mathbb{Z}^2$ ובהסתברות $p \in (0,1)$ כל אחת מהצלעות בשריג נמחקת או נשארת בהסתברות $(1-p)$. נשאלת השאלה מה קורה לשריג. למשל ניתן לשאול כתלות ב־ p האם יהיה רכיב קשירות אינסופי בגרף שנשאר. ניתן להראות שעבור $p < \frac{1}{2}$ בהסתברות 1 לא נשאר רכיב קשירות אינסופי ועבור $p > \frac{1}{2}$ בהסתברות 1 נשאר רכיב קשירות אינסופי והוא יחיד. בנוסף עבור $p = \frac{1}{2}$ מתברר שאין רכיב קשירות אינסופי.

דוגמה 1.4 השרדות של שם משפחה:

נניח שמספר הצאצאים הזכרים של אדם מפולג בהתפלגות F כלשהי (למשל $U\{1, 2, \ldots, n\}$) והנ"ל נכון גם עבור הצאצאים וצאצאי הצאצאים וכן הלאה באופן בלתי תלוי. נשאלת השאלה מה הסיכוי שקיים דור שבו לא יוולדו צאצאים זכרים בכלל או לחילופין הסיכוי ששם המשפחה ישרוד לעד. אפשר להראות שאם התוחלת של ההתפלגות גדולה מ־ 1 אז בהסתברות חיובית אז שם המשפחה ישרוד לעד ואם התוחלת קטנה מ־ 1 אז השם לא שורד. כמו כן אם התוחלת היא 1 אז יש שרידות רק אם בכל שלב יש צאצא אחד בדיוק.

הערה 1.5 ניתן לייצג את הבעיה הזו בתור עץ אינסופי שמבצעים עליו פרקולציה ונשאלת השאלה האם נשאר מסלול אינסופי.

דוגמה 1.6 ערבוב של חפיסת קלפים: נניח שיש לנו חבילת כלפים מסודרת (בת 52 קלפים) ואנחנו חותכים את החפיסה במקום מקרי לפי התפלגות מסוימת ואז מערבבים באופן מסוים שגם נתון על ידי חוקיות מסוימת. נשאלת כמה ערבובים צריך לבצע כדי ששיטת הערבוב תהיה שקולה לבחירה מקרית ואחידה מ־ S_{52} (פרמוטציות של 52 איברים). בפועל באף שיטה סבירה לא נקבל אף פעם באמת פילוג אחיד מ־ S_{52} אבל אפשר לשאול עד כמה שיטת הערבוב מתקרבת לכך אחרי מספר מסוים של ערבובים וכמה צעדים נדרשים כדי להתקרב מספיק (כאשר מגדירים מרחק בין התפלגויות באופן מסוים).

דוגמה 1.7 הקלדה של רצף אותיות מסוים: קופים מקלידים אותיות ב־ ABC באופן אחיד ואקראי וכבר ידוע לנו שכל רצף של אותיות יוקלד בשלב מסוים בהסתברות 1. נשאל לדוגמה כמה זמן צריך לחכות (בתוחלת) עד שנראה את הרצף ABRAKEDABRA. לכל רצף של 11 אותיות מתוך 26 אותיות יש הסתברות של $\frac{1}{26^{11}}$ לצאת אבל מאחר ואנחנו מקלידים ברצף (בניגוד לבחירת רצף כל פעם מחדש ללא המשכיות) וגם מאחר ויש חפיפה בין הסוף וההתחלה של הרצף יקח דווקא יותר זמן עד שנראה את הרצף הזה. באופן כללי לרצפים שיש בהם פחות חפיפה יקח פחות זמן להופיע מאשר לרצפים שיש בהם יותר חפיפה.

דוגמה 1.8 חידה: נתונה חפיסת קלפים מעורבבת ושולפים קלפים זה אחר זה מראש הערימה ומסתכלים בהם ובכל שלב ניתן להחליט שעוצרים ולא שולפים יותר. כאשר עוצרים מקבלים ± 1 לפי צבע הקלף הבא. נשאלת השאלה מה האסטרטגיה האופטימלית כדי למקסם את הרווח.

1.1 סילבוס הקורס:

1. תהליכים סטוכסטיים (סדרות של משתנים מקריים) והתכנסויות.
2. פונקציות אופיניות ומשפט הגבול המרכזי.
3. מרטינגלים (שם קוד לסדרה של הימורים הוגנים).
4. שרשראות מרקוב.
5. סטיות גדולות.
6. ריכוז מידה (Concentration of Measure).
7. תנועה בראונית.

2 חזרה על תורת המידה:

הגדרה 2.1 אלגברה על קבוצה:
בהנתן קבוצה Ω משפחת קבוצות $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ תקרא אלגברה אם היא מקיימת:

1. $\emptyset \in \mathcal{F}$.
2. אם $A \in \mathcal{F}$ אז $A^c \in \mathcal{F}$.
3. אם $A, B \in \mathcal{F}$ אז $(A \cup B) \in \mathcal{F}$.

לפעמים דורשים סגירות לחיתוך במקום לאיחוד והנ"ל שקול שכן יש סגירות למשלים. כמו כן באינדוקציה מקבלים שאלגברה סגורה לאיחודים וחיתוכים סופיים.

דוגמה 2.2 דוגמאות:

1. אוסף האיחודים של קטעים חצי פתוחים $(a, b] \subseteq [0, 1]$ הוא אלגברה.
2. בהנתן $\Omega = \{0, 1\}^{\mathbb{N}}$ נגדיר קבוצת צילינדר בסיסית ע"י:

$$C_{b_1, \ldots, b_k} = \left\{ (a_n)_{n=1}^{\infty} \in \{0, 1\}^{\mathbb{N}} \mid a_i = b_i \, \forall 1 \leq i \leq k \right\}$$

כלומר קבוצת כל הסדרות ש־ k המקומות הראשונים שלהן הם $b_1, \ldots, b_k$ עבור $b_1, \ldots, b_k \in \{0, 1\}$ נתונים. אוסף כל האיחודים הסופיים של כל קבוצות הצילנידר הוא אלגברה על $\{0, 1\}^{\mathbb{N}}$.

3. אוסף כל תת־הקבוצות הסופיות\קו־סופיות של Ω הוא אלגברה.

הגדרה 2.3 סיגמה אלגברה:
בהנתן קבוצה Ω משפחת קבוצות $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ תקרא סיגמה־אלגברה אם היא אלגברה והיא סגורה גם לאיחודים בני־מנייה.

הגדרה 2.4 סיגמה אלגברה נוצרת:
בהנתן $S \subseteq \Omega$ ה־ σ־אלגברה הנוצרת על ידי S שתסומן $\sigma(\mathcal{S})$ היא ה־ σ־אלגברה המינימלית שמכילה את S ובאופן שקול חיתוך כל ה־ σ־אלגבראות המכילות את S.

הגדרה 2.5 סיגמה אלגברת בורל:
בהנתן מרחב טופולוגי $(X, \mathcal{T})$ נגדיר את σ־אלגברת בורל על X שתסומן $\mathcal{B}(X)$ בתור $\mathcal{B}(X) := \sigma(\mathcal{T})$. כלומר ה־ σ־אלגברה שנוצרת על ידי כל הקבוצות הפתוחות (באופן שקול אפשר להגדיר גם באמצעות סגורות).

הגדרה 2.6 מרחב מדיד:
מרחב מדיד הוא זוג $(\Omega, \mathcal{F})$ כאשר Ω היא קבוצה ו־ $\mathcal{F}$ היא σ־אלגברה על Ω.

דוגמה 2.7 נגדיר את $S \subseteq \mathcal{P}(\mathbb{N})$ להיות אוסף הקבוצות $A \subseteq \mathbb{N}$ שיש להן צפיפות אסימפטוטית, כלומר קיים הגבול:

$$\lim_{n \to \infty} \frac{|A \cap \{1, \ldots., n\}|}{n}$$

בפרט S איננה σ־אלגברה שכן ניתן לבחור קבוצה בת־מנייה שאין לה צפיפות אסימפטוטית והיא תהיה איחוד בן־מנייה של יחידונים שכמובן יש להם צפיפות אסימפטוטית ולכן S לא סגורה לאיחודים בני־מנייה בבירור. אפשר להראות גם ש־ $\mathcal{S}$ איננה אלגברה ולשם כך מספיק למצוא שתי קבוצות בעלות צפיפות אסימפטוטית שלאיחוד או לחיתוך שלהם אין צפיפות אסימפטוטית.

פתרון: נקח את A להיות כל הזוגיים ואת B להיות הקבוצה שמכילה את כל הזוגיים עד מיליון, כל האי־זוגיים בין מיליון ומיליארד, כל הזוגיים בין מיליארד למיליארד ומאה מיליון וכן הלאה. לשתי הקבוצות הללו תהיה צפיפות אסימפטוטית אולם לחיתוך יהיו פלקטואציות בצפיפות שתשתנה בין 0 לבערך $\frac{1}{2}$ ולכן לא תהיה צפיפות אסימפטוטית.

הגדרה 2.8 מידה על מרחב מדיד:
בהנתן מרחב מדיד $(\Omega,\mathcal{F})$ פונקציה $\mu:\mathcal{F}\to[0,\infty]$ תקרא מידה אם היא מקיימת את התכונות הבאות:

1. $\mu(\emptyset)=0$.
2. $\mu(A)\geq 0$ לכל $A\in\mathcal{F}$.
3. **סיגמה אדיטיביות:** בהנתן קבוצות זרות $A_1,A_2,..\in\mathcal{F}$ מתקיים $\mu\left(\dot{\bigcup}_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}\mu(A_n)$.

בפרט אם $\mu(\Omega)<\infty$ נאמר כי μ היא מידה סופית ואם $\mu(\Omega)=1$ אז זוהי מידת הסתברות.

הגדרה 2.9 מרחב מידה\הסתברות:
מרחב מידה הוא שלשה $(\Omega,\mathcal{F},\mu)$ כאשר $(\Omega,\mathcal{F})$ הוא מרחב מדיד ו־ μ היא מידה על $(\Omega,\mathcal{F})$.

הערה 2.10 בפרט אם μ היא מידת הסתברות נאמר כי המרחב הוא מרחב הסתברות.

תזכורת 2.11 כמה תכונות של מידה:
יהא $(\Omega,\mathcal{F},\mu)$ מרחב מידה, מתקיימים הדברים הבאים:

1. **מונוטוניות:** עבור $A,B\in\mathcal{F}$ כך ש־ $B\subseteq A$ מתקיים $\mu(A)\leq\mu(B)$.
2. **תת־אדיטיביות:** בהנתן $A_1,A_2,...\in\mathcal{F}$ מתקיים $\mu\left(\bigcup_{n=1}^{\infty}A_n\right)\leq\sum_{n=1}^{\infty}\mu(A_n)$.
3. **רציפות מלמטה:** בהנתן סדרה עולה $A_n\uparrow\in\mathcal{F}$ מתקיים:
$$\lim_{n\to\infty}\mu(A_n)=\mu\left(\bigcup_{n=1}^{\infty}A_n\right)$$
4. **רציפות מלמעלה:** בהנתן סדרה יורדת $A_n\downarrow\in\mathcal{F}$ כך שקיים N שעבורו $\mu(A_N)<\infty$ מתקיים:
$$\lim_{n\to\infty}\mu(A_n)=\mu\left(\bigcap_{n=1}^{\infty}A_n\right)$$

הערה 2.12 פונקציית קבוצות $\mu:\mathcal{F}\to[0,\infty]$ שהינה אדיטיבית סופית ומקיימת $\mu(\emptyset)=0$ היא מידה אמ"מ היא מקיימת את תכונה 3 או 4.

הגדרה 2.13 מידה על אלגברה (Pre-Measure):
בהנתן אלגברה $\mathcal{A}$ פונקציית קבוצות $\mu:\mathcal{A}\to[0,\infty]$ תקרא מידה על האלגברה אם $\mu(\emptyset)=0$ וגם לכל $A_1,A_2,...\in\mathcal{A}$ זרות כך ש־ $\dot{\bigcup}_{n=1}^{\infty}A_n\in\mathcal{A}$ מתקיים $\mu\left(\dot{\bigcup}_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}\mu(A_n)$.

משפט 2.14 משפט ההרחבה של *Caratheodory*:

תהא $\mathcal{F}_0$ היא אלגברה ותהא μ_0 מידה עליה. אזי קיימת מידה μ על $\mathcal{F}:=\sigma(\mathcal{F}_0)$ כך ש־ $\mu|_{\mathcal{F}_0}=\mu_0$. הנ"ל מכונה הרחבה של המידה μ_0 מהאלגברה $\mathcal{F}_0$ ל־ σ־אלגברה הנוצרת ממנה. בפרט אם μ_0 היא מידה סופית (למעשה גם σ־סופית) אז ההרחבה הנ"ל יחידה.

הערה 2.15 באמצעות המשפט הנ"ל ניתן להרחיב מידות שמוגדרות על אלגברה לסיגמה אלגברה הנוצרת. לדוגמה:

1. הרחבת מידת האורך $\mu_0([a,b])=b-a$ מהאלגברה שנוצרת על ידי תת־קטעים ל־ σ־אלגברה שהם יוצרים שהיא $\mathcal{B}(\mathbb{R})$.
2. הרחבת מידת הטלת מטבעות מהאלגברה הנוצרת על ידי צילינדרים ל־ σ־אלגברה הנוצרת.

משפט 2.16 מסקנה ממשפט $\pi-\lambda$ (משפט דינקין):

יהא $(\Omega,\mathcal{F})$ מרחב מדיד כך ש־ $\mathcal{F}=\sigma\left(\mathcal{F}_0\right)$ ויהיו μ_1,μ_2 מידות סופיות על $(\Omega,\mathcal{F})$. אזי אם $\mathcal{F}_0$ סגורה לחיתוכים וגם $\mu_1\left(E\right)=\mu_2\left(E\right)$ לכל $E\in\mathcal{F}_0$ אז $\mu_1\left(E\right)=\mu_2\left(E\right)$ לכל $E\in\mathcal{F}$.

משפט 2.17 משפט ־ תנאי למדידות (תרגיל):

יהא $(\Omega,\mathcal{F},\mu)$ מרחב מידה כך ש־ $\mathcal{F}=\sigma\left(\mathcal{F}_0\right)$ כאשר $\mathcal{F}_0$ היא אלגברה. אזי לכל $A\in\mathcal{F}$ ולכל $\varepsilon>0$ קיימת $A_0\in\mathcal{F}_0$ כך ש:

$$\mu\left(A\triangle A_0\right)=\mu\left(\left(A\backslash A_0\right)\cup\left(A_0\backslash A\right)\right)<\varepsilon$$

המשמעות היא שבמובן מסוים ניתן לקרב קבוצות ב־ σ־אלגברה ע"י קבוצות באלגברה.

3 הסתברות בסיסית (יחסית):

3.1 כמה טענות בסיסיות בהסתברות:

הגדרה 3.1 ה ־ Limsup/Liminf של סדרת מאורעות:

יהא $(\Omega,\mathcal{F})$ מרחב מדיד, בהנתן סדרה $E_1,E_2,...\in\mathcal{F}$ נגדיר:

$$\limsup_{n\to\infty}E_n=\bigcap_{m=1}^{\infty}\bigcup_{n>m}E_n$$

$$\liminf_{n\to\infty}E_n=\bigcup_{m=1}^{\infty}\bigcap_{n>m}E_n$$

בפרט ה־ Liminf של הסדרה הוא אוסף ה־ $\omega\in\Omega$ כך ש־ $\omega\in E_n$ פרט למספר סופי של ערכי n וה־ Limsup הוא אוסף ה־ $\omega\in\Omega$ כך ש־ $\omega\in E_n$ עבור אינסוף ערכי n. בפרט אם $\mathbb{P}\left(\limsup\limits_{n\to\infty}E_n\right)=\infty$ אז אינסוף מהמאורעות E_n קורים כמעט תמיד ואם $\mathbb{P}\left(\liminf\limits_{n\to\infty}E_n\right)=1$ אז כל המאורעות E_n פרט למספר סופי קורים כמעט־תמיד.

משפט 3.2 הלמה של *Fatou*:

יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ויהיו $E_1,E_2,...\in\mathcal{F}$, אזי:

$$\mathbb{P}\left(\liminf_{n\to\infty}E_n\right)\leq\liminf_{n\to\infty}\mathbb{P}\left(E_n\right)$$

$$\mathbb{P}\left(\limsup_{n\to\infty}E_n\right)\geq\limsup_{n\to\infty}\mathbb{P}\left(E_n\right)$$

הוכחה:

1. נגדיר $A_m:=\bigcap_{n>m}E_n$ ונקבל כי $A_m\uparrow\liminf\limits_{n\to\infty}E_n$, לכן מרציפות של מידה נקבל כי:

$$\liminf_{n\to\infty}\mathbb{P}\left(E_n\right)\geq\mathbb{P}\left(A_m\right)\overset{m\to\infty}{\longrightarrow}\mathbb{P}\left(\liminf_{n\to\infty}E_n\right)$$

כאשר א"ש המסומן נובע מכך ש־ $A_m\subseteq E_n$ לכל $n>m$ ולכן $\mathbb{P}\left(A_m\right)\leq\mathbb{P}\left(E_n\right)$ לכל $n>m$.

2. נשים לב כי $\left(\limsup\limits_{n\to\infty}E_n\right)^c=\liminf\limits_{n\to\infty}E_n^c$ ומכך על סמך הטענה הראשונה נקבל כי:

$$1-\mathbb{P}\left(\limsup_{n\to\infty}E_n\right)=\mathbb{P}\left(\left(\liminf_{n\to\infty}E_n\right)^c\right)\leq\liminf_{n\to\infty}\left(1-\mathbb{P}\left(E_n^c\right)\right)=1-\liminf_{n\to\infty}\mathbb{P}\left(E_n^c\right)=1-\limsup_{n\to\infty}\mathbb{P}\left(E_n\right)$$

מעבירים אגפים ומקבלים את הנדרש.

■

למה 3.3 הלמה הראשונה של בורל קנטלי:

יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ותהא $E_1,E_2,...\in\mathcal{F}$ כך ש־ $\sum_{n=1}^{\infty}\mathbb{P}(E_n)<\infty$ אזי $\mathbb{P}\left(\limsup_{n\to\infty}E_n\right)=0$.

הוכחה: נשים לב כי לכל $m\in\mathbb{N}$ מתקיים $\limsup_{n\to\infty}E_n\subseteq\bigcup_{n>m}E_n$, מכך נקבל כי:

$$\mathbb{P}\left(\limsup_{n\to\infty}E_n\right)\leq\mathbb{P}\left(\bigcup_{n>m}E_n\right)\leq\sum_{n>m}\mathbb{P}(E_n)\overset{m\to\infty}{\longrightarrow}0$$

■

הערה 3.4 נשים לב כי הטענה הזו איננה אמ"מ, כלומר אם $\sum_{n=1}^{\infty}\mathbb{P}(E_n)=\infty$ לא בהכרח נובע ש־ $\mathbb{P}\left(\limsup_{n\to\infty}E_n\right)>0$. למשל עבור $E_n=\left(0,\frac{1}{n}\right]$ ו־ $\Omega=(0,1]$ עם מידת הסתברות אחידה נקבל כי $\limsup_{n\to\infty}E_n=\emptyset$ אולם $\sum_{n=1}^{\infty}\mathbb{P}(E_n)=\sum_{n=1}^{\infty}\frac{1}{n}=\infty$.

הגדרה 3.5 אי־תלות של סדרת σ־אלגבראות ושל סדרת מאורעות:
יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ותהא תהא $\{\mathcal{F}_i\}_{i=1}^{\infty}$ סדרה של תת־σ־אלגבראות $\mathcal{F}_i\subseteq\mathcal{F}$:

1. נאמר ש־ $\{\mathcal{F}_i\}_{i=1}^{\infty}$ בלתי תלויה אם לכל $n\in\mathbb{N}$ ולכל בחירה של $A_i\in\mathcal{F}_i$ ($i=1,...,n$) מתקיים $\mathbb{P}\left(\bigcap_{i=1}^{n}A_i\right)=\prod_{i=1}^{n}\mathbb{P}(A_i)$.
2. נאמר שסדרת מאורעות $A_1,A_2,...\in\mathcal{F}$ היא בלתי תלויה אם הסדרה $\sigma(\{A_i\})=\{\emptyset,\Omega,A_i,A_i^c\}$ בלתי־תלויה.

הערה 3.6 בשני המקרים הסדרות יכולות להיות סופיות בשינוי מתאים של ההגדרה.

טענה 3.7 הגדרה שקולה לאי־תלות של אוסף וסדרת מאורעות (תוספת):

יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות:

1. מאורעות $A_1,...,A_N\in\mathcal{F}$ הם ב"ת אם לכל $1\leq k_1<...<k_m\leq N$ מתקיים $\mathbb{P}\left(\bigcap_{j=1}^{m}A_{k_j}\right)=\prod_{j=1}^{m}\mathbb{P}\left(A_{k_j}\right)$.
2. מאורעות $A_1,A_2,...\in\mathcal{F}$ הם ב"ת אם לכל $N\in\mathbb{N}$ המאורעות $A_1,...,A_N$ בלתי תלויים.

הגדרה 3.8 אי־תלות בזוגות של סדרת σ־אלגבראות וסדרת מאורעות:
יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות נאמר שסדרה $\{\mathcal{F}_i\}_{i=1}^{\infty}$ של תת־σ־אלגבראות $\mathcal{F}_i\subseteq\mathcal{F}$ בלתי־תלויה בזוגות אם לכל $i\neq j$ ו־ $A\in\mathcal{F}_i$ ו־ $B\in\mathcal{F}_j$ מתקיים $\mathbb{P}(A\cap B)=\mathbb{P}(A)\cdot\mathbb{P}(B)$. באופן מתאים נאמר שסדרת מאורעות $E_1,E_2,...\in\mathcal{F}$ בלתי תלויים בזוגות אם אם הסדרה $\sigma(\{E_i\})=\{\emptyset,\Omega,E_i,E_i^c\}$ בלתי־תלויה בזוגות.

הערה 3.9 אי־תלות כמובן תמיד גוררת אי־תלות בזוגות אבל ההפך לא נכון.

למה 3.10 תנאי לאי־תלות של שתי תת־σ־אלגבראות במרחב הסתברות:

יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ויהיו $S_1,S_2\subseteq\mathcal{F}$ תת־קבוצות סגורות לחיתוכים כך שלכל $A\in S_1$ ו־ $B\in S_1$ מתקיים

$$\mathbb{P}(A\cap B)=\mathbb{P}(A)\,\mathbb{P}(B)$$

אזי $\sigma(S_1)$ ו־ $\sigma(S_2)$ הן ב"ת.

הערה 3.11 באופן כללי יותר אם $S_1, S_2, ..., S_n \subseteq \mathcal{F}$ סגורות לחיתוכים ומתקיים:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} \mathbb{P}\left(A_i\right)$$

לכל $A_i \in \mathcal{S}_i$ אז $\sigma\left(S_1\right), ..., \sigma\left(S_n\right)$ ב"ת.

הוכחה: נקבע $A \in S_1$ ונגדיר $\mathbb{P}_1\left(B\right) = \mathbb{P}\left(B\right)\mathbb{P}\left(A\right)$ ו־ $\mathbb{P}_2\left(B\right) = \mathbb{P}\left(B \cap A\right)$ עבור כל $B \in \sigma\left(S_2\right)$. נשים לב כי $\mathbb{P}_1, \mathbb{P}_2$ הן מידות סופיות על $\left(\Omega, \sigma\left(S_2\right)\right)$ ומתקיים $\mathbb{P}_2\left(B\right) = \mathbb{P}_1\left(B\right)$ לכל $B \in S_2$. לכן ממשפט דינקין מאחר ש־ S_2 סגורה לחיתוכים נובע שהשוויון מתקיים לכל $B \in \sigma\left(S_2\right)$. כעת נסמן $T_1 = \sigma\left(S_1\right)$ ו־ $T_2 = \sigma\left(S_2\right)$, גם אלו הן קבוצות שסגורות לחיתוכים, נקבע $B \in T_2$ ונגדיר $\mathbb{P}_1\left(A\right) = \mathbb{P}\left(B\right)\mathbb{P}\left(A\right)$ ו־ $\mathbb{P}_2\left(A\right) = \mathbb{P}\left(B \cap A\right)$ עבור כל $A \in \sigma\left(S_1\right)$. אלו הן מידות סופיות על $\left(\Omega, \sigma\left(S_1\right)\right)$ ועל סמך מה שהוכחנו מתקיים $\mathbb{P}_2\left(A\right) = \mathbb{P}_1\left(A\right)$ לכל $A \in S_1$. לכן שוב ממשפט דינקין מתקיים השוויון הנ"ל לכל $A \in \sigma\left(S_1\right)$. סה"כ ניתן לראות שלכל $B \in \sigma\left(S_2\right)$ ולכל $A \in \sigma\left(S_1\right)$ מתקיים $\mathbb{P}\left(A \cap B\right) = \mathbb{P}\left(A\right)\mathbb{P}\left(B\right)$ ומכך ישנה אי־תלות כנדרש. ∎

למה 3.12 הלמה השנייה של בורל־קנטלי:

יהא $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ מרחב הסתברות ותהא $E_1, E_2, ... \in \mathcal{F}$ סדרת מאורעות ב"ת כך ש־ $\sum_{n=1}^{\infty} \mathbb{P}\left(E_n\right) = \infty$ אזי $\mathbb{P}\left(\limsup_{n\to\infty} E_n\right) = 1$.

הוכחה: מספיק שנראה כי $\mathbb{P}\left(\liminf_{n\to\infty} E_n^c\right) = 0$. נשים לב כי:

$$\mathbb{P}\left(\bigcap_{n>m} E_n^c\right) = \prod_{n>m}\left(1 - \mathbb{P}\left(E_n\right)\right) \leq \prod_{n>m} e^{-\mathbb{P}\left(E_n\right)} = e^{-\sum\limits_{n>m}\mathbb{P}\left(E_n\right)} = 0$$

כאשר השתמשנו בכך ש־ $1 - \mathbb{P}\left(E_n\right) \leq e^{-\mathbb{P}\left(E_n\right)}$ וגם $\sum_{n>m} \mathbb{P}\left(E_n\right) \overset{n\to\infty}{\longrightarrow} \infty$. בפרט מכך נסיק כי:

$$\mathbb{P}\left(\liminf_{n\to\infty} E_n^c\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{n>m} E_n^c\right) = 0$$

∎

הגדרה 3.13 σ־אלגברת זנב של סדרת σ־אלגבראות:
יהא $\left(\Omega, \mathcal{F}\right)$ מרחב מדיד ותהא $\left\{\mathcal{F}_n\right\}_{n=1}^{\infty}$ סדרת σ־אלגבראות חלקיות של $\mathcal{F}$, נגדיר:

$$\mathcal{T}_n = \sigma\left(\left\{\mathcal{F}_n, \mathcal{F}_{n+1}, ...\right\}\right) = \sigma\left(\bigcup_{n>m} \mathcal{F}_n\right)$$

כמו כן נגדיר $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$, הנ"ל מכונה σ־אלגברת הזנב של הסדרה $\left\{\mathcal{F}_n\right\}_{n=1}^{\infty}$ (נשים לב כי $\left\{\emptyset, \Omega\right\} \subseteq \mathcal{T}$ ולכן הנ"ל לא ריקה).

הערה 3.14 כל מאורע של $\mathcal{T}$ מכונה גם "מאורע זנב".

תרגיל: אם נקח סדרת מאורעות $\left\{E_n\right\}_{n=1}^{\infty}$ כך ש־ $E_n \in \mathcal{F}_n$ לכל n אז $\liminf_{n\to\infty} E_n \in \mathcal{T}$ וגם $\limsup_{n\to\infty} E_n \in \mathcal{T}$.

משפט 3.15 משפט 0־1 של קולמוגורוב:

יהא $\left(\Omega, \mathcal{F}, \mathbb{P}\right)$ מרחב הסתברות ותהא $\left\{\mathcal{F}_n\right\}_{n=1}^{\infty} \subseteq \mathcal{P}\left(\mathcal{F}\right)$ סדרה ב"ת של σ־אלגבראות אזי לכל $A \in \mathcal{T}$ מתקיים $\mathbb{P}\left(A\right) \in \left\{0, 1\right\}$.

דוגמה 3.16 דוגמה קנונית שבד"כ משתמשים בה כדי להתבונן בהתנהגות של סדרה ב"ת של σ־אלגבראות ומאורעות זנב בה היא הדוגמה שבה $\Omega = \left\{0, 1\right\}^{\mathbb{N}}$ ו־ $\mathcal{F}_n = \sigma\left(\left\{\omega : \omega_n = 0\right\}\right)$.

הוכחה: (חסרים פרטים) ראשית נראה ש־ $\mathcal{T},\mathcal{F}_1,\mathcal{F}_2,...$ ב"ת. בהנתן $A\in\mathcal{T}$ לכל $n\in\mathbb{N}$ מתקיים $A\in\mathcal{T}_n$ ולכן בהנתן מאורעות $A_i\in\mathcal{F}_i$ נקבל כי:

$$\mathbb{P}\left(A\cap\bigcap_{i=1}^{n-1}A_i\right)=\mathbb{P}\left(A\right)\prod_{i=1}^{n-1}\mathbb{P}\left(A_i\right)$$

כאשר השוויון הנ"ל נובע מכך ש־ $\mathcal{T}_n,\mathcal{F}_1,...,\mathcal{F}_{n-1}$ הן ב"ת על סמך הגרסה הכללית יותר של למה 3.10. כעת שהראינו ש־ $\mathcal{T},\mathcal{F}_1,\mathcal{F}_2,...$ ב"ת נסיק ש־ $\mathcal{T}$ ב"ת ב־ $\sigma\left(\{\mathcal{F}_1,\mathcal{F}_2,...\}\right)$ כאשר הנ"ל נובע שוב מלמה 3.10 כאשר משתמשים באי־תלות, בכך ש־ $\mathcal{T}$ ו־ $\bigcup_{n=1}^{\infty}\sigma\left(\bigcup_{i=1}^{n}\mathcal{F}_i\right)$ (זוהי איננה σ־אלגברה אך כן אלגברה) סגורות לחיתוכים ובכך שמתקיים:

$$\sigma\left(\{\mathcal{F}_1,\mathcal{F}_2,...\}\right)=\sigma\left(\bigcup_{n=1}^{\infty}\sigma\left(\bigcup_{i=1}^{n}\mathcal{F}_i\right)\right)$$

לסיום מאחר ש־ $\mathcal{T}\subseteq\sigma\left(\{\mathcal{F}_1,\mathcal{F}_2,...\}\right)$ ומאי־התלות הנ"ל נקבל שלכל $A\in\mathcal{T}$ מתקיים $\mathbb{P}\left(A\right)=\mathbb{P}\left(A\cap A\right)=\mathbb{P}\left(A\right)\cdot\mathbb{P}\left(A\right)$ ולכן $\mathbb{P}\left(A\right)\in\{0,1\}$. ■

תרגיל: האם הטענה נכונה גם במקרה שהסדרה ב"ת רק בזוגות?

3.2 משתנים מקריים:

הגדרה 3.17 פונקציה מדידה:
יהיו $(X,\mathcal{F})$ ו־ $(Y,\mathcal{G})$ מרחבים מדידים. העתקה $g:X\to Y$ תקרא $(\mathcal{F},\mathcal{G})$־מדידה אם $g^{-1}\left(B\right)\in\mathcal{F}$ לכל $B\in\mathcal{G}$.

הערה 3.18 בד"כ ה־ σ־אלגברה בטווח פחות רלוונטית ולכן נרשום באופן מקוצר שההעתקה היא $\mathcal{F}$־מדידה.

הערה 3.19 העתקה מדידה $f:\left(\mathbb{R}^n,\mathcal{B}^n\right)\to\left(\mathbb{R}^m,\mathcal{B}^m\right)$ תקרא מדידה־בורל. בפרט כל רציפה היא מיידית מדידה־בורל.

הגדרה 3.20 משתנה מקרי:
יהיו $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות, משתנה מקרי (מ"מ) הוא פונקציה $\mathcal{F}$־מדידה מ־ $(\Omega,\mathcal{F})$ למרחב מדיד $(Y,\mathcal{G})$ כלשהו. משתנה מקרי ממשי הוא פונקציה $\mathcal{F}$־מדידה מ־ $(\Omega,\mathcal{F})$ ל־ $(\mathbb{R},\mathcal{B})$.

הערה 3.21 כאשר נרשום ש־ X הוא מ"מ הכוונה תמיד תהיה שהתחום הוא מרחב הסתברות והטווח הוא $(\mathbb{R},\mathcal{B})$.

תזכורת 3.22 תנאי מספיק למדידות (ראינו במידה):
יהיו $(X,\mathcal{F})$ ו־ $(Y,\mathcal{G})$ מרחבים מדידים ותהא $g:X\to Y$. תהא $S\subseteq\mathcal{G}$ כך ש־ $g^{-1}\left[A\right]\in\mathcal{F}$ לכל $A\in S$ אז g היא $(\mathcal{F},\sigma\left(S\right))$־מדידה. בפרט אם $\sigma\left(S\right)=\mathcal{G}$ אז g היא $\mathcal{F}$־מדידה.

מסקנה 3.23 תנאי מספיק למדידות של מ"מ ממשי:

יהיו $(\Omega,\mathcal{F})$ מרחב מדיד ותהא $f:(\Omega,\mathcal{F})\to(\mathbb{R},\mathcal{B})$. אזי f היא $\mathcal{F}$־מדידה אמ"מ $f^{-1}\left[(-\infty,t]\right]\in\mathcal{F}$ לכל $t\in\mathbb{R}$.

הוכחה: מיידי מהטענה הקודמת ומכך ש־ $\{(-\infty,t]\mid t\in\mathbb{R}\}$ יוצרת את $\mathcal{B}$. ■

הערה 3.24 מתקיים $\mathcal{B}^n=\sigma\left(\mathcal{B}\times\cdots\times\mathcal{B}\right)$ ובפרט:

$$\mathcal{B}^n=\sigma\left(\{(-\infty,t_1]\times\cdots\times(-\infty,t_n]\mid t_1,...,t_n\in\mathbb{R}\}\right)$$

כלומר σ־אלגברת בורל ב־ $\mathbb{R}^n$ היא ה־ σ־אלגברה שנוצרת על ידי מכפלות של קבוצות בורל ובפרט של קרנות.

מסקנה 3.25 וקטור מקרי הוא משתנה מקרי מדיד $\Omega\to\mathbb{R}^2$:

יהא $(\Omega,\mathcal{F})$ מרחב מדיד ויהיו X,Y משתנים מקריים. אזי $(X,Y):\Omega\to\mathbb{R}^2$ הוא וקטור מקרי למרחב $\left(\mathbb{R}^2,\mathcal{B}^2\right)$.

הערה 3.26 כמובן שהנ"ל נכון גם עבור וקטור מקרי $\left(X_1,...,X_n\right):\Omega\to\mathbb{R}^n$ והנ"ל מדיד כפונקציה $(\Omega,\mathcal{F})\to(\mathbb{R}^n,\mathcal{B}^n)$.

הוכחה: מספיק להראות ש- $\left(X,Y\right)^{-1}\left[(-\infty,t]\times(-\infty,s]\right]$ היא $\mathcal{F}$-מדידה לכל $t,s\in\mathbb{R}$ וזה נובע מיידית מכך ש:

$$\left(X,Y\right)^{-1}\left[(-\infty,t]\times(-\infty,s]\right]=X^{-1}\left[(-\infty,t]\right]\cap Y^{-1}\left[(-\infty,s]\right]$$

■

טענה 3.27 הרכבת פונקציות מדידות היא מדידה:

יהיו $(X,\mathcal{F}),(Y,\mathcal{G}),(Z,\mathcal{H})$ מרחבים מדידים. תהא $f:X\to Y$ העתקה $\mathcal{F}$-מדידה ותהא $g:Y\to Z$ העתקה $\mathcal{G}$-מדידה. אזי $g\circ f:X\to Z$ היא $\mathcal{F}$-מדידה.

הוכחה: תהא $A\in\mathcal{H}$, מההנחה כי g היא $\mathcal{G}$-מדידה נקבל כי $g^{-1}(A)\in\mathcal{G}$. מההנחה כי f היא $\mathcal{F}$-מדידה נובע כי:

$$\left(g\circ f\right)^{-1}(A)=f^{-1}\left(g^{-1}(A)\right)\in\mathcal{F}$$

מכך ניתן לראות כי $g\circ f$ היא $\mathcal{F}$-מדידה כנדרש. ■

מסקנה 3.28 מסקנה מיידית מהטענה הקודמת:

יהא $(\Omega,\mathcal{F})$ מרחב מדיד ויהיו X,Y משתנים מקריים אזי $X+Y$ הוא משתנה מקרי (כלומר הוא $\mathcal{F}$-מדיד).

הוכחה: מיידי מכך שהעתקת החיבור $+:\mathbb{R}^2\to\mathbb{R}$ היא רציפה ולכן מדידה-בורל. ■

טענה 3.29 אריתמטיקה כללית יותר של פונקציות מדידות (ללא הוכחה):

יהיו $f,g:(X,\mathcal{F})\to(\mathbb{R},\mathcal{B})$ פונקציות $\mathcal{F}$-מדידות ויהיו $\alpha,\beta\in\mathbb{R}$ אזי:

1. הקבוצה $\{x\in X\mid f(x)>g(x)\}$ היא $\mathcal{F}$-מדידה (כנ"ל עבור $\leq,\geq,<,=,\neq$).
2. הפונקציות $f+\alpha$, αf ו- $\alpha f+\beta g$ הן $\mathcal{F}$-מדידות.
3. הפונקציה $f\cdot g$ היא $\mathcal{F}$-מדידה ואם $g\neq 0$ אז $\frac{f}{g}$ היא $\mathcal{F}$-מדידה.
4. הפונקציות $\max\{f,g\}$ ו- $\min\{f,g\}$ הן $\mathcal{F}$-מדידות.
5. הפונקציה f^2 היא $\mathcal{F}$-מדידה.
6. הפונקציות f^+ ו- f^- הן $\mathcal{F}$-מדידות.
7. הפונקציה $|f|$ היא $\mathcal{F}$-מדידה.

3.3 סדרות של משתנים מקריים:

את חלק מההגדרות בקטע הבא אני הוספתי מאחר והן רלוונטיות לדיון שהתקיים בהרצאה.

טענה 3.30 כמה תכונות בסיסיות של סדרות משתנים מקריים:

יהא $(X,\mathcal{F})$ מרחב מדיד ותהא $f_n:(X,\mathcal{F})\to(\mathbb{R},\mathcal{B})$ סדרת פונקציות $\mathcal{F}$-מדידות אזי הפונקציות:

$$\sup_{n\in\mathbb{N}}f_n(x),\quad\inf_{n\in\mathbb{N}}f_n(x),\quad\limsup_{n\to\infty}f_n(x),\quad\liminf_{n\to\infty}f_n(x)$$

הן פונקציות $\mathcal{F}$-מדידות (יש לשים לב כי הפונקציות הללו יכולות לקבל גם את הערכים $\pm\infty$).

הוכחה: נשים לב כי לכל $a\in\mathbb{R}$ מתקיים:

$$\left(\sup_{n\in\mathbb{N}}f_n\right)^{-1}[(-\infty,a]]=\left\{x\in X\mid \sup_{n\in\mathbb{N}}f_n(x)<a\right\}=\bigcap_{n=1}^{\infty}\{x\in X\mid f_n(x)<a\}=\bigcap_{n=1}^{\infty}f_n^{-1}[(-\infty,a)]\in\mathcal{F}$$

$$\left(\inf_{n\in\mathbb{N}}f_n\right)^{-1}[(-\infty,a]]=\left\{x\in X\mid \inf_{n\in\mathbb{N}}f_n(x)<a\right\}=\bigcup_{n=1}^{\infty}\{x\in X\mid f_n(x)<a\}=\bigcup_{n=1}^{\infty}f_n^{-1}[(-\infty,a)]\in\mathcal{F}$$

מכך $\sup,\inf$ הם $\mathcal{F}$־מדידים על סמך למה 3.23 ולכן מיידית גם $\limsup$ וגם $\liminf$ הן $\mathcal{F}$־מדידות שכן:

$$\limsup_{n\to\infty}f_n(x)=\inf_{n\in\mathbb{N}}\left(\sup_{k\geq n}f_k(x)\right)$$

$$\liminf_{n\to\infty}f_n(x)=\sup_{n\in\mathbb{N}}\left(\inf_{k\geq n}f_k(x)\right)$$

■

טענה 3.31 אוסף נקודות ההתכנסות של סדרת משתנים מקריים היא קבוצה מדידה:

יהא $(\Omega,\mathcal{F})$ מרחב מדיד ותהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ אזי $\left\{\omega\in\Omega\mid \lim_{n\to\infty}X_n(\omega)\ exists\right\}$ היא $\mathcal{F}$־מדידה.

הוכחה: אם $X_n(\omega)$ לא מתכנסת ל־ $\pm\infty$ באף $\omega\in\Omega$ אז פשוט נקבל כי:

$$\left\{\omega\in\Omega\mid \lim_{n\to\infty}X_n(\omega)\text{ exists}\right\}=\left\{\omega\in\Omega\mid \liminf_{n\to\infty}X_n(\omega)=\limsup_{n\to\infty}X_n(\omega)\right\}$$

מאחר ש־ $\limsup X_n$ ו־ $\liminf X_n$ הן $\mathcal{F}$־מדידות ידוע שקבוצת הנקודות שעליהן הן משתוות היא קבוצה מדידה, כנדרש.

הערה 3.32 אם מאפשרים לסדרה להתכנס ל־ $\pm\infty$ צריך לעבוד קצת יותר אבל הטענה נשארת נכונה.

■

הגדרה 3.33 התכנסות כמעט תמיד של סדרת משתנים מקריים:
יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ותהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ. נאמר כי הנ"ל מתכנסת כמעט־תמיד (כ"ת) למ"מ X אם:

$$\mathbb{P}\left(\left\{\omega\in\Omega\mid \lim_{n\to\infty}X_n(\omega)\neq X(\omega)\right\}\right)=0$$

כלומר X_n מתכנסת נקודתית ל־ X פרט אולי על תת־קבוצה מהסתברות אפס. נסמן זאת $X_n\xrightarrow{\text{a.s}}X$.

הגדרה 3.34 סיגמה אלגברה הנוצרת ע"י משתנה מקרי וע"י סדרת משתנים מקריים:
יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ויהא X משתנה מקרי:

1. נגדיר את $\sigma(X)$ להיות ה־ σ אלגברה המינימלית על Ω שעבורה X מדיד. באופן שקול $\sigma(X)=\sigma\left(\left\{X^{-1}(B)\mid B\in\mathcal{B}\right\}\right)$.
2. עבור סדרת מ"מ $\{X_n\}_{n=1}^{\infty}$ נגדיר את $\sigma\left(\{X_n\}_{n=1}^{\infty}\right)$ להיות ה־ σ־אלגברה המינימלית על Ω כך ש־ X_n מדיד לכל n.

הגדרה 3.35 אי־תלות של סדרת משתנים מקריים:
יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות נאמר כי סדרת מ"מ $\{X_n\}_{n=1}^{\infty}$ היא ב"ת אם $\{\sigma(X_n)\}_{n=1}^{\infty}$ היא סדרה ב"ת של σ־אלגבראות.

הערה 3.36 גם כאן אפשר לדבר כמובן על אי־תלות של מספר סופי של מ"מ בשינוי מיידי של ההגדרה.

טענה 3.37 הגדרה שקולה לאי־תלות של אוסף וסדרת משתנים מקריים (תוספת):

יהא $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות:

1. משתנים מקריים $X_1, ..., X_N$ הם ב"ת אם לכל $B_1, ..., B_N \in \mathcal{F}$ המאורעות $X_1^{-1}[B_1], ..., X_N^{-1}[B_N]$ ב"ת.
2. משתנים מקריים $X_1, X_2, ...$ הם ב"ת אם לכל $N \in \mathbb{N}$ המשתנים $X_1, ..., X_N$ ב"ת.

הגדרה 3.38 אי־תלות של מ"מ בסדרת מ"מ:
תהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ. נאמר כי מ"מ Y ב"ת בסדרה אם $\sigma(Y)$ ו־ $\sigma(\{X_n\}_{n=1}^{\infty})$ ב"ת.

הגדרה 3.39 אי־תלות של מאורע בסדרת מ"מ:
תהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ. נאמר כי $A \in \mathcal{F}$ ב"ת בסדרה אם $\sigma(\{A\})$ ו־ $\sigma(\{X_n\}_{n=1}^{\infty})$ ב"ת.

הגדרה 3.40 סיגמת־אלגברת זנב של סדרת משתנים מקריים:
תהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ נגדיר $\mathcal{T}_n := \sigma(\{X_n, X_{n+1}, ...\})$ ונגדיר $\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{T}_n$, הנ"ל היא σ־אלגברת הזנב של הסדרה. בפרט כל מאורע $A \in \mathcal{T}$ יכונה מאורע זנב של הסדרה.

הערה 3.41 נשים לב שאם Y הוא מ"מ $\mathcal{T}$־מדיד אז הוא מיידית גם $\sigma(\{X_n\}_{n=1}^{\infty})$־מדיד שכן $\mathcal{T} \subseteq \sigma(\{X_n\}_{n=1}^{\infty})$.

למה 3.42 משתנים מקריים $\mathcal{T}$־מדידים הם ב"ת בסדרה ב"ת (ללא הוכחה):

תהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ ב"ת ותהא $\mathcal{T}$ σ־אלגברת הזנב של הסדרה. אזי אם Y הוא מ"מ $\mathcal{T}$־מדיד אזי Y ב"ת ב־ $\{X_n\}_{n=1}^{\infty}$.

הערה 3.43 אפשר להראות גם שכל מאורע זנב של סדרה הוא ב"ת בסדרה (ללא צורך שהסדרה תהיה ב"ת).

משפט 3.44 גרסה של חוק 0־1 של קולמוגורוב (תרגיל):

תהא $\{X_n\}_{n=1}^{\infty}$ סדרת מ"מ ב"ת ויהא Y משתנה מקרי $\sigma(\{X_n\}_{n=1}^{\infty})$־מדיד כך ש־ $Y, X_1, X_2, ...$ הם בלתי־תלויים. אזי Y קבוע כ"ת (כלומר קיים $c \in \mathbb{R}$ כך ש־ $\mathbb{P}(\{\omega \in \Omega \,|\, Y(\omega) = c\}) = 1$ ובפרט ניתן להראות ש:

$$c = \sup\left\{a \in \mathbb{R} \,:\, \mathbb{P}\left(Y^{-1}[(-\infty, a]]\right) = 0\right\}$$

הערה 3.45 נזכיר שבמשפט 0-1 הקודם הראינו שמאורע זנב של סדרת מאורעות הוא ב"ת בסדרה ומכך הסקנו שהוא ב"ת תלוי בעצמו ולכן טריוויאלי. במשפט הזה התנאים קצת שונים ונתון לנו מ"מ שהוא ב"ת בסדרה שהיא בעצמה ב"ת וצריך להסיק שהמשתנה הנ"ל טריוויאלי.

הערה 3.46 להיות מדיד ביחס ל־ $\sigma(\{X_n\}_{n=1}^{\infty})$ היא דרישה חלשה בהרבה מאשר להיות מדיד ביחס ל־ $\sigma(X_n)$ לכל n.

דוגמה 3.47 דוגמאות:

1. בהנתן סדרת מ"מ $\{X_n\}_{n=1}^{\infty}$ הקבוצה:

$$A := \left\{\omega \in \Omega \,\middle|\, \sum_{n=1}^{\infty} X_n(\omega) < \infty\right\}$$

היא $\mathcal{F}$־מדידה וגם מהווה מאורע זנב של הסדרה. בנוסף $\mathbb{1}_A(\omega)$ הוא משתנה מקרי $\sigma(\{X_n\}_{n=1}^{\infty})$־מדיד וב"ת ב־ $X_1, X_2, ...$.
לכן אם $X_1, X_2, ...$ היא סדרה ב"ת אז $\mathbb{1}_A$ הוא קבוע כ"ת. הנ"ל שקול לכך שהמאורע A מתרחש בהסתברות 0 או 1 (הטור מתכנס על קבוצה ממידה 1 או לא מתכנס על קבוצה ממידה 1).

2. בהנתן סדרת מ"מ $\{X_n\}_{n=1}^{\infty}$ הקבוצה:

$$A:=\left\{\omega\in\Omega\mid\lim_{n\to\infty}\frac{1}{n}\sum_{n=1}^{\infty}X_i\left(\omega\right)\text{ exists}\right\}$$

היא מאורע זנב. לכן אם הסדרה ב"ת אז $\mathbb{P}\left(A\right)=0,1$ וכאשר $\mathbb{P}\left(A\right)=1$ אפשר להגדיר מ"מ $Y\left(\omega\right)=\lim_{n\to\infty}\frac{1}{n}\sum_{n=1}^{\infty}X_i\left(\omega\right)$.
בפרט הנ"ל הוא $\mathcal{T}$־מדיד ולכן ב"ת ב־ $X_1,X_2,...$ ומכך קבוע כ"ת.

תרגיל: נתבונן בסדרה המוגדרת ע"י:

$$X_n=\begin{cases}\frac{1}{n} & \mathbb{P}=\frac{1}{2}\\ -\frac{1}{n} & \mathbb{P}=\frac{1}{2}\end{cases}$$

נניח ש־ $\{X_n\}_{n=1}^{\infty}$ היא סדרה ב"ת מהו הערך של $\mathbb{P}\left(\left\{\omega\in\Omega\mid\sum_{n=1}^{\infty}X_n\left(\omega\right)<\infty\right\}\right)$ (האפשרויות הן $0,1$). מתברר שהתשובה היא 1 והדרך להראות זאת משתמשת בא"ש צ'ביש ובכך שסכום השונויות של המשתנים המקריים הללו מתכנס.

תרגיל: משתנה מקרי שמדיד ביחס למשתנה מקרי אחר הוא פונקציה שלו:
יהא $\left(\Omega,\mathcal{F},\mathbb{P}\right)$ ויהיו $X,Y:\left(\Omega,\mathcal{F}\right)\to\left(\mathbb{R},\mathcal{B}\right)$ מ"מ כך ש־ Y הוא $\sigma\left(X\right)$־מדיד אזי קיימת $f:\mathbb{R}\to\mathbb{R}$ מדידה כך ש־ $Y=f\left(X\right)$.

דוגמה 3.48 דוגמה מעניינת לשאלות על מאורעות זנב ־ פרקולציה בגרף אינסופי:
נתבונן בשריג אינסופי בן־מנייה $G=\left(V,E\right)$ (גם E וגם V בני־מנייה) שבו דרגת הקודקודים היא חסומה. בהנתן מנייה $E=\{e_1,e_2,..\}$ נבצע פרקולציה על השריג עם הסתברות p (נמחק כל קשת בהסתברות $1-p$ ונשאיר אותה בהסתברות p באופן בלתי תלוי) ונקבל גרף חדש $G^{'}=\left(V,E^{'}\right)$ כאשר $E^{'}\subseteq E$ כמובן. כעת נגדיר סדרת מ"מ $X_n=\mathbb{1}_{\left\{e_n\in E^{'}\right\}}$ ונגדיר סדרת σ־אלגבראות על ידי $\mathcal{F}_n=\sigma\left(X_1,...,X_n\right)$, נתבונן בזנב $\mathcal{T}=\bigcap_{n=1}^{\infty}\mathcal{F}_n$. בפרט עבור כל קשת e_n מתקיים:

$$\sigma\left(X_n\right)=\left\{\Omega,\emptyset,\left\{E^{'}\mid e_n\in E^{'}\right\},\left\{E^{'}\mid e_n\notin E^{'}\right\}\right\}$$

הערה 3.49 נשים לב שעבור קודקוד $v\in V$ כלשהו המאורע "הקודקוד שייך לרכיב קשירות אינסופי בגרף $G^{'}$" הוא מאורע מדיד ביחס לזנב שכן הוא המשלים של איחוד המאורעות "v שייך לרכיב קשירות מגודל k עבור $k\in\mathbb{N}$ כלשהו". לכן המאורע "קיים רכיב קשירות אינסופי" הוא גם כן מדיד ביחס לזנב כאיחוד על פני כל $v\in V$ של המאורעות הללו. (למעשה נרצה לחשוב על רכיבי הקשירות כמיוצגים על ידי צלעות ולא על ידי קודקודים שכן כל הפורמליזם של הבעיה מנוסח על פי מצב הקשתות בגרף).

לאחר שהשתכנענו שהמאורע A = "קיים רכיב קשירות אינסופי" הוא מדיד ביחד לזנב נקבל ממשפט קולמוגורוב שהוא בעל הסתברות $\mathbb{P}\left(A\right)\in\{0,1\}$. כעת אפשר להראות ש־ $\mathbb{P}\left(A\right)$ יכולה רק לעלות כאשר p גדל. כלומר $\varphi\left(p\right)=\mathbb{P}_p\left(A\right)$ היא פונקציה עולה ולכן מכך ש־ $\mathbb{P}_p\left(A\right)\in\{0,1\}$ לכל p נסיק שקיים ערך קריטי $p_c\in\left[0,1\right]$ כך ש־ $\varphi_p\left(A\right)=\mathbb{1}_{\{p>p_c\}}$. אפשר להראות שתמיד מתקיים $\frac{1}{3}\leq p_c\leq\frac{2}{3}$. נתבונן במקרה הפרטי של השריג $\mathbb{Z}^2$ (במקרה הזה אפשר להראות ש־ $p_c=\frac{1}{2}$).

- נבצע פרקולציה עם $p<\frac{1}{3}$. נשים לב שמספר המסלולים הפשוטים מאורך n ב־ $\mathbb{Z}^2$ שמתחילים בראשית הוא לכל היותר $4\cdot3^{n-1}$. הסיכוי של מסלול כנ"ל לשרוד את הפרקולציה הוא כמובן p^n ולכן תוחלת מספר המסלולים מאורך n שמתחילים בראשית ושורדים את הפרקולציה היא לכל היותר $4\cdot3^{n-1}\cdot p^n$. עבור $p<\frac{1}{3}$ נקבל כי $4\cdot3^{n-1}\cdot p^n\overset{n\to\infty}{\longrightarrow}0$ ומכך בהכרח ההסתברות שהראשית תהיה שייכת לרכיב קשירות אינסופי היא אפס (אחרת התוחלת הזו הייתה שואפת למשהו גדול מאפס). מאחר והנ"ל נכון לא רק לראשית אלא עבור כל קודקוד נסיק שהסיכוי של כל קודקוד להופיע ברכיב קשירות אינסופי הוא אפס ולכן כאשר $p<\frac{1}{3}$ ההסתברות שקיים רכיב קשירות אינסופי היא אפס.
- עבור $p>\frac{2}{3}$ נראה שבהסתברות חיובית קיים רכיב קשירות שמכיל את הראשית. נתבונן בגרף הדואלי ל־ $\mathbb{Z}^2$ (הגרף שבו הפאות הן קודקודים וקשרתות עוברות בין פאות סמוכות בגרף המקורי). רכיב הקשירות של הראשית הוא סופי אמ"מ קיים מסלול בגרף הדואלי מאורך שמקיף את הראשית. אפשר להראות שמספר המסלולים הדואליים מאורך n שמקיפים את הראשית הוא לכל היותר $\frac{n}{2}\cdot3^{n-1}$ והסיכוי של מסלול כזה לחסום את הראשית הוא $\left(1-p\right)^n$. עבור $p>\frac{2}{3}$ נקבל כי:

$$\sum_{n=1}^{\infty}\frac{n}{2}3^{n-1}\cdot\left(1-p\right)^n<\infty$$

לכן מבורל קנטלי יש רק מספר סופי של ערכי n ככה שיש מסלול דואלי מאורך n סביב הראשית. מהמסקנה הזו אפשר להראות שבהכרח קיים רכיב קשירות אינסופי (לא בהכח שמכיל את הראשית) עבור כל $p > \frac{2}{3}$.

- התובנה העיקרית שצריך כדי להראות ש־ $p_c = \frac{1}{2}$ היא שהגרף הדואלי ל־ $\mathbb{Z}^2$ הוא גם כן $\mathbb{Z}^2$ אבל ההוכחה מסובכת.

3.4 התפלגויות של משתנים מקריים:

הגדרה 3.50 התפלגות של משתנה מקרי:
יהא $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהא X מ"מ ממשי, אזי X משרה מידת הסתברות על $\mathbb{R}$ שנתונה ע"י $\mu(A) = \mathbb{P}\left(X^{-1}(A)\right)$ לכל $A \in \mathcal{B}$. בפרט מוגדרת פונקציית ההתפלגות המצטברת (CDF) של X שנתונה ע"י:

$$F_X(t) \overset{\text{def}}{=} \mu((-\infty, t]) = \mathbb{P}\left(X^{-1}(-\infty, t]\right) = \mathbb{P}(X \leq t)$$

טענה 3.51 תכונות של ה־ CDF של משתנה מקרי ממשי:

יהא X מ"מ ממשי עם התפלגות מצטברת F_X אזי:

1. F_X היא מונוטונית עולה (חלש).
2. $F_X(t) \overset{t\to\infty}{\longrightarrow} 1$ ו־ $F_X(t) \overset{t\to\infty}{\longrightarrow} -1$.
3. F רציפה מימין $F_X(t) \overset{t\downarrow s}{\longrightarrow} F_X(s)$.

בנוסף כל פונקציה F שמקיימת את תכונות $1-3$ היא CDF של משתנה מקרי ממשי. כלומר קיימת מידת הסתברות μ_F כך ש:

$$\forall x \in \mathbb{R} \ : \mu_F((-\infty, x]) = F(x)$$

ובפרט לכל $[a,b] \subseteq \mathbb{R}$ מתקיים $\mu_F([a,b]) = F(b) - F(a)$.

הוכחה: להלן ההוכחה מהקורס בתורת המידה:

1. יהיו $-\infty < a < b < \infty$, נשים לב כי:

$$F_X(b) - F_X(a) = \mu((-\infty, b]) - \mu((-\infty, a]) = \mu((a, b]) \geq 0$$

לכן F מונוטונית לא יורדת. נראה רציפות מימין, יהא $x \in \mathbb{R}$ ותהא $\{x_n\} \subseteq \mathbb{R}$ סדרה כך ש־ $x_n \downarrow x$. מכך נקבל כי $(-\infty, x] = \bigcap_{n=1}^{\infty}(-\infty, x_n]$ ומכך מרציפות מלמטה של מידות (כאשר כאן משתמשים בסופיות של המידה בפרט) נקבל:

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mu((-\infty, x_n]) = \mu((-\infty, x]) = F_X(x)$$

כעת תהא $\{x_n\}_{n=1}^{\infty} \subseteq \mathbb{R}$ כך ש־ $x_n \downarrow -\infty$. מכך נקבל כי $\bigcap_{n=1}^{\infty}(-\infty, x_n] = \emptyset$ ולכן:

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} \mu((-\infty, x_n]) = \mu(\emptyset) = 0$$

מאחר והנ"ל נכון לכל סדרה כך ש־ $x_n \downarrow -\infty$ נסיק כי $\lim_{x\to-\infty} F_X(x) = 0$ ובאופן דומה מראים את הגבול השני.

2. נגדיר $\mathcal{S} := \{(a,b] \mid -\infty \leq a < b \leq \infty\}$ ונגדיר $F(\infty) = 1$ ו־ $F(-\infty) = 0$. לכל $(a,b] \in \mathcal{S}$ נגדיר:

$$\mu_F((a,b]) = F(b) - F(a)$$

זוהי מידה σ־אדיטיבית על $\mathcal{S}$ ולכן נוכל להרחיב את המידה הזו ל־ $\mathcal{B}$ באמצעות משפט Caratheodory ובכך נסיים.

■

תרגיל: **חידה:** יהיו X, Y, Z מ"מ ב"ת אחידים בקטע $[0,1]$ הראו ש־ $(XY)^Z \sim U[0,1]$.

הגדרה 3.52 התכנסות בהתפלגות (Convergence in Distribution/Law):
תהא X_n סדרת מ"מ (לא בהכרח באותו מרחב הסתברות) נאמר ש־ X_n מתכנסת בהתפלגות למ"מ X ונסמן $X_n \xrightarrow{D} X$ או $\mathcal{L}(X_n) \overset{n\to\infty}{\longrightarrow} \mathcal{L}(X)$ אם $F_n(t) \overset{n\to\infty}{\longrightarrow} F(t)$ בכל t שהיא נקודת רציפות של F.

הערה 3.53 הגדרה חילופית כאשר כל המשתנים מוגדרים באותו מרחב הסתברות היא שיש התכנסות חלשה של סדרת המידות המושרות $\mu_n(A) = \mathbb{P}\left(X_n^{-1}(A)\right)$ למידה $\mu(A) = \mathbb{P}\left(X^{-1}(A)\right)$.

3.5 תוחלת של משתנה מקרי וכמה תכונות:

הגדרה 3.54 תוחלת של משתנה מקרי (אינטגרציה):
יהא $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהא X מ"מ, נגדיר את התוחלת של X ע"י:

$$\mathbb{E}[X] = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} x d\mu_X$$

כאשר האינטגרלים הם אינטגרלי לבג ו־ μ_X היא המידה שמשרה X על $\mathbb{R}$.

הערה 3.55 בפרט נאמר ש־ X בעל תוחלת אם $\mathbb{E}[|X|] < \infty$ או לחילופין $X \in L^1(\Omega, \mathbb{P})$.

הערה 3.56 התוחלת של משתנה מקרי X תלויה רק בהתפלגות של X ולא במרחב ההסתברות.

הגדרה 3.57 מומנט של משתנה מקרי:
יהא X מ"מ נגדיר את המומנט ה־ $k \in \mathbb{N}$ של X ע"י $\mathbb{E}\left[X^k\right]$ כאשר נאמר שהנ"ל קיים כאשר $\mathbb{E}\left[|X|^k\right] < \infty$.

הערה 3.58 אפשר להראות שאם ל־ X קיים מומנט k אז קיימים כל המומנטים עבור $l \leq k$.

הגדרה 3.59 שונות של משתנה מקרי:
יהא X מ"מ נגדיר את השונות של X על ידי:

$$\operatorname{Var}(X) \overset{\text{def}}{=} \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$$

בפרט הנ"ל קיימת אמ"מ X בעל מומנט שני.

משפט 3.60 אי־שוויון ינסן (*Jensen Inequality*):

תהא φ פונקציה ממשית קמורה בקטע (a,b) ויהא X מ"מ בעל תוחלת שמקבל את ערכיו ב־ (a,b) אזי:

$$\varphi(\mathbb{E}(X)) \leq \mathbb{E}[\varphi(X)]$$

בתנאי ש־ $\mathbb{E}[\varphi(X)]$ קיים (אפילו במובן הרחב) ובפרט אם $\mathbb{E}[|\varphi(X)|] < \infty$.

הערה 3.61 ברור שעבור פונקציה קעורה אי־השוויון מתהפך.

הוכחה: מההנחה כי $a < X < b$ ומכך ש־ $\mathbb{P}$ היא מידת הסתברות נקבל כי:

$$a = \mathbb{E}(a) < \mathbb{E}[X] < \mathbb{E}[b] = b$$

מכך ש־ φ קמורה קיים $\beta_x \in \mathbb{R}$ כך שעבור $x := \mathbb{E}[X] \in (a,b)$ לכל $X(z) \in (a,b)$ מתקיים:

$$\varphi(X(z)) \geq \varphi(\mathbb{E}[X]) + \beta_x(X(z) - \mathbb{E}[X])$$

נקח את האינטגרל $d\mathbb{P}$ על שני האגפים ונקבל:

$$\begin{aligned}\mathbb{E}\left[\varphi\left(X\right)\right] &= \int\limits_{\Omega}\left(\varphi\circ X\right)d\mathbb{P} \geq \int\limits_{\Omega}\left[\varphi\left(\mathbb{E}\left[X\right]\right)\right]d\mathbb{P} + \beta_x\left(\mathbb{E}\left[X\right]-\mathbb{E}\left[\mathbb{E}\left[X\right]\right]\right)\\ &= \varphi\left(\mathbb{E}\left[X\right]\right)\int\limits_{\Omega}d\mathbb{P} + \beta_x\left(\mathbb{E}\left[X\right]-\mathbb{E}\left[\mathbb{E}\left[X\right]\right]\right)\\ &= \varphi\left(\mathbb{E}\left[X\right]\right)\cdot 1 + \beta\left(x\right)\cdot 0 = \varphi\left(\mathbb{E}\left[X\right]\right)\end{aligned}$$

■

הערה 3.62 בפרט $|x|, x^2, e^x$ כולן קמורות ולכן מקבלים את אי־השוויונות המתאימים. כמו כן $\sqrt{x}$ קעורה ולכן $\mathbb{E}\left[\sqrt{X}\right] \leq \sqrt{\mathbb{E}\left[X\right]}$.

תרגיל: האם אפשר לקבל חסם תחתון מעניין על $\mathbb{E}\left[\sqrt{X}\right]$ במקרה ש־ X הוא מ"מ פואסוני עם פרמטר λ?

משפט 3.63 אי־שוויון הלדר:

יהיו $p,q>1$ אקספוננטים צמודים ($\frac{1}{p}+\frac{1}{q}=1$) ויהיו X,Y מ"מ אזי $\mathbb{E}\left[|X\cdot Y|\right] \leq \sqrt[p]{\mathbb{E}\left[|X|^p\right]}\cdot\sqrt[q]{\mathbb{E}\left[|Y|^q\right]}$.

הוכחה: בה"כ אפשר להניח ש־ $X,Y\geq 0$ ומהומוגניות בשני האגפים אפשר להצטמצם למקרה שבו $\mathbb{E}\left[|X|^P\right]=1$ וגם $\mathbb{E}\left[|Y|^q\right]=1$. מכך מספיק להראות ש־ $\mathbb{E}\left[|X\cdot Y|\right]\leq 1$. מא"ש Young ידוע לנו ש־ $xy\leq\frac{x^p}{p}+\frac{y^q}{q}$ ומכך נקבל כי:

$$\mathbb{E}\left(XY\right)\leq\frac{\mathbb{E}\left[X^p\right]}{p}+\frac{\mathbb{E}\left[Y^q\right]}{q}=\frac{1}{p}+\frac{1}{q}=1$$

עבור הוכחת א"ש Young נשים לב כי עבור $y\geq 0$ קבוע הפונקציה $f\left(x\right)=\frac{x^p}{p}+\frac{y^q}{q}-xy$ מקיימת $f'\left(x\right)=x^{p-1}-y$ וגם $f''\left(x\right)=\left(p-1\right)x^{p-2}>0$ (לכל $x>0$) מכך הנקודה $x=y^{\frac{1}{p-1}}=y^{q-1}$ היא נקודת מינימום של הפונקציה שבה ערך הפונקציה הוא 0 ומהעברת אגפים מקבלים את הנדרש. ■

תרגיל: א"ש מינקובסקי: יהיו $p,q>1$ אקספוננטים צמודים ($\frac{1}{p}+\frac{1}{q}=1$) ויהיו X,Y מ"מ אזי:

$$\sqrt[p]{\mathbb{E}\left(|X+Y|^p\right)}\leq\sqrt[p]{\mathbb{E}\left[|X|^p\right]}+\sqrt[p]{\mathbb{E}\left[|Y|^p\right]}$$

ובפרט $\|x\|=\sqrt[p]{\mathbb{E}\left[|X|^p\right]}$ מגדירה נורמה על $L^p\left(\Omega,\mathbb{P}\right)$ (מרחב המשתנים המקריים $X:\Omega\to\mathbb{R}$ כך ש־ $\mathbb{E}\left[|X|^p\right]<\infty$).

3.6 התכנסות בהסתברות של משתנים מקריים:

הגדרה 3.64 התכנסות בהסתברות:
יהא $\left(\Omega,\mathcal{F},\mathbb{P}\right)$ מרחב הסתברות. נאמר שסדרת מ"מ X_n מתכנסת בהסתברות ונסמן $X_n\xrightarrow{\mathbb{P}}X$ אם לכל $\varepsilon>0$:

$$\mathbb{P}\left(|X_n-X|>\varepsilon\right)\xrightarrow{n\to\infty}0$$

או לחילופין לכל $\varepsilon>0$ קיים $N\in\mathbb{N}$ כך שלכל $n\geq N$ מתקיים $\mathbb{P}\left(|X_n-X|>\varepsilon\right)<\varepsilon$.

למה 3.65 התכנסות בהסתברות גוררת התכנסות כ"ת של תת־סדרה:

תהא X_n סדרת מ"מ כך ש־ $X_n\xrightarrow{\mathbb{P}}X$ אזי קיימת ת"ס X_{n_k} כך ש־ $X_{n_k}\xrightarrow{a.s}X$.

הוכחה: מהתכנסות בהסתברות בהנתן $k \in \mathbb{N}$ קיים N_k כך שלכל $n \geq N_k$ מתקיים:

$$\mathbb{P}\left(\left\{|X(x) - X_n(x)| > 2^{-k}\right\}\right) < 2^{-k}$$

נגדיר $E_k := \left\{|X(\omega) - X_{N_k}(\omega)| > 2^{-k}\right\}$ ונקבל כי $\mathbb{P}(E_k) < 2^{-k}$. נשים לב כי אם $\omega \notin \bigcup_{i=k}^{\infty} E_i$ אז $\omega \in \bigcap_{i=k}^{\infty} E_i^c$ ולכן $|X(\omega) - X_{N_i}(\omega)| < 2^{-i}$ לכל $i \geq k$ כלומר $X_{N_i}(\omega) \overset{i\to\infty}{\longrightarrow} X(\omega)$. כעת נגדיר $A := \limsup\limits_{k\to\infty} E_k$, על סמך מה שראינו אם $\omega \notin A$ אז $X_{N_i}(\omega) \overset{i\to\infty}{\longrightarrow} X(\omega)$ וכמו כן ניתן לראות כי לכל $k \in \mathbb{N}$ מתקיים:

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcap_{k=1}^{\infty}\bigcup_{i=k}^{\infty} E_i\right) \leq \mathbb{P}\left(\bigcup_{i=k}^{\infty} E_i\right) \leq \sum_{i=k}^{\infty} 2^{-i} = 2^{-k+1}$$

לכן בהכרח $\mathbb{P}(A) = 0$ ומכך $X_{N_k} \overset{\text{a.s}}{\longrightarrow} X$, כנדרש. ■

משפט 3.66 התכנסות כ"ת גוררת התכנסות בהסתברות גוררת התכנסות בהתפלגות:

יהא $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ותהא X_n סדרת מ"מ ו־ X מ"מ. אזי:

$$X_n \overset{\text{a.s}}{\longrightarrow} X \Longrightarrow X_n \overset{\mathbb{P}}{\longrightarrow} X \Longrightarrow \mathcal{L}(X_n) \overset{n\to\infty}{\longrightarrow} \mathcal{L}(X)$$

הוכחה: ניתן סקיצה של ההוכחות:

1. נראה שהתכנסות כ"ת גוררת התכנסות בהסתברות. נגדיר $A_n^{\varepsilon} = \{|X_n - X| < \varepsilon\}$ ונקבל שהתכנסות כ"ת שקולה לכך שעבור כל $\varepsilon > 0$ מתקיים $\mathbb{P}\left(\liminf\limits_{n\to\infty} A_n^{\varepsilon}\right) = 1$. כעת מהלמה של Fatou נקבל כי:

$$1 = \mathbb{P}\left(\liminf_{n\to\infty} A_n^{\varepsilon}\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n^{\varepsilon}) \Longrightarrow \liminf_{n\to\infty} \mathbb{P}(A_n^{\varepsilon}) = 1 \Longrightarrow X_n \overset{\mathbb{P}}{\longrightarrow} X$$

2. נראה שהתכנסות בהסתברות גוררת התכנסות בהתפלגות. נשים לב כי:

$$X_n \leq a \iff |X_n - X| > \varepsilon \wedge X \leq a + \varepsilon$$

מכך נקבל כי:

$$\mathbb{P}(X_n \leq a) \leq \mathbb{P}(X \leq a + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$$

מכך מהתכנסות בהסתברות לכל $\varepsilon > 0$ קיים N מספיק גדול כך שלכל $n \geq N$ מתקיים:

$$F(a + \varepsilon) - \varepsilon \leq F_n(a) \leq F(a + \varepsilon) + \varepsilon$$

בפרט נקבל כי:

$$\limsup_{n\to\infty} F_n(a) \leq \lim_{\varepsilon\to 0} F(a + \varepsilon) = F(a)$$
$$\liminf_{n\to\infty} F_n(a) \geq \lim_{\varepsilon\to 0} F(a + \varepsilon) \underbrace{=}_{\text{continuity points}} F(a)$$

לכן נקבל כי בנקודות רציפות של F מתקיים $\lim\limits_{n\to\infty} F_n(a) = F(a)$.

■

הערה 3.67 התכנסות כ"ת איננה תכונה טופולוגית (במרחב של משתנים מקריים):

- **עובדה:** במרחב טופולוגי X $\{x_n\} \subseteq X$ מתכנסת ל- $x \in X$ אמ"מ לכל ת"ס $\{x_{n_k}\}$ קיימת ת"ס $\left\{x_{n_{k_l}}\right\}$ כך ש- $x_{n_{k_l}} \to x$.
- ניתן למצוא סדרה $\{X_n\}_n$ שמתכנסת בהסתברות ולא מתכנסת כ"ת. בהנתן ת"ס $\{X_{n_k}\}$ גם היא מתכנסת בהסתברות ולכן מהלמה שהוכחנו יש לה ת"ס $\left\{X_{n_{k_l}}\right\}$ שמתכנסת כ"ת. כלומר הסדרה $\{X_n\}$ מקיימת את התכונה שלכל ת"ס שלה יש ת"ס מתכנסת כ"ת אבל היא בעצמה לא מתכנסת כ"ת ולכן לא ייתכן שהתכנסות כ"ת מגיעה מטופולוגיה.

משפט 3.68 משפט ההתכנסות החסומה (*Bounded Convergence Theorem*):

תהא X_n סדרת מ"מ כך ש- $X_n \xrightarrow{a.s} X$ או $X_n \xrightarrow{\mathbb{P}} X$ וגם $|X_n| \leq M$ לכל n אזי $\mathbb{E}[X_n] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X]$.

משפט 3.69 משפט ההתכנסות המונוטונית (*Monotone Convergence Theorem*):

תהא X_n סדרה עולה של מ"מ כך ש- $X_n \xrightarrow{a.s} X$ או $X_n \xrightarrow{\mathbb{P}} X$ אזי $\mathbb{E}[X_n] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X]$.

משפט 3.70 משפט ההתכנסות הנשלטת (*Dominated Convergence Theorem*):

תהא X_n סדרת מ"מ כך ש- $X_n \xrightarrow{a.s} X$ או $X_n \xrightarrow{\mathbb{P}} X$. יהא Y מ"מ בעל תוחלת כך ש- $|X_n| \leq Y$ לכל n אזי $\mathbb{E}[X_n] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X]$.

4 ריכוז מידה של סדרות משתנים מקריים:

4.1 הקדמה ודוגמאות ראשונות:

ריכוז מידה של מ"מ מתעניין בכמה משתנה מקרי מפוזר סביב התוחלת שלו, כלומר בחסמים על $\mathbb{P}\left(\left|X-\mathbb{E}\left[X\right]\right|>\varepsilon\right)$.

תזכורת 4.1 א"ש מרקוב וצ'בישב:
עבור משתנה מקרי X מתקיימים הדברים הבאים:

1. **א"ש מרקוב:** לכל $a>0$ מתקיים $\mathbb{P}\left(\left|X\right|\geq a\right)\leq\frac{\mathbb{E}(\left|X\right|)}{a}$
2. **א"ש צ'בישב:** לכל $a>0$ מתקיים $\mathbb{P}\left(\left|X\right|\geq a\right)\leq\frac{\mathbb{E}\left[X^{2}\right]}{a^{2}}$ ובפרט אם $\mathbb{E}\left[X^{2}\right]<\infty$ אז:
$$\mathbb{P}\left(\left|X-\mathbb{E}\left[X\right]\right|\geq a\right)\leq\frac{\mathrm{Var}\left(X\right)}{a^{2}}$$
3. עבור פונקציה $g:\mathbb{R}\to\mathbb{R}$ לא יורדת ואי־שלילית בטווח של X ולכל $a>0$ מתקיים:
$$\mathbb{P}\left(\left|X\right|\geq a\right)\leq\frac{\mathbb{E}\left[g\left(X\right)\right]}{g\left(a\right)}$$
בפרט אי השוויון הקודם הוא מקרה פרטי של כך עבור $g\left(t\right)=t^{2}$.

תזכורת 4.2 משפט פוביני (בגרסת הסתברותית):
אם X,Y הם משתנים מקריים ב"ת ובעלי תוחלת אזי $\mathbb{E}\left[X\cdot Y\right]=\mathbb{E}\left[X\right]\cdot\mathbb{E}\left[Y\right]$.

הערה 4.3 הניסוח הזה לא נראה כמו הניסוח הקלאסי של משפט פוביני בתורת המידה אבל הוא נובע מהמשפט הכללי יותר.

הגדרה 4.4 שונות משותפת ומ"מ בלתי־מתואמים:
יהיו X,Y מ"מ בעלי תוחלת נגדיר את השונות המשותפת של X,Y ע"י:
$$\mathrm{Cov}\left(X,Y\right)=\mathbb{E}\left[\left(X-\mathbb{E}\left[X\right]\right)\cdot\left(Y-\mathbb{E}\left[Y\right]\right)\right]$$
אפשר בקלות להראות שמתקיים:
$$\mathrm{Cov}\left(X,Y\right)=\mathbb{E}\left[X\cdot Y\right]-\mathbb{E}\left[X\right]\cdot\mathbb{E}\left[Y\right]$$
ובפרט $\mathrm{Cov}\left(X,Y\right)=0$ אמ"מ $\mathbb{E}\left[X\cdot Y\right]=\mathbb{E}\left[X\right]\cdot\mathbb{E}\left[Y\right]$, במקרה זה נאמר ש־ X ו־ Y בלתי־מתואמים (Uncorrelated).

הערה 4.5 בפרט ניתן לראות שאי־תלות גוררת חוסר־תיאום אולם ההפך לא בהכרח נכון.

דוגמה 4.6 יהיו $X_{1},X_{2},...$ מ"מ כך ש־ $\mathbb{P}\left(X_{i}=1\right)=\mathbb{P}\left(X_{i}=-1\right)=\frac{1}{2}$, נגדיר $S_{n}=\frac{1}{n}\sum_{i=1}^{n}X_{i}$ ונשים לב כי $\mathbb{E}\left[S\right]=0$ וגם:
$$\mathbb{E}\left[S^{2}\right]=\sum_{i=1}^{n}\mathbb{E}\left[X_{i}^{2}\right]+2\sum_{1\leq i<j<n}\mathbb{E}\left[X_{i}X_{j}\right]\overbrace{=}^{\text{Independence}}\sum_{i=1}^{n}1+2\sum_{1\leq i<j<n}\overbrace{\mathbb{E}\left[X_{i}\right]}^{=0}\overbrace{\mathbb{E}\left[X_{j}\right]}^{=0}=n$$
מכך מא"ש צ'בישב נקבל כי:
$$\mathbb{P}\left(\left|S\right|>\lambda\right)\leq\frac{n}{\lambda^{2}}$$

בפרט ברור שעבור $\sqrt{n} > \lambda$ החסם הנ"ל חסר תועלת והוא משתפר ככל ש־ λ גדל. כמו כן עבור $\lambda = n$ נקבל כי:

$$\mathbb{P}\left(|S| = n\right) = \mathbb{P}\left(|S| \geq n\right) \leq \frac{1}{n}$$

כאשר בפרט $|S| = n$ רק כאשר $X_1 = ... = X_n$.

שאלה: השתמשנו כאן באי־תלות כדי לקבל את החסם, נבחן מה קורה אם יש רק אי־תלות בזוגות.

תרגיל: שיטה לבניית מ"מ ב"ת בזוגות:
יהיו $Y_1, ..., Y_m$ מ"מ ב"ת כך ש־ $Y_i \in \{\pm 1\}$ בהסתברות $\frac{1}{2}$. עבור כל $\emptyset \neq A \in \mathcal{P}(\{1, ..., m\})$ נגדיר $X_A := \prod_{j \in A} Y_j$. אזי עבור כל $A \neq B$ המ"מ X_A ו־ X_B הם ב"ת כלומר האוסף $\{X_A \mid A \subseteq \{1, ..., m\}\}$ ב"ת בזוגות.

הערה 4.7 מאידך כבר לא מתקיימת אי־תלות בשלשות שכן למשל $X_{\{1,2\}} = X_{\{1\}} X_{\{2\}}$.

הערה 4.8 באמצעות וריאציה על השיטה הזו אפשר לבנות אוספים של משתנים מקריים ב"ת בשלשות, רבעיות וכו'.

מסקנה 4.9
בהנחה שהוכחנו את התרגיל הנ"ל נשים לב כי:

- ניתן להגדיר $2^m - 1$ משתנים מקריים שונים מהצורה X_A שכולם ב"ת בזוגות.
- מתקיים $X_A = 1$ לכל $A \subseteq \{1, ..., m\}$ רק אם $Y_1 = ... = Y_m = 1$ והנ"ל קורה בהסתברות 2^{-m}.
- במקרה שבו $Y_1 = ... = Y_m = 1$ נקבל כי:

$$S = \sum_{A \subseteq \{1, ..., m\}} X_A = 2^m - 1$$

- קיבלנו שעבור $n = 2^m - 1$ אי־תלות בזוגות הספיקה לנו כדי לקבל ש:

$$\mathbb{P}(|S| = n) = \frac{1}{n+1}$$

הנ"ל מראה לנו שהחסם $\mathbb{P}(|S| = n) \leq \frac{1}{n}$ שראינו בדוגמה הקודמת הוא כמעט הדוק אפילו אם דורשים רק אי־תלות בזוגות (כי יש דוגמה שמשיגה את החסם ולכן לא יכול להיות חסם הרבה יותר טוב עבור אי־תלות בזוגות).

דוגמה 4.10 נשים לב כי עבור מ"מ $X_1, ..., X_n$ כך ש־ $X_i \in \{\pm 1\}$ עם הסתברות $\frac{1}{2}$ ו־ $S_n = \sum_{i=1}^{n} X_i$ נקבל כי אם יש אי־תלות או אפילו אי־תלות ברביעיות (מספיק אפילו אי־תיאום ברביעיות) אז מתקיים:

$$\mathbb{E}\left[S^4\right] \overbrace{=}^{\text{Independence}} \sum_{i=1}^{n} \mathbb{E}\left[X_i^4\right] + \binom{4}{2} \sum_{1 \leq i < j \leq n} \overbrace{\mathbb{E}\left[X_i^2 X_j^2\right]}^{=1} = n + 6\frac{n(n-1)}{2} \leq 3n^2$$

מכך נקבל כי $\mathbb{P}(|S| > \lambda) \leq \frac{3n^2}{\lambda^4}$ ובפרט:

$$\mathbb{P}(|S| = n) \leq \frac{3}{n^2}$$

לכן ניתן לראות שקיבלנו חסם יותר קטן שדועך יותר מהר על ההסתברות ש־ $\mathbb{P}(|S| = n)$ מאשר במקרה שהנחנו אי־תלות רק בזוגות. אפשר להראות שבמקרה של אי־תלות בשלשות מקבלים ש־ $\mathbb{E}\left[S^3\right] = 0$ ולכן לא ניתן להשיג חסם מעניין בשיטה הזו.

שאלה: מה ההתנהגות של $\mathbb{P}(|S| = n)$ במקרה זה, האם היא דומה לזו של אי־תלות בזוגות או לזו של אי־תלות ברביעיות?

4.2 החוק החלש והחזק של המספרים הגדולים:

תרגיל: עבור זוג מאורעות A,B מתקיים ש־ A,B ב"ת אמ"מ המ"מ המציינים $\mathbb{1}_A,\mathbb{1}_B$ הם בלתי־מתואמים.

למה 4.11 הלמה השנייה של בורל קנטלי עבור מ"מ ב"ת בזוגות:

יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות ויהיו $A_1,A_2,...\in\mathcal{F}$ מאורעות ב"ת בזוגות כך ש־ $\sum_{i=1}^{\infty}\mathbb{P}(A_i)=\infty$ אזי $\mathbb{P}\left(\limsup_{n\to\infty}A_n\right)=1$.

הוכחה: נגדיר $X_i:=\mathbb{1}_{A_i}$ ו־ $S_n=\sum_{i=1}^{n}X_i$. מאחר שהנחנו אי־תלות בזוגות נקבל כי $\text{Cov}(X_i,X_j)=\delta_{ij}$ ומכך נקבל כי:

$$\text{Var}(S_n)=\sum_{i=1}^{n}\text{Var}(X_i)+\overbrace{\frac{1}{2}\sum_{i\neq j}\text{Cov}(X_i,X_j)}^{=0}=\sum_{i=1}^{n}\overbrace{\mathbb{P}(A_i)(1-\mathbb{P}(A_i))}^{\text{Var}(X_i)}\leq\sum_{i=1}^{n}\mathbb{P}(A_i)$$

נשים לב כי

$$\mathbb{E}[S_n]=\mathbb{E}\left[\sum_{i=1}^{n}X_i\right]=\sum_{i=1}^{n}\mathbb{E}[X_i]=\sum_{i=1}^{n}\mathbb{P}(A_i)$$

מכך מא"ש צ'בישב נקבל כי:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{\sum_{i=1}^{n}\mathbb{P}(A_i)}-1\right|>\varepsilon\right)&=\mathbb{P}\left(\left|S_n-\sum_{i=1}^{n}\mathbb{P}(A_i)\right|>\varepsilon\sum_{i=1}^{n}\mathbb{P}(A_i)\right)\\&\leq\frac{\text{Var}(S_n)}{\left(\varepsilon\sum_{i=1}^{n}\mathbb{P}(A_i)\right)^2}\leq\frac{\sum_{i=1}^{n}\mathbb{P}(A_i)}{\left(\varepsilon\sum_{i=1}^{n}\mathbb{P}(A_i)\right)^2}=\frac{1}{\varepsilon^2\sum_{i=1}^{n}\mathbb{P}(A_i)}\overset{n\to\infty}{\longrightarrow}0\end{aligned}$$

מאחר והנ"ל נכון לכל $\varepsilon>0$ קיבלנו כי $\frac{S_n}{\mathbb{E}[S_n]}\overset{\mathbb{P}}{\longrightarrow}1$ ומכך בהנתן $\varepsilon>0$ קיים n כך ש:

$$\mathbb{P}\left(\left|\frac{S_n}{\mathbb{E}[S_n]}-1\right|<\varepsilon\right)\leq 1-\varepsilon$$

הנ"ל בשילוב עם כך ש־ $\mathbb{E}[S_n]\overset{n\to\infty}{\longrightarrow}\infty$ גורר ש־ $S_n\overset{\mathbb{P}}{\longrightarrow}\infty$ ולכן עבור כל $\varepsilon>0$ וכל $M>0$ קיים n גדול מספיק כך ש־ $\mathbb{P}(|S_n|>M)\leq 1-\varepsilon$ כלומר למעשה $S_n\overset{\text{a.s}}{\longrightarrow}\infty$ ומהגדרת S_n המשמעות של כך היא שאינסוף A_i קורים כמעט תמיד. ■

למה 4.12

יהא $X\geq 0$ מ"מ אי־שלילי בעל תוחלת אזי לכל $\varepsilon>0$ קיים $K\in\mathbb{N}$ כך ש:

$$\mathbb{E}\left(X\cdot\mathbb{1}_{\{X\geq K\}}\right)<\varepsilon$$

בנוסף $X\cdot\mathbb{1}_{\{X<K\}}$ הוא בעל מומנט שני סופי.

הוכחה: נתבונן ב־ $Y_K=X\mathbb{1}_{\{x<K\}}$, מאחר ש־ X אי־שלילי ברור כי $Y_k\uparrow X$ כאשר $K\uparrow\infty$ ולכן ממשפט ההתכנסות המונוטונית:

$$\mathbb{E}[Y_K]\overset{K\to\infty}{\longrightarrow}\mathbb{E}[X]\Longrightarrow\mathbb{E}\left(X\cdot\mathbb{1}_{\{X\geq K\}}\right)=\mathbb{E}[X-Y_K]=\mathbb{E}[X]-\mathbb{E}[Y_K]\overset{K\to\infty}{\longrightarrow}0$$

בנוסף ניתן לראות כי:

$$\mathbb{E}\left[\left(X\cdot\mathbb{1}_{\{X<K\}}\right)^2\right]=\mathbb{E}\left[X^2\mathbb{1}_{\{X<K\}}\right]\leq\mathbb{E}\left[K^2\right]=K^2<\infty$$

■

משפט 4.13 החוק החלש של המספרים הגדולים:

תהא X_i סדרת מ"מ שווי התפלגות, ב"ת בזוגות ובעלי תוחלת $\mathbb{E}[X_i]=\mu$. נסמן $S_n=\sum_{i=1}^{n}X_i$, אזי $\frac{1}{n}S_n\xrightarrow{\mathbb{P}}\mu$.

הוכחה: **מקרה ראשון:** ראשית נניח ש־ X_i בעלי מומנט שני סופי כך ש־ $\mathrm{Var}(X_i)<\infty$, מכך מא"ש צ'בישב נקבל כי:

$$\mathbb{P}(|S_n-\mu n|>\varepsilon n)\le\frac{n\cdot\mathrm{Var}(X_1)}{\varepsilon^2 n^2}=\frac{\mathrm{Var}(X_1)}{\varepsilon^2 n}\xrightarrow{n\to\infty}0$$

כאשר השתמשנו בכך שמאי־תלות בזוגות של X_i נובע כי:

$$\mathrm{Var}(S_n)=\sum_{i=1}^{n}\mathrm{Var}(X_i)=n\mathrm{Var}(X_1)$$

כעת באופן כללי יותר: מאחר ו־ $X_i=X_i^+-X_i^-$ בה"כ נניח כי $X_i\ge 0$. יהא $\varepsilon>0$, על סמך הלמה הקודמת קיים $K\in\mathbb{N}$ כך ש:

$$\mathbb{E}\left(X_i\cdot\mathbb{1}_{\{X_i\ge K\}}\right)<\varepsilon^2$$

נשים לב כי $X_i=X_i\cdot\mathbb{1}_{\{X_i\ge K\}}+X_i\cdot\mathbb{1}_{\{X_i<K\}}$ ומכך:

$$S_n=\overbrace{\sum_{i=1}^{n}X_i\cdot\mathbb{1}_{\{X_i<K\}}}^{S_n^{'}}+\overbrace{\sum_{i=1}^{n}X_i\cdot\mathbb{1}_{\{X_i\ge K\}}}^{S_n^{"}}$$

נסמן $\mu^{'}=\mathbb{E}\left[X_i\cdot\mathbb{1}_{\{X_i<K\}}\right]$ ו־ $\mu^{"}=\mathbb{E}\left[X_i\cdot\mathbb{1}_{\{X_i\ge K\}}\right]$. מכך ש־ $\mu=\mu^{'}+\mu^{"}$ נקבל כי:

$$\mathbb{P}(|S_n-\mu n|>\varepsilon n)\le\mathbb{P}\left(\left|S_n^{'}-\mu^{'}n\right|>\frac{\varepsilon}{2}n\right)+\mathbb{P}\left(\left|S_n^{"}-\mu^{"}n\right|>\frac{\varepsilon}{2}\right)$$

כעת מאחר ש־ $X_i^{'}$ הוא בעל מומנט שני ובפרט $\mathrm{Var}\left(S_n^{'}\right)=n\mu^{'}$ נקבל מא"ש צ'בישב כי:

$$\mathbb{P}\left(\left|S_n^{'}-\mu^{'}n\right|>\frac{\varepsilon}{2}n\right)\overbrace{\le}^{\text{Chebyshev}}\frac{n\mu^{'}}{\left(\frac{\varepsilon}{2}n\right)^2}\le\frac{4\mu}{\varepsilon^2 n}$$

כמו כן מאחר ו־ $\mu^{"}=\mathbb{E}\left[X_i\cdot\mathbb{1}_{\{X_i\ge K\}}\right]<\varepsilon^2$ נקבל כי $\mathbb{E}\left[S_n^{"}\right]\le\varepsilon^2 n$ ולכן:

$$\mathbb{P}\left(\left|S_n^{"}-\mu^{"}n\right|>\frac{\varepsilon}{2}n\right)\le\mathbb{P}\left(S_n^{"}>\frac{\varepsilon}{2}n\right)\overbrace{\le}^{\text{Markov}}\frac{\varepsilon^2 n}{\frac{\varepsilon}{2}n}=2\varepsilon$$

מכך ניתן לראות כי:

$$\mathbb{P}(|S_n-\mu n|>\varepsilon n)\le\frac{4\mu}{\varepsilon^2 n}+2\varepsilon\xrightarrow{n\to\infty}2\varepsilon$$

מאחר והנ"ל נכון לכל $\varepsilon>0$ נסיק את השאיפה בהסתברות, כנדרש. ∎

הערה 4.14 נשים לב שכאשר X_i הם מ"מ בעלי מומנט ראשון ושני חסומים באופן אחיד (אבל לא בהכרח זהים) החוק החלש נובע מיידית מא"ש צ'בישב ולא צריך לדרוש שוויון התפלגות.

משפט 4.15 החוק החזק של המספרים הגדולים:

תהא X_i סדרת משתנים מקריים ש"ה, ב"ת בזוגות ובעלי תוחלת $\mathbb{E}\left[X_i\right]=\mu$ אזי $\frac{1}{n}S_n\xrightarrow{\text{a.s}}\mu$.

הערה 4.16 אם מניחים אי־תלות ברביעיות וקיום של מומנט רביעי אז אפשר להוכיח את המשפט באותם כלים שהשתמשנו בהם בחוק החלש באמצעות פירוק של המשתנים המקריים ושימוש בחסמים. בפרט מקבלים חסם מהצורה של $\frac{K}{n^2}$ שמסתכם לערך סופי ולכן מהלמה הראשונה של מבורל־קנטלי תהיה סטייה של $\frac{1}{n}S_n$ מ־ μ רק מספר סופי של פעמים.

הוכחה: מקרה ראשון: ראשית נניח כי $\left|X_i\right|\le 1$ ולכן כל המומנטים סופיים ובנוסף בה"כ $\mathbb{E}\left[X_i\right]=0$. באופן מיידי כך ש־ $\left|X_i\right|\le 1$ נובע כי $\text{Var}\left(X_i\right)\le\mathbb{E}\left[X_i^2\right]\le 1$ ומכך מא"ש צ'בישב נקבל כי:

$$\mathbb{P}\left(\left|S_n\right|>\varepsilon n\right)\le\frac{\text{Var}\left(S_n\right)}{\varepsilon^2n^2}\overbrace{=}^{\dagger}\frac{n\text{Var}\left(X_i\right)}{\varepsilon^2n^2}\le\frac{n}{\varepsilon^2n^2}=\frac{1}{\varepsilon^2n}$$

כאשר המעבר המסומן נובע מא"ת בזוגות. נתבונן בת"ס $n_k=k^2$ עבורה $\mathbb{P}\left(\frac{\left|S_{k^2}\right|}{k^2}>\varepsilon\right)\le\frac{1}{\varepsilon^2k^2}$. מאחר ש־ $\sum_{k=1}^{\infty}\frac{1}{\varepsilon^2k^2}<\infty$ נסיק מהלמה הראשונה של בורל קנטלי שהמאורע $\left\{\frac{\left|S_n\right|}{k^2}>\varepsilon\right\}$ בהסתברות 1 קורה רק מספר סופי של פעמים ומאחר והנ"ל נכון לכל $\varepsilon>0$ נסיק כי $\frac{S_{k^2}}{k^2}\xrightarrow{\text{a.s}}0$. כעת נשים לב כי אם $k^2\le n\le(k+1)^2$ נקבל כי:

$$\left|S_n-S_{k^2}\right|=\left|\sum_{i=k^2+1}^{n}X_i\right|\le\sum_{i=k^2+1}^{n}\left|X_i\right|\overbrace{\le}^{\left|X_i\right|\le 1}(k+1)^2-k^2+1=2k$$

מכך נקבל כי:

$$\frac{\left|S_n\right|}{n}\le\frac{\left|S_{k^2}\right|}{n}+\frac{2k}{n}\le\frac{\left|S_{k^2}\right|}{n}+\frac{2\sqrt{n}}{n}=\frac{\left|S_{k^2}\right|}{k^2}+\frac{2}{\sqrt{n}}\xrightarrow{\text{a.s}}0$$

מקרה שני: בה"כ אפשר להניח כי $X\ge 0$, נסמן $\mu=\mathbb{E}\left[X_i\right]$ ונראה כי $\limsup\limits_{n\to\infty}\frac{S_n}{n}\le 4\mu$ בהסתברות 1. בהנתן $k\in\mathbb{N}$ נסמן $X_i=X_i^{'}+X_i^{"}$ כאשר $X_i^{'}=X_i\mathbb{1}_{\{X_i\le 2^k\}}$ ו־ $X_i^{"}=X_i\mathbb{1}_{\{X_i>2^k\}}$. כמו כן נסמן:

$$S_{2^k}=S_{2^k}^{'}+S_{2^k}^{"}=\sum_{i=1}^{2^k}X_i^{'}+\sum_{i=1}^{2^k}X_i^{"}$$

מאחר ש־ $S_{2^k}=S_{2^k}^{'}+S_{2^k}^{"}$ נקבל כי:

$$\mathbb{P}\left(\frac{S_{2^k}}{2^k}>2\mu\right)=\mathbb{P}\left(S_{2^k}>2\mu 2^k\right)\le\mathbb{P}\left(S_{2^k}^{'}\ge 2\mu 2^k\right)+\mathbb{P}\left(S_{2^k}^{"}>0\right)$$

כמו כן $\mathbb{E}\left[S_{2^k}^{'}\right]=\sum_{i=1}^{2^k}\mathbb{E}\left[X_i^{'}\right]=2^k\mu^{'}$ ולכן מא"ש צ'בישב נקבל כי:

$$\begin{aligned}\mathbb{P}\left(S_{2^k}^{'}\ge 2\mu 2^k\right)&=\mathbb{P}\left(S_{2^k}^{'}-\mu 2^k\ge\mu 2^k\right)\overbrace{\le}^{\mu^{'}\le\mu}\mathbb{P}\left(\left|S_{2^k}^{'}-\mu^{'}2^k\right|\ge\mu 2^k\right)\\&\le\frac{\text{Var}\left(S_{2^k}^{'}\right)}{\mu^2 2^{2k}}\overbrace{=}^{\dagger}\frac{2^k\text{Var}\left(X_i^{'}\right)}{\mu^2 2^{2k}}\le\frac{\mathbb{E}\left[\left(X_i^{'}\right)^2\right]}{\mu^2 2^k}:=\alpha_k\end{aligned}$$

כאשר המעבר המסומן נובע מא"ת בזוגות. כמו כן ניתן לראות כי:

$$\mathbb{P}\left(S_{2^k}^{"}>0\right)=\mathbb{P}\left(\sum_{i=1}^{2^k}X_i^{"}>0\right)\le\sum_{i=1}^{2^k}\mathbb{P}\left(X_i^{"}>0\right)=2^k\mathbb{P}\left(X_i^{"}>0\right)=2^k\mathbb{P}\left(X_i\ge 2^k\right)$$

נסמן $X_i := X$ (הנחנו שוויון התפלגות) ונסמן $p_i = \mathbb{P}\left(2^i \le X \le 2^{i+1}\right)$. נשים לב כי:

$$\sum_{k=1}^{\infty} 2^k \mathbb{P}\left(X_i \ge 2^k\right) = \sum_{k=1}^{\infty} 2^k \sum_{i=k}^{\infty} p_i = \sum_{i=1}^{\infty} p_i \sum_{k=1}^{i} 2^k \approx \sum_{i=1}^{\infty} p_i 2^i \approx \mathbb{E}\left[X\right] < \infty$$

כמו כן ניתן לראות כי:

$$\sum_{k=1}^{\infty} \alpha_k = \sum_{k=1}^{\infty} \frac{\mathbb{E}\left[X_i^2 \mathbb{1}_{\{X_i \le 2^k\}}\right]}{\mu^2 2^k} \approx \frac{1}{\mu^2} \sum_{k=1}^{\infty} \sum_{i=1}^{k} \frac{p_i 2^{2i}}{2^k} = \frac{1}{\mu^2} \sum_{k=1}^{\infty} 2^{-k} \sum_{i=1}^{k} p_i 2^{2i}$$

$$= \frac{1}{\mu^2} \sum_{i=1}^{\infty} p_i 2^i \overbrace{\sum_{k=i}^{\infty} 2^{-k}}^{\le 1} \le \frac{1}{\mu^2} \sum_{i=1}^{\infty} p_i 2^i \approx \frac{1}{\mu^2} \mathbb{E}\left[X\right] < \infty$$

מכך נסיק כי $\sum_{k=1}^{\infty} \mathbb{P}\left(\frac{S_{2^k}}{2^k} > 2\mu\right) < \infty$ ולכן מהלמה הראשונה של בורל קנטלי בהסתברות 1 קיים k_0 כך שלכל $k > k_0$ מתקיים $\frac{S_{2^k}}{2^k} \le 2\mu$. מכך עבור כל $k \ge k_0$ ו־ $2^k \le n \le 2^{k+1}$ מתקיים $\frac{S_n}{n} \le \frac{S_{2^{k+1}}}{2^k} \le 4\mu$ ולכן $\limsup_{n\to\infty} \frac{S_n}{n} \le 4\mu$ בהסתברות 1. לסיום בהנתן $\varepsilon > 0$ נבחר K כך ש־ $\mathbb{E}\left[X_i \mathbb{1}_{\{K>k\}}\right] < \varepsilon$ ונסמן $X_i = X_i^{'} + X_i^{"}$ ו־ $S_n = S_n^{'} + S_n^{"}$ כמו קודם. על סמך המקרה הראשון נקבל כי $\frac{S_n^{'}}{n} \xrightarrow{\text{a.s}} \mu^{'}$ ועל סמך מה שראינו כעת נקבל כי בהסתברות 1 מתקיים:

$$\limsup_{n\to\infty} \frac{\left|S_n^{"}\right|}{n} \le 4\mu^{"} \le 4\varepsilon$$

מכך נקבל כי בהסתברות 1 מתקיים:

$$\limsup_{n\to\infty} \left|\frac{S_n}{n} - \mu\right| \le 5\varepsilon$$

מאחר והנ"ל נכון לכל $\varepsilon > 0$ נסיק כי $\frac{S_n}{n} \xrightarrow{\text{a.s}} \mu$, כנדרש. ■

תרגיל: נניח ש־ X_i ב"ת וש"ה (אך לא בהכרח בעלי תוחלת סופית) האם $\frac{S_n}{n} \xrightarrow{\text{a.s}} \mu$ אמ"מ $\mathbb{E}\left[X_i\right] = \mu$?

הערה 4.17 החוק החלש\חזק של המשתנים הגדולים מאששים את האינטואיציה שמשחק שבו תוחלת הרווח היא פונקציה של התפלגות מסוימת נעדיף לשחק משחק שבו ההתפלגות היא בעלת תוחלת גדולה יותר במיוחד אם נעשה חזרות רבות על המשחק.

4.2.1 כמה דוגמאות משעשעות:

דוגמה 4.18 פרדוקס שתי המעטפות:

- נניח שישנן שתי מעטפות שבאחת יש X כסף ובשנייה יש $2X$ כסף (X משתנה מקרי מפולג בצורה כלשהי). נותנים לנו מעטפה באקראי (הסתברות $\frac{1}{2}$) שבה יש Y כסף (X או $2X$ בהסתברות $\frac{1}{2}$). אנחנו מסתכלים בתוכן של המעטפה ושואלים אותנו האם נרצה לשמור את המעטפה או להחליף למעטפה השנייה. לכאורה בהסתברות $\frac{1}{2}$ מתקיים $Y = X$ ואז החלפה תתן לנו $Y^{'} = 2Y$ ובהסתברות $\frac{1}{2}$ מתקיים $Y = 2X$ ואז החלפה תתן לנו $Y^{'} = \frac{Y}{2}$, מכך נקבל כי לכאורה אם נחליף אז מתקיים:

$$\mathbb{E}\left[Y^{'} | Y\right] = \frac{1}{2} 2Y + \frac{1}{2}\frac{Y}{2} = 1\frac{1}{4} Y \Longrightarrow \mathbb{E}\left[Y^{'}\right] = \mathbb{E}\left[\mathbb{E}\left[Y^{'} | Y\right]\right] = 1\frac{1}{4} \mathbb{E}\left[Y\right]$$

אולם $Y^{'}$ היא גם הסכום של מעטפה מקרית ולכן צריך להתקיים $\mathbb{E}\left[Y^{'}\right] = \mathbb{E}\left[Y\right]$ ובפרט נגיע למסקנה הלא הגיונית שתמיד צריך להחליף. הסיבה שאין כאן פרדוקס היא שהחישוב שעשינו כאן איננו תקין מאחר ולא לקחנו בחשבון את האינפורמציה שגלומה בכך שראינו את Y (כלומר את הפילוג של X).

- נתבונן במקרה שבו $X=3^N$ כאשר $\mathbb{P}(N=n)=2^{-n}$ ונשים במעטפות X ו־ $3X$. נניח שקיבלנו מעטפה וראינו $Y=3^k$, הנ"ל קורה בשני מקרים:
  - אם $X=3^k$ וקיבלנו את המעטפה עם X ־ הנ"ל קורה בהסתברות $p=2^{-k}\cdot 2^{-1}$.
  - אם $X=3^{k-1}$ וקיבלנו את המעטפה עם $3X$ ־ הנ"ל קורה בהסתברות $p=2^{-k-1}\cdot 2^{-1}$.
  - בפרט ניתן לראות שהמאורע הראשון סביר פי 2 מהמאורע השני.

כעת אם נחליף ונסמן ב־ $Y^{'}$ את הסכום שנקבל אז מנוסחת Bayes נקבל כי:

$$\mathbb{P}\left[Y^{'}=3Y\,|\,Y=3^k\right]=\frac{2^{-k-1}\cdot 2^{-1}}{2^{-k-1}\cdot 2^{-1}+2^{-k}\cdot 2^{-1}}=\frac{1}{3}$$
$$\mathbb{P}\left[Y^{'}=\frac{Y}{3}\,|\,Y=3^k\right]=\frac{2^{-k}\cdot 2^{-1}}{2^{-k-1}\cdot 2^{-1}+2^{-k}\cdot 2^{-1}}=\frac{2}{3}$$

מכך נקבל כי:

$$\mathbb{E}\left[Y^{'}|Y\right]=\frac{1}{3}\cdot 3Y+\frac{2}{3}\cdot\frac{Y}{3}=1\frac{2}{9}Y\Longrightarrow\mathbb{E}\left[Y^{'}\right]=\mathbb{E}\left[\mathbb{E}\left[Y^{'}|Y\right]\right]=1\frac{2}{9}\mathbb{E}\left[Y\right]$$

לכן ניתן לראות שלכאורה כדאי לנו תמיד להחליף לא משנה מה ראינו. מצד שני Y ו־ $Y^{'}$ הם בעלי אותה התפלגות ולכן צריך להתקיים $\mathbb{E}\left[Y^{'}\right]=\mathbb{E}\left[Y\right]$. הסיבה שאין כאן פרדוקס למרות שכאן בניגוד למקרה קודם כן ידועה לנו ההתפלגות של X היא פשוט שמתקיים $\mathbb{E}\left[Y\right]=\infty$ ולכן השוויון $\mathbb{E}\left[Y^{'}\right]=1\frac{2}{9}\mathbb{E}\left[Y\right]$ הוא תקין לחלוטין שכן לא נובע ממנו ש־ $\mathbb{E}\left[Y^{'}\right]>\mathbb{E}\left[Y\right]$.

תרגיל: נניח שישנו משחק שבו בכל שלב אם מהמרים על סכום x אז בהסתברות חצי נפסיד את הכל ובהסתברות $\frac{1}{2}$ נכפיל את מה שהימרנו. כעת בהנתן שיש לנו סכום התחלתי x_0 ומשחקים k סיבובים על המשחק שבכל סיבוב $1\le j\le k$ נהמר על x_j (שהוא חלקי ל־ x_{j-1} שהוא הסכום שהיה לנו בסוף הסיבוב הקודם). נשאלת השאלה מהי אסטרטגיית ההימור (החלק היחסי מ־ x_{j-1} שנרצה להמר עליה במשחק ה־ j) כך שנמקסם את חציון הסכום שיהיה לנו אחרי k סיבובים (שאותו נסמן x_k).

4.3 חזרה לריכוז מידה:

הגדרה 4.19 פונקציה יוצרת מומנטים של משתנה מקרי:
יהא X מ"מ הפונקציה יוצרת מומנטים של X היא $\varphi_X(t):=\mathbb{E}\left[e^{tX}\right]$ כאשר הנ"ל מוגדרת בכל נקודה שבה $\mathbb{E}\left[\left|e^{tX}\right|\right]<\infty$.

הערה 4.20 אפשר לראות ש־ $\varphi_X(0)=1$ בהכרח ולכן φ_X תמיד מוגדרת באפס.

הערה 4.21 עבור משתנה מקרי X אפשר להגדיר $e^{tX}=\sum_{k=0}^{\infty}\frac{t^kX^k}{k!}$ ומכך בתנאי התכנסות מסוימים נקבל כי:

$$\mathbb{E}\left[e^{tX}\right]=\sum_{k=0}^{\infty}\frac{t^k\mathbb{E}\left[X^k\right]}{k!}$$

מכך ניתן לראות ש φ_X היא הפונקציה היוצרת שמוגדרת על ידי הסדרה $\{\mathbb{E}\left[X^n\right]\}_{n=1}^{\infty}$ ובפרט לכל $k\in\mathbb{N}$ מתקיים $\varphi_X^{(k)}(0)=\mathbb{E}\left[X^k\right]$.

הערה 4.22 ישנם מקרים שבהם ל־ X יש את כל המומנטים שלו והם סופיים אך φ_X לא מוגדרת באף נקודה פרט לאפס.

למה 4.23
יהיו $X_1,...,X_n$ מ"מ ב"ת, נסמן $\varphi_i(t)=\mathbb{E}\left[e^{tX_i}\right]$ ו־ $S_n=\sum_{i=1}^{n}X_i$, אזי $\varphi_{S_n}(t)=\prod_{i=1}^{n}\varphi_i(t)$.

הוכחה: באינדוקציה פשוטה. ■

דוגמה 4.24 נניח ש- X_n היא סדרת מ"מ ב"ת המקבלים את הערכים ±1 בהסתברות $\frac{1}{2}$, מכך נקבל כי:

$$\mathbb{E}\left[e^{tX}\right]=\frac{e^{-t}+e^{t}}{2}=\frac{1}{2}\sum_{k=0}^{\infty}\left(\frac{(-1)^{k}t^{k}+t^{k}}{k!}\right)=\frac{1}{2}\sum_{k=0}^{\infty}\frac{t^{2k}}{(2k)!}\leq\sum_{k=0}^{\infty}\frac{t^{2k}}{2^{k}k!}=e^{\frac{t^{2}}{2}}$$

נסמן $S_n=\sum_{i=1}^{n}X_i$ ונקבל כי לכל $t>0$ מתקיים:

$$\varphi_{S_n}(t)\overbrace{=}^{\text{Independence}}\prod_{i=1}^{n}\varphi_i(t)=\prod_{i=1}^{n}\mathbb{E}\left[e^{tX_i}\right]\leq e^{\frac{t^{2}}{2}\cdot n}$$

לכן נקבל כי:

$$\mathbb{P}(S_n>\lambda)=\mathbb{P}\left(e^{tS_n}>e^{t\lambda}\right)\overbrace{\leq}^{\text{Markov}}\frac{\mathbb{E}\left[e^{tS_n}\right]}{e^{t\lambda}}\leq\frac{e^{\frac{t^{2}}{2}\cdot n}}{e^{t\lambda}}=e^{\frac{t^{2}}{2}\cdot n-t\lambda}$$

הפונקציה $f(t)=e^{\frac{t^{2}}{2}\cdot n-t\lambda}$ מקבלת מינימום בנקודה $t=\frac{\lambda}{n}$ שערכו $e^{-\frac{\lambda^{2}}{n^{2}}}$ ומכך נסיק כי:

$$\mathbb{P}(S_n>\lambda)\leq e^{-\frac{\lambda^{2}}{n^{2}}}$$

נכליל את הדוגמה הזו:

דוגמה 4.25 נניח ש- X_i היא סדרת מ"מ ב"ת כך ש- $|X_i|\leq1$ וגם $\mathbb{E}[X_i]=0$.

- נסמן $S_n=\sum_{i=1}^{n}X_i$ ועבור $\lambda\geq0$ כלשהו נרצה לחסום את ההסתברות $\mathbb{P}(S_n\geq\lambda)$.
- ראשית ניתן לראות כי לכל $t\geq0$ מתקיים:

$$\mathbb{P}(S_n\geq\lambda)=\mathbb{P}\left(e^{tS_n}\geq e^{t\lambda}\right)\overbrace{\leq}^{\text{Markov}}\frac{\mathbb{E}\left[e^{tS_n}\right]}{e^{t\lambda}}=\varphi_{S_n}(t)\cdot e^{-t\lambda}$$

- מאחר ש- e^{tx} קמורה ב- t ומכך שלכל $|x|\leq1$ מתקיים $\frac{1-x}{2}\in[0,1]$ נקבל כי:

$$e^{tx}=e^{\frac{1-x}{2}(-t)+\frac{1+x}{2}(t)}\leq\frac{1-x}{2}e^{-t}+\frac{1+x}{2}e^{t}$$

מכך נקבל שלכל i מתקיים:

$$\mathbb{E}\left[e^{tX_i}\right]\leq\frac{1-\mathbb{E}[X_i]}{2}e^{-t}+\frac{1+\mathbb{E}[X_i]}{2}e^{t}=\frac{e^{-t}+e^{t}}{2}+\overbrace{\mathbb{E}[X_i]}^{=0}\frac{e^{-t}+e^{t}}{2}$$

- כבר ראינו ש- $\frac{1}{2}\left(e^{-t}+e^{t}\right)\leq e^{\frac{t^{2}}{2}}$ ולכן קיבלנו שלכל i מתקיים $\varphi_{X_i}(t)=\mathbb{E}\left[e^{tX_i}\right]\leq e^{\frac{t^{2}}{2}}$ ומכך:

$$\mathbb{P}(S_n\geq\lambda)\leq\varphi_{S_n}(t)\cdot e^{-t\lambda}=\left(\prod_{i=1}^{n}\varphi_{X_i}(t)\right)\cdot e^{-t\lambda}\leq\prod_{i=1}^{n}e^{\frac{t^{2}}{2}}\cdot e^{-t\lambda}=e^{\frac{n}{2}t^{2}-t\lambda}$$

- בפרט $t=\frac{\lambda}{n}$ היא נקודת מינימום של $e^{\frac{n}{2}t^{2}-t\lambda}$ ומכך נקבל כי:

$$\mathbb{P}(S_n\geq\lambda)\leq e^{\frac{n}{2}\cdot\frac{\lambda^{2}}{n^{2}}-\frac{\lambda^{2}}{n}}=e^{-\frac{\lambda^{2}}{2n}}$$

הערה 4.26 נשים לב שחייבים לדרוש ש- $\lambda\geq0$ שכן האנליזה שעשינו הייתה עבור $t\geq0$.

הערה 4.27 בפרט ניתן לראות שהמקרה שבו $X_i=\pm1$ בהסתברות $\frac{1}{2}$ משיג את החסם הנ"ל.

על סמך הדוגמה הזו ננסח את המשפט הבא (שכרגע הוכחנו למעשה):

משפט 4.28 משפט חסם הופדינג (*Hoeffding Bound*):

תהא X_i סדרת מ"מ ב"ת כך ש־ $|X_i|\le 1$ וגם $\mathbb{E}[X_i]=0$ אזי לכל $\lambda\ge 0$ מתקיים $\mathbb{P}\left(\sum_{i=1}^{n}X_i\ge\lambda\right)\le e^{-\frac{\lambda^2}{2n}}$.

הערה 4.29 נשים לב שהדעיכה של $e^{-\frac{\lambda^2}{2n}}$ כפונקציה של λ מתחילה רק ב־ $\lambda=\sqrt{n}$ ולכן עבור $\lambda<\sqrt{n}$ החסם הנ"ל נותן ש־ $\mathbb{P}(S_n\ge\lambda)=1$ וזה כמובן טריוויאלי.

הערה 4.30 אם במקום ההנחה ש־ $|X_i|\le 1$ נניח ש־ $|X_i|\le C_i$ לכל i אז אפשר להראות באמצעות אותה אנליזה שמתקיים:

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

דוגמה 4.31 נתבונן במקרה של $\lambda=an$, במקרה הנ"ל נקבל כי:

$$\mathbb{P}\left(\frac{S_n}{n}\ge a\right)\le e^{-\frac{a^2}{2}n}$$

אפשר להראות שקיים הגבול $I(a):=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left(\frac{S_n}{n}\ge a\right)$ (מכונה ה־ Rate Function).

דוגמה 4.32 במובן מסוים חסם הופדינג עשוי להיות רחוק מאופטימלי, נסתכל למשל במקרה שבו:

$$X_i=\begin{cases}1 & \mathbb{P}=p\\ -1 & \mathbb{P}=p\\ 0 & \mathbb{P}=1-2p\end{cases}$$

עבור $p=n^{-\frac{2}{3}}$ נקבל ש־ $\mathrm{Var}(S_n)=2n^{\frac{1}{3}}$ וסטיית התקן היא $\sqrt{2}\cdot n^{\frac{1}{6}}$. מאידך הדעיכה של חסם הופדינג מתחילה רק ב־ $\lambda=n^{\frac{1}{2}}$ שזה הרבה מעבר לסדר של סטיית תקן אחת של S_n כאשר n גדול. (היינו רוצים לקבל אומדן טוב לריכוז המידה סביב התוחלת באזור של כמה סטיות תקן בודדות מהתוחלת).

משפט 4.33 משפט חסם צ'רנוף (*Chernoff Bound*):

תהא X_i סדרת מ"מ ב"ת כך ש־ $|X_i|\le 1$, $\mathbb{E}[X_i]=0$ ו־ $Var(X_i)=V_i<\infty$. נסמן:

$$V:=\mathrm{Var}(S_n)=\sum_{i=1}^{n}V_i$$

אזי לכל $\lambda\ge 0$ מתקיים:

$$\mathbb{P}\left(\sum_{i=1}^{n}X_i\ge\lambda\right)\le\max\left\{e^{-\frac{\lambda^2}{4V}},e^{-\frac{\lambda}{2}}\right\}$$

הערה 4.34 בפרט בדוגמה הקודמת עבור λ מסדר של $\sqrt{2}K\cdot n^{\frac{1}{6}}$ כדי לקבל חסם על ריכוז המידה באזור של K סטיות תקן סביב התוחלת.

הוכחה: ראשית על סמך הפיתוח של e^{tx} לטור חזקות נקבל כי לכל i מתקיים:

$$\mathbb{E}\left[e^{tX_i}\right]=1+t\overbrace{\mathbb{E}[X_i]}^{=0}+\sum_{k=2}^{\infty}\frac{t^k\mathbb{E}\left[X_i^k\right]}{k!}$$

מכך ש־ $|X_i| \leq 1$ ו־ $\mathbb{E}[X_i]=0$ נקבל כי לכל $k \geq 2$ מתקיים $V_i=\mathbb{E}\left[X_i^2\right] \geq \mathbb{E}\left[X_i^k\right]$ ולכן:

$$\mathbb{E}\left[e^{tX_i}\right]=1+\sum_{k=2}^{\infty}\frac{t^k\mathbb{E}\left[X_i^k\right]}{k!} \leq 1+V_i\sum_{k=2}^{\infty}\frac{t^k}{k!}=1+V_i\left(e^t-1-t\right)\overbrace{\leq}^{1+x\leq e^x} e^{V_i\left(e^t-1-t\right)}$$

כעת לא קשה להראות שעבור $0 \leq t \leq 1$ מתקיים $\left(e^{t-1-t}\right) \leq t^2$ ומכך עבור t כנ"ל נקבל כי:

$$\mathbb{E}\left[e^{tX_i}\right] \leq e^{V_i\left(e^t-1-t\right)} \leq e^{V_it^2}$$
$$\Downarrow$$
$$\mathbb{E}\left[e^{tS_n}\right]=\prod_{i=1}^{n}\mathbb{E}\left[e^{tX_i}\right] \leq \prod_{i=1}^{n}e^{V_it^2}=e^{\sum\limits_{i=1}^{n}V_it^2}=e^{Vt^2}$$

לכן על סמך פיתוחים שעשינו קודם נקבל שאם $0 \leq t \leq 1$ אז לכל $\lambda \geq 0$ מתקיים:

$$\mathbb{P}\left(S_n \geq \lambda\right) \leq \mathbb{E}\left[e^{tS_n}\right]\cdot e^{-t\lambda} \leq e^{Vt^2}\cdot e^{-t\lambda}=e^{Vt^2-t\lambda}$$

כמו כן הפונקציה $e^{Vt^2-t\lambda}$ מקבלת מינימום כאשר $t=\frac{\lambda}{2V}$ שערכו הוא $e^{\frac{\lambda^2}{4V}-\frac{\lambda^2}{2V}}=e^{-\frac{\lambda^2}{4V}}$. לסיום נפריד לשני מקרים:

- אם $\lambda \leq 2V$ אז $0 \leq t \leq 1$ ולכן על סמך האנליזה הזו נקבל כי $\mathbb{P}\left(S_n \geq \lambda\right) \leq e^{-\frac{\lambda^2}{4V}}$.
- אם $\lambda \geq 2V$ אז נקח $t=1$ ונקבל כי $\mathbb{P}\left(S_n \geq \lambda\right) \leq e^{-\frac{\lambda}{2}}$.

מכך סה"כ נקבל את המסקנה הנדרשת.

■

הערה 4.35 נשים לב שללא הניתוח שעשינו כדי לקבל ש־ $\left(e^t-1-t\right) \leq t^2$ עדיין נקבל שבכל מקרה מתקיים:

$$\mathbb{P}\left(S_n \geq \lambda\right) \leq e^{V\left(e^t-1-t\right)-\lambda t}$$

ניתן לראות כי:

$$\frac{d}{dt}\left(e^{V\left(e^t-1-t\right)-\lambda t}\right)=V\left(e^t-1\right)-\lambda$$

הנ"ל מקבל מינימום ב־ $t=\log\left(\frac{\lambda}{V}+1\right)$. אם נציב זאת נקבל כי:

$$\mathbb{P}\left(S_n \geq \lambda\right) \leq e^{V\left(\frac{\lambda}{V}+1-1-\log\left(\frac{\lambda}{V}+1\right)\right)-\lambda\log\left(\frac{\lambda}{V}+1\right)}=e^{\lambda-(\lambda+V)\log\left(\frac{\lambda}{V}+1\right)}$$

אם $\lambda < 2V$ נוכל להשתמש באנליזה שעשינו קודם. מאידך אם $\lambda > 2V$ אז נקבל כי:

$$\mathbb{P}\left(S_n \geq \lambda\right) \leq e^{\lambda-(\lambda+V)\log\left(\frac{\lambda}{V}+1\right)} \leq e^{\lambda-\lambda\log\left(\frac{\lambda}{V}+1\right)}=e^{-\lambda\left(\log\left(\frac{\lambda}{V}+1\right)-1\right)}$$

אפשר להראות שכאשר $\lambda > 2V$ אז מתקבל החסם הבא:

$$\mathbb{P}\left(S_n \geq \lambda\right) \leq e^{-\lambda\left(\log\left(\frac{\lambda}{V}+1\right)-1\right)} \leq e^{\frac{-\lambda\log\left(\frac{\lambda}{V}\right)}{8}}$$

לכן סה"כ נקבל את החסם:

$$\mathbb{P}\left(S_n \geq \lambda\right) \leq \max\left\{e^{-\frac{\lambda^2}{4V}},e^{\frac{-\lambda\log\left(\frac{\lambda}{V}\right)}{8}}\right\}$$

דוגמה 4.36 נרצה להראות שבמובן מסוים החסם הנ"ל הוא הדוק. נתבונן במקרה שבו $\frac{1}{n} - \text{Ber}\left(\frac{1}{n}\right) \sim X_i$:

- ניתן לראות כי $\mathbb{E}\left[X_i\right] = 0$ וגם $\text{Var}\left(X_i\right) = \frac{1}{n}\left(1 - \frac{1}{n}\right)$.
- כמו כן $S_n \sim \text{Bin}\left(n, \frac{1}{n}\right) - 1 \approx \text{Pois}\left(1\right) - 1$ ובפרט $\text{Var}\left(S_n\right) = 1 - \frac{1}{n}$.

על סמך כך ש־ S_n מפולג בקירוב $\text{Pois}\left(1\right) - 1$ נקבל כי לכל $0 \le k \le n$ מתקיים:

$$\mathbb{P}\left(S_n \ge \lambda\right) \ge \mathbb{P}\left(S_n = \lambda\right) = \frac{e^{-1}}{(\lambda+1)!} \sim e^{-1}e^{-(\lambda+1)\log((\lambda+1)-1)} = e^{-(\lambda+1)\log(\lambda)-1}$$

מאחר ש־ $\text{Var}\left(S_n\right) \overset{n\to\infty}{\longrightarrow} 1$ אז החל מ־ $\lambda = 2$ מתקיים $\lambda > 2V$ ולכן על סמך החסם שפיתחנו נקבל כי:

$$e^{-(\lambda+1)\log(\lambda)-1} \le \mathbb{P}\left(S_n \ge \lambda\right) \le e^{\frac{-\lambda\log\left(\frac{\lambda}{V}\right)}{8}}$$

הנ"ל מראה ש־ $\log\left(\mathbb{P}\left(S_n \ge \lambda\right)\right)$ דועך בסדר גודל של $-\lambda\log\left(\frac{\lambda}{V}\right)$.

דוגמה 4.37 מודל הצפרדעים: יש גרף אינסופי שעל כל קודקוד בו יושבת צפרדע, בהתחלה כולן ישנות, מעירים צפרדע כלשהי והיא מתחילה לעשות הילוך מקרי על הגרף (מעבר לקודקוד סמוך בהסתברות אחידה). כל פעם שצפרדע מגיעה לקודקוד שיש בו צפרדע שעוד לא התעוררה היא מתעוררת ומתחילה מהלך מקרי זהה. אפשר לשאול כתלות בגרף האם כל הצפרדעים מתעוררות. כמו כן אפשר לעשות וריאציה של השאלה שבה יש יותר מצפרדע אחת בכל קודקוד. בפרט ידוע שעבור עץ 100־רגולרי (שבו יש הרבה התפצלויות) כאשר יש צפרדע אחת בכל קודקוד לא כל הצפרדעים מתעוררות. נעשה קצת אנליזה פשוטה של הבעיה:

- נשים לב כי בכל שלב לכל היותר מספר הצפרדעים הערות מוכפל במקרה שכל הצפרדעים העירו צפרדע ישנה.
- לכן לאחר n צעדים של הבעיה מספר הצפרדעים שהתעוררו הוא לכל היותר 2^n.
- נתבונן בעץ 100 רגולרי שבו 99 קשתות עולות למעלה וקשת אחת יורדת למטה בכל קודקוד.
- אפשר לראות שלאחר n צעדים תוחלת הגובה של הצפרדע שהתחילה את התהליך ביחס לגובה שבו היא התחילה היא $\mathbb{E}\left[S_n\right] = 0.98n$ כאשר:

$$S_n = \sum_{i=1}^{n} \mathbb{1}_{\{\text{The frog went up on the i'th step}\}}$$

- מכך על סמך חסם צ'רנוף אפשר להסיק שעבור הצפרדע המקורית צפרדע הסיכוי להיות בגובה 0 (הגובה שבו היא התחילה) לאחר n צעדים דועך בצורה אקספוננציאלית כפונקציה של n. כלומר:

$$\mathbb{P}\left(S_n = 0\right) \le e^{-Kn}$$

עבור קבוע K שנתון מהחסם והוא פונקציה של $d = 100$ (דרגת כל הקודקודים) והעובדה ש־ 99 קשתות עולות. אפשר להראות שעבור דרגת קודקודים $d = 99$ מתקיים ש־ $e^{-Kn} \le 3^{-n}$.

- נשים לב שעבור כל צפרדע שמתעוררת ההתפלגות של הגובה שלה לאחר שהתהליך עבור n צעדים היא זהה להתפלגות של S_n (שכן אנחנו מעירים כל צפרדע בגובה התחלתי מתאים) .
- מאחר שיש 2^n צפרדעים ערות לכל היותר לאחר n צעדים וההסתברות לרדת לגבהים נמוכים היא כל כך נמוכה כך שלמשל שההסתברות לרדת לגובה $-n$ שואפת לאפס כאשר $n \to \infty$. לכן לא נכסה את כל עץ במקרה הנ"ל.

5 מרטינגלים:

5.1 מרטינגלים בדידים:

תזכורת 5.1 בהנתן מרחב הסתברות $\left(\Omega, 2^{\Omega}, \mathbb{P}\right)$ עם Ω סופית\בת־מנייה ומאורע $B \in 2^{\Omega}$ כך ש־ $\mathbb{P}(B) \neq 0$ ניתן להגדיר מידת הסתברות מותנית על $\left(\Omega, 2^{\Omega}\right)$ ע"י:

$$\mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

בפרט לכל מ"מ $X : \left(\Omega, 2^{\Omega}\right) \to \mathbb{R}$ מוגדרת התוחלת של X בהנתן B שתסומן $\mathbb{E}[X|B]$ בתור התוחלת ביחס למידה $\mathbb{P}_B$.

דוגמה 5.2 נתבונן בהסתברות הבאה המוגדרת על $\Omega : \{1,2,3\}^2$ כאשר נייצג את הקוארדינטות כמ"מ X, Y:

$Y \backslash X$	1	2	3
1	0	$\frac{1}{6}$	$\frac{1}{9}$
2	$\frac{1}{3}$	0	$\frac{1}{9}$
3	0	$\frac{1}{6}$	$\frac{1}{9}$

ניתן לראות ש־ $X \sim U\{1,2,3\}$ עם $\mathbb{E}[X] = 2$ וניתן לשאול מהי $\mathbb{E}[Y|X=a]$ (כלומר לכל $a \in \{1,2,3\}$ ניתן לשאול מה התוחלת של Y בהנתן המאורע $X = a$). אפשר לחשב ולהראות ש־ $\mathbb{E}[Y|X=a] = 2$ לכל $a \in \{1,2,3\}$. כמו כן אפשר להראות ש־ $\mathbb{E}[X \cdot Y] = 4$ ומתקיים:

$$\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[Y|X] \cdot X]$$

כאשר $\mathbb{E}[Y|X] : \left(\Omega, 2^{\Omega}\right) \to \mathbb{R}$ היא הפונקציה כך ש:

$$\mathbb{E}[Y|X](a) = \mathbb{E}[Y|X=a]$$

כאשר נחשוב עליה כפונקציה שתלויה רק בקוארדינטה הראשונה והיא מדידה ביחס ל־ σ־אלגברה על $\left(\Omega, 2^{\Omega}\right)$ שנוצרת ע"י X

תזכורת 5.3 משתנה מקרי בדיד אם התמונה שלו בת־מנייה (בפרט סופית) או באופן שקול:

- $\mathrm{Im}(X) \subseteq \mathbb{R}$ היא קבוצה בת־מנייה כך שלכל $a \in \mathrm{Im}(X)$ מתקיים $\mathbb{P}(X = a_i) > 0$.
- קיימת קבוצה בת־מנייה $A \subseteq \mathbb{R}$ כך ש־ $\mathbb{P}(X \in A) = 1$.

נזכיר שאם התמונה של משתנה מקרי איננה בת־מנייה אז לא ייתכן שלכל $a \in \mathrm{Im}(X)$ מתקיים $\mathbb{P}(X = a) > 0$.

דוגמה 5.4 בהנתן מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ וסדרת מ"מ בדידים $X_1, X_2, ...$ נניח שמתקיים:

$$\star \quad \mathbb{E}[X_{k+1}|X_1 = a_1, ..., X_k = a_k] = 0$$

עבור כל $k \in \mathbb{N}$ ועבור כל $a_1, ..., a_k$ בטווח של $X_1, ..., X_k$. נתבונן בסכום $S_k = \sum_{i=1}^{k} X_i$ ונתבונן בתוחלת המותנית הבאה:

$$\mathbb{E}[S_{k+1}|\forall i \leq k \; S_i = t_i]$$

נשים לב כי:

$$\begin{aligned}
\mathbb{E}[S_k + X_{k+1}|\forall i \leq k \; S_i = t_i] &= \mathbb{E}[S_k + X_{k+1}|\forall i \leq k \; S_i = t_i] \\
&= \mathbb{E}[S_k|\forall i \leq k \; S_i = t_i] + \mathbb{E}[X_{k+1}|\forall i \leq k \; S_i = t_i] \\
&= t_k + \mathbb{E}[X_{k+1}|\forall i \leq k \; S_i = t_i] = t_k + 0 = t_k
\end{aligned}$$

כאשר המעבר לפני האחרון נובע מכך שמהסדרה $\{S_1, ..., S_k\}$ ניתן לשחזר בדיוק את $\{X_1, ..., X_k\}$ ומההנחה $\star$.

הערה 5.5 בקונטקסט של החלק הבא שנדבר על סדרות של משתנים מקריים המשתנים שבסדרה כולם יהיו באותו מרחב הסתברות.

הגדרה 5.6 מרטינגל (הגדרה ראשונית ולא מלאה):
סדרת משתנים מקריים בדידים $\{M_i\}_{i=1}^{\infty}$ תקרא **מרטינגל** אם לכל $n \in \mathbb{N}$ ולכל $a_1, ..., a_n$ בטווח של $X_1, ..., X_n$ מתקיים:

$$\mathbb{P}\left(M_1 = a_1, ..., M_n = a_n\right) > 0 \Longrightarrow a_n = \mathbb{E}\left[M_{n+1} | \forall i \leq n \; M_i = a_i\right]$$

כאשר ההסתברויות והתוחלות נלקחות ביחס למרחב ההסתברות שבו מוגדרים M_i.

דוגמה 5.7 מהלך מקרי סטנדרטי: עבור $X_i \in \{\pm 1\}$ בהסתברות $\frac{1}{2}$ ובלתי תלויים סדרת המשתנים המקריים $S_n = \sum_{i=1}^{n} X_i$ היא מרטינגל.

דוגמה 5.8 בהנתן סדרת מ"מ בדידים וב"ת X_i כך ש־ $X_i \geq 0$ ו־ $\mathbb{E}\left[X_i\right] = 1$ נתבונן בסדרה $S_n = \prod_{i=1}^{n} X_i$. נשים לב כי:

$$\mathbb{E}\left[S_{n+1} | \forall i \leq n \; S_i = t_i\right] = \mathbb{E}\left[S_n \cdot X_{n+1} | \forall i \leq n \; S_i = t_i\right] = t_n \mathbb{E}\left[X_{n+1} | \forall i \leq n \; S_i = t_i\right] = t_n \mathbb{E}\left[X_{n+1}\right] = t_n$$

לכן ניתן לראות שזהו גם כן מרטינגל.

דוגמה 5.9 הכד של פוליה: בכד בשלב ראשון יש שני כדורים, שחור ולבן. בכל שלב מוציאים כדור באקראי (באופן אחיד) ומחזירים אותו ועוד כדור באותו צבע. נגדיר $X_n = \mathbb{1}_{\{\text{The n'th ball is black}\}}$ ונגדיר $B_n = \sum_{i=1}^{n} X_i$ שהוא מספר הכדורים השחורים בשלב ה־ n. נראה ש־ $M_n = \frac{B_n}{n}$ הוא מרטינגל, ניתן לראות כי:

$$\begin{aligned}
\mathbb{E}\left[\frac{B_{n+1}}{n+1} | \, B_1 = a_1, ..., B_n = a_n\right] &= \mathbb{E}\left[\frac{B_n}{n+1} + \frac{X_{n+1}}{n+1} | B_1 = a_1, ..., B_n = a_n\right] \\
&= \mathbb{E}\left[\frac{B_n}{n+1} | B_1 = a_1, ..., B_n = a_n\right] + \mathbb{E}\left[\frac{X_{n+1}}{n+1} | B_1 = a_1, ..., B_n = a_n\right] \\
&= \frac{a_n}{n+1} + \frac{1}{n+1} \mathbb{E}\left[X_{n+1} \, | B_1 = a_1, ..., B_n = a_n\right] \\
&\overbrace{=}^{\dagger} \frac{a_n}{n+1} + \frac{1}{n+1} \frac{a_n}{n} = \frac{a_n}{n+1}\left(1 + \frac{1}{n}\right) = \frac{a_n}{n}
\end{aligned}$$

כאשר המעבר המסומן נובע מכך שההסתברות לשלוף כדור שחור בשלב ה־ n בהנתן ש־ $B_n = a_n$ היא בדיוק $\frac{a_n}{n}$ (פרופרוציית הכדורים השחורים בשלב ה־ n).

דוגמה 5.10 יש חפיסת קלפים ושולפים קלפים בלי החזרה עד שעוצרים וצריך להחליט האם הקלף הבא הוא אדום או שחור. שאלנו את השאלה האם יש אסטרטגיה מועדפת שתעלה את תוחלת ההצלחה בניחוש. נסמן ב־ R_n את מספר הקלפים האדומים שנותרו בחפיסה לאחר n סיבובים (בפרט $R_0 = 26$). נתבונן בסדרה $M_n = \frac{R_n}{52-n}$ עבור $n = 1, ..., 51$ ונראה שזהו מרטינגל. נסמן $X_n = \mathbb{1}_{\{\text{the n'th card was red}\}}$ ונקבל:

$$R_n = 26 - \sum_{i=1}^{n} X_n$$

כעת ניתן לראות כי:

$$\begin{aligned}
\mathbb{E}\left[\frac{R_{n+1}}{52-(n+1)} | \, R_0 = a_0, ..., R_n = a_n\right] &= \mathbb{E}\left[\frac{R_n - X_{n+1}}{52-(n+1)} | R_0 = a_0, ..., R_n = a_n\right] \\
&\frac{a_n}{52-(n+1)} - \frac{1}{52-(n+1)} \mathbb{E}\left[X_{n+1} | R_0 = a_0, ..., R_n = a_n\right] \\
&= \frac{a_n}{52-(n+1)} - \frac{1}{52-(n+1)} \mathbb{P}\left[X_{n+1} = \text{red} | \, R_0 = a_0, ..., R_n = a_n\right] \\
&= \frac{a_n}{52-(n+1)} - \frac{1}{52-(n+1)} \frac{52-a_n}{52-n} = \frac{a_n}{52-n}
\end{aligned}$$

בקרוב נראה איך מסיקים מכך שזהו מרטינגל שלכל אסטרטגיה שנבחר נקבל שהסיכוי שהקלף הבא יהיה אדום הוא $\frac{1}{2}$. באופן שקול אם נרוויח 1 כל פעם שננחש נכון ונפסיד 1 כל פעם שנטעה עבור כל אסטרטגיית עצירה נקבל שתוחלת הרווח היא אפס.

דוגמה 5.11 נזכיר שפונקציה $f:\mathbb{R}^2\to\mathbb{R}$ תקרא הרמונית אם $\Delta f=\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}=0$ בכל נקודה.

הערה 5.12 זה שקול לכך שהאינטגרל המסילתי של f על פני מעגלים שווה לערך הפונקציה במרכז המעגל.

בהנתן גרף $G=(V,E)$ פונקציה $f:V\to\mathbb{R}$ תקרא הרמונית אם ממוצע הערכים מסביב לכל קודקוד הוא ערך הקודקוד.

תרגיל: עבור X_n הילוך מקרי על $\mathbb{Z}^2$ ו־ f הרמונית על $\mathbb{Z}^2$ הסדרה $f\left(X_n\right)$ היא מרטינגל.

עוד נראה בהמשך באמצעות מרטינגלים שפונקציה הרמונית וחסומה על $\mathbb{Z}^2$ היא בהכרח קבועה (אנלוגי למשפט ליוויל).

5.2 תוחלת מותנית:

היינו רוצים שבאופן כללי יותר גם במקרה הלא בדיד וגם כאשר יש מאורעות עם הסתברות אפס נוכל איכשהו להתנות בהם, לקחת תוחלת מותנית ולקבל את אותן תוצאות שיש במקרה שתיארנו עד עכשיו. במציאות לצערנו יש דוגמאות שעבורן גישה נאיבית כזו לא עובדת וההגדרות שנתנו לא עובדות.

דוגמה 5.13 נניח שיש לנו זוג מ"מ (X,Y) שההתפלגות המשותפת שלהם היא התפלגות אחידה על כדור היחידה $B^1\subseteq\mathbb{R}^2$.

הערה 5.14 כלומר לכל $A\subseteq B^1$ מתקיים $\mathbb{P}\left((X,Y)\in A\right)=\mathcal{V}\left(A\right)$ כאשר $\mathcal{V}$ הוא הנפח לפי מידת לבג.

היינו מצפים שאם נתנה במאורע $Y=0$ (שהוא מאורע מהסתברות אפס) אז ההתפלגות של $X|Y=0$ תהיה $U\left[-1,1\right]$. מאידך, אם נעבור לקואורדינטות פולריות נקבל מהזוג (X,Y) את הזוג (θ,r) כאשר θ,r בלתי תלויים, $\theta\sim U\left[0,2\pi\right]$ ו־ r הוא בעל צפיפות $f\left(r\right)=2r$ על הקטע $[0,1]$. לאחר מעבר הקואורדינטות המאורע $Y=0$ שקול למאורע $\theta\in\{0,\pi\}$ ולכן מאחר ש־ $X=r$ כאשר $\theta=0$ ו־ $X=-r$ כאשר $\theta=\pi$ נצפה שערכי $X|Y=0$ יהיו בעלי צפיפות בקטע $[-1,1]$ שנתונה על ידי:

$$f\left(X=x|Y=0\right)=x\cdot\operatorname{sgn}\left(x\right)$$

הדוגמה הזו מראה לנו שיש שני פילוגים שנראים לכאורה הגיוניים עבור $X|Y=0$ והנ"ל נובע מכך שאנחנו למעשה רוצים להתנות במאורע שההסתברות שלו היא אפס.

המטרה שלנו: בהנתן מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$ ועבור כל $A\in\mathcal{F}$ נרצה להגדיר את ההסתברות $\mathbb{P}\left(A|\mathcal{G}\right)$ כאשר $\mathcal{G}\subseteq\mathcal{F}$ היא תת־σ־אלגברה כלשהי.

הערה 5.15 למעשה נרצה להגדיר פונקציה $\mathcal{G}$־מדידה $\mathbb{P}\left(\cdot|\mathcal{G}\right):\mathcal{F}\to[0,1]$.
באופן הנ"ל נוכל להגדיר לכל $A\in\mathcal{F}$ את התוחלת המותנית $\mathbb{E}\left[\mathbb{1}_A|\mathcal{G}\right]=\mathbb{P}\left(A|\mathcal{G}\right)$ ואם ההגדרה הזו היא לינארית נוכל להרחיב אותה באופן הסטנדרטי ולהגדיר את $\mathbb{E}\left[X|\mathcal{G}\right]$ לכל $X:(\Omega,\mathcal{F})\to\mathbb{R}$ כך ש־ $\mathbb{E}\left[|X|\right]<\infty$. בפרט בהנתן מ"מ $Y:(\Omega,\mathcal{F})\to\mathbb{R}$ אם נקח את $\mathcal{G}=\sigma\left(Y\right)$ אז נקבל הגדרה של תוחלת מותנית במשתנה מקרי ע"י $\mathbb{E}\left[X|Y\right]=\mathbb{E}\left[X|\sigma\left(Y\right)\right]$.

הגדרה 5.16 תוחלת מותנית (הגדרה מלאה):
יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות, תהא $\mathcal{G}\subseteq\mathcal{F}$ תת־σ־אלגברה ויהא X מ"מ $\mathcal{F}$־מדיד כך ש־ $\mathbb{E}\left[|X|\right]<\infty$ ($X\in L^1\left(\Omega,\mathcal{F},\mathbb{P}\right)$).
נאמר שמ"מ Y הוא תוחלת מותנית של X בהנתן $\mathcal{G}$ ונסמן $Y=\mathbb{E}\left[X|\mathcal{G}\right]$ אם:

1. $Y\in L^1\left(\Omega,\mathcal{G},\mathbb{P}\right)$ ובפרט Y הוא $\mathcal{G}$־מדיד.
2. לכל $A\in\mathcal{G}$ מתקיים $\mathbb{E}\left[Y\cdot\mathbb{1}_A\right]=\mathbb{E}\left[X\cdot\mathbb{1}_A\right]$ (לכל $A\in\mathcal{G}$ מתקיים $\int_A Yd\mathbb{P}=\int_A Xd\mathbb{P}$).

מיד נוכיח קיום של Y כנ"ל ויחידות שלו עד כדי שינוי על קבוצה עם הסתברות אפס (Y הוא נציג של מחלקת השקילות ב־ $L^1\left(\Omega,\mathcal{G}\right)$).

הערה 5.17 כמקרה פרטי עבור $A=\Omega$ נקבל שמתקיים:

$$\mathbb{E}\left[X\right]=\mathbb{E}\left[X\cdot\mathbb{1}_\Omega\right]=\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]\cdot\mathbb{1}_\Omega\right]=\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{G}\right]\right]$$

תזכורת 5.18 יהא $(\Omega,\mathcal{F})$ מרחב מדיד ויהיו μ,ν מידות עליו:

1. נאמר ש־ ν רציפה בהחלט ביחס ל־ μ ונסמן $\nu \ll \mu$ אם לכל $A\in\mathcal{F}$ מתקיים $\nu(A)=0 \Longrightarrow \mu(A)=0$.
2. **משפט רדון־ניקודים:** אם ν,μ הן מידות σ־סופיות ו־ $\mu \ll \nu$ אז קיימת $f\in L^1(\mathcal{F},\mu)$ כך שלכל $A\in\mathcal{F}$ מתקיים $\int_A fd\mu=\nu(A)$. בנוסף f הנ"ל מוגדרת ביחידות עד כדי שינוי על קבוצה עם μ־מידה אפס.

הוכחה: הוכחת קיום של תוחלת מותנית: בהנתן $X\in L^1(\Omega,\mathcal{F})$ בה"כ נניח כי $X\geq 0$ (שכן $X=X^+-X^-$) ונגדיר $\nu(A)=\int_A Xd\mathbb{P}$. זוהי מידת הסתברות על $(\Omega,\mathcal{F})$ והיא באופן מיידי רציפה בהחלט ביחס ל־ $\mathbb{P}$. כעת מאחר ש־ $\mathcal{G}\subseteq\mathcal{F}$ נקבל ש־ $\mathbb{P},\nu$ הן מידות על $(\Omega,\mathcal{G})$ כך ש־ $\nu\ll\mathbb{P}$ ולכן ממשפט רדון ניקודים קיימת $f\in L^1(\Omega,\mathcal{G})$ כך ש:

$$\int_A fd\mathbb{P}=\nu(A)=\int_A Xd\mathbb{P}$$

לכן $f=\mathbb{E}[X|\mathcal{G}]$ והנ"ל מוגדרת ביחידות עד כדי שינוי על קבוצה עם הסתברות אפס. ■

הוכחה: נביא הוכחה אלטרנטיבית לקיום של תוחלת מותנית (ההוכחה המלאה בספר של וויליאמס):

תזכורת 5.19 נזכיר שאם X הוא מ"מ בעל מומנט שני אז $\mathbb{E}[X]=\inf_{a\in\mathbb{R}}\mathbb{E}\left[(X-a)^2\right]$.

כעת בהנתן $X\in L^1(\Omega,\mathcal{F},\mathbb{P})$ ותת־σ־אלגברה $\mathcal{G}\subseteq\mathcal{F}$ נחפש מ"מ Y שהינו $\mathcal{G}$־מדיד וממזער את $\mathbb{E}\left[(X-Y)^2\right]$. נשים לב שאם $X\in L^2(\Omega,\mathcal{F},\mathbb{P})$ אז קיום של Y כנ"ל הוא מיידי מאחר ש־ $L^2(\Omega,\mathcal{G},\mathbb{P})$ הוא מרחב הילברט עם מכפלה פנימית $\langle X,Y\rangle=\mathbb{E}[X\cdot Y]$. כעת מאחר ש־ $Y\perp X$ נקבל כי לכל $Z\in L^2(\Omega,\mathcal{G},\mathbb{P})$ מתקיים:

$$\langle X-Y,Z\rangle=\mathbb{E}[(X-Y)\cdot Z]=0$$

בפרט עבור $Z=\mathbb{1}_A$ כאשר $A\in\mathcal{G}$ נקבל כי:

$$\mathbb{E}[(X-Y)\mathbb{1}_A]=0\Longrightarrow\mathbb{E}[X\mathbb{1}_A]=\mathbb{E}[Y\mathbb{1}_A]$$

לכן Y הנ"ל שממזער את $\mathbb{E}\left[(X-Y)^2\right]$ הוא התוחלת המותנית של X. לסיום גם במקרה שבו X איננו בעל מומנט שני אפשר להרחיב את ההוכחה הזו ולקבל את אותה תוצאה.

■

נביא שתי דוגמאות פשוטות שבהן ניתן למצוא בקלות את התוחלת המותנית במפורש:

דוגמה 5.20 עבור $\mathcal{G}=\{\emptyset,\Omega\}$ פונקציה היא $\mathcal{G}$־מדידה אמ"מ היא קבועה ולכן ברור ש־ $\mathbb{E}[X|\mathcal{G}]=\mathbb{E}[X]$.

דוגמה 5.21 עבור $A\in\mathcal{F}$ כך ש־ $\mathbb{P}(A)>0$ ועבור $\mathcal{G}=\sigma(A)=\{\emptyset,\Omega,A,A^c\}$ לא קשה להראות ש:

$$\mathbb{E}[X|\mathcal{G}]=\begin{cases}\frac{1}{\mathbb{P}(A)}\mathbb{E}[X\mathbb{1}_A] & x\in A\\ \frac{1}{\mathbb{P}(A^c)}\mathbb{E}[X\mathbb{1}_{A^c}] & x\in A^c\end{cases}$$

במובן מסוים אפשר לחשוב על חישוב התוחלת המותנית של מ"מ בהנתן σ־אלגברה כעל התוחלת שתתקבל בהינתן "אינפורמציה" שניתנת לנו מכך שאנחנו מסתכלים רק על מאורעות שנמצאים באותה σ־אלגברה. אולם צריך להזהר מנקודת המבט הזו שכן לא תמיד ברור מהי האינפורמציה הזו שתמונה ב־ σ־אלגברה שבה אנו מתנים.

דוגמה 5.22 נקח $\Omega=\{0,1\}^{\mathbb{Z}}$ ונגדיר $\mathcal{F}_n=\sigma(\{\omega\in\Omega\,|\,a_n=1\})$ לכל $n\in\mathbb{Z}$. נגדיר σ־אלגבראות זנב עבור הסדרה הזו:

$$\mathcal{T}_R=\bigcap_{n=1}^{\infty}\sigma(F_n,F_{n+1},...)$$

$$\mathcal{T}_L=\bigcap_{n=1}^{\infty}\sigma(F_{-n},F_{-n-1},...)$$

אפשר להראות שמתקיים:

$$\sigma\left(\mathcal{T}_L,\mathcal{T}_R\right)\subsetneq\mathcal{T}_{RL}=\bigcap_{n=1}^{\infty}\sigma\left(F_n,F_{-n},F_{n+1},F_{-n-1}\right)$$

כלומר ה־σ־אלגברה שנוצרת על ידי שתי הזנבות לא זהה לזנב הדו־צדדי.

טענה 5.23 שלל תכונות של תוחלת מותנית:

יהא $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות, תהא $\mathcal{G}\subseteq\mathcal{F}$ ויהיו $X,Y\in L^1(\Omega,\mathcal{F},\mathbb{P})$, אזי:

1. מתקיים $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]]=\mathbb{E}[X]$.
2. אם X הוא $\mathcal{G}$־מדיד אז $\mathbb{E}[X|\mathcal{G}]=X$.
3. **לינאריות:** $\mathbb{E}[X+Y|\mathcal{G}]=\mathbb{E}[X|\mathcal{G}]+\mathbb{E}[Y|\mathcal{G}]$
4. **חיוביות:** אם $X\geq 0$ אז $\mathbb{E}[X|\mathcal{G}]\geq 0$.
5. **א"ש ינסן:** אם $\varphi:\mathbb{R}\to\mathbb{R}$ היא פונקציה קמורה אז $\varphi(\mathbb{E}[X|\mathcal{G}])\leq\mathbb{E}[\varphi(X)|\mathcal{G}]$ כ"ת.
6. **תכונת ההרכבה:** עבור $\mathcal{H}\subseteq\mathcal{G}\subseteq\mathcal{F}$ מתקיים $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]=\mathbb{E}[X|\mathcal{H}]$.
7. **תכונות *TOWIK*:** אם Z הוא $\mathcal{G}$־מדיד אז $\mathbb{E}[Z\cdot X|\mathcal{G}]=Z\cdot\mathbb{E}[X|\mathcal{G}]$.
8. **אי־תלות:** אם X ב"ת ב־ $\mathcal{G}$ אז $\mathbb{E}[X|\mathcal{G}]=\mathbb{E}[X]$.

עבור סדרת מ"מ $X_n\in L^1(\Omega,\mathcal{F},\mathbb{P})$ מתקיים:

1. **התכנסות מונוטונית:** אם $X_n\geq 0$ ו־ $X_n\uparrow X$ כ"ת אז $\mathbb{E}[X_n|\mathcal{G}]\uparrow\mathbb{E}[X|\mathcal{G}]$.
2. **התכנסות נשלטת:** אם $X_n\xrightarrow{\text{a.s}}X$ וגם קיים $Z\in L^1(\Omega,\mathcal{F},\mathbb{P})$ כך ש־ $|X_n|\leq Z$ אז $\mathbb{E}[X_n|\mathcal{G}]\xrightarrow{\text{a.s}}\mathbb{E}[X|\mathcal{G}]$.
3. **הלמה של פאטו:** אם $X_n\geq 0$ אז $\mathbb{E}[\liminf X_n|\mathcal{G}]\leq\liminf(\mathbb{E}[X_n|\mathcal{G}])$.

5.3 פילטרציות ומרטינגלים:

הגדרה 5.24 פילטרציה:
בהנתן מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$ פילטרציה היא סדרה עולה $\mathcal{F}_0\subseteq\mathcal{F}_1\subseteq\ldots$ של תת־σ־אלגבראות של $\mathcal{F}$.

הגדרה 5.25 תהליך סטוכסטי מותאם (Adapted Process):
תהליך סטוכסטי X_n יקרא מותאם ביחס לפילטרציה $\mathcal{F}_n$ אם X_n הוא $\mathcal{F}_n$־מדיד לכל n.

דוגמה 5.26 עבור $\Omega=\{0,1\}^{\mathbb{N}}$ נגדיר סדרת פונקציות $f_n:\Omega\to\mathbb{R}$ ע"י $f_n(\omega)=\omega_n$ ונגדיר את הפילטרציה $\mathcal{F}_n=\sigma(f_1,...,f_n)$.
באופן הנ"ל נקבל שהסדרה $X_n=\sum_{i=1}^n f_i$ היא $\mathcal{F}_n$־מותאמת.

הערה 5.27 לכל תהליך X_n יש פילטרציה טבעית שמתאימה לו וביחס אליה הוא מתואם והיא נתונה פשוט ע"י $\mathcal{F}_n=\sigma(X_1,...,X_n)$.
זוהי כמובן הפילטרציה המינימלית שעבורה התהליך הנ"ל מתואם.

הגדרה 5.28 מרטינגל (הגדרה מלאה):
תהליך סטוכסטי $X_n\in L^1(\Omega,\mathcal{F},\mathbb{P})$ יקרא מרטינגל ביחס לפילטרציה $\mathcal{F}_n$ אם X_n הוא $\mathcal{F}_n$־מותאם ובנוסף:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n]\overset{\text{a.s}}{=}X_n$$

הערה 5.29 הנ"ל מתלכד עם ההגדרה שנתנו עבור מ"מ בדידים כאשר X_n בדידים ו־ $\mathcal{F}_n$ היא הפילטרציה הטבעית.

דוגמה 5.30 הצטברות מידע: יהא X מ"מ עם תוחלת ותהא $\mathcal{F}_n$ פילטרציה כלשהי. נגדיר $X_n:=\mathbb{E}[X|\mathcal{F}_n]$, זהו מרטינגל שכן:

$$\mathbb{E}[X_{n+1}|\mathcal{F}_n]=\mathbb{E}[\mathbb{E}[X|\mathcal{F}_{n+1}]|\mathcal{F}_n]\overbrace{=}^{\mathcal{F}_n\subseteq\mathcal{F}_{n+1}}\mathbb{E}[X_n|\mathcal{F}_n]=X_n$$

הערה 5.31 אפשר להראות שכל מרטינגל X_n ביחס לפילטרציה $\mathcal{F}_n$ שהינו חסום ($|X_n| \leq M$) קיים מ"מ X כך ש- $X_n = \mathbb{E}\left[X|\mathcal{F}_n\right]$.

תזכורת 5.32 צביעה של קודקודי גרף תקרא צביעה חוקית אם כל שני קודקודים המחוברים בקשת צבועים בצבעים שונים. מספר הצביעה של גרף סופי הוא מספר הצבעים המינימלי שנדרשים כדי לצבוע את הגרף באופן חוקי.

דוגמה 5.33 נסמן ב- Ω את כל הגרפים בעלי n קודקודים $\{v_1,...,v_k\}$ ונסמן ב- $G\left(n,\frac{1}{2}\right)$ **גרף מקרי** ב- Ω שבו כל קשת אפשרית נמצאת בהסתברות $\frac{1}{2}$ (הנ"ל שקול לכך לבחירה מקרית אחידה של גרף ב- Ω). בהנתן $G\sim G\left(n,\frac{1}{2}\right)$ נסמן ב- $X=\chi\left(G\right)$ את מספר הצביעה של G (זהו מ"מ). נגדיר פילטרציה על ידי:

$$\mathcal{F}_k=\sigma\left(\text{Induced graph on }v_1,...,v_k\right)$$

נגדיר תהליך סטוכסטי $X_1,...,X_n$ על ידי $X_k=\mathbb{E}\left[X|\mathcal{F}_k\right]$ (בפרט $X_n=X$) . ידוע שמתקיים:

$$\mathbb{E}\left[X\right]=\frac{n}{2\log_2 n}\left(1+o\left(1\right)\right)$$

כפי שראינו בדוגמה הקודמת X_k הוא מרטינגל ביחס ל- $\mathcal{F}_k$, אפשר להראות **(תרגיל)** שמכך נובע כי $|X_{k+1}-X_k|\leq 1$(באופן אינטואטיבי בהנתן שראינו את מה שקורה על הקודקודים $v_1,...,v_k$ השינוי בערך הצפוי של X לאחר שראינו מה שקורה גם על $v_1,...,v_{k+1}$ הוא לכל היותר 1, כלומר או שנוסף צבע אחד או שלא). בפרט מא"ש אזומה (מופיע בהמשך) אפשר להסיק כי:

$$\mathbb{P}\left(|X-\mathbb{E}\left[X\right]|\geq\lambda\right)\leq 2e^{-\frac{\lambda^2}{2n}}$$

דוגמה 5.34 הכד של פוליה - תיאור אלטרנטיבי:

מגרילים $X\sim U\left[0,1\right]$ ואחר כך מגרילים סדרה $\{Y_n\}\in\left[0,1\right]^{\mathbb{N}}$ באמצעות הגרלה של $Y_n\sim U\left[0,1\right]$ ב"ת. נגדיר:

$$a_n=\begin{cases}\text{black ball} & Y_n\leq X\\ \text{white ball} & Y_n>X\end{cases}$$

אפשר להראות שבהנתן X הסדרה $\{a_n\}$ היא סדרה ב"ת. נגדיר $\mathcal{F}_n=\sigma\left(a_1,...,a_n\right)$ ונגדיר מרטינגל ע"י $X_n=\mathbb{E}\left[X|\mathcal{F}_n\right]$.

תרגיל: המרטינגל הנ"ל זהה למרטינגל של כפרופורציית הכדורים השחורים בתיאור המקורי של הכד של פוליה. כלומר מתקיים:

$$X_n=\frac{1}{n+2}\left(\sum_{i=1}^{n}\mathbb{1}_{\{a_i=\text{black}\}}+1\right)$$

בנוסף המודל כולו זהה למודל הקודם של הכד של פולייה במובן שהפילוג של סדרת הצבעים של הכדורים שנגריל\נשלוף היא זהה.

הגדרה 5.35 תהליך סטוכסטי צפוי (previsible):

תהליך סטוכסטי $\{C_n\}$ יקרא **צפוי** ביחס ל- $\mathcal{F}_n$ אם C_n הוא $\mathcal{F}_{n-1}$-מדיד לכל n.

הגדרה 5.36 טרנספורם-מרטינגל (Martingale Transform):

יהא X_n מרטינגל ביחס לפילטרציה $\mathcal{F}_n$ ויהא C_n תהליך צפוי ביחס ל- $\mathcal{F}_n$. נגדיר תהליך $\{\left(C\bullet X\right)_n\}_{n=1}^{\infty}$ ע"י:

$$\left(C\bullet X\right)_n=\sum_{k=1}^{n}C_k\left(X_k-X_{k-1}\right)$$

התהליך הנ"ל מכונה ה- Martingale Transform.

הערה 5.37 למשל X_n יכול להיות מרטינגל כלשהו שמייצג משחק ו- C_n יכול להיות "הימור" על הערך של X_n.

טענה 5.38 טרנספורם-מרטינגל של תהליך צפוי וחסום הוא מרטינגל:

אם C_n צפוי וחסום אז התהליך $Y_n := (C \bullet X)_n$ הוא מרטינגל שמתחיל באפס.

הוכחה: נשתמש בתכונות של תוחלת מותנית ונקבל:

$$\mathbb{E}[Y_{n+1}|\mathcal{F}_n] = \mathbb{E}[Y_n + C_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] = \mathbb{E}[Y_n|\mathcal{F}_n] + \mathbb{E}[C_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n]$$

$$\overbrace{=}^{Y_n \text{ is } \mathcal{F}_n-\text{measurable}} Y_n + \mathbb{E}[C_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] \overbrace{=}^{C_n \text{ is } \mathcal{F}_n-\text{measurable}} Y_n + C_{n+1}\mathbb{E}[X_{n+1} - X_n|\mathcal{F}_n]$$

$$= Y_n + C_{n+1}(\mathbb{E}[X_{n+1}|\mathcal{F}_n] - \mathbb{E}[X_n|\mathcal{F}_n]) \overbrace{=}^{X_n=\text{martingale}} Y_n + C_{n+1}(X_n + \mathbb{E}[X_n|\mathcal{F}_n])$$

$$\overbrace{=}^{X_n \text{ is } \mathcal{F}_n-\text{measurable}} Y_n + C_{n+1}(X_n - X_n) = Y_n$$

■

הערה 5.39 נשים לב שכאשר X_n הוא מרטינגל עבור $m < n$ מתקיים:

$$\mathbb{E}[X_n|\mathcal{F}_m] \overbrace{=}^{\mathcal{F}_m \subseteq \mathcal{F}_{n-1}} \mathbb{E}[\mathbb{E}[X_n|\mathcal{F}_{n-1}]|\mathcal{F}_m] \overbrace{=}^{\mathbb{E}[X_n|\mathcal{F}_{n-1}]=\text{martingale}} \mathbb{E}[X_{n-1}|\mathcal{F}_m] = \ldots = \mathbb{E}[X_{m+1}|\mathcal{F}_m] = X_m$$

כלומר תוחלת הערך של X_n בהנתן שראינו עד השלב ה- m הוא פשוט הערך של X_m.

5.4 זמני עצירה:

הגדרה 5.40 זמן עצירה ביחס לפילטרציה:
בהנתן מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ ופילטרציה $\mathcal{F}_n$ משתנה מקרי $T : \Omega \to \mathbb{N} \cup \{\infty\}$ יקרא **זמן עצירה** אם $\{T \le n\} \in \mathcal{F}_n$ לכל n.

הערה 5.41 אפשר להראות שמספיק לדרוש ש- $\{T = n\} \in \mathcal{F}_n$.

הערה 5.42 לחילופין אפשר להגדיר זמן עצירה בכך שנקח סדרת פונקציות $f_n : \Omega \to \{0, 1\}$ שהינה $\mathcal{F}_n$-מותאמת ונגדיר:

$$T = \min\{n \mid f_n = 1\}$$

למה 5.43 מינימום ומקסימום של זמני עצירה הם זמן עצירה (ללא הוכחה):

אם T_1, T_2 הם זמני עצירה ביחס לפילטרציה $\mathcal{F}_n$ אז $\max\{T_1, T_2\}$ ו- $\min\{T_1, T_2\}$ הם גם כן זמני עצירה.

הגדרה 5.44 בהנתן X_n מרטינגל ו- T זמן עצירה (שניהם ביחס לפילטרציה $\mathcal{F}_n$) נגדיר את התהליך "שעוצר בזמן T" ע"י:

$$X_{T\wedge n} := X_{\min(T,n)}$$

כלומר כאשר $T \ge n$ מתקיים $X_{T\wedge n} = X_n$ וכאשר $T \le n$ מתקיים $X_{T\wedge n} = X_T$.

טענה 5.45

התהליך $X_{T\wedge n} := X_{\min(T,n)}$ הוא מרטינגל.

הוכחה: נגדיר $C_n = \mathbb{1}_{\{T \ge n\}} = 1 - \mathbb{1}_{\{T \le n-1\}}$, ברור שזהו תהליך צפוי וחסום ולכן $(C \bullet X)_n$ הוא מרטינגל כפי שראינו. אולם ניתן לראות כי:

$$(C \bullet X)_n = \ldots = X_{T\wedge n}$$

■

למה 5.46

אם T הוא זמן עצירה כך ש־ $\infty \overset{\text{a.s}}{>} T$ אזי מתקיים $X_T \xrightarrow[n\to\infty]{\text{a.s}} X_{\min(T,n)}$.

הוכחה: מכך ש־ $\infty \overset{\text{a.s}}{>} T$ נקבל שכ"ת קיים N כך שלכל $n \geq N$ מתקיים $X_{\min(T,n)} = X_T$. ■

משפט 5.47 משפט ־ *Doob's Optional Stopping Theorem* (*OST*):

יהיו X_n מרטינגל ו־ T זמן עצירה ביחס לפילטרציה $\mathcal{F}_n$ כך שמתקיים אחד מהתנאים הבאים:

1. $K \overset{a.s}{\geq} T$ (כלומר $\exists K\, \mathbb{P}(T \leq K) = 1$).
2. $\infty \overset{\text{a.s}}{>} T$ ו־ $X_{\min(T,n)}$ חסום יוניפורמית כ"ת (כלומר $\exists K\, \forall n\, \mathbb{P}\left(\left|X_{\min(T,n)}\right| \leq K\right) = 1$).
3. $\mathbb{E}[T] < \infty$ וההפרשים $\left|X_{\min(T,n+1)} - X_{\min(T,n)}\right|$ חסומים יוניפורמית כ"ת.

אזי X_T מוגדר היטב כמעט תמיד ומתקיים $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.

הערה 5.48 דוגמה נגדית פשוטה לזמן עצירה כך ש־ $\infty \overset{\text{a.s}}{>} T$ אולם $X_{\min(T,n)}$ לא חסום יוניפורמית הוא מהלך מקרי פשוט שעוצר כאשר מגיעים חזרה לאפס (המאורע הזה מתרחש בהסתברות 1).

הוכחה:

1. מכך ש־ $K \overset{\text{a.s}}{\geq} T$ נקבל ש־ $X_{\min(T,K)} \overset{\text{a.s}}{=} X_T$ ומכך:

$$\mathbb{E}[X_T] = \mathbb{E}\left[X_{\min(T,K)}\right]$$

כעת מאחר ש־ $X_{\min(T,n)}$ הוא מרטינגל נקבל כפי שראינו בהערה 5.39 שמתקיים $\mathbb{E}\left[X_{\min(T,n)}|\mathcal{F}_0\right] = X_0$ ומכך:

$$\mathbb{E}\left[X_{\min(T,K)}\right] = \mathbb{E}\left[\mathbb{E}\left[X_{\min(T,n)}|\mathcal{F}_0\right]\right] = \mathbb{E}[X_0]$$

לכן סה"כ מתקבל:

$$\mathbb{E}[X_T] = \mathbb{E}\left[X_{\min(T,K)}\right] = \mathbb{E}[X_0]$$

2. ראינו בלמה 5.46 שכאשר $\infty \overset{\text{a.s}}{>} T$ מתקיים $X_T \xrightarrow[n\to\infty]{\text{a.s}} X_{\min(T,n)}$. לכן מאחר והנחנו ש־ $X_{\min(T,n)}$ חסום יוניפרומית נוכל להשתמש במשפט ההתכנסות החסומה ונקבל כי:

$$\mathbb{E}[X_0] = \mathbb{E}\left[X_{\min(T,n)}\right] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X_T] \Longrightarrow \mathbb{E}[X_T] = \mathbb{E}[X_0]$$

3. מההנחה ש־ $K \overset{\text{a.s}}{\geq} |X_{n+1} - X_n|$ נקבל כי:

$$\left|X_{\min(T,n)} - X_0\right| \leq \sum_{i=1}^{\min(T,n)} |X_i - X_{i-1}| \leq K \cdot \min(T,n) \leq K \cdot T$$

לכן $\left|X_{\min(T,n)} - X_0\right|$ חסום ע"י מ"מ אינטגרבילי $K \cdot T$ (שכן הנחנו ש־ $\mathbb{E}[T] < \infty$) וממשפט ההתכנסות הנשלטת נקבל כי:

$$\mathbb{E}\left[\left|X_{\min(T,n)} - X_0\right|\right] \overset{n\to\infty}{\longrightarrow} 0 \Longrightarrow \mathbb{E}\left[X_{\min(T,n)}\right] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X_0]$$

מצד שני מאחר ש־ $\mathbb{E}[T] < \infty$ בהכרח $\mathbb{P}(T < \infty) = 1$ ולכן:

$$\mathbb{E}\left[X_{\min(T,n)}\right] \overset{n\to\infty}{\longrightarrow} \mathbb{E}[X_T]$$

מכך נסיק כי $\mathbb{E}[X_T] = \mathbb{E}[X_0]$, כנדרש.

■

5.5 על\תת־מרטינגלים:

הגדרה 5.49 על\תת־מרטינגל:
תהליך סטוכסטי X_n יקרא:

1. על־מרטינגל ביחס לפילטרציה $\mathcal{F}_n$ אם $\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \leq X_n$ לכל n.
2. תת־מרטינגל ביחס לפילטרציה $\mathcal{F}_n$ אם $\mathbb{E}\left[X_{n+1}|\mathcal{F}_n\right] \geq X_n$ לכל n.

הערה 5.50 נשים לב שאם X_n הוא על־מרטינגל אז לכל $n < m$ מתקיים:

$$\mathbb{E}\left[X_n|\mathcal{F}_m\right] = \mathbb{E}\left[\overbrace{\mathbb{E}\left[X_n|\mathcal{F}_{n-1}\right]}^{\leq X_{n-1}}|\mathcal{F}_m\right] \leq \ldots \leq X_m$$

טענה 5.51

אם X_n הוא על־מרטינגל ו־ $0 \leq C_n \leq K$ הוא תהליך חסום אז התהליך $Y_n := \left(C \bullet X\right)_n$ הוא על־מרטינגל.

הוכחה: אותה הוכחה כמו במקרה של מרטינגל רגיל. ■

מסקנה 5.52

אם T הוא זמן עצירה ו־ X_n הוא על־מרטינגל אז $X_{\min(T,n)}$ הוא על־מרטינגל.

הוכחה: אותה הוכחה כמו במקרה של מרטינגל רגיל. ■

מסקנה 5.53 משפט OST עבור על־מרטינגלים:

אם T הוא זמן עצירה ו־ X_n הוא על־מרטינגל אזי:

- התנאים של משפט OST גוררים ש־ $\mathbb{E}\left[X_T\right] \leq \mathbb{E}\left[X_0\right]$.
- במקום תנאי 2 מספיק לדרוש ש־ $T \overset{a.s}{<} \infty$ ו־ $X_n \overset{a.s}{\geq} 0$ לכל n.

הוכחה: נוכיח את החיזוק:

$$\mathbb{E}\left[X_T\right] = \mathbb{E}\left[\lim_{n\to\infty} X_{\min(T,n)}\right] = \mathbb{E}\left[\liminf_{n\to\infty} X_{\min(T,n)}\right] \overbrace{\leq}^{\text{Fatou}} \liminf_{n\to\infty} \mathbb{E}\left[X_{\min(T,n)}\right] \leq \mathbb{E}\left[X_0\right]$$

כאשר המעבר האחרון נובע מכך שלכל n מתקיים $\mathbb{E}\left[X_{\min(T,n)}\right] \leq \mathbb{E}\left[X_0\right]$ מאחר ש־ $X_{\min(T,n)}$ הוא על־מרטינגל. ■

הערה 5.54 כל הדברים האלה נכונים באופן סימטרי בשינוי מתאים של התנאים עבור תת־מרטינגלים.

5.6 מרטינגלים וריכוז מידה:

נזכיר שהחסמים שפיתחנו בנושא הקודם השתמשו באי־תלות רק כדי לחסום את הפונקציה יוצרת מומנטים של $S_n = \sum_{i=1}^{n} X_n$ באמצעות העובדה שהפונקציה הנ"ל שווה למכפלת היוצרות של $X_i, \ldots, X_n$. באופן כללי יותר גם במקרה שקיימת תלות אם היינו יודעים את התוחלות המותנות של כל X_n בהנתן $X_1, \ldots, X_{n-1}$ היינו יכולים לקבל פירוק דומה ואולי להשיג חסם דומה. המשפט הבא נותן תנאי שבו אכן ניתן לקבל תוצאה כזו:

משפט 5.55 אי־שוויון הופדינג־אזומה:

אם $\{M_i\}_{i=1}^{\infty}$ הוא על־מרטינגל כך ש $|M_{i+1}-M_i|\le K$ כ"ת. אז לכל $k\in\mathbb{N}$ ולכל $\lambda>0$ מתקיים:

$$\mathbb{P}\left(M_k-M_0\ge\lambda\right)\le e^{-\frac{\lambda^2}{2k}}$$

אם $\{M_i\}_{i=1}^{\infty}$ הוא תת־מרטינגל כך ש $|M_{i+1}-M_i|\le K$ כ"ת. אז לכל $k\in\mathbb{N}$ ולכל $\lambda>0$ מתקיים:

$$\mathbb{P}\left(M_k-M_0\le-\lambda\right)\le e^{-\frac{\lambda^2}{2kK^2}}$$

כתוצאה אם $\{M_i\}_{i=1}^{\infty}$ הוא מרטינגל כך ש $|M_{i+1}-M_i|\le K$ כ"ת מתקבל חסם דו־צדדי מהצורה:

$$\mathbb{P}\left(\left|M_k-\mathbb{E}\left[M_k\right]\right|\ge\lambda\right)=\mathbb{P}\left(\left|M_k-M_0\right|\ge\lambda\right)\le 2e^{-\frac{\lambda^2}{2kK^2}}$$

כאשר $M_0=\mathbb{E}\left[M_k\right]$ מאחר ומדובר במרטינגל שהפרשיו חסומים כ"ת.

הערה 5.56 א"ש הופדינג הוא מקרה פרטי שכן עבור $|X_i|\le 1$ ב"ת $S_n=\sum_{i=1}^{n}X_i$ הוא מרטינגל שמקיים את התנאים.

5.7 דוגמאות משעשעות עם מרטינגלים:

דוגמה 5.57 מהלך מקרי פשוט ומאוזן:

נתבונן ב־ X_n שמתאר מהלך מקרי פשוט על $\mathbb{Z}$ שמתחיל ב־ a כאשר $0<a<b\in\mathbb{Z}$:

$$X_n=a+\sum_{k=1}^{n}Y_k\quad;\ Y_k\overset{\text{i.i.d}}{\sim}Y=\begin{cases}+1 & \frac{1}{2}\\ -1 & \frac{1}{2}\end{cases}$$

ראינו ש־ X_n הוא מרטינגל. נסמן ב־ $\mathcal{F}_n$ את הפילטרציה הטבעית ביחס ל־ X_n (שהיא גם הפילטרציה הטבעית ביחס לסדרה Y_n) ונגדיר:

$$T=\min\left\{n\,|\,X_n=0\ \vee\ X_n=b\right\}$$

זהו זמן עצירה שכן $T_c=\min\left\{n\,|\,X_n=c\right\}$ הוא זמן עצירה לכל $c\in\mathbb{Z}$ ובפרט מתקיים $T=\min\left\{T_0,T_b\right\}$.

- **נשאל מהי** $\mathbb{P}_a\left(T_b<T_0\right)$**:** מאחר ש־ $T\overset{\text{a.s}}{<}\infty$ ו־ $\left|X_{\min(T,n)}\right|\le b$ מתקיים תנאי 2 של משפט OST ומכך נקבל כי:

$$a=\mathbb{E}\left[X_0\right]\overbrace{=}^{\text{Doob}}\mathbb{E}\left[X_T\right]=b\cdot\mathbb{P}\left(T_b<T_0\right)+0\cdot\mathbb{P}\left(T_0\le T_b\right)=b\mathbb{P}\left(T_b<T_0\right)$$

לכן ניתן לראות כי $\mathbb{P}\left(T_b<T_0\right)=\frac{a}{b}$.

- **נשאל מהו** $\mathbb{E}\left[T\right]$ ונסיק ש־ $\mathbb{P}\left(T_0<\infty\right)=1$. ראשית נגדיר $M_n=X_n^2-n$ ונראה שזהו מרטינגל:

$$\begin{aligned}\mathbb{E}\left[X_{n+1}^2-(n+1)\,|\mathcal{F}_n\right]&=\mathbb{E}\left[X_{n+1}^2|\mathcal{F}_n\right]-(n+1)=\mathbb{E}\left[\left(X_n+Y_{n+1}\right)^2|\mathcal{F}_n\right]-(n+1)\\ =\mathbb{E}\left[X_n^2+2X_nY_n+Y_{n+1}^2|\mathcal{F}_n\right]-(n+1)&=\mathbb{E}\left[X_n^2|\mathcal{F}_n\right]+2\mathbb{E}\left[X_nY_{n+1}|\mathcal{F}_n\right]+\mathbb{E}\left[Y_{n+1}^2|\mathcal{F}_n\right]\\ &=X_n^2+2X_n\underbrace{\mathbb{E}\left[Y_{n+1}|\mathcal{F}_n\right]}_{=\mathbb{E}[Y_{n+1}]=0}+1-(n+1)=X_n^2-n\end{aligned}$$

כאשר השתמשנו בכך ש־ Y_{n+1} ב"ת ב־ $\mathcal{F}_n$. כמו כן M_n מקיים את תנאי (3) של משפט OST שכן:

$$\left|M_{\min(T,n+1)}-M_{\min(T,n)}\right|=\left|X_{\min(T,n+1)}^2-(n+1)-\left(X_{\min(T,n)}^2-n\right)\right|=\left|X_{\min(T,n+1)}^2-X_{\min(T,n)}^2-1\right|\le 1$$

מכך $\mathbb{E}\left[X_T^2-T\right]=\mathbb{E}\left[M_T\right]=\mathbb{E}\left[M_0\right]=a^2$. כמו כן ניתן לראות כי:

$$\mathbb{E}\left[X_T^2\right]=b^2\cdot\mathbb{P}\left(T_b<T_0\right)+0^2\cdot\mathbb{P}\left(T_0\le T_b\right)=b^2\mathbb{P}\left(T_b<T_0\right)$$

מכך נקבל כי:

$$\mathbb{E}\left[T\right]=\mathbb{E}\left[X_{T}^{2}\right]-\mathbb{E}\left[X_{T}^{2}-T\right]=b^{2}\overbrace{\mathbb{P}\left(T_{b}<T_{0}\right)}^{=\frac{a}{b}}-a^{2}=ab-a^{2}=a\left(b-a\right)$$

מכך שהתוחלת הזו סופית נסיק ש־ $\mathbb{P}\left(T<\infty\right)=1$. כמו כן

$$\mathbb{P}\left(T<\infty\right)\leq\mathbb{P}\left(T_{0}<\infty\right)+\mathbb{P}\left(T_{b}<T_{0}\right)$$

מכך ניתן לראות כי:

$$\mathbb{P}\left(T_{0}<\infty\right)\geq\mathbb{P}\left(T<\infty\right)-\mathbb{P}\left(T_{b}<T_{0}\right)=1-\frac{a}{b}$$

אולם הנ"ל נכון לכל b ומכך אפשר להסיק $\mathbb{P}\left(T_{0}<\infty\right)=1$. מכך אפשר למעשה להסיק שהתהליך יבקר בנקודה 0 (ובכל נקודה בין a ל־ b) אינסוף פעמים בהסתברות 1.

תרגיל: הוכח ש־ $\mathbb{E}_{a}\left[T_{b}|T_{b}<T_{0}\right]=\frac{b^{2}-a^{2}}{3}$ ע"י הוכחה ש־ $M_{n}:=X_{n}^{3}-3nX_{n}$ הוא מרטינגל ושימוש בכך ש־ $\left|M_{\min(T_{b},n)}\right|\leq b^{3}+3bT_{b}$ ובטריק דומה להוכחה של תנאי 3 במשפט OST.

דוגמה 5.58 מהלך מקרי לא מאוזן:

נשתמש באותם הגדרות וסימונים של הדוגמה הקודמת בשינוי יחיד:

$$X_{n}=a+\sum_{k=1}^{n}Y_{k}\quad;\,Y_{k}\overset{\text{i.i.d}}{\sim}Y=\begin{cases}+1 & p\\-1 & q=1-p\end{cases}$$

לא קשה לראות ש־ X_{n} איננו מרטינגל ומאידך $M_{n}:=X_{n}-n\left(p-q\right)$ הוא כן.

- **נשאל מהו** $\mathbb{P}_{a}\left(T_{b}<T_{0}\right)$**:** התהליך M_{n} לא יועיל לנו עבור החישוב הנ"ל ולכן נגדיר תהליך $N_{n}=\left(\frac{q}{p}\right)^{X_{n}}$, זהו מרטינגל שכן:

$$\mathbb{E}\left[N_{n+1}|\mathcal{F}_{n}\right]=\mathbb{E}\left[\left(\frac{q}{p}\right)^{X_{n+1}}|\mathcal{F}_{n}\right]=\mathbb{E}\left[\left(\frac{q}{p}\right)^{X_{n}+Y_{n+1}}|\mathcal{F}_{n}\right]=\mathbb{E}\left[\left(\frac{q}{p}\right)^{X_{n}}\cdot\left(\frac{q}{p}\right)^{Y_{n+1}}|\mathcal{F}_{n}\right]$$

$$=\left(\frac{q}{p}\right)^{X_{n}}\mathbb{E}\left[\left(\frac{q}{p}\right)^{Y_{n+1}}|\mathcal{F}_{n}\right]=\left(\frac{q}{p}\right)^{X_{n}}\mathbb{E}\left[\left(\frac{q}{p}\right)^{Y_{n+1}}\right]=\left(\frac{q}{p}\right)^{X_{n}}\left(p\frac{q}{p}+q\frac{p}{q}\right)=\left(\frac{q}{p}\right)^{X_{n}}\left(p+q\right)=\left(\frac{q}{p}\right)^{X_{n}}=N_{n}$$

כמו כן $N_{\min(T,n)}$ חסום יוניפורמית כ"ת ולכן ממשפט OST נקבל כי:

$$\left(\frac{p}{q}\right)^{a}=\mathbb{E}\left[N_{0}\right]\overset{\text{OST}}{\overbrace{=}}\mathbb{E}\left[N_{T}\right]=\mathbb{E}\left[\left(\frac{p}{q}\right)^{X_{T}}\right]=\mathbb{P}\left(X_{T}=b\right)\left(\frac{p}{q}\right)^{b}+\mathbb{P}\left(X_{T}=0\right)\left(\frac{p}{q}\right)^{0}=\underbrace{\mathbb{P}\left(T_{b}<T_{0}\right)}_{p_{b}}\left(\frac{p}{q}\right)^{b}+\underbrace{\mathbb{P}\left(T_{0}<T_{b}\right)}_{1-p_{b}}$$

ומכך לסיכום:

$$\mathbb{P}\left(T_{b}<T_{0}\right)=\frac{\left(\frac{q}{p}\right)^{a}-1}{\left(\frac{q}{p}\right)^{b}-1}$$

תרגיל: להשתמש ב־ M_{n} כדי לחשב את $\mathbb{E}\left[T\right]$.

הערה 5.59 ניתן לראות שכאשר $p \neq q$ מספר הפעמים שנגיע לכל נקודה הוא סופי (כלומר התהליך הוא חולף). מאידך כאשר $p = \frac{1}{2}$ ההילוך יבקר בכל נקודה אינסוף פעמים.

דוגמה 5.60 ממתינים להופעת תבנית:
נתבונן ברצפים אינסופים המגיעים מ־ $\{A, ..., Z\}^{\mathbb{N}}$ ונדגמים עם התפלגות אחידה וב"ת להופעה של כל אות. נסמן ב־ L_i את האות ה־ i ברצף ונשאל מהי תוחלת הזמן להופעה של רצף אותיות מסוים.

- עבור רצף מאורך 1 ידוע לנו שזמן ההמתנה מתפלג גיאומטרית עם פרמטר $p = \frac{1}{26}$ ולכן התוחלת היא 26.
- בפרט עבור האות "I" נראה זאת גם באמצעות מרטינגל, נגדיר $Y_n = 26\mathbb{1}_{\{L_n = I\}}$ ונגדיר מרטינגל עם הפרשים חסומים ע"י $X_n = \sum_{k=1}^{n} Y_k - n$. כמו כן נגדיר זמן עצירה $T = \min\{n \mid L_n = I\}$, הנ"ל מקיים $\mathbb{E}[T] < \infty$ ולכן ממשפט OST נקבל:

$$0 = \mathbb{E}[X_0] \overbrace{=}^{\text{OST}} \mathbb{E}[X_T] = \mathbb{E}[26 - T] \Longrightarrow \mathbb{E}[T] = 26$$

- כעת עבור מילה כלשהי $\lambda_1\lambda_2...\lambda_k$ נגדיר:

$$Y_n^m = \begin{cases} 26^{n+1-m} & \text{if } n \geq m \wedge n - m \leq k \wedge L_m...L_n = \lambda_1...\lambda_{n+1-m} \\ 26^k & \text{if } n \geq m \wedge n - m > k \wedge L_m...L_{m+k} = \lambda_1....\lambda_k \\ 0 & \text{Otherwise} \end{cases}$$

נטען שעבור m כלשהו התהליך:

$$Y_n^m - \mathbb{1}_{\{n \geq m\}}$$

הא מרטינגל ביחס לפילטרציה $\mathcal{F}_n = \sigma(L_1...L_n)$. כמו כן סכום סופי של מרטינגלים הוא גם כן מרטינגל ולכן:

$$X_n = \sum_{m=1}^{\infty} Y_n^m - n = \sum_{m=1}^{n} Y_n^m + \sum_{m=n+1}^{\infty} \overbrace{Y_n^m}^{=0} - n = \sum_{m=1}^{n} Y_n^m - m$$

הוא גם כן מרטינגל. נגדיר זמן עצירה על ידי:

$$T = \min\{n \mid L_{n-k}...L_n = \lambda_1...\lambda_k\}$$

הנ"ל מקיים $\mathbb{E}[T] < \infty$ וכמו כן $X_{\min(T,n)}$ הוא בעל הפרשים חסומים יוניפורמית ולכן ממשפט OST:

$$0 = \mathbb{E}[X_0] = \mathbb{E}[X_T] = \mathbb{E}\left[\sum_{m=1}^{T} Y_T^m - T\right] \Longrightarrow \mathbb{E}[T] = \sum_{m=1}^{T} \mathbb{E}[Y_T^m] = \sum_{i=1}^{T} \mathbb{E}\left[Y_T^{T-i}\right]$$

מההגדרה $\mathbb{E}\left[Y_T^{T-i}\right] = 0$ כאשר $i > k$ וכאשר $i \leq k$ נקבל כי:

$$Y_T^{T-i} = \begin{cases} 26^{i+1} & \lambda_1...\lambda_i = \lambda_{k-i+1}...\lambda_k \\ 0 & \text{otherwise} \end{cases}$$

לכן סה"כ נקבל כי:

$$\mathbb{E}[T] = \sum_{i=1}^{k} 26^{i+1}\mathbb{1}_{\{\lambda_1...\lambda_i = \lambda_{k-i+1}...\lambda_k\}} = 26^k + \sum_{i=1}^{k-1} 26^{i+1}\mathbb{1}_{\{\lambda_1...\lambda_i = \lambda_{k-i+1}...\lambda_k\}}$$

הערה 5.61 כדי להראות ש־ $\mathbb{E}[T] < \infty$ מספיק להבחין שיש פילוג גיאומטרי על הזמן עד שאחד מהרצפים הב"ת מאורך k (כלומר $L_1....L_k$, $L_{k+1}....L_{2k+1}$ וכן הלאה) יהיה זהה למילה $\lambda_1...\lambda_k$. לכן מאחר שתוחלת הזמן עד שרצף כלשהו מאורך k יהיה זהה למילה המבוקשת קטנה מתוחלת הזמן (הסופית) עד שרצף בלתי תלוי יהיה שווה למילה נסיק ש־ $\mathbb{E}[T] < \infty$.

5.8 התכנסות של מרטינגלים:

נרצה להוכיח את המשפט הבא:

משפט 5.62 מרטינגל חסום מתכנס כמעט תמיד:

יהא X_n מרטינגל חסום ($\exists K\,:\,\forall n\,\mathbb{P}\left(\left|X_n\right|\leq K\right)=1$) אזי קיים מ"מ Y כך ש־ $Y\xrightarrow{\text{a.s}}X_n$.

הערה 5.63 יש תנאים הרבה יותר חלשים להתכנסות של מרטינגלים כגון מומנט שני חסום.

הוכחה: נוכיח טענה יותר חזקה לפיה על־מרטינגל אי־שלילי מתכנס כ"ת:

למה 5.64

יהא X_n על־מרטינגל אי־שלילי כך ש־ $X_0\equiv a>0$. בהנתן $b>a$ נגדיר $T_b=\min\left\{n\,|\,X_n\geq b\right\}$ אזי $\mathbb{P}_a\left(T_b<\infty\right)\leq\frac{a}{b}$.

הוכחה: בהנתן $N\in\mathbb{N}$ נתבונן ב־ $X_{\min(T_b,n)}$ שהינו על־מרטינגל אי־שלילי עבור $n=0,...,N$. עבור הנ"ל מתקיים:

$$b\mathbb{P}\left(T_b\leq N\right)\leq\mathbb{E}\left[X_{\min(T_b,N)}\right]\leq\mathbb{E}\left[X_0\right]=a$$

כאשר אי־השוויון השמאלי נובע מכך ש־ $T_b\leq N$ אמ"מ $X_{\min(T_b,N)}\geq b$ ומכך ש־ $X_n\geq0$ () לכן $\mathbb{P}\left(T_b\leq N\right)\leq\frac{a}{b}$ לכל $N\in\mathbb{N}$ ולכן $\mathbb{P}_a\left(T_b<\infty\right)\leq\frac{a}{b}$ כנדרש. ■

על סמך הלמה הזו נסיק ש־ X_n בהסתברות 1 לא שואף לאינסוף שכן כאשר $b\to\infty$ מתקבל ש־ $\mathbb{P}_a\left(T_b<\infty\right)\to0$. לכן כדי להראות ש־ X_n מתכנס מספיק להראות שהוא לא עושה פלקטואציות. עבור $0<a<b$ כלשהם נגדיר את מספר החציות מעלה (upcreasing) של $[a,b]$ על ידי:

$$u\left(a,b\right)=\max\left\{n\,|\,\exists\,t_1<s_1<t_2<...<s_n\,\text{s.t}\,\forall\,i=1,...,n\;X_{t_i}\leq a,\,X_{s_i}\geq b\right\}$$

טענה 5.65

מתקיים $\mathbb{P}\left(u\left(a,b\right)\geq k\right)\leq\left(\frac{a}{b}\right)^k$.

הוכחה: באינדוקציה. עבור $k=1$ המסקנה היא תוצאה של הלמה הקודמת התרחשה חצייה אמ"מ $T_b<\infty$ והנ"ל מתרחש בהסתברות הקטנה מ־ $\frac{a}{b}$. כעת בהנתן ש־ $u\left(a,b\right)\geq k$ אם המאורע שירדנו שוב מתחת ל־ a התרחש אז ההסתברות שעלינו מחדש מעל b חסומה שוב על סמך הלמה הקודמת על ידי $\left(\frac{a}{b}\right)$ ולכן מתקבלת המסקנה הנדרשת. ■

על סמך הטענה הזו $u\left(a,b\right)$ סופי כ"ת לכל $a<b$ שכן ההסתברות ש־ $u\left(a,b\right)\geq k$ שואפת אקספוננציאלית לאפס עם k. מכך נסיק ש־ X_n מתכנס כ"ת שכן אם הוא לא היה מתכנס אז למשל עבור $a=\liminf X_n<\limsup X_n=b$ היינו מקבלים שלכל $a<c<d<b$ המרטינגל חייב לחצות את (c,d) אינסוף פעמים אולם כפי שראינו הנ"ל לא קורה. לכן בהכרח $\liminf X_n=\limsup X_n$ ויש התכנסות כ"ת כנדרש.

■

דוגמה 5.66 הכד של פולייה: אם X_n הוא המרטינגל שמוגדר ע"י פרופורציית הכדורים השחורים אז מהיותו חסום המשפט הקודם נותן לנו את זה ש־ $X\xrightarrow{\text{a.s}}X_n$ למ"מ X כלשהו. בפרט מתקבל ש־ $X\sim U\left[0,1\right]$.

דוגמה 5.67 יהא X מ"מ חסום ותהא $\mathcal{F}_n$ פילטרציה. נתבונן במרטינגל $X_n=\mathbb{E}\left[X|\mathcal{F}_n\right]$, זהו מרטינגל חסום ולכן קיים $\lim\limits_{n\to\infty}X_n$ כ"ת על סמך המשפט הקודם. נשאלת השאלה האם $\lim\limits_{n\to\infty}X_n=X$. התשובה היא "לא בהכרח" אולם בתנאי מסוים הנ"ל כן מתקיים כמו שמעיד המשפט הבא:

משפט 5.68 משפט - *Levy's Upward Theorem*:

יהא X מ"מ חסום ותהא $\mathcal{F}_n$ פילטרציה אזי $\lim\limits_{n\to\infty}\mathbb{E}\left[X|\mathcal{F}_n\right]=\mathbb{E}\left[X|\mathcal{F}_\infty\right]$ כאשר $\mathcal{F}_\infty=\sigma\left(\bigcup_{n\in\mathbb{N}}\mathcal{F}_n\right)$.

הוכחה: נסמן $X_n=\mathbb{E}\left[X|\mathcal{F}_n\right]$ ו־ $X_{\infty}=\lim\limits_{n\to\infty}X_n$ (קיים שכן X_n הוא מרטינגל חסום). בה"כ אפשר להניח ש־ X $\mathcal{F}_{\infty}$-מדיד שכן אחרת נקח $X=\mathbb{E}\left[X|\mathcal{F}_{\infty}\right]$ ונקבל ש־ X_n נשארים אותו דבר לפי תכונת המגדל. כעת בהנתן $n\in\mathbb{N}$ ו־ $A\in\mathcal{F}_n$ נקבל כי:

$$\mathbb{E}\left[X_{\infty}\mathbb{1}_A\right]\overbrace{=}^{(1)}\mathbb{E}\left[X_n\mathbb{1}_A\right]\overbrace{=}^{(2)}\mathbb{E}\left[X\mathbb{1}_A\right]$$

המעבר השני נובע פשוט מהגדרת התוחלת המותנית ומכך ש־ $X_n=\mathbb{E}\left[X|\mathcal{F}_n\right]$ והמעבר הראשון נובע מכך שאם נסתכל על $X_m\mathbb{1}_A$ עבור $m=n,n+1,...$ נקבל כי זהו מרטינגל חסום המתכנס כ"ת ל־ $X_{\infty}\mathbb{1}_A$ ולכן ממשפט ההתכנסות החסומה נובעת גם התכנסות של התוחלות. בכך הראינו ש־ X_{∞} ו־ X "מסכימים" בתוחלת עבור כל $A\in\bigcup_{n\in\mathbb{N}}\mathcal{F}_n$ שהיא קבוצה סגורה לחיתוכים. העובדה הזו מספיקה כדי להסיק ש־ $X_{\infty}\overset{\text{a.s}}{=}X$ על סמך הלמה הבאה שאותה לא נוכיח:

למה 5.69

יהיו X,Y מ"מ חסופים ותהא $\mathcal{S}$ קבוצה סגורה לחיתוכים היוצרת את הסיגמה אלגברה $\mathcal{F}$ אזי $Y\overset{\text{a.s}}{=}X$ אמ"מ:

$$\forall A\in\mathcal{S}\,:\,\mathbb{E}\left[X\mathbb{1}_A\right]=\mathbb{E}\left[Y\mathbb{1}_A\right]$$

■

דוגמה 5.70 נגדיר את המשתנים המקריים הבאים:

$$X=\begin{cases}0 & \frac{1}{2}\\ 1 & \frac{1}{2}\end{cases}\,,\,Y_n=\begin{cases}1 & p_n\\ 0 & 1-p_n\end{cases}$$

כמו כן נגדיר $Z_n=X+Y_n \mod 2$. אפשר לחשוב על המודל הנ"ל כעל מצב שבו X מייצג ביט של אינפורמציה ו־ Y_n מייצג רעש מקרי שמפריע לנו לקלוט את האינפורמציה הרלוונטית כאשר Z_n הוא הקלט שמתקבל. נגדיר $\mathcal{F}_n=\sigma\left(Z_1,...,Z_n\right)$ ו־ $X_n=\mathbb{E}\left[X|\mathcal{F}_n\right]$ (כלומר X_n הוא הצפייה שלנו ביחס לערך של X בהנתן הקלט שהתקבל ב־ n המקרים הראשונים).

תרגיל: אם $p_n\equiv p\neq\frac{1}{2}$ אז $X_n\overset{\text{a.s}}{\longrightarrow}X$ והנ"ל לא קורה אם p_n מתכנס מהר ל־ $\frac{1}{2}$ (באופן מדויק $X_n\overset{\text{a.s}}{\longrightarrow}X$ אמ"מ $\sum\left(p_n-\frac{1}{2}\right)^2=\infty$).

6 שרשראות מרקוב:

ספר מומלץ - Markov Chains and Mixing Times by Yuval Peres.

הגדרה 6.1 שרשרת מרקוב (בזמן בדיד):
סדרת משתנים מקריים בדידים $X_n : \Omega \to \mathcal{S}$ המקבלים ערכים בקבוצה $\mathcal{S}$ תקרא **שרשרת מרקוב** אם ההתפלגות של X_{n+1} בהנתן $X_0, ..., X_n$ היא פונקציה של X_n בלבד במובן שקיימת מטריצה\פונקציה $P : \mathcal{S} \times \mathcal{S} \to [0,1]$ כך ש:

$$\mathbb{P}\left(X_{n+1} = y \,|\, X_0 = a_0, X_1 = a_1, ..., X_n = x\right) = P\left(x, y\right)$$

התכונה הזו מכונה **תכונת מרקוב**.

הערה 6.2 הקבוצה $\mathcal{S}$ מכונה מרחב המצבים ו- P מכונה מטריצת\פונקציית המעבר ולמעשה הזוג $(\mathcal{S}, P)$ מאפיין את השרשרת בלי צורך להתייחס להתפלגויות של המשתנים המקריים X_n עצמם.

הערה 6.3 אפשר לראות ש- P היא **מטריצה סטוכסטית** כלומר:

$$\forall\, x, y \in \mathcal{S}\, P\left(x, y\right) \geq 0 \,\wedge\, \forall x \in \mathcal{S} \,:\, \sum_{y \in \mathcal{S}} P\left(x, y\right) = 1$$

הערה 6.4 כרגע אנחנו עוסקים בשרשראות מרקוב עם מרחב מצבים סופי\בן־מנייה וסט אינדקסים בן־מנייה, הנ"ל מכונות שרשראות מרקוב עם מרחב מצבים דיסקרטי בזמן דיסקרטי. יש גם הכללה לשרשראות מרקוב עם קבוצת אינדקסים שאיננה בת מנייה אשר מכונות שרשראות מרקוב בזמן רציף.

דוגמה 6.5 דוגמאות פשוטות:

1. סדרת מ"מ בלתי תלויים היא מיידית שרשרת מרקוב.
2. סדרת הסכומים\מכפלות החלקיים של סדרת מ"מ בלתי תלויים. למעשה הנ"ל נכון עבור כל פעולה מצטברת ולמשל עבור הרכבת פרמוטציות כאשר X_n הן פרמוטציות ב"ת על N איברים ונגדיר $S_n = X_0 \circ X_1 \circ ... \circ X_n$.

דוגמה 6.6 נשאל האם הכד של פולייה הוא שרשרת מרקוב:

- אם מסתכלים על תיאור הבעיה כך ש- X_n מכיל בכל שלב את סדרת הצבעים שנשלפו עד כה אז כמובן שמתקבלת שרשרת מרקוב.
- אם נסתכל על תיאור הבעיה כך ש- X_n מכיל בכל שלב כמה כדורים יש מכל צבע ($X_n = (\#\text{black}, \#\text{white})$) או כך ש- X_n היא פרופורציית הכדורים השחורים שנשלפו אז לא מתקבלת שרשרת מרקוב.

הערה 6.7 לפעמים מסתכלים על שרשראות מרקוב שבהן מטריצת המעבר P יכולה להיות תלויה ב- n (מכונות שרשראות מרקוה לא אחידות בזמן) ובמקרה הנ"ל גם התיאור השני מהווה שרשרת מרקוב.

דוגמה 6.8 צפרדע יושבת על אחד משני עלים W, E ובכל יום היא בוחרת האם לעבור מהעלה שהיא נמצאת בו לעלה השני כפי שמתואר במטריצת המעברים הבאה:

$P =$

	E	W
E	$1-p$	p
W	q	$1-q$

נתבונן בסדרת המשתנים המקריים $\{X_t\}_{t=0}^{\infty} \in \{E,W\}^{\mathbb{N}}$ אשר מתארת את מיקום הצפרדע בזמן t. נניח שבזמן t ההתפלגות של הצפרדע להיות במיקום מסוים נתונה על ידי:

$$\mu_t = (\mu_t(E), \mu_t(W)) = (\mathbb{P}(X_t = E), \mathbb{P}(X_t = W))$$

נשים לב שמנוסחת ההסתברות השלמה מתקיים:

$$\begin{aligned}
\mu_{t+1}(E) = \mathbb{P}(X_{t+1} = E) &= \sum_{a_0,a_1,\dots,a_t \in \{E,W\}} \mathbb{P}(X_0 = a_0, \dots, X_t = a_t, X_{t+1} = E) \\
&= \sum_{a_0,a_1,\dots,a_t \in \{E,W\}} \mathbb{P}(X_0 = a_0, \dots, X_t = E, X_{t+1} = E) + \sum_{a_0,a_1,\dots,a_t \in \{E,W\}} \mathbb{P}(X_0 = a_0, \dots, X_t = W, X_{t+1} = E) \\
&= \sum_{a_0,a_1,\dots,a_t \in \{E,W\}} \mathbb{P}(X_0 = a_0, \dots, X_t = E) \cdot \mathbb{P}(X_{t+1} = E | X_0 = a_0, \dots, X_t = E) \\
&+ \sum_{a_0,a_1,\dots,a_t \in \{E,W\}} \mathbb{P}(X_0 = a_0, \dots, X_t = W) \cdot \mathbb{P}(X_{t+1} = E | X_0 = a_0, \dots, X_t = E) \\
&= \mathbb{P}(X_t = E)\,\mathbb{P}(X_{t+1} = E | X_0 = a_0, \dots, X_t = E) + \mathbb{P}(X_t = W)\,\mathbb{P}(X_{t+1} = E | X_0 = a_0, \dots, X_t = E) \\
&\underbrace{=}_{\text{Markov Property}} \mu_t(E) \cdot P(E,E) + \mu_t(W) \cdot P(W,E) = \mu_t(E) \cdot (1-p) + \mu_t(W) \cdot q
\end{aligned}$$

מכך ניתן לראות שבזמן $t+1$ ההתפלגות כולה מתקבלת על ידי המכפלה:

$$\mu_{t+1} = (\mathbb{P}(X_{t+1} = E), \mathbb{P}(X_{t+1} = W)) = \begin{pmatrix} \mu_t(E)(1-p) + \mu_t(W)q \\ \mu_t(E)p + \mu_t(W)(1-q) \end{pmatrix}^{\top} = \mu_0 P$$

באופן כללי יותר נקבל באינדוקציה שההתפלגות בזמן $t+1$ נתונה על ידי:

$$\mu_{t+1} = \mu_t P = \mu_0 P^{t+1}$$

כעת נניח שסדרת הפילוגים μ_t מתכנסת לפילוג π כאשר $t \to \infty$. מכך ש־ $\mu_t P = \mu_{t+1}$ נסיק שגם $\mu_t P \overset{t\to\infty}{\longrightarrow} \pi$ ולכן מכך ש־ P היא פונקציה רציפה על מרחב קומפקטי (כי $\mathcal{S}$ סופית או בת־מנייה עם טופולוגיה דיסקרטית) נקבל כי $\pi = \pi P$.

הערה 6.9 בפרט נאמר במקרה כנ"ל ש־ π היא מידה סטציונרית וש־ μ_t מתכנסת למידה סטציונרית.

נשים לב שפתרון של המשוואה $\pi = \pi P$ מתקבל רק עבור π שנתונה על ידי $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ (פרט למקרה שבו $p = q = 0$ שבו כל וקטור הוא פתרון). אם כן הראינו שבמקרה שבו לא מתקיים $p = q = 0$ אז אם יש התכנסות בהכרח הגבול הוא $\pi = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$, נראה מתי באמת מתקיימת התכנסות כנ"ל. נסמן $\mu_t = (x_t, 1-x_t)$, נגדיר $\Delta_t = x_t - \frac{p}{p+q}$ ונבדוק מתי - $\Delta_t \overset{t\to\infty}{\longrightarrow} 0$. מאחר ש־ $\mu_{t+1} = \mu_t P$ נקבל כי:

$$x_{t+1} = x_t(1-p) + (1-x_t)q = x_t(1-p-q) + q$$

לכן נקבל כי:

$$\begin{aligned}
\Delta_{t+1} = x_{t+1} - \frac{q}{p+q} &= x_t(1-p-q) + q - \frac{q}{p+q} \\
= \Delta_t(1-p-q) + &\underbrace{\frac{q}{p+q}(1-p-q) + q - \frac{q}{p+q}}_{=0} = \Delta_t(1-p-q)
\end{aligned}$$

מאחר ש־ $\Delta_t \overset{t\to\infty}{\longrightarrow} 0$ אמ"מ $\mu_t \overset{t\to\infty}{\longrightarrow} \pi$ נסיק שמתקיימת התכנסות אלא אם $p = q = 0$ או $p = q = 1$, נבחן את המקרים הללו:

1. אם $p = q = 0$ אז $\mu_{t+1} = \mu_t = \dots = \mu_0$ אולם כל התפלגות היא סטציונרית שכן $P = I$.

2. אם $p=q=1$ אז $\Delta_t=\Delta_0\left(-1\right)^t$ ולכן אין התכנסות אלא אם μ_0 עצמה סטציונרית.

נגדיר כעת את המושגים מהדוגמה הקודמת בצורה פורמלית:

הגדרה 6.10 מידה סטציונרית:
בהנתן שרשרת מרקוב $(\mathcal{S},P)$ נאמר ש־ π היא מידה סטציונרית על $\mathcal{S}$ אם $\pi=\pi P$.

הערה 6.11 בפרט אם π היא מידת הסתברות נאמר ש־ π היא התפלגות סטציונרית.

משפט 6.12 לשרשרת מרקוב עם מרחב מצבים סופי קיימת התפלגות סטציונרית:

בהנתן שרשרת מרקוב $(\mathcal{S},P)$ כך ש־ $\mathcal{S}$ סופית קיימת התפלגות סטציונרית על $\mathcal{S}$.

הוכחה: הוכחה ראשונה: במקרה שבו $\mathcal{S}$ סופי הכפלה מימין ב־ P היא פונקציה רציפה מהסימפלקס (שקול טופולוגית לכדור):

$$\Delta\left(S\right)=\left\{\mu:S\to[0,1]\mid\sum_{s\in\mathcal{S}}\mu\left(s\right)=1\right\}$$

לעצמו. לכן על פי משפט נקודת השבת של ברואר יש לה נקודת שבת ולכן קיימת $\pi\in\Delta\left(S\right)$ כך ש־ $\pi=\pi P$.

הוכחה שנייה: נתבונן במידות מהצורה $\nu_n:=\frac{\mu_0}{n+1}\left(I+P+...+P^n\right)$. אם הסדרה הזו הייתה מתכנסת אז היינו מקבלים שהיא מתכנסת למידה סטציונרית שכן:

$$\nu_n\left(P-I\right)=\mu_0\left(\frac{P^{n+1}-I}{n+1}\right)\overset{n\to\infty}{\longrightarrow}\vec{0}$$

כאשר עושים שימוש בכך ש־ P^n הן מטריצות סטוכסטיות שערכיהן חסומים ע"י 1. בכל מקרה מדובר בסדרה חסומה ולכן קיימת לה תת־סדרה מתכנסת לאיזושהי מידה π וזוהי מידה סטציונרית שכן מרציפות של $(P-I)$ מתקבל ש:

$$\pi\left(P-1\right)=\lim_{n\to\infty}\nu_{n_k}\left(P-I\right)=0$$

הוכחה שלישית (חלקית): מאחר ש־ P היא מטריצה סטוכסטית 1 הוא ערך עצמי ימני שלה עם וקטור עצמי $\vec{1}$. לכן יש ל־ P וקטור עצמי שמאלי המקיים $\pi=\pi P$ אולם לא ברור באופן מיידי ש־ $\pi\in\Delta\left(\mathcal{S}\right)$. מאידך משפט פרון־פורבניוס נותן את התוצאה הזו.

■

תרגיל: נתבונן במהלך מקרי פשוט $\{X_t\}_t$ על גרף לא מכוון וסופי $G=(V,E)$ כאשר הערך של X_{t+1} נבחר בהתפלגות אחידה מבין השכנים של X_t. אזי $\pi\left(x\right)=\frac{\deg(x)}{2|E|}$ מהווה מידת הסתברות סטציונרית כאשר ערכי מטריצת המעברים נתונים על ידי:

$$P\left(x,y\right)=\frac{1}{\deg\left((x,y)\right)}\mathbb{1}_{\{x\sim y\}}$$

אם הגרף איננו סופי אך קשיר ועם דרגות קודקודים סופיות $\pi\left(x\right)=\deg\left(x\right)$ היא מידה סטציונרית אך איננה מידה סופית.

דוגמה 6.13 נתבונן בשרשרת מרקוב עם מרחב מצבים $\mathbb{N}$ ופונקציית מעברים $P\left(n,n+1\right)=1$. במקרה הנ"ל לא קיימת התפלגות סטציונרית שכן אם הייתה קיימת התפלגות אז היה מתקיים $\pi\left(0\right)=\pi P\left(0\right)=0$ וכנ"ל $\pi\left(1\right)=\pi P^2\left(1\right)=0$ וכן הלאה $\pi\left(n\right)=\pi P^n\left(n\right)=0$ ומכך $\pi\equiv0$ וזוהי כמובן איננה מידת הסתברות.

הגדרה 6.14 שרשרת מרקוב אי־פריקה:
שרשרת מרקוב $(\mathcal{S},P)$ תקרא אי־פריקה אם לכל $x,y\in\mathcal{S}$ קיים $n\in\mathbb{N}$ כך ש־ $P^n\left(x,y\right)>0$.

הערה 6.15 אי־פריקות לא תלויה בהסתברויות המעברים אלא רק בגרף המעברים של השרשרת (הגרף המכוון שבו קיימת קשת $x\to y$ אמ"מ $P\left(x,y\right)>0$) ולמעשה אי־פריקות שקולה לקשירות חזקה של הגרף הנ"ל.

הגדרה 6.16 פונקציה הרמונית על שרשרת מרקוב:
בהנתן שרשרת מרקוב $(\mathcal{S},P)$ פונקציה $h:\mathcal{S}\to\mathbb{R}$ תקרא הרמונית אם $h=Ph$.

הערה 6.17 בפועל נתייחס ל־ h בתור וקטור ב־ $\mathbb{R}^{\mathcal{S}}$.

הערה 6.18 אם μ היא התפלגות על $\mathcal{S}$ ו־ $f:\mathcal{S}\to\mathbb{R}$ היא פונקציה אז מתקיים:

$$\mathbb{E}_{\mu}\left(f\left(Z\right)\right)=\sum_{x\in\mathcal{S}}\mu\left(x\right)f\left(x\right)=\mu\cdot f$$

כאשר Z הוא מ"מ עם התפלגות μ.

דוגמה 6.19 בהנתן שרשרת מרקוב $(\mathcal{S},P)$ ו־ $x\in\mathcal{S}$ נתבונן במידת ההסתברות $e_x:\mathcal{S}\to[0,1]$ הנתונה ע"י:

$$e_{x}\left(z\right)=\begin{cases}1 & z=x\\0 & z\neq x\end{cases}$$

נשים לב שבהגדרה הזו בהנתן פונקציה $h:\mathcal{S}\to\mathbb{R}$ (וקטור $h\in\mathbb{R}^{\mathcal{S}}$) נקבל כי $e_x\cdot h=h\left(x\right)$ וכמו כן:

$$\left(e_{x}P\right)\cdot h=\sum_{x\in\mathcal{S}}\left(e_{x}P\right)\left(x\right)h\left(x\right)=\mathbb{E}_{\mu_{0}}\left(h\left(X_{1}\right)\right)$$

כאשר $X_0,X_1,..$ היא שרשרת מרקוב עם מטריצת מעברים P המתחילה ב־ $X_0\equiv x$ ו־ μ_0 הוא הפילוג של X_0.

הערה 6.20 במילים אחרות ניתן לראות ש־ $h\left(X_n\right)$ הוא מרטינגל (לכל תנאי התחלה):

$$h\left(X_{n}\right)=\sum_{y\in\mathcal{S}}P\left(X_{n},y\right)h\left(y\right)=\sum_{y\in\mathcal{S}}\mathbb{P}\left(X_{n+1}=y|X_{1}=x_{1},...,X_{n}=x_{n}\right)h\left(y\right)=\mathbb{E}\left[h\left(X_{n+1}\right)|\,X_{1}=x_{1},...,X_{n}=x_{n}\right]$$

כאשר זוהי התוחלת המותנית על סמך תכונת מרקוב של השרשרת:

משפט 6.21 פונקציה הרמונית על שרשרת מרקוב סופית ואי פריקה היא קבועה:

תהא $(\mathcal{S},P)$ שרשרת מרקוב אי־פריקה עם מרחב מצבים סופי ותהא $h:\mathcal{S}\to\mathbb{R}$ פונקציה הרמונית על P אזי h קבועה.

הוכחה: יהא M המקסימום של h ונניח כי $h\left(x\right)=M$ עבור $x\in\mathcal{S}$ כלשהו. כעת ניתן לראות כי:

$$M=h\left(x\right)=\sum_{y\in\mathcal{S}}P\left(x,y\right)h\left(y\right)\leq\sum_{y\in\mathcal{S}}P\left(x,y\right)\cdot M=M\sum_{y\in\mathcal{S}}P\left(x,y\right)=M$$

בפרט שוויון מתקיים רק אם לכל y כך ש־ $P\left(x,y\right)>0$ מתקיים $h\left(y\right)=M$ ולכן מאי־פריקות $h\left(y\right)=M$ לכל $y\in\mathcal{S}$ (יש צורך להשתמש בטיעון אינדוקטיבי כדי להגיע למסקנה הסופית). ■

מסקנה 6.22 שרשרת מרקוב אי־פריקה עם מרחב מצבים סופי היא בעלת התפלגות סטציונרית יחידה:

לשרשרת מרקוב $(\mathcal{S},P)$ אי־פריקה ובעלת מרחב מצבים סופי קיימת התפלגות סטציונרית יחידה.

הוכחה: נתבונן במרחב הוקטורים העצמיים הימניים עם ערך עצמי 1 שהוא בדיוק מרחב הפונקציות ההרמונית $\{h\,|\,h=Ph\}$, הראינו במשפט הקודם שהמרחב הנ"ל הוא חד־ממדי שכן כל הפונקציות ההרמוניות הללו הן קבועות. לכן גם המרחב של הוקטורים העצמיים השמאליים עם ערך עצמי 1 הוא חד ממדי והמרחב הנ"ל $\{\pi\,|\,\pi=\pi P\}$ הוא בדיוק מרחב המידות הסטציונריות המסומנות. מאחר שהראינו כבר קיום של התפלגות סטציונרית עבור שרשרת מרקוב סופית נסיק שמידת ההסתברות הזו היא יחידה (שכן כפל שלה בקבוע לא יהיה מידת הסתברות). ■

מסקנה 6.23

במקרה של מהלך מקרי פשוט על גרף לא מכוון וסופי G אי־פריקות של השרשרת שקולה לקשירות של G עצמו ולכן על סמך המסקנה

הערה 6.24 שתי הערות אגב:

- באופן כללי עבור שרשרת מרוקב סופית התפלגות סטציונרית יחידה לא גוררת אי־פריקות של השרשרת ואפשר להביא דוגמה פשוטה עם שרשרת בעלת גרף מעברים עם שני קודקודים בלבד.
- אפשר להראות שלשרשרת מרקוב סופית יש התפלגות סטציונרית יחידה רק אם בגרף רכיבי הקשירות החזקים של גרף המעברים של השרשרת יש רק עלה יחיד (עבור כל עלה קיימת התפלגות סטציונרית על רכיב הקשירות שמתאים לו וכל צירוף קמור של התפלגויות הללו מגדיר התפלגות סטציונרית על השרשרת כולה ־ כל מה שהוא לא עלה יהיה בגבול אפס).

6.1 מחזוריות והתכנסות לסטציונריות:

הגדרה 6.25 בהנתן שרשרת מרקוב $(\mathcal{S},P)$ ו־ $x\in\mathcal{S}$ נסמן:

$$A_x=\{n\geq 0\,|\,P^n(x,x)>0\}$$

נגדיר $a_x=\gcd(A_x)$ הנ"ל מכונה המחזור של x.

טענה 6.26

תרגיל בתורת המספרים: אם $A\subseteq\mathbb{N}$ סגורה לחיבור אז $|A\triangle(gcd(A)\,\mathbb{N})|<\infty$.

למה 6.27 בשרשרת מרקוב אי־פריקה כל המחזורים שווים:

אם $(\mathcal{S},P)$ היא שרשרת מרקוב אי־פריקה אז כל המחזורים שווים ($a_x=a_y$ לכל $x,y\in\mathcal{S}$).

הוכחה: בהנתן $x,y\in\mathcal{S}$ מאי־פריקות קיימים m,l כך ש־ $P^m(x,y)>0$ ו־ $P^l(y,x)>0$. מכך נקבל כי $A_y+m+l\subseteq A_x$ ולכן על סמך טענה 6.26 מקבלים כי $\gcd(A_x)\leq\gcd(A_y)$. מטעמי סימטריה מתקיים גם השוויון ההפוך ולכן מתקיימת המסקנה הנדרשת. ■

הגדרה 6.28 שרשרת מרקוב לא מחזורית:
שרשרת מרקוב $(\mathcal{S},P)$ תקרא לא־מחזורית אם $a_x=1$ לכל $x\in\mathcal{S}$.

תרגיל: מהלך מקרי פשוט על גרף לא מכוון הוא מחזורי אמ"מ הגרף הוא דו־צדדי.

טענה 6.29 תכונה של שרשרת מרקוב אי־פריקה ולא מחזורית:

תהא $(\mathcal{S},P)$ שרשרת מרקוב אי־פריקה סופית ולא־מחזורית אזי קיים $r>0$ כך שלכל $x,y\in\mathcal{S}$ מתקיים $P^r(x,y)>0$.

הוכחה: מאי־מחזוריות לכל $x\in\mathcal{S}$ מתקיים $\gcd(A_x)=a_x=1$ ולכן טענה 6.26 $|\mathbb{N}\backslash A_x|<\infty$. כעת בהנתן $x,y\in\mathcal{S}$ נתבונן ב:

$$A_{xy}=\{n\geq 0\,|\,P^n(x,y)>0\}$$

מאי־פריקות קיים $m\in A_{xy}$ ולכן $m+A_x\subseteq A_{xy}$ ומכך $|\mathbb{N}\backslash A_{xy}|<\infty$ ובפרט נוכל לקחת את $r=\max\bigcap_{x,y}A_{xy}$. ■

הערה 6.30 הקשר המשמעותי בין מחזוריות והתכנסות לסטציונריות הוא שלא תתכן התכנסות לסטציונריות כאשר יש מחזוריות שכן אם למשל יש מחזור 3 ב־ $x\in\mathcal{S}$ אז ידוע לנו שבכל הזמנים שלא מתחלקים ב־ 3 לא נוכל להגיע ל־ x. אולם ראינו שבמקרה של אי־פריקות אם קיימת התפלגות הסטציונרית היא נותנת הסתברות חיובית לכל המעברים האפשריים ולא ייתכן ש־ $\mu_0P^n\overset{n\to\infty}{\longrightarrow}\pi$ כאשר $\pi_i>0$ לכל i אם $(p^n)_{ii}=0$ לכל n שמתחלק ב־ 3. המשפט הבא מפרמל את הקשר הנ"ל.

משפט 6.31 עבור שרשרת מרקוב סופית, אי־פריקה ולא מחזורית קיימת התכנסות להתפלגות סטציונרית:

תהא $(\mathcal{S},P)$ שרשרת מרקוב אי־פריקה ולא מחזורית עם $\mathcal{S}$ סופית. אזי קיים התפלגות π כך שלכל התפלגות μ מתקיים $\mu P^n\overset{n\to\infty}{\longrightarrow}\pi$.
ליתר דיוק קיים $0<\alpha<1$ כך שלכל μ_0 מתקיים $\|\mu P^n-\pi\|_{L_1}\leq 2\alpha^n$.

הוכחה: על סמך טענות קודמות ידוע לנו כבר שקיימת התפלגות סטציונרית יחידה על $(\mathcal{S},P)$ שנסמנה π. על סמך טענה 6.29 קיים $r>0$ כך שלכל $x,y\in\mathcal{S}$ מתקיים $P^{r}(x,y)>0$ ונניח ראשית ש־ $r=1$. מאחר ש־ $\mathcal{S}$ סופית קיים $\beta>0$ כך שלכל $x,y\in\mathcal{S}$ מתקיים $P(x,y)\geq\beta\pi(y)$. כעת לכל x ניתן להציג את $P(x,y)$ בצורה:

$$P(x,\cdot)=\beta\pi(\cdot)+\alpha\nu_{x}(\cdot)$$

כאשר $\alpha=1-\beta$ ו־ ν_{x} היא התפלגות על $\mathcal{S}$. כנ"ל לכל μ ניתן לרשום $\mu P=\beta\pi+\alpha\nu_{\mu}$ כאשר ν_{μ} היא גם כן התפלגות על $\mathcal{S}$. נתחיל ב־ μ_{0} ונקבל ש:

$$\mu_{1}=\mu_{0}P=\beta\pi+\alpha\nu_{1}$$

כאשר $\nu_{1}=\nu_{\pi_{0}}$. נמשיך ונקבל:

$$\mu_{2}=\mu_{1}P=(\beta\pi+\alpha\nu_{1})P=\beta\pi P+\alpha\nu_{1}P=\beta\pi+\alpha\nu_{1}P$$

כעת ניתן לפרק באופן דומה את $\nu_{1}P$ ולקבל $\nu_{1}P=\beta\pi+\alpha\nu_{2}$ ולכן נקבל:

$$\mu_{2}=\beta\pi+\alpha(\beta\pi+\alpha\nu_{2})=(\beta+\alpha\beta)\pi+\alpha^{2}\nu_{2}=\left(1-\alpha^{2}\right)\pi+\alpha^{2}\nu_{2}$$

נמשיך באופן הנ"ל ונקבל כי:

$$\mu_{3}=\mu_{2}P=\left(1-\alpha^{2}\right)\pi+\alpha^{2}\nu_{2}P=\left(1-\alpha^{3}\right)\pi+\alpha^{3}\nu_{3}$$

ובאינדוקציה:

$$\mu_{n}=(1-\alpha^{n})\pi+\alpha^{n}\nu_{n}\ ;\quad \nu_{n}=\frac{\nu_{n-1}P-\beta\pi}{\alpha}$$

מכך נקבל כי:

$$\|\mu_{n}-\pi\|_{L_{1}}=\|(1-\alpha^{n})\pi+\alpha^{n}\nu_{n}-\pi\|_{L_{1}}=\|-\alpha^{n}\pi+\alpha^{n}\nu_{n}\|_{L_{1}}=\alpha^{n}\|\nu_{n}-\pi\|_{L_{1}}\leq 2\alpha^{n}$$

כאשר $\|\nu_{n}-\pi\|_{L_{1}}$ שכן ν_{n},π הן מידות הסתברות. במקרה בו $r\neq 1$ נתבונן בשרשרת $(\mathcal{S},P^{r})$ שהינה גם שרשרת אי־פריקה ולא־מחזורית ונקבל על סמך מה שהוכחנו ש־ $\mu P^{rn}\overset{n\to\infty}{\longrightarrow}\pi$ (עבור אותה π שכן π סטציונרית ביחס ל־ P^{r} ויש יחידות). באותו אופן לכל $0\leq k<r$ נקבל ש $\mu P^{rn+k}\overset{n\to\infty}{\longrightarrow}\pi$ במובן שבו:

$$\left\|\mu P^{rn+k}-\pi\right\|_{L_{1}}\leq 2\alpha^{n}=2\left(\alpha^{'}\right)^{rn+k}$$

עבור $0<\alpha^{'}<1$ ולכן $\mu P^{n}\overset{n\to\infty}{\longrightarrow}\pi$, כנדרש. ■

הגדרה 6.32 בהנתן שתי מידות הסתברות μ,ν על קבוצה סופית $\mathcal{S}$ נוכל להתבונן ב־ Total Variation Distance המוגדר ע"י:

$$d_{TV}(\mu,\nu)=\max_{A\subseteq\mathcal{S}}|\mu(A)-\nu(A)|$$

אפשר להראות כי $d_{TV}(\mu,\nu)=\frac{1}{2}\|\mu-v\|_{L_{1}}$ באמצעות התבוננות ב־ $A=\{x\,|\,\mu(x)>\nu(x)\}$.

על סמך ההגדרה הזו אפשר לראות שנהיה יותר ויותר קשה להבחין בין המידות μP^{n} והמידה הסטציונרית π.

הערה 6.33 מאחר ש־ $\mathcal{S}$ היא קבוצה סופית התכנסות של ההתפלגויות היא למעשה התכנסות של סדרות וקטורים ב־ $\mathbb{R}^{\mathcal{S}}$ ומבחינה טופולוגית לא חשוב באיזה נורמה נבחר שכן המרחב הוא סוף ממדי.

6.2 נשנות של שרשראות מרקוב:

הגדרה 6.34 מצב נשנה בשרשרת מרקוב:
בהנתן שרשרת מרקוב $(\mathcal{S},P)$ מצב $s\in\mathcal{S}$ יקרא **נשנה** אם $\mathbb{P}_s\left(T_s^+<\infty\right)=1$ כאשר:

$$T_s^+=\min\left\{n\geq 1\,|\,X_n=s\right\}$$

הערה 6.35 משמעות הסימון $\mathbb{P}_s$ הוא שהשרשרת מתחילה ב־ s ($X_0=s$).

הערה 6.36 מצב שאיננו נשנה מכונה מצב **חולף**. כמו כן מאופן ההגדרה ברור שמצב נשנה הוא מצב שבו נבקר אינסוף פעמים ומצב חולף הוא מצב שבו נבקר רק מספר סופי של פעמים.

תרגיל: בשרשרת מרקוב אי־פריקה כל המצבים נשנים או חולפים ביחד. מעבר לכך לכל $x,y\in\mathcal{S}$ מתקיים:

$$\mathbb{P}_x\left(T_y^+<\infty\right)=1$$

כלומר ההסתברות להגיע ל־ y בהנתן שהתחלנו ב־ x בזמן סופי היא 1 לכל $x,y\in\mathcal{S}$.

משפט 6.37 אפיון של המצבים בשרשרת מרקוב אי־פריקה:

בהנתן $(\mathcal{S},P)$ שרשרת מרקוב אי־פריקה לכל $s\in\mathcal{S}$ מתקיים:

$$\mathbb{E}_s\left(\#\text{visits } to\ s\right)=\left(\mathbb{P}_s\left(T_s^+=\infty\right)\right)^{-1}$$

הוכחה: המשפט הוא תוצאה של כך שמספר הביקורים ב־ $s\in\mathcal{S}$ הוא גיאומטרי עם פרמטר $\mathbb{P}_s\left(T_s^+=\infty\right)$ (כאשר הצלחה משמעותה שלא חזרנו ל־ s יותר). זוהי תוצאה של **תכונת מרקוב החזקה** לפיה אם T הוא זמן עצירה אז $Y_k=X_{T+k}$ היא שרשרת מרקוב עם אותן הסתברויות מעבר, כלומר:

$$\mathbb{P}\left(\left(X_T,X_{T+1},...,X_{T+k}\right)\in A\,|\,X_1,...,X_T\right)=\mathbb{P}_\mu\left(\left(X_0,...,X_k\right)\in A\right)$$

כאשר μ היא ההתפלגות של X_T. ∎

דוגמה 6.38 נתבונן במהלך מקרי פשוט על $\mathbb{Z}$, ראינו כבר משיקולים של מרטינגלים שזוהי שרשרת מרקוב נשנית וזוהי כמובן גם שרשרת אי־פריקה. נשים לב כי:

$$\mathbb{E}_0\left(\#\text{visits to }0\right)=\sum_{n=0}^{\infty}\mathbb{1}_{\{X_n=0\}}=\sum_{n=0}^{\infty}\mathbb{P}_0\left(X_n=0\right)$$

מאחר שזוהי שרשרת עם מחזור 2 לא ניתן לחזור ל־ 0 בזמנים אי־זוגיים ולכן נקבל:

$$\mathbb{E}_0\left(\#\text{visits to }0\right)=\sum_{n=1}^{\infty}\mathbb{P}_0\left(X_{2n}=0\right)=\sum_{n=1}^{\infty}\binom{2n}{n}2^{-2n}$$
$$\approx\sum_{n=1}^{\infty}\frac{\sqrt{4\pi n}\left(\frac{2n}{e}\right)^{2n}2^{-2n}}{\left(\sqrt{2\pi n}\left(\frac{n}{e}\right)^n\right)^2}=\sum_{n=1}^{\infty}\frac{\sqrt{4\pi n}}{2\pi n}=\sum_{n=1}^{\infty}\frac{1}{\sqrt{\pi n}}=\infty$$

בכך קיבלנו הוכחה נוספת לכך שזוהי שרשרת מרקוב נשנית.

דוגמה 6.39 נתבונן במהלך מקרי פשוט על $\mathbb{Z}^2$. אם ציר X ו־ Y היו בלתי־תלויים אז על סמך הדוגמה הקודמת היינו מקבלים ש:

$$\mathbb{P}_0\left(X_{2n}=0\right)\approx\frac{1}{\sqrt{\pi n}}\cdot\frac{1}{\sqrt{\pi n}}=\frac{1}{\pi n}$$

מכך היינו מסיקים ש־ $\mathbb{E}_0\left(\#\text{visits to }0\right)=\infty$ ולכן השרשרת נשנית. נתבונן בתהליך אחר Y_n שבו בכל שלב תמיד עושים צעד אחד בציר ה־ X וצעד אחד בציר ה־ Y, בתהליך הנ"ל נקבל שזזים על האלכסונים בגריד של $\mathbb{Z}^2$ ולמעשה מתקבל שיקוף ב־ 45 מעלות של הגריד $\mathbb{Z}^2$ אולם בגריד הנ"ל ציר ה־ X וציר ה־ Y כן בלתי־תלויים ולכן ההסתברות שנגיע לקודקוד מסוים שווה למכפלת ההסתברויות שנגיע לשיעורי ה־ X,Y שלו ובפרט $\mathbb{P}\left(Y_{2n}=0\right)=\frac{1}{\pi n}$. מאחר שכיוון הגריד לא משנה נסיק שגם מתקיים $\mathbb{P}_0\left(X_{2n}=0\right)\approx\frac{1}{\pi n}$ ולכן אפס הוא מצב נשנה של השרשרת המקורית.

דוגמה 6.40 באופן כללי יותר במהלך מקרי פשוט על $\mathbb{Z}^d$ אפשר להראות שכאשר מבצעים n צעדים מספר הצעדים בכל ציר מרוכז סביב $\frac{n}{d}$ והסיכוי לעשות פחות מ־ $\frac{n}{2d}$ צעדים בציר מסוים קטן אקספוננציאלית ב־ n. בפרט עבור $d=3$ בהנתן שעשינו $2n_1$ צעדים בציר ה־ X, $2n_2$ צעדים בציר ה־ Y ו־ $2n_3$ צעדים בציר ה־ Z (שזהו המקרה הסביר) נקבל כי:

$$\binom{2n_1}{n_1}\binom{2n_2}{n_2}\binom{2n_3}{n_3}\cdot2^{-2(n_1+n_2+n_3)}\approx\frac{1}{\pi^{\frac{3}{2}}\sqrt{n_1n_2n_3}}\leq\frac{C}{n^{\frac{3}{2}}}$$

לכן נקבל כי:

$$\mathbb{P}\left(X_{2n}=0\right)\leq\frac{C}{n^{\frac{3}{2}}}\Longrightarrow\mathbb{E}_0\left(\#\text{visits to }0\right)\leq\sum_{n=1}^{\infty}\frac{C}{n^{\frac{3}{2}}}<\infty$$

לכן נסיק שההילוך המקרי הפשוט על $\mathbb{Z}^d$ עבור $d\geq3$ הוא חולף.

תזכורת 6.41 בהנתן שרשרת מרקוב $(\mathcal{S},P)$ פונקציה $h:\mathcal{S}\to\mathbb{R}$ תקרא הרמונית אם $h=Ph$ או לחילופין לכל $x\in\mathcal{S}$:

$$h\left(x\right)=\sum_{y\in\mathcal{S}}P\left(x,y\right)h\left(y\right)$$

הגדרה 6.42 תכונת ליוביל של שרשרת מרקוב:
שרשרת מרקוב $(\mathcal{S},P)$ תקרא **ליוביל** אם הפונקציות ההרמוניות החסומות היחידות הן הקבועות.

הערה 6.43 ראינו במשפט 6.21 ששרשרת מרקוב סופית ואי־פריקה היא ליוביל.

משפט 6.44 שרשרת מרקוב אי־פריקה ונשנית היא ליוביל:

אם $(\mathcal{S},P)$ היא שרשרת מרקוב אי־פריקה ונשנית אז היא ליוביל.

הוכחה: נקבע $x,y\in\mathcal{S}$, מההנחה שהשרשרת אי־פריקה ונשנית ידוע לנו על סמך תרגיל ש־ $\mathbb{P}_x\left(T_y^+<\infty\right)=1$. נתחיל שרשרת מרקוב ב־ $X_0=x$, ידוע לנו ש־ $f\left(X_n\right)$ הוא מרטינגל חסום. לכן מאחר ש־ T_y^+ הוא זמן עצירה סופי נקבל ממשפט OST שמתקיים:

$$f\left(x\right)=f\left(X_0\right)=\mathbb{E}\left[f\left(X_{T_y^+}\right)\right]=\mathbb{E}\left[f\left(y\right)\right]=f\left(y\right)$$

■

מסקנה 6.45

ראינו כבר שמהלך מקרי פשוט על $\mathbb{Z}$ ו־ $\mathbb{Z}^2$ הוא נשנה ואי־פריק ולכן הן ליוביל. מאידך ראינו שמהלך מקרי פשוט על $\mathbb{Z}^d$ עבור $d\geq3$ איננו נשנה אולם מתברר שהוא עדיין ליוביל (הוכחה קשה).

תרגיל: אם $(\mathcal{S},P)$ היא שרשרת מרקוב אי־פריקה ונשנית אז לא קיימות פונקציות הרמוניות **חיוביות** ולא קבועות על $\mathcal{S}$.

הערה 6.46 הטענה הזו נכונה גם עבור $\mathbb{Z}^d$ לכל d גם כאשר השרשרת לא נשנית אבל ההוכחה לא פשוטה.

7 משפט הגבול המרכזי:

7.1 התכנסות חלשה:

תזכורת 7.1 בהנתן סדרת מ"מ $X_n \sim F_n$ (כלומר F_n היא ה־ CDF של X_n) נאמר ש־ X_n **מתכנסת בהתפלגות** למ"מ $X \sim F$ ונסמן $X_n \xrightarrow{D} X$ או $F_n \xrightarrow{D} F$ אם $F_n(t) \xrightarrow{n\to\infty} F(t)$ בכל t שהיא נקודת רציפות של F.

הגדרה 7.2 התכנסות חלשה:
נאמר שסדרת מידות μ_n (המוגדרות באותו מרחב) מתכנסת חלש למידה μ ונסמן $\mu_n \xrightarrow{W} \mu$ אם לכל $h:\mathbb{R}\to[0,1]$ רציפה חסומה:

$$\mathbb{E}\left[h\left(X_n\right)\right] \xrightarrow{n\to\infty} \mathbb{E}\left[h\left(X\right)\right]\ ;\quad X_n\sim\mu_n\quad X\sim\mu$$

כאשר משמעות הסימון $X_n \sim \mu_n$ היא ש־ X_n הוא מ"מ כך שלכל $A\in\mathcal{F}$ מתקיים $\mathbb{P}\left(X_n\in A\right)=\mu_n\left(A\right)$.

הערה 7.3 סימון: אם $X_n \sim \mu_n$ ו־ $X_n \sim \mu$ ו־ $\mu_n \xrightarrow{W} \mu$ אז נסמן גם $X_n \xrightarrow{W} X$.

הערה 7.4 עבור מ"מ ממשי $X\sim F$ ישנה מידת הסתברות $\mu:\mathbb{R}\to[0,1]$ המוגדרת ע"י $\mu\left((-\infty,x]\right)=F\left(x\right)$ והרחבה על סמך משפט קרטיאודורי. לחילופין כל מידת הסתברות $\mu:\mathcal{B}\left(\mathbb{R}\right)\to[0,1]$ מגדירה משתנה מקרי ממשי עם התפלגות מצטברת הנתונה על ידי $F\left(x\right)=\mu\left((-\infty,x]\right)$ (תמיד אפשר למצוא מרחב הסתברות מתאים ככה שמוגדר משתנה מקרי כנ"ל).

משפט 7.5 התכנסות חלשה שקולה להתכנסות בהתפלגות:

תהא $X_n \sim F_n$ סדרת מ"מ אזי $X_n \xrightarrow{D} X\sim F$ אמ"מ $X_n \xrightarrow{W} X$.

הוכחה: $\Longleftarrow$ נניח כי $\mu_n \xrightarrow{W} \mu$ ויהיו $X_n\sim\mu_n$, $X\sim\mu$ ו־ x נקודת רציפות של F, נראה כי $F_n\left(x\right)\xrightarrow{n\to\infty}F\left(x\right)$. נתבונן בפונקציה:

$$h_\delta\left(y\right)=\begin{cases}1 & y\le x\\ 1-\frac{y-x}{\delta} & x<y<x+\delta\\ 0 & y\ge x+\delta\end{cases}$$

עבור $\delta>0$ כלשהו. זוהי פונקציה רציפה וחסומה ולכן מהתכנסות חלשה נקבל כי:

$$F_n\left(x\right)=\mathbb{E}\left[\mathbb{1}_{\{X_n\le x\}}\right]\le\mathbb{E}\left[h_\delta\left(X_n\right)\right]\xrightarrow{n\to\infty}\mathbb{E}\left[h_\delta\left(X\right)\right]\le\mathbb{E}\left[\mathbb{1}_{\{X\le x+\delta\}}\right]=F\left(x+\delta\right)$$

מכך מרציפות מימין של F ומכך שהנ"ל נכון לכל $\delta>0$ נסיק כי:

$$\limsup_{n\to\infty}F_n\left(x\right)\le F\left(x+\delta\right)\xrightarrow{\delta\to0}F\left(x\right)$$

מאידך נתבונן בפונקציה:

$$g_\delta\left(y\right)=\begin{cases}1 & y\le x-\delta\\ 1-\frac{y-(x-\delta)}{\delta} & x-\delta<y<x\\ 0 & y\ge x\end{cases}$$

מרציפות של F ב־ x ושיקולים דומים לכיוון השני נקבל כי:

$$F_n\left(x\right)\ge\mathbb{E}\left[g_\delta\left(X_n\right)\right]\xrightarrow{n\to\infty}\mathbb{E}\left[g_\delta\left(X\right)\right]\ge F\left(x-\delta\right)$$
$$\Longrightarrow\liminf_{n\to\infty}F_n\left(x\right)\ge F\left(x-\delta\right)\xrightarrow{\delta\to0}F\left(x\right)$$

לכן קיבלנו את הנדרש:

$$\liminf_{n\to\infty}F_n\left(x\right)=F\left(x\right)=\limsup_{n\to\infty}F_n\left(x\right)\Longrightarrow\lim_{n\to\infty}F_n\left(x\right)=F\left(x\right)$$

$\Longrightarrow$ נניח ש־ $F_n \xrightarrow{D} F$, נראה שקיימים מ"מ X_n, X כך ש־ $X_n \sim F_n$, $X \sim F$ ובנוסף $X_n \xrightarrow{\text{a.s}} X$. נגדיר:

$$X^{+}(w) = \inf\{z \in \mathbb{R} \,|\, F(z) > w\}\ ,\ w \in [0,1]$$
$$X_n^{+}(w) = \inf\{z \in \mathbb{R} \,|\, F_n(z) > w\}\ ,\ w \in [0,1]$$

אלו הן ה־ Psuedo-Inverse של F ו־ F_n בהתאמה. נקבע $w \in [0,1]$ ונשים לב שעבור z כלשהו מתקיים $z > X^{+}(w)$ אמ"מ $F(z) < w$. כעת אם z היא לא אטום (נקודת רציפות) של F אז מההנחה $F_n(z) \xrightarrow{n\to\infty} F(z)$ ולכן עבור n מספיק גדול $w < F_n(z)$ (כאשר מסתמכים על כך ש־ F_n הן פונקציות עולות חלש) ומכך $X_n^{+}(w) \le z$. מאחר שיש לכל היותר מספר בן־מנייה של אטומים של F אפשר לבחור סדרה $z_k \searrow X^{+}(w)$ שכולם לא אטומים ולקבל כי:

$$\limsup_{n\to\infty} X_n^{+}(w) \le z_k \searrow X^{-1}(w)$$

מאידך נוכל להגדיר:

$$X^{-}(w) = \inf\{z \in \mathbb{R} \,|\, F(z) \ge w\}\ ,\ w \in [0,1]$$
$$X_n^{-}(w) = \inf\{z \in \mathbb{R} \,|\, F_n(z) \ge w\}\ ,\ w \in [0,1]$$

באמצעות שיקולים דומים נקבל כי:

$$\liminf_{n\to\infty} X_n^{-}(w) \ge X^{-}(w)$$

אולם נשים לב ש־ $X_n^{-}(w) = X_n^{+}(w)$ פרט למספר בן־מנייה של נקודות שכן עבור כל w כנ"ל יש קטע $[X_n^{-}(w), X_n^{+}(w)]$ שבו F_n שקבועה והקטעים האלו הם זרים. לכן $X_n^{+} \stackrel{\text{a.s}}{=} X_n^{-}$ ומכך $X_n^{+} \xrightarrow{\text{a.s}} X$. כעת בהנתן $X_n \sim F_n$ ו־ $X \sim F$ כנ"ל ופונקציה רציפה $h : \mathbb{R} \to [0,1]$ מרציפות נקבל כי $h(X_n) \xrightarrow{\text{a.s}} h(X)$ ומכך ש־ h חסומה נקבל ממשפט ההתכנסות החסומה ש:

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

■

הערה 7.6 אם נתבונן ב־ $\overline{\mathbb{R}} = [-\infty, \infty]$ ונתבונן במרחב $\mathcal{C}(\overline{\mathbb{R}})$ של פונקציות רציפות (ולכן חסומות) משפט ריס אומר לנו שהמרחב הדואלי הוא מרחב המידות הסופיות המסומנות על $\overline{\mathbb{R}}$. מידות הסתברות הן תת־קבוצה סגורה של כדור היחידה במרב הזה ולכן משפט בנך־אלאוגלו אומר שזוהי קבוצה קומפקטית ביחס לטופולוגיה החלשה־* שהיא הטופולוגיה שבה התכנסות נתונה על ידי $\int h d\mu_n \xrightarrow{n\to\infty} \int h d\mu$ לכל $h \in \mathcal{C}(\overline{\mathbb{R}})$.

למה 7.7 הלמה של *Helly-Bray*:

תהא $\{F_n\}_{n=1}^{\infty}$ סדרה של CDF אז קיימת ת"ס n_k ופונקציה עולה חלש ורציפה מימין F כך שלכל x שהיא נקודת רציפות של F מתקיים $F_{n_k}(x) \xrightarrow{k\to\infty} F(x)$.

הוכחה: נקבע מנייה של הרציונליים $\{r_i\}_{i=1}^{\infty}$ ונקח תת־סדרה n_k^1 שעבורה $F_{n_k^1}(r_1)$ מתכנס ואז בטיעון אלכסון סטנדרטי תת־סדרה $n_k^2 \subseteq n_k^1$ שעבורה $F_{n_k^2}(r_2)$ מתכנס. לבסוף נקח את האלכסון $n_k := n_k^k$ ועבורו נגדיר לכל $r \in \mathbb{Q}$ פונקציה G על ידי:

$$G(r) = \lim_{k\to\infty} F_{n_k}(r)$$

זוהי פונקציה עולה חלש כגבול של פונקציות עולות חלש. נגדיר:

$$F(x) = \inf\{G(r) \,|\, r > x,\, r \in \mathbb{Q}\}$$

זוהי פונקציה עולה חלש ורציפה מימין. נניח ש־ x היא נקודת רציפות של F, מכך לכל $\varepsilon > 0$ קיימים $r_1 < x < r_2$ רציונליים כך ש־ $|F(r_1) - F(r_2)| < \varepsilon$. מהגדרת F קיימים $r_3 < x < r_4$ רציונליים כך ש־ $|G(r_3) - G(r_4)| < 2\varepsilon$ ומכך מהגדרת G נקבל:

$$\left|\lim_{k\to\infty} F_{n_k}(r_3) - \lim_{k\to\infty} F_{n_k}(r_r)\right| < 2\varepsilon$$

מכך נסיק כי:

$$\left|\limsup_{k\to\infty} F_{n_k}(x) - \liminf_{k\to\infty} F_{n_k}(x)\right| < 2\varepsilon$$

מאחר והנ"ד נכון לכל $\varepsilon > 0$ נקבל כי הגבולות שווים ולכן $F(x) \overset{k\to\infty}{\longrightarrow} F_{n_k}(x)$. ■

הגדרה 7.8 סדרת הדוקה של פונקציות CDF:
סדרת פונקציות CDF $\{F_n\}_{n=1}^{\infty}$ תקרא **הדוקה** אם לכל $\varepsilon > 0$ קיים K כך שלכל n מתקיים:

$$F_n(k) - F_n(-K) = \mu([-K, K]) \geq 1 - \varepsilon$$

מסקנה 7.9 מסקנה מהלמה של *Helly-Bray* (ללא הוכחה):

אם $\{F_n\}_{n=1}^{\infty}$ היא סדרה הדוקה של פונקציות CDF אז קיימת CDF F ותת־סדרה n_k כך ש־ $F_{n_k} \xrightarrow{D} F$. מאידך אם קיימת ת"ס F_{n_k} כך ש־ $F_{n_k} \xrightarrow{D} F$ אז $\{F_n\}_{n=1}^{\infty}$ היא הדוקה.

7.2 משפט הגבול המרכזי:

משפט 7.10 משפט הגבול המרכזי:

בהנתן סדרת משתנים מקריים $X_n \overset{i.i.d}{\sim} F$ עם $\mathbb{E}[X_n] = 0$ ו־ $\mathbb{E}\left[X_n^2\right] = 1$ אזי:

$$\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) \xrightarrow{D} N(0,1)$$

הערה 7.11 המשפט הנ"ל לא נותן לנו אינפורמציה על קצב ההתכנסות. בתוספת של הנחה נוספת שקיים מומנט שלישי סופי מתקבל משפט Berry-Esseen לפיו בתנאים של המשפט הקודם אם בנוסף $\mathbb{E}\left[|X_n|^3\right] = M_3 < \infty$ אז קיים $C \in \mathbb{R}$ כך ש:

$$\forall n \in \mathbb{N} \,:\, |F_n(x) - F(x)| \leq \frac{CM_3}{\sqrt{n}}$$

כאשר F היא ה־ CDF של מ"מ $N(0,1)$.

7.2.1 קצת הקדמה להוכחה:

נזכיר את הפונקציה יוצרת מומנטים של מ"מ X הנתונה ע"י $\varphi_X(t) = \mathbb{E}\left[e^{tX}\right]$ (אם יש התכנסות) ובפרט:

- כאשר X, Y ב"ת מתקבל ש־ $\varphi_{X+Y}(t) = \varphi_X(t) \cdot \varphi_Y(t)$.
- לכל $a, b \in \mathbb{R}$ מתקיים $\varphi_{aX+b}(t) = e^{tb}\varphi_x(at)$.

נניח ש- $F \overset{\text{i.i.d}}{\sim} X_n$ עם תוחלת 0 ושונות 1, במקרה שישנה התכנסות נקבל כי:

$$\varphi_{X_1}(t)=\mathbb{E}\left[e^{tX}\right]=\mathbb{E}\left[1+tX+\frac{t^2X^2}{2}+...\right]=1+t\mathbb{E}[X]+\frac{t^2}{2}\mathbb{E}\left[X^2\right]+...=1+\frac{t^2}{2}+o\left(t^2\right)$$

מכך נקבל כי עבור $S_n=\sum_{i=1}^n X_i$ מתקיים:

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t)=\varphi_{S_n}\left(\frac{t}{\sqrt{n}}\right)=\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n=\left(1+\frac{t^2}{2}+o\left(t^2\right)\right)^n\overset{n\to\infty}{\longrightarrow}e^{\frac{t^2}{2}}$$

כאשר ידוע ש- $e^{\frac{t^2}{2}}$ היא הפונקציה יוצרת מומנטים של מ"מ $N(0,1)$. שתי שאלות שראוי לשאול הן:

1. האם התכנסות של פונקציות יוצרות מומנטים גוררת התכנסות בהתפלגות?
2. האם שוויון של כל המומנטים של שתי התפלגויות גורר שוויון של ההתפלגויות?

כמה עובדות:

- במידה והיא קיימת הפונקציה היוצרת מומנטים כן קובעת את ההתפלגות של משתנה מקרי.
- קיימות התפלגויות שיש להן מומנטים מכל סדר אבל הפונקציה יוצרת מומנטים לא מוגדרת לכל $t\neq 0$ (תרגיל).
- באופן כללי המומנטים של ההתפלגות לא קובעים את ההתפלגויות (קיימים מ"מ עם התפלגויות שונות ומומנטים זהים).
- עבור מ"מ חסום המומנטים כן קובעים את ההתפלגות של X (רמז - להשתמש במשפט ויירשטראס על צפיפות הפולינומים).

7.2.2 פונקציות אופייניות:

הגדרה 7.12 פונקציה אופיינית (Characteristic Function):
עבור מ"מ X הפונקציה האופיינית שלו $\varphi:\mathbb{R}\to\mathbb{C}$ מוגדרת ע"י $\varphi_X(t)=\mathbb{E}\left[e^{itX}\right]$. כמה תכונות שלה הן:

1. מתקיים $|\varphi_X(t)|\leq 1$ לכל t ובפרט הנ"ל מוגדרת היטב תמיד עבור כל מ"מ.
2. עבור X,Y ב"ת מתקיים $\varphi_{X+Y}(t)=\varphi_X(t)\cdot\varphi_Y(t)$.
3. לכל a מתקיים $\varphi_{aX+b}(t)=e^{itb}\varphi_X(at)$ ובפרט $\varphi_{-X}(t)=\varphi_X(-t)=\overline{\varphi_X(t)}$.
4. φ_X רציפה במ"ש בכל $\mathbb{R}$.
5. φ_X היא פונקציה ממשית אמ"מ X הוא מ"מ סימטרי ($\mathbb{P}(X>x)=\mathbb{P}(X<-x)$ לכל x).

תרגיל: הוכח שלא קיימים מ"מ X,Y ב"ת כך ש- $X-Y\sim U[-1,1]$.

הערה 7.13 עבור פונקציה $f:\mathbb{R}\to\mathbb{R}$ טרנספורם פורייה שלה נתון ע"י:

$$\hat{f}\left(t\right)=\int\limits_{-\infty}^{\infty}f\left(x\right)e^{itx}dx$$

אם F היא ה־ CDF של מ"מ X אז באינטגרציית לבג־סטליג'ס מתקבל ש:

$$\varphi_{X}\left(t\right)=\int\limits_{-\infty}^{\infty}e^{itx}dF$$

אם יש ל־ X יש צפיפות $f=dF$ אז הפונקציה האופיינית היא ממש טרנספורם פורייה של הצפיפות:

$$\varphi_{X}\left(t\right)=\int\limits_{-\infty}^{\infty}f\left(x\right)e^{itx}dx$$

דוגמה 7.14 כמה דוגמאות:

1. עבור מ"מ $X=\pm1$ עם הסתברות $\frac{1}{2}$ נקבל כי:

$$\varphi_{X}\left(t\right)=\frac{e^{it}+e^{-it}}{2}=\cos t$$

2. עבור מ"מ $X\sim\exp\left(1\right)$ עם צפיפות $f\left(x\right)=e^{-x}\mathbb{1}_{\{x\geq0\}}$ נקבל כי:

$$\varphi_{X}\left(t\right)=\int\limits_{0}^{\infty}e^{-x}e^{itx}dx=\frac{e^{x(it-1)}}{it-1}\int\limits_{0}^{\infty}=\frac{1}{1-it}$$

3. עבור מ"מ אקספוננציאלי כפול המוגדר ע"י:

$$X=\frac{1}{2}Y_{1}Z-\frac{1}{2}Y_{2}\left(1-Z\right)$$

כאשר $Z=0,1$ עם הסתברות $\frac{1}{2}$ ו־ $Y_{1},Y_{2}\sim\exp\left(1\right)$ ב"ת מתקבלת צפיפות $f\left(x\right)=\frac{1}{2}e^{-|x|}$ ומקבלים כי:

$$\varphi_{X}\left(t\right)=\frac{1}{2}\left(\frac{1}{1-it}+\frac{1}{1+it}\right)=\frac{1}{1+t^{2}}$$

4. עבור מ"מ נורמלי $X\sim N\left(0,1\right)$ נקבל כי:

$$\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^{2}}{2}}e^{itx}dx=e^{-\frac{t^{2}}{2}}\underbrace{\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{\frac{(x-it)^{2}}{2}}dx}_{=1}=e^{-\frac{t^{2}}{2}}$$

כאשר האינטגרל המסומן שווה ל־ 1 משיקולים של פונקציות מרוכבות.

הערה 7.15 ניתן לראות בפרט כי הצפיפות של התפלגות נורמלית מהווה עד כדי נרמול נקודת שבת של טרנספורם פורייה.

7.2.3 התכנסות של פונקציות אופייניות:

טענה 7.16

יהא $X\sim\mu$ מ"מ (μ מידת הסתברות על $\mathbb{R}$) אזי לכל $a\in\mathbb{R}$ מתקיים:

$$\mu\left(\{a\}\right)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}e^{-ita}\varphi_X\left(t\right)dt$$

הוכחה: על סמך משפט פוביני נקבל כי:

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}e^{-ita}\varphi_X\left(t\right)dt=\lim_{T\to\infty}\left(\frac{1}{2T}\int_{-T}^{T}e^{-ita}\int_{-\infty}^{\infty}e^{itX}d\mu\left(x\right)dt\right)$$
$$=\lim_{T\to\infty}\int_{-\infty}^{\infty}\left(\frac{1}{2T}\int_{-T}^{T}e^{-ita}e^{itX}dt\right)d\mu\left(x\right)=\lim_{T\to\infty}\int_{-\infty}^{\infty}\left(\frac{1}{2T}\int_{-T}^{T}e^{it(X-a)}dt\right)d\mu\left(x\right)$$

ממשפט ההתכנסות החסומה ניתן להכניס את הגבול לתוך האינטגרל החיצוני (התוחלת) שכן $\frac{1}{2T}\int_{-T}^{T}e^{it(X-a)}dt$ חסומה ומכך:

$$\lim_{T\to\infty}\int_{-\infty}^{\infty}\left(\frac{1}{2T}\int_{-T}^{T}e^{it(X-a)}dt\right)d\mu\left(x\right)=\int_{-\infty}^{\infty}\left(\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}e^{it(X-a)}dt\right)dtd\mu\left(x\right)$$

נשים לב כי:

$$\left(\frac{1}{2T}\int_{-T}^{T}e^{it(X-a)}dt\right)|_{X=a}\equiv1$$

מאידך אם $X\neq a$ נקבל משיקולי מחזוריות וחסימות של הפונקציה מתחת לאינטגרל שמתקיים:

$$\lim_{T\to\infty}\frac{1}{2T}\overbrace{\int_{-T}^{T}e^{it(X-a)}dt}^{\text{bounded and periodic}}=0$$

לכן סה"כ נקבל כי:

$$\int_{-\infty}^{\infty}\left(\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}e^{it(X-a)}dt\right)dtd\mu=\int_{-\infty}^{\infty}\left(\mathbb{1}_{\{X=a\}}\left(x\right)\right)d\mu\left(x\right)=\mathbb{E}_{\mu}\left(\mathbb{1}_{\{X=a\}}\left(x\right)\right)=\mu\left(\{a\}\right)$$

■

משפט 7.17 נוסחת ההיפוך של לוי:

בהנתן מ"מ $X\sim\mu$ ו־ $a<b\in\mathbb{R}$ מתקיים:

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

הוכחה: ראשית נשים לב כי:

$$b-a \geq \left|\int_a^b e^{-ity}dy\right| = \left|\frac{e^{-ita}-e^{-itb}}{it}\right|$$

משיקולים דומים להוכחה הקודמת באמצעות משפט פוביני וחסימות נקבל כי:

$$\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}\varphi_X(t)\,dt = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-\infty}^{\infty}\left(\underbrace{\int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itX}dt}_{=I_T(x)}\right)d\mu(x)$$

$$= \frac{1}{2\pi}\lim_{T\to\infty}\mathbb{E}_\mu\left(I_T(x)\right) = \frac{1}{2\pi}\mathbb{E}_\mu\left(\lim_{T\to\infty}I_T(x)\right)$$

כעת נשים לב כי:

$$I_T(x) = \int_{-T}^{T}\frac{e^{-ita}-e^{-itb}}{it}e^{itX}dt = \int_{-T}^{T}\frac{e^{-it(X-a)}-e^{-it(X-b)}}{it}dt$$

$$= \int_{-T}^{T}\frac{\sin(t(X-a))-\sin(t(X-b))}{t}dt = 2\left(S(T(X-a))-S(T(X-b))\right)$$

כאשר $S(T)=\int_0^T\frac{\sin t}{t}dt$ היא פונקציה המקיימת $S(s)\xrightarrow{s\to\pm\infty}\pm\frac{\pi}{2}$ ו־ $S(0)=0$ ומכך נובע כי:

- אם $a<b<X$ אז:

$$I_T(x) = 2\left(S(T(X-a))-S(T(X-b))\right)\xrightarrow{T\to\infty}2\left(\frac{\pi}{2}-\frac{\pi}{2}\right)=0$$

- אם $X<a<b$ אז:

$$I_T(x) = 2\left(S(T(X-a))-S(T(X-b))\right)\xrightarrow{T\to\infty}2\left(-\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)\right)=0$$

- אם $a<X<b$ אז:

$$I_T(x) = 2\left(S(T(X-a))-S(T(X-b))\right)\xrightarrow{T\to\infty}2\left(\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)\right)=2\pi$$

- אם $X=a,b$ אז:

$$I_T(x) = 2\left(S(T(X-a))-S(T(X-b))\right)\xrightarrow{T\to\infty}\pi$$

מכך סה"כ נקבל כי:

$$\begin{aligned}\frac{1}{2\pi}\mathbb{E}_{\mu}\left(\lim_{T\to\infty}I_{T}\left(x\right)\right) &= \frac{1}{2\pi}\mathbb{E}_{\mu}\left(2\pi\mathbb{1}_{\{X\in(a,b)\}}\left(x\right)+\pi\mathbb{1}_{\{X=a,b\}}\left(x\right)\right)\\ &= \frac{1}{2\pi}\left[2\pi\mu\left((a,b)\right)+\pi\mu\left(a,b\right)\right]=\mu\left((a,b)\right)+\frac{1}{2}\mu\left(\{a,b\}\right)\end{aligned}$$

הערה 7.18 הפונקציה $S\left(T\right)$ ידועה בתור אינטגרל דירכלה.

■

מסקנה 7.19

אם בתנאים של המשפט הקודם מתקיים $\int_{-\infty}^{\infty}|\varphi_X(t)|\,dt<\infty$ (כלומר φ אינטגרבילית לבג) אז אפשר ישירות לקחת $T\to\infty$ ולקבל שלכל a,b שהם לא אטומים של X מתקיים:

$$F_X(b)-F_X(a)=\mu((a,b])=\mu((a,b))=\frac{1}{2\pi}\int_{-\infty}^{\infty}\underbrace{\frac{e^{-ita}-e^{itb}}{it}}_{\geq b-a}\varphi_X(t)\,dt$$

על ספך כך לכל x_0 ניתן לבחור $a<x_0<b$ ולקבל שכאשר $b\to x_0\leftarrow a$ האינטגרל שואף לאפס ולכן x_0 הוא לא אטוס. כמו כן:

$$\frac{F_X(b)-F_X(a)}{b-a}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-ita}-e^{itb}}{it(b-a)}\varphi_X(t)\,dt\xrightarrow{b\to a}\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ita}\varphi_X(t)\,dt$$

בפרט מקבלים ש־ F, כלומר יש צפיפות הנתונה על ידי:

$$F^{'}(a)=\lim_{b\to a}\frac{F_X(b)-F_X(a)}{b-a}=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-ita}\varphi_X(t)\,dt$$

תרגיל: עבור X בעל צפיפות $f(x)=\frac{1}{\pi(1+x^2)}$ (מ"מ מפולג קושי) מתקיים $\varphi_X(t)=e^{-|t|}$ אולם ל־ X אין תוחלת. כמו כן אם $X_1,...,X_n\overset{\text{i.i.d}}{\sim}\text{Cauchy}$ אז $\frac{1}{n}\sum_{i=1}^{n}X_i\sim\text{Cauchy}$.

משפט 7.20 משפט הרציפות של לוי:

בהנתן F_n סדרה הדוקה של פונקציות CDF נסמן ב־ φ_n את סדרת הפונקציות האופייניות המתאימות להן ($\varphi_n(t)=\int_{-\infty}^{\infty}e^{itx}dF_n$). נניח כי $\varphi_n(t)\xrightarrow{t\to\infty}\varphi(t)$, אזי קיימת פונקציית CDF F כך ש־ $F_n\xrightarrow{D}F$ (באופן שקול $F_n\xrightarrow{W}F$) כאשר בפרט φ_F (הפונקציה האופיינית המתאימה ל־ F) היא φ.

הוכחה: מההנחה כי הסדרה הדוקה על סמך מסקנה 7.9 קיימת פונקציית CDF F ותת־סדרה F_{n_k} כך ש $F_{n_k}\xrightarrow{D}F$ ולכן $F_{n_k}\xrightarrow{W}F$. מאחר שלכל t הפונקציה e^{it} היא פונקציה רציפה וחסומה נקבל מההתכנסות החלשה שמתקיים:

$$\varphi_{n_k}(t)=\mathbb{E}\left(e^{itX_{n_k}}\right)\xrightarrow{n\to\infty}\mathbb{E}\left[e^{itX}\right]=\varphi_F(t)$$

כאשר $X_{n_k}\sim F_{n_k}$ ו־ $X\sim F$ הם מ"מ עם ההתפלגויות המתאימות. מאחר שיש ת"ס של φ_n שמתכנס ל־ φ_F נסיק ש־ $\varphi_F=\varphi$.■

הערה 7.21 הערות:

1. הדרישה של הדיקות של הסדרה איננה הכרחית והמשפט נכון גם בלעדיה.
2. המשפט הנ"ל מראה שהתכנסות נקודתית של פונקציות אופייניות גוררת התכנסות חלשה של פונקציות ההתפלגות.

7.2.4 הוכחת משפט הגבול המרכזי:

משפט 7.22 משפט הגבול המרכזי:

בהנתן סדרת משתנים מקריים $X_n\overset{i.i.d}{\sim}F$ עם $\mathbb{E}[X_n]=0$ ו־ $\mathbb{E}\left[X_n^2\right]=1$ מתקיים

$$\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\right)\xrightarrow{D}N(0,1)$$

הוכחה: נשים לב כי על סמך תכונות של פונקציות אופייניות מתקיים:

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t)=\varphi_{(X_1+\ldots+X_n)}\left(\frac{t}{\sqrt{n}}\right)\overbrace{=}^{\text{I.I.D}}\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n$$

בנוסף נעזר בעובדה שלכל $x\in\mathbb{R}$ מתקיים:

$$\left|e^{ix}-\left(1+ix-\frac{x^2}{2}\right)\right|\le\min\left(|x|^2,\frac{|x|^3}{6}\right)$$

בהנתן העובדה הזו עבור $X\sim F$ אם נקח תוחלת על אי השוויון הנ"ל נקבל כי:

$$\left|\varphi_X(t)-\left(1-\frac{t^2}{2}\right)\right|=\left|\mathbb{E}\left(e^{itX}-1+itX-\frac{(tX)^2}{2}\right)\right|$$
$$\le\mathbb{E}\left[\min\left(|t|^2|X|^2,\frac{|t|^3|X|^3}{6}\right)\right]=\left|t^2\right|\mathbb{E}\left[\min\left(|X|^2,\frac{|t||X|^3}{6}\right)\right]\xrightarrow{t\to 0}0$$

כעת נשים לב כי $Y:=\min\left(|X|^2,\frac{|t||X|^3}{6}\right)$ הוא מ"מ החסום על ידי המשתנה המקרי האינטגרבילי $|X|^2$ וכמו כן לכל ω מתקיים $\frac{|t||X(\omega)|^3}{6}\xrightarrow{t\to\infty}0$ שכן . לכן $Y\xrightarrow{t\to\infty}0$ ומכך ממשפט ההתכנסות החסומה:

$$\mathbb{E}\left[\min\left(|X|^2,\frac{|t||X|^3}{6}\right)\right]\xrightarrow{t\to 0}\mathbb{E}[0]=0$$

מכך נסיק כי $\varphi_X(t)=1-\frac{t^2}{2}+o\left(t^2\right)$ ולכן:

$$\varphi_{\frac{S_n}{\sqrt{n}}}(t)=\left(\varphi_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n=\left(1-\frac{t^2}{2n}+o\left(\frac{t^2}{n}\right)\right)^n\xrightarrow{n\to\infty}e^{-\frac{t^2}{2}}=\varphi_{N(0,1)}(t)$$

בכך הראינו התכנסות של הפונקציות האופייניות ולכן על סמך משפט הרציפות גם התכנסות חלשה כאשר הנ"ל נובע מכך שסדרת פונקציות ההתפלגות של הסדרה $\frac{S_n}{\sqrt{n}}$ היא סדרה הדוקה על סמך א"ש צ'בישב והעובדה שלכל n למשתנה המקרי $\frac{S_n}{\sqrt{n}}$ יש שונות השווה ל- 1. ∎

8 תנועה בראונית:

הגדרה 8.1 תנועת בראון:
תנועת בראון היא תהליך סטוכסטי בזמן רציף $\{B(t)\}_{t\ge 0}$ המקיים את התנאים הבאים:

1. לכל $0\le s<t$ מתקיים $B(t)-B(s)\sim N(0,t-s)$.
2. לכל $0\le t_1<t_2<\ldots<t_k$ המ"מ $\{B(t_{i+1})-B(t_i)\}_{1\le i\le k-1}$ בלתי תלויים.
3. הפונקציה $t\mapsto B(t)$ היא רציפה כ"ת (כלומר לכמעט כל $\omega\in\Omega$ הפונקציה $B(t)(\omega)$ רציפה כפונקציה של t).

הערה 8.2 אם $B(0)=0$ אז נאמר שזוהי תנועת בראון סטנדרטית.

הערה 8.3 בשלב זה לא ברור שקיים בכלל תהליך כנ"ל וזה עיקר מה שנרצה להראות.

נוכיח קיום של תנועת בראון:

תהא $\{X_i\}_{i=1}^{\infty}$ סדרת מ"מ ב"ת מפולגים $N(0,1)$. נגדיר את B על קבוצת השברים הדיאדיים כך שראשית:

$$B(0)=0\,,\,B(1)=X_1,\;B\left(\frac{1}{2}\right)=\frac{B(1)}{2}+\frac{X_2}{2}$$

על סמך תכונות של ההתפלגות הרב נורמלית נקבל שהמ"מ $B\left(\frac{1}{2}\right)-B(0)$ ו־ $B(1)-B\left(\frac{1}{2}\right)$ הם מ"מ ב"ת המפולגים $N\left(0,\frac{1}{2}\right)$ (כטרנספורמציה לינארית של הוקטור הרב־נורמלי (X_1,X_2)). נמשיך ונגדיר:

$$B\left(\frac{1}{4}\right)=\frac{B(0)+B\left(\frac{1}{2}\right)}{2}+\frac{X_3}{\sqrt{8}}$$
$$B\left(\frac{3}{4}\right)=\frac{B\left(\frac{1}{2}\right)+B(1)}{2}+\frac{X_4}{\sqrt{8}}$$

מכך נקבל כי:

$$B\left(\frac{1}{2}\right)-B\left(\frac{1}{4}\right)=\frac{B\left(\frac{1}{2}\right)-B(0)}{2}-\frac{X_3}{\sqrt{8}}\sim N\left(0,\frac{1}{4}\right)$$
$$B\left(\frac{1}{4}\right)-B(0)=\frac{-B(0)+B\left(\frac{1}{2}\right)}{2}-\frac{X_3}{\sqrt{8}}\sim N\left(0,\frac{1}{4}\right)$$

כאשר שוב משתמשים בכך שזוהי טרנפורמציה לינארית של הוקטור הרב־נורמלי $\left(B\left(\frac{1}{2}\right),X_3\right)$. כעת באופן כללי נגדיר:

$$B\left(\frac{2k-1}{2}\right)=\frac{B\left(\frac{k-1}{2^{n-1}}\right)+B\left(\frac{k}{2^{n-1}}\right)}{2}+\frac{X_{2^{n-1}+k}}{2^{\frac{n+1}{2}}}$$

באופן זה הגדרנו את B על הרציונליים הדיאדיים בקטע $[0,1]$ באמצעות הסדרה $\{X_i\}_{i=1}^{\infty}$. נגדיר סדרת פונקציות:

$$F_n(x)=\begin{cases}\frac{X_{2^n+k}}{2^{\frac{n+1}{2}}} & x=\frac{2k-1}{2^n}\\ 0 & x=\frac{2k}{2^n}\\ \text{piecewise linear} & \text{otherwise}\end{cases}$$

זוהי סדרת פונקציות מקריות אשר כל אחת מהן היא רציפה בקטע $[0,1]$ בהסתברות 1 ובפרט לכל שבר דיאדי מתקיים:

$$B\left(\frac{2k+1}{2^n}\right)=\sum_{i=0}^{n}F_i\left(\frac{2k+1}{2^n}\right)$$

בפרט אם הטור הנ"ל מתכנס במידה שווה (בנורמת סופרמום) בהסתברות 1 אז נקבל כי עבור כל שבר דיאדי מתקיים:

$$B\left(\frac{2k+1}{2^n}\right)=\sum_{i=0}^{\infty}F_i\left(\frac{2k+1}{2^n}\right)\overbrace{=}^{\text{def}}F\left(\frac{2k+1}{2^n}\right)$$

מכך נסיק ש־ F היא פונקציה רציפה כגבול במ"ש של פונקציות רציפות והיא גם עבור השברים הדיאדיים בעלת יתר התכונות שרצינו מתנועת בראון . אם כן כדי להראות את ההתכנסות נשים לב כי:

$$\|F_n\|_\infty=\frac{\max\limits_{k=1,\ldots,2^n}|X_{2^n+k}|}{2^{\frac{n+1}{2}}}$$

אם נסמן ב־ $f_X(x)$ את הצפיפות הנורמלית הסטנדרטית נקבל כי לכל $C>0$ מתקיים:

$$\mathbb{P}\left(|X_i|>C\sqrt{n}\right)=\int\limits_{-\infty}^{-\lambda}f_X(x)\,dx+\int\limits_{\lambda}^{\infty}f_X(x)\,dx\le e^{-\frac{C^2n}{2}}$$

מכך באמצעות חסם איחוד נקבל כי:

$$\mathbb{P}\left(\max_{k=1,...,2^n}|X_{2^n+k}|>C\sqrt{n}\right)\leq 2^n e^{-\frac{C^2 n}{2}}$$

הנ"ל דועך אקספוננציאלית ולכן:

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\max_{k=1,...,2^n}|X_{2^n+k}|>C\sqrt{n}\right)<\infty$$

מכך מהלמה הראשונה של בורל קנטלי נסיק כי בהסתברות 1 קיים N כך שלכל $n\geq N$ מתקיים:

$$\|F_n\|_\infty\leq\frac{C\sqrt{n}}{2^{\frac{n+1}{2}}}\overset{n\to\infty}{\longrightarrow}0$$

לכן מקריטריון ויירשטראס הטור מתכנס במ"ש ובפרט F היא פונקציה רציפה. באופן כללי יותר אם $t_1<t_2<...<t_k$ אז נקח סדרות של דיאדיים $t_j^i\overset{i\to\infty}{\longrightarrow}t_j$ ונקבל על סמך מה שהוכחנו שההפרשים באינדקסים הללו הם ב"ת ומפולגים:

$$B\left(t_{j+1}^i\right)-B\left(t_j^i\right)\sim N\left(0,t_{j+1}^i-t_j^i\right)$$

ואפשר להראות שמתקיימת התכנסות של הוקטור הרב־נורמלי

$$\left(B\left(t_2^i\right)-B\left(t_1^i\right),...,B\left(t_k^i\right)-B\left(t_{k-1}^i\right)\right)\overset{i\to\infty}{\longrightarrow}\left(B\left(t_2\right)-B\left(t_1\right),...,B\left(t_k\right)-B\left(t_{k-1}\right)\right)$$

לוקטור רב־נורמלי. כך ש:

$$B\left(t_j\right)-B\left(t_{j-1}\right)\sim N\left(0,t_j-t_{j-1}\right)$$

הערה 8.4 זה בקירוב מוכיח את הקיום של תנועת בראון על $[0,1]$ ואפשר להרחיב את זה לכל הישר על ידי שרשור של תהליכים זהים המוגדרים על כל הקטעים מהצורה $[n,n+1]$. (בנוסף כדי להיות פורמלי לחלוטין צריך לדבר גם על ה־ σ־אלגברה שעובדים איתה במרחב הפונקציות הרציפות וכו').

נתעניין כעת בהתנהגות של $\frac{B(t)}{\sqrt{t}}\sim N\left(0,1\right)$ כאשר t גדול מאוד או קטן מאוד.

טענה 8.5

מתקיים $\limsup_{t\to\infty}\frac{B(t)}{\sqrt{t}}=\infty$ כ"ת.

הוכחה: נתבונן בסדרה $t_n=2^{2^{2^n}}$ אשר עבורה מתקיים $\frac{t_{n+1}}{t_n^2}\overset{n\to\infty}{\longrightarrow}\infty$. אפשר להראות שקיים N כך שלכל $n\geq N$ מתקיים $|B\left(t_n\right)|\leq t_n$ שכן:

$$\mathbb{P}\left(B\left(t_n\right)\geq\lambda\sqrt{t_n}\right)\leq e^{-\frac{\lambda^2}{2}}\Longrightarrow\mathbb{P}\left(B\left(t_n\right)\geq t_n\right)\leq e^{-\frac{t_n}{2}}$$

והמסקנה נובעת מהלמה הראשונה של בורל קנטלי. כעת ידוע לנו כי:

$$B\left(t_{n+1}\right)-B\left(t_n\right)\sim N\left(0,t_{n+1}-t_n\right)$$

לכן לכל $K>0$ קיים a_K קבוע כך ש:

$$\mathbb{P}\left(B\left(t_{n+1}\right)-B\left(t_n\right)>K\sqrt{t_{n+1}}\right)>a_K>0$$

מאחר וזוהי סדרה של מ"מ בלתי־תלויים נסיק מהלמה הראשונה של בורל־קנטלי שעבור אינסוף ערכי n מתקיים:

$$B\left(t_{n+1}\right)=B\left(t_{n+1}\right)-B\left(t_n\right)+B\left(t_n\right)\geq K\sqrt{t_{n+1}}-t_n\geq\frac{K}{2}\sqrt{t_{n+1}}\overset{n\to\infty}{\longrightarrow}\infty$$

■

נשאל כעת מה ההתפלגות של $M:=\max_{t\in[0,1]}B\left(t\right)$, נתבונן בזמן עצירה (בהגדרה מתאימה לתהליך רציף) שנתון על ידי:

$$T_{\lambda}=\min\left\{ t\,|\,B\left(t\right)=\lambda\right\}$$

באופן זה נקבל כי:

$$\mathbb{P}\left(M\geq\lambda\right)=\mathbb{P}\left(T_{\lambda}\leq1\right)$$

מאחר ש־ $B\left(1\right)-B\left(T_{\lambda}\right)\sim N\left(0,1-T_{\lambda}\right)$ וזוהי התפלגות סימטרית נסיק כי:

$$\begin{aligned}\mathbb{P}\left(B\left(1\right)\geq\lambda\,|\,T_{\lambda}\leq1\right)=\frac{1}{2} & \Longrightarrow\frac{\mathbb{P}\left(B\left(1\right)\geq1\right)}{\mathbb{P}\left(T_{\lambda}\leq1\right)}=\frac{\mathbb{P}\left(B\left(1\right)\geq\lambda\,|\,T_{\lambda}\leq1\right)}{\mathbb{P}\left(T_{\lambda}\leq1\right)}=\frac{1}{2}\\
 & \Longrightarrow\mathbb{P}\left(T_{\lambda}\leq1\right)=2\mathbb{P}\left(B\left(1\right)\geq\lambda\right)=\mathbb{P}\left(\left|B\left(1\right)\right|\geq\lambda\right)
\end{aligned}$$

מכך ניתן לראות כי $\max_{t\in[0,1]}B\left(t\right)\sim\left|Z\right|$ כאשר $Z\sim N\left(0,1\right)$.

טענה 8.6

נראה באמצעות מה שהראינו ש־ $\limsup_{t\to\infty}\frac{\left|B\left(t\right)\right|}{\sqrt{t\log t}}=0$.

הוכחה: נגדיר $M\left(t\right):=\max_{s\in[0,t]}\left|B\left(s\right)\right|$ ונקבל באופן דומה למה שעשיה ש־ $M\left(t\right)\sim N\left(0,t\right)$ ומכך:

$$\mathbb{P}\left(M\left(t\right)>K\sqrt{t\log t}\right)\leq e^{-\frac{K^{2}\log t}{2}}=t^{-\frac{K^{2}}{2}}$$

בפרט עבור $t_{n}=n$ נקבל כי:

$$\mathbb{P}\left(M\left(n\right)>K\sqrt{n\log n}\right)=n^{-\frac{K^{2}}{2}}$$

הטור של איברי אגף ימין מתכנס ולכן מבורל קנטלי המאורע ש־ $M\left(n\right)>K\sqrt{n\log n}$ קורה רק מספר סופי של פעמים עבור כל K ומכל המאורע ש־ $\frac{M(n)}{\sqrt{n\log n}}>K$ קורה מספר סופי של פעמים לכל $K>0$ ומכך אפשר להסיק ש־ $\limsup_{t\to\infty}\frac{\left|B\left(t\right)\right|}{\sqrt{t\log t}}=0$. כמו כן אפשר לקחת סדרה עולה אקספוננציאלית ולקבל ש:

$$\limsup_{t\to\infty}\frac{B\left(t\right)}{\sqrt{t\log\log t}}=\frac{1}{\sqrt{2}}<\infty$$

■

עוד כמה תכונות מעניינות של תנועת בראון:

1. התהליך $B^{'}\left(t\right)=tB\left(\frac{1}{t}\right)$ הוא תנועת בראון.
2. התהליך $\frac{1}{a}B\left(a^{2}t\right)$ הוא תנועת בראון לכל a.
3. בכל נקודה t ההסתברות ש־ B גזיר בנקודה t היא אפס.
4. $B\left(t\right)$ לא גזירה באף נקודה בהסתברות 1 (לא נובע מהטענה הקודמת).