פרק 1

מבוא לסטטיסטיקה

1.1 סטטיסטיקה מהי?

הסטטיסטיקה היא מדע העוסק בנתונים כמותיים, איסופם, עיבודם, הצגתם והסקת מסקנות מהם וזאת כדי לסייע בפתרון בעיות מסוגים שונים. בימינו, קשה להעלות על הדעת איזה תחום בחיינו, שאין לו היבט סטטיסטי ואשר יד הסטטיסטיקאים לא נגעה בה. החלטות הנוגעות אישית לכל אחד ואחת מאתנו מבוססות על סטטיסטיקה, כגון: קביעת מחירי מצרכי המכולת הבסיסיים שאנו קונים, הנעשית לפי מדדים וחישובים שונים: מדידת חום, לחץ דם, תכנון התקציב החודשי, וחישוב כמות הקלוריות. כל הפעולות הרגילות האלה מערבות שימוש בסטטיסטיקה באופן אינטואיטיבי. חשוב לציין, שתחומים שלמים של מדע כמו גנטיקה, מדעי התנהגות, סוציולוגיה, רפואה וכו' לא היו קיימים בלי סטטיסטיקה כי היא מהווה שפה טבעית שלהם. השימוש בנתונים סטטיסטיים בעיתונות ותקשורת מסמן גישה מודרנית ורצינית. אבל כאן מתחילים הסיבוכים כי הנתונים המופיעים בתקשורת הם לעיתים קרובות לא מדויקים, לא שלמים ולפעמים אף שגויים או מעוותים, אשר הביא לכך שבציבור הרחב יש שתי דעות מנוגדות לגבי סטטיסטיקה:

- ה"**תמימה**"; אשר מבוססת על כבוד רב למדע כולו ולסטטיסטיקה בפרט, מה שגורם לקבלת הנתונים הסטטיסטים כאל אמת מוחלטת. הדעה הזאת מבוטאת בבדיחה על גבר שקורא בעיתון שכל ילד רביעי בעולם הוא סיני ופונה לאישתו ואומר "איזה מזל שיש לנו רק 3 ילדים". הגישה הזאת מביאה לשימוש בסטטיסטיקה בתחומים שבכלל לא שייכים לה, כמו הוכחות סטטיסטיות לקיום אלוהים וכו'.

- ה"**צינית**"; שמאמינה שבעזרת סטטיסטיקה אפשר להוכיח כל דבר, לכן לנתונים סטטיסטים אין שום משקל. חסידים של הגישה הזאת נוהגים להכריז כי יש 3 סוגים של שקר: שקר, שקר גס וסטטיסטיקה. הגישה הזאת מביאה להזמנות מחקרים סטטיסטיים שלא רק המטרות אך גם התוצאות הסופיות מוזמנות מראש, גישה זו פורחת בתחומי פרסום ופוליטיקה.

שתי הגישות שגויות כי הן מבוססות על אי־ידיעה ואי הבנה של מטרות, הגבלות, גבולות ודרישות של סטטיסטיקה עיונית ושימושית כמדע. כעת עולה השאלה, מה זה סטטיסטיקה? סטטיסטיקה היא המדע והאומנות העוסקת, באיסוף,עיבוד וניתוח נתונים כמותיים על מנת להסיק

מסקנות מהימנות בנוכחות אי־ודאות. המילה "אי־ ודאות" היא מלת מפתח בתורת ההסתברות וסטטיסטיקה. הבדל בין התחומים האלה לתחומים אחרים של מתמטיקה הוא שכאן מסיבות שונות יש לנו רק מידע חלקי על פרמטר שאנו חוקרים. לפעמים זה קורה כי האוכלוסייה יותר מדי גדולה למחקר כולל (למשל, אם המטרה של המחקר היא בדיקת איכות ביצור של מוצר מסיים כאשר אין זמן ותקציב לערוך בדיקה של כל מוצר ומוצר) ולפעמים זה קורה כי האוכלוסייה היא אינסופית (למשל במחקר בתחומי גנטיקה או משחקי הימורים). לשם כך, אנו עורכים מדגם ועל סמכו מנסים לאמוד את הפרמטר הנדרש.

מאילו שלבים מורכב מחקר סטטיסטי? כל מחקר סטטיסטי מורכב משלבים הבאים:

1. הגדרה מדויקת של מטרות המחקר: ראשית כל, מטרתו של המחקר חייבת להיות בעלת משמעות, כי אחרת אף מחקר סטטיסטי לא יעזור. שנית, המטרה יכול להיות: מאוד עיונית (כמו למשל, חישוב מסלולי חלקיקים בפלזמה או הגדרת חוקי תורשה בגנטיקה). או מאוד מעשית (כמו בקרת איכות ביצור או בדיקת שוק לפני שיווק מוצר חדש). העיקר שפרמטרים של המחקר צריכים להיות נתנים למדידה ביחידות מסוימות (הגדרת מטרה היא לא חלק מסטטיסטיקה).

2. בחירת המדגם המייצג: שאלה עיקרית וחשובה אשר עולה בשלב זה היא: איך לבחור את המדגם המייצג? במדעי טבע כאשר אנחנו עורכים סדרת ניסיונות , כל ניסיון הוא חלק מהמדגם המייצג בתנאי שכל המכשירים פועלים נכון והניסוי נערך באופן מדויק. הצפייה באוכלוסייה מסוימת במדעי טבע כבר מחייבת רמה מסוימת של זהירות: עד כמה תוצאות שנתקבלו מתצפית בקבוצה מסוימת נכונות לגבי כל האוכלוסייה? כאן מתחיל תחום נפרד, מרתק ומסובך של סטטיסטיקה שנקרא "תורת הניסיונות". מבחר נכון של מדגם מייצג בפסיכולוגיה וסוציולוגיה זה אומנות בפני עצמה. בקורס זה אין לנו אפשרות לגעת בתחום של תורת הניסיונות, ולשם כך תמיד נניח שהמדגם הוא מייצג ונתעסק רק בעיבוד הנתונים של המדגם.

3. חישוב הפרמטרים הנדרשים במדגם: חישוב פרמטרים של מדגם הוא בד"כ השלב הכי טכני בכל המחקר. בהמשך נגדיר נוסחאות חישוב של הפרמטרים הנפוצים.

4. הערכת משמעות הפרמטר האמיתי מתוצאות המדגם: על תוצאות המדגם מראש אפשר להכריז רק דבר אחד: הערכים של הפרמטרים שנתקבלו מהמדגם הם שונים מהערכים האמיתיים. בשאלה עד כמה הם קרובים לתוצאת אמת אנחנו נתעסק בפרק זה ובהמשך הקורס.

1.2 סטטיסטיקה תיאורית

אנחנו נתחיל ממחקר קצר בתחום משחקי הימורים. זה היה התחום שממנו צמחה תורת ההסתברות במאה השבעה עשרה, כך שאנחנו חוזרים לשורשים של המדע. עד היום משחקי ההימורים משמשים כשדה הנרחב והנוח ביותר למחקר בתורת ההסתברות. במשחקי הימורים רואים את היופי של תורת ההסתברות. הולדות הרמוניה מתוך כאוס. דוגמא: אם מטילים מטבע הוגן אחד אין שום אפשרות לחזות מראש האם נקבל "עץ" או "פלי", אבל אם נטיל 2 טונות של מטבעות הוגנים, אז נקבל טונה אחת של "עצים" וטונה אחת של "פלים". הניסוי שאנו נערוך

הוא הטלת 3 קוביות משחק 60 פעם כאשר אנחנו מעוניינים בסכום שלושת הקוביות בכל הטלה. התוצאות הן:

2,3,6	1,3,5	4,4,5	2,5,6	3,3,6	1,1,5	2,4,5	4,5,6	3,4,5	5,6,6
1,3,4	2,3,3	5,5,6	2,4,5	2,2,5	1,1,5	3,3,3	4,5,6	1,1,4	2,4,5
2,3,4	3,3,6	1,5,6	3,4,6	3,3,3	2,3,5	1,2,4	1,5,5	4,5,6	1,2,3
5,6,6	2,5,6	3,3,4	2,2,5	1,4,5	4,5,6	2,4,6	4,5,5	1,4,5	1,2,6
2,3,4	1,4,4	2,3,4	3,4,5	1,3,4	1,4,5	2,2,3	2,5,6	1,5,6	1,1,4
6,6,6	2,3,6	1,3,4	1,1,2	1,3,5	2,2,6	3,3,4	1,1,4	3,5,6	5,6,6

קשה להסיק מסקנות כאשר התוצאות מוצגות בצורה אקראית. לכן קודם כל, נסדר את הנתונים לפי סכום בסדר עולה, נחשב כמה פעמים מופיע כל סכום ונקבל טבלת שכיחות:

סכום	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	סך־הכל
שכיחות	1	0	4	4	4	11	7	6	7	5	2	4	1	3	1	60

נגדיר מושגים בסיסיים של סטטיסטיקה תיאורית ונדגים אותם על־פי הדוגמא הנ"ל.

- מספר הערכים במדגם נקרא גודל המדגם. בדוגמא שלנו, גודל המדגם הוא 60.
- הפרש בין הערך מקסימלי והערך מינימלי נקרא טווח המדגם. בדוגמא שלנו, הטווח הוא $18-4=14$.
- מספר הפעמים שערך הנתון מופיע במדגם נקרא שכיחות שלו. בדוגמא שלנו, שכיחות של הסכום השווה ל־8 היא 4.
- הערך שחצי מהנתונים גדולים ממנו וחצי־קטנים ממנו נקרא חציון. נכתוב את הערכים שנתקבלו במדגם מגודל n בסדר לא יורד ונסמן אותם $x_1,\dots,x_n$, כאשר $x_1\le\dots\le x_n$:

  1. אם גודל המדגם $n=2k+1$ (ז"א אי־זוגי) אז החציון הוא x_{k+1}.
  2. אם גודל המדגם $n=2k$ (ז"א זוגי) אז החציון הוא $\frac{1}{2}(x_k+x_{k+1})$.

  בדוגמא שלנו החציון הוא $10=\frac{1}{2}(10+10)$.
- הערך (או הערכים) ששכיחותו מקסימלית נקרא הערך השכיח. בדוגמא שלנו הערך השכיח הוא 9.
- הממוצע הוא סכום של כל הערכים מחולק לגודל המדגם. כותבים

$$\overline{x}=\frac{1}{n}(x_1+x_2+\cdots+x_n)=\frac{1}{n}\sum_{i=1}^{n}x_i.$$

כאשר n הוא גודל המדגם ו־ $x_1,\dots,x_n$ הם הערכים של המדגם. שימו לב שמספר הפעמים שכל ערך מופיע בסכום שווה לשכיחות שלו. לכן אם כל ערך y_i נתון עם השכיחות

שלו f_i ומספר ערכים שונים הוא k אז הנוסחה הופכת ל־ $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} f_i y_i$. בדוגמא שלנו מקבלים:

$$\overline{x} = \frac{1}{60}(4+24+28+32+99+70+66+84+65+28+60+16+51+18) = 10.75.$$

- שונות המדגם מוגדר על־ידי הנוסחה:

$$Var = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} f_i(y_i - \overline{x})^2.$$

כאשר n הוא גודל המדגם, $x_1, \ldots, x_n$ הם הערכים של המדגם, $\overline{x}$ הוא הממוצע של המדגם, ו־ y_i הם הערכים השונים בקבוצה $\{x_1, \ldots, x_n\}$. מסקנה, ניתן לראות ששונות היא ממוצע של ריבועי המרחקים בין הערכים לממוצע. בדוגמא שלנו:

$$Var = \frac{1}{60}((4-10.75)^2 + 4(6-10.75)^2 + \cdots + (18-10.75)^2) = 9.7875.$$

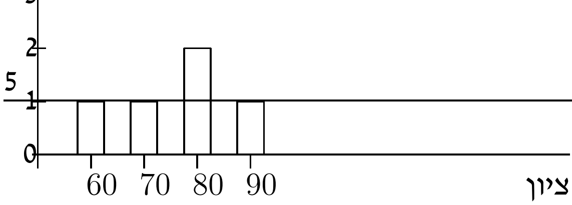
- סטיית התקן של המדגם מוגדרת על־ידי $s = \sqrt{Var}$. בדוגמא שלנו: $s = \sqrt{9.7875} \approx 3.1285$.

הממוצע, החציון והשכיח הם מידות מרכז של המדגם. המידות האלה מראות משהו המדגם אבל צריך להיזהר: התמונה מיוצגת רק על ידי מידות מרכז יכולה להיות שונה מאוד מהאמת, כגון: אם בחדר בית חולים לאחד יש חום $42C$ ולשני $31.2C$ אז החום הממוצע הוא $36.6C$ שזה מצוין למרות שאחד גוסס והשני מת. לכן לא פחות חשובות מידות הפיזור שמראות את פיזור הנתונים סביב המרכז. טווח, ערכים מינימליים ומקסימליים שייכים למידות הפיזור. המידות החשובות בין מידות הפיזור הן שונות וסטיית תקן. הסיבה שאנו מגדירים סטיית תקן היא לחזור ליחידות מידה שמהם התחלנו.

בדרך כלל ממוצע שונה מהחציון. הם תמיד שווים במקרה של התפלגות סימטרית המוגדרת באופן הבא. לסדרה $x_1, x_2, \ldots, x_n$ של מדגם מגודל n יש התפלגות סימטרית אם לכל x_i קטן מהממוצע ומקיים $\overline{x} - x_i = d$ קיים x_j גדול מהממוצע ומקיים $x_j - \overline{x} = d$. אם בסדרה יש k ערכים קטנים מהממוצע ומקיימים $\overline{x} - x_i = d$ אז יש בדיוק k מספרים גדולים מהממוצע ומקיימים $x_j - \overline{x} = d$.

את טבלת השכיחות ניתן לתאר באופן גרפי ע"י היסטוגרמה. היסטוגרמה מכינים בצורה הבא: לכל ערך X מטבלת השכיחות מציירים מלבן שמרכז בסיסו על x וגובהו הוא שכיחות של X. דוגמא:

ציון	שכיחות
60	1
70	1
80	2
90	1
סה"כ	5



כמה מילים ביחס לסימונים, כאשר המדגם הוא האוכלוסייה כולה, כמו למשל מבחני כניסה לאוניברסיטאות נהוג לסמן את הממוצע ב μ וסטיית תקן ב־ σ.

- μ מסמנת את התוחלת או במילים אחרות הממוצע האמיתי של האוכלוסייה
- σ מסמנת את סטיית התקן האמיתית של האוכלוסייה.

לכן כאשר מדגם הוא רק חלק של האוכלוסייה הממוצע מסומן ב־ $\overline{x}$ וסטיית התקן מסומנת ב־ s וזאת כדי להדגיש שהם רק אומדנים של פרמטרים μ ו־ σ.

1.3 תרגילים

תרגיל 1.1 יהיו $x_1,\ldots,x_n$ ו־ $y_1,\ldots,y_m$ שני מדגמים. נניח שהממוצע של המדגם הראשון קטן מהממוצע של המדגם השני: $\overline{x} < \overline{y}$. הוכח או הפרך (על ידי דוגמה נגדית מתאימה) את הטענות הבאות:

(א) קיים x_i וקיים y_j כך ש־ $x_i < y_j$.
(ב) לכל x_i ולכל y_j מתקיים $x_i < y_j$.
(ג) לכל x_i קיים y_j כך ש־ $x_i < y_j$.
(ד)קיים x_i כך שלכל y_j מתקיים $x_i < y_j$.

פתרון:
(א) נכון. אם זה היה לא כך, היה מתקיים לכל x_i ולכל y_j כך ש־ $x_i \geq y_j$. בפרט $\min x_i \geq \max y_j$. אבל $\overline{x} \geq \min x_i$ ו־ $\max y_j \geq \overline{y}$. ולכן $\overline{x} \geq \overline{y}$ בסתירה לנתון.
(ב) לא נכון, למשל נקח $x = (10,0)$ ו־ $y = (7,9)$.
(ג) לא נכון, למשל כמו ב־ (ב).
(ד) לא נכון, למשל נקח $x = (-5,0,5)$ ו־ $y = (-6,0,12)$. □

תרגיל 1.2 ענה על העיפים הבאים:
(א) כתוב מדגם בן 2 מספרים שהממוצע שלו הוא 0 והשונות שלו היא 5.
(ב)הוכח או הפרך: אם למדגם א' סטיית תקן גדולה מסטיית התקן של מדגם ב', אז הטווח של מדגם א' גדול מהטווח של מדגם ב'.
(ג)הוכח או הפרך: אם למדגם א' טווח גדול מהטווח של מדגם ב', אז סטיית התקן של מדגם א' גדולה מסטיית התקן של מדגם ב'.

פתרון:
(א) אם המדגם הוא x,y אז $\frac{x+y}{2} = 0$ ולכן $y = -x$. השונות $\frac{x^2+x^2}{2} = 5$ כלומר $x = \pm\sqrt{5}$. המדגם הוא לכן $\sqrt{5}, -\sqrt{5}$.
(ב) לא נכון, למשל נקח מדגם א' $\{0,1\}$ ומדגם ב' $\{-1,0,0,0,0,0,0,1\}$. אז טווח $2 > 1$ וסטיית תקן $\sqrt{1/4} < \sqrt{2/9}$.
(ג) לא נכון, למשל נקח מדגמים בהפוך מ־ (ב). □

תרגיל 1.3 חברת טלויזיה בלויין עורכת סקר שביעות רצון מסדרת דרמה חדשה. מכלל המנויים נבחר מדגם מייצג של 100 איש, אשר ענו על שביעות רצונם כמספר מ־1 (בכלל לא מעניין) ועד 5 (מעניין ביותר) (אם כך, בידנו מדגם בן 100 מספרים). הנה התוצאות (בעזרת אישור בלעדי מהחברה):

התשובה	5	4	3	2	1
מספר האנשים שבחרובתשובה זו	5	15	30	45	5

הוחלט בדירקטוריון החברה שאם ממוצע המדגם קטן מ־2.5, הסדרה היא כישלון, ואין היא ראויה להצגה בטלויזיה. הלאה, אם תנאי זה לא מתקיים, וסטיית התקן של המדגם תהיה גדולה או שווה ל־2, הסדרה תוצג בערוץ בחירה (בתשלום נוסף). אם גם זה לא מתקיים, הסדרה תוצג בערוץ ללא תשלום. ערכו חישובים, ואמרו מה תהיה החלטת החברה.

פתרון: ממוצע המדגם $\frac{1}{100}(5\cdot 1+45\cdot 2+30\cdot 3+15\cdot 4+5\cdot 5)=2.7$. לכן הסדרה תוצג בטלויזיה. השונות $\frac{1}{100}(5\cdot 1.7^2+45\cdot 0.7^2+30\cdot 0.3^2+15\cdot 1.3^2+5\cdot 2.3^2)=0.91$. סטיית התקן, שורש ריבועי מהשונות, קטנה מ־ 1, ולכן מ־ 2, ולכן הסדרה תוצג בערוץ ללא תשלום. $\square$

תרגיל 1.4 סטודנטים נתבקשו לנתח קבוצה של 50 ערכים אי־שליליים שלא כולם זהים. הקבוצה כללה בדיוק שלושה אפסים ושני ערכים זהים לממוצע.

(א) סטודנט A החליט לא לכלול את האפסים בחישוביו. לכל פרמטר ציין האם התוצאה שתתקבל גדולה, קטנה, שווה או לא ניתן לקבוע ללא מידע נוסף יחסית לפרמטר הנתון: (א) ממוצע, (ב) טווח, (ג) שונות.

(ב) סטודנט B החליט לא לכלול את שני הערכים השווים לממוצע. לכל פרמטר ציין האם התוצאה שתתקבל גדולה, קטנה, שווה או לא ניתן לקבוע ללא מידע נוסף יחסית לפרמטר הנתון: (א) ממוצע, (ב) טווח, (ג) שונות.

פתרון:

(א) הממוצע המקורי היה גדול מאפס, כלומר סטודנט A גרע שלושה ערכים קטנים מהממוצע ובכך הגדיל אותו. הטווח קטן. לגבי השונות אי אפשר להכריע. למשל, עבור $\{0,0,0,4\}$ השונות היא 3, ואילו השונות של $\{4\}$ היא 0, כלומר במקרה זה השונות קטנה. אולם, עבור $\{0,0,0,1,9\}$ השונות היא $\frac{62}{5}$, ואילו השונות של $\{1,9\}$ היא 16, כלומר במקרה זה השונות גדלה.

(ב) הממוצע נשאר זהה. הטווח לא ישתנה: הרי ברור שיש ערכים קטנים מהממוצע (0 למשל) ולכן גם ערכים גדולים מהממוצע, וכאשר נוריד את הממוצע הטווח יישאר זהה. השונות תגדל: אם $\overline{x}$ הממוצע, אז הוא גם הממוצע של הסדרה החדשה ומתקיים

$$Var_{old}=\frac{(\overline{x}-\overline{x})^2+(\overline{x}-\overline{x})^2+\sum^*(x_i-\overline{x})^2}{50}=\frac{\sum^*(x_i-\overline{x})^2}{50}\leq\frac{\sum^*(x_i-\overline{x})^2}{48}=Var_{new}.$$

$\square$

תרגיל 1.5 נניח שלמדגם $x_1,\dots,x_n$ יש ממוצע 0. הראו שהממוצע של המדגם $x_1x_n,\dots,x_2x_{n-1}$, x_nx_1 קטן או שווה לשונות של המדגם $x_1,\dots,x_n$.

פתרון: צריך להראות $\frac{x_1x_n+x_2x_{n-1}+\cdots+x_nx_1}{n} \le \frac{x_1^2\cdots+x_n^2}{n}$. וזה נובע ישירות מאי־שוויון קושי־שוורץ עבור הוקטורים $(x_1,x_2,\ldots,x_n)$ ו־ $(x_n,x_{n-1},\ldots,x_1)$. □

תרגיל 1.6 יהיו $X_1,X_2,\ldots,X_n$ ו־ $Y_1,Y_2,\ldots,Y_m$ שני מדגמים. הוכח או הפרך:

(א) $\overline{X_i^2} \le \overline{Y_i^2}$ גורר ש־ $\overline{\overline{X}-X} = \overline{\overline{Y}-Y}$.

(ב) אם לכל $i=1,2,\ldots,n$, $X_i \ge 0$, אזי $\overline{X} \ge \overline{X^2}$.

(ג) שונושת של X שווה לשונות של $\overline{X}-X$.

(ד) אם לכל $i=1,2,\ldots,m$, $Y_i \ge 0$, אזי $\overline{Y} \ge \frac{1}{\overline{\frac{1}{Y}}}$.

פתרון:

(א) נכון מכיוון ש־

$$\overline{\overline{X}-X} = \frac{1}{n}\sum_{i=1}^{n}(\overline{X}-X_i) = \frac{1}{n}n\overline{X} - \frac{1}{n}\sum_{i=1}^{n}X_i = \overline{X}-\overline{X} = 0,$$

נקבל שמתקיים $\overline{\overline{Y}-Y} = \overline{\overline{x}-x}$.

(ב) לא נכון כי למשל אם $X_1=1$ ו־ $X_2=2$, אז $\overline{X}=\frac{1+2}{2}=1.5$ וגם $\overline{X^2}=\frac{1^2+2^2}{2}=2.5$.

(ג) נכון. שונשות של $\overline{X}-X$ שווה ל־ $\overline{\overline{\overline{X}-X}-(\overline{X}-X)}$, אבל $\overline{\overline{X}-X}=0$ אז

$$\overline{(\overline{\overline{X}-X}-(\overline{X}-X))^2} = \overline{(\overline{X}-X)^2}.$$

(ד) נכון. נוכיח את זה באינדוקציה על m: עבור $m=1$ הטענה שקולה ל־ $\frac{y_1}{1} \ge \frac{1}{\frac{1}{y_1}}$ שהיא טענה נכונה. נניח נכונות הטענה עבור m, כלומר

$$\frac{y_1+\cdots+y_m}{m} \ge \frac{m}{\frac{1}{y_1}+\cdots+\frac{1}{y_m}},$$

ששקול ל־

$$(y_1+\cdots+y_m)\left(\frac{1}{y_1}+\cdots+\frac{1}{y_m}\right) \ge m^2,$$

ואז נקבל ש־

$$\begin{aligned}
&(y_1+\cdots+y_{m+1})\left(\tfrac{1}{y_1}+\cdots+\tfrac{1}{y_{m+1}}\right)\\
&\quad= (y_1+\cdots+y_m)\left(\tfrac{1}{y_1}+\cdots+\tfrac{1}{y_m}\right)+\tfrac{y_1+\cdots+y_m}{y_{m+1}}+y_{m+1}\left(\tfrac{1}{y_1}+\cdots+\tfrac{1}{y_m}\right)+1\\
&\quad\ge m^2+\underbrace{\left(\left(\frac{y_1}{y_{m+1}}+\frac{y_{m+1}}{y_1}\right)+\cdots+\left(\frac{y_m}{y_{m+1}}+\frac{y_{m+1}}{y_m}\right)\right)}_{n}+1\\
&\quad\ge m^2+2m+1=(m+1)^2,
\end{aligned}$$

כלומר

$$(y_1 + \cdots + y_{m+1})\left(\frac{1}{y_1} + \cdots + \frac{1}{y_{m+1}}\right) \geq (m+1)^2,$$

ששקול ל־

$$\frac{y_1 + \cdots + y_{m+1}}{m+1} \geq \frac{m+1}{\frac{1}{y_1} + \cdots + \frac{1}{y_{m+1}}}.$$

□

תרגיל 1.7 יהיו $x_1, x_2, \ldots, x_n$ ו־ $y_1, y_2, \ldots, y_m$ שני מדגמים. הוכח או הפרך:

(א) אם שונות של X גדול משונות של Y אזי קיימים x_i ו־ y_j כך ש־ $x_i \geq y_j$.

(ב) אם ממוצע של X גדול ממוצע של Y אזי קיימים x_i ו־ y_j כך ש־ $x_i \geq y_j$.

(ג) אם שכיח של X גדול משכיח של Y אזי קיימים x_i ו־ y_j כך ש־ $x_i \geq y_j$.

תרגיל 1.8 יהיו $x_1, x_2, \ldots, x_n$ ו־ $y_1, y_2, \ldots, y_m$ שני מדגמים ונניח ש־

$$A = [\min_{1 \leq i \leq n} x_i, \max_{1 \leq i \leq n} x_i]$$

ו־

$$B = [\min_{1 \leq i \leq m} y_i, \max_{1 \leq i \leq m} y_i].$$

הוכח או הפרך:

(א) אם ממוצע של X שווה לממוצע של Y, אז $A \cap B \neq \emptyset$.

(ב) אם ממוצע של X שווה לממוצע של Y וגם $|A \cap B| = 1$, אז

$$x_1 = \cdots = x_n = y_1 = \cdots = y_m.$$

פרק 2

סיגמא אלגברה ומרחב המאורעות

2.1 מבוא

הבה נסקור תחילה את ההגדרות המקובלות במתמטיקה למושגיה היסודיים של תורת הקבוצות, ואחר־כך נסקור ונביא ההגדרות הבסיסות למרחב ההסתברות. לפיכך נפתח בסימונים וטענות מתורת הקבוצות כדלקמן:

- קבוצה אוניברסלית Ω: כל הקבוצות הן תת־קבוצות של Ω. סימון: $A \subset \Omega$ לכל A.
- הקבוצה הריקה $\emptyset$: לכל קבוצה A מתקיים $\emptyset \subseteq A$.
- המשלים $\overline{A}$ של קבוצה : לכל תת־קבוצה נגדיר $\overline{A} = \{x \in \Omega | x \notin A\}$.
- איחוד של שתי קבוצות A ו B נגדיר על ידי $A + B = \{x \in \Omega | x \in A$ או $x \in B\}$.
- חיתוך של שתי קבוצות A ו B נגדיר על ידי $AB = A \cdot B = \{x \in \Omega | x \in A$ גם $x \in B\}$. נאמר ששתי הקבוצות הללו הן קבוצות זרות אם ורק אם $A \cdot B = \emptyset$.
- הפרש קבוצות A ו B, $A - B$, נגדיר על ידי $A - B = \{x \in \Omega | x \in A$ גם $x \notin B\}$.
- קבוצה של תת־קבוצות של Ω נקראת משפחה של קבוצות $\mathbb{A}$ אם $A \in \mathbb{A}$ תת־קבוצה מ $\mathbb{A}$.
- משפחה של תת־קבוצות של Ω, $\{U_i\}_{i \in I}$, היא פירוק של Ω אם מתקיים $\sum_{i \in I} U_i = \Omega$ וגם $U_i U_j = \emptyset$ לכל $i \neq j$ ו $i, j \in I$.

נסמן תת קבוצה של הטבעיים $\mathbb{N}$ ב I. באמצעות הגדרות הללו ניתן להראות את קיום התכונות הבאות. לכל שלושת קבוצות A, B, C ב Ω מתקיים:

$$\begin{array}{lll}
(1)\ A + \emptyset = A & (2)\ A \cdot \emptyset = \emptyset & (3)\ A + \Omega = \Omega \\
(4)\ A \cdot \Omega = A & (5)\ A + \overline{A} = \Omega & (6)\ A \cdot \overline{A} = \emptyset \\
(7)\ A + A = A & (8)\ A \cdot A = A & (9)\ A + B = B + A \\
(10)\ A \cdot B = B \cdot A & (11)\ A + (B + C) = (A + B) + C & (12)\ A \cdot (B \cdot C) = (A \cdot B) \cdot C \\
 & (13)\ (A + B) \cdot C = A \cdot C + B \cdot C & \\
 & (14)\ A \cdot B + C = (A + C) \cdot (B + C) &
\end{array}$$

(2.1)

וגם מתקיימים חוקי דה־מורגן:

$$\overline{A+B} = \overline{A}\cdot\overline{B}, \quad \overline{A\cdot B} = \overline{A}+\overline{B}. \tag{2.2}$$

2.2 סיגמא אלגברה

סיגמא אלגברה או σ־אלגברה היא אחת המושגים המודרניים במתמטיקה שמתאר אוסף של תת־קבוצות של קבוצה מסויימת. במקור, מושג זה הוגדר באנליזה שהוא אחד הענפים הראשיים של המתמטיקה. הוא כלי הכרחי למען הגדרת מידה שהוא בפני עצמו כלי חשוב מאוד באנליזה. ענף אחר של המתמטיקה, הוא תורת ההסתברות שבו משתמשים במושג זה ככלי להגדרת מרחב הסתברות, כפי שנראה. ולפיכך, אנו נתחיל את סעיף זה בהגדרת מושג זה.

הגדרה 2.1 בהינתן קבוצה $\Omega \neq \emptyset$ משפחה לא ריקה $\mathbb{A}$ של תת־קבוצות של Ω. $\mathbb{A}$ נקראת σ־אלגברה מעל Ω אם מתקיים

(i) לכל $A \in \mathbb{A}$, $\overline{A} \in \mathbb{A}$.

(ii) לכל $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathbb{A}$, $\sum_{i\in\mathbb{N}} A_i \in \mathbb{A}$ (האיחודים נמצאים ב ־ $\mathbb{A}$).

במילים אחרות, σ־אלגברה היא משפחה לא ריקה של תת־קבוצות של Ω הסגורה לגבי פעולת ההשלמה ופעולת האיחוד הבן מניה.

טענה 2.1 אם $\Omega \neq \emptyset$, $\mathbb{A}$ σ־אלגברה מעל Ω. אזי

(i) הקבוצות $\emptyset$ ו־ Ω שייכות ל־ $\mathbb{A}$.

(ii) אם $\{A_i\}_{i=1}^n \subseteq \mathbb{A}$, אזי $\sum_{i=1}^n A_i \in \mathbb{A}$ וגם $\prod_{i=1}^n A_i \in \mathbb{A}$.

(iii) אם $A, B \in \mathbb{A}$, אזי $A - B \in \mathbb{A}$.

(iv) אם $\{A_i\}_{i\in I} \subseteq \mathbb{A}$, אזי $\sum_{i\in I} A_i \in \mathbb{A}$ וגם $\prod_{i\in I} A_i \in \mathbb{A}$.

הוכחה: (i) מכיוון $\mathbb{A}$ לא ריקה אז קיימת קבוצה $A \in \mathbb{A}$. לכן שימוש בתכונות של σ־אלגברה ו־ (2.1) נקבל ש־ $A + \overline{A} = \Omega \in \mathbb{A}$ וגם $A \cdot \overline{A} = \emptyset \in \mathbb{A}$.

(ii) מתכונות של σ־אלגברה נותר להראות ש $\prod_{i=1}^n A_i \in \mathbb{A}$. מכיוון שנתון $\{A_i\}_{i=1}^n \subseteq \mathbb{A}$ אז נקבל $\{\overline{A_i}\}_{i=1}^n \subseteq \mathbb{A}$, וגורר $\sum_{i=1}^n \overline{A_i} = \overline{\prod_{i=1}^n A_i}$ על פי חוקי דה־מורגן (2.2), לכן מתכונות של σ־אלדברה נקבל הדרוש.

(iii) מתכונות של σ־אלדברה נקבל גם $B, \overline{A} \in \mathbb{A}$ ואז

$$A - B = A\overline{B} = \overline{\overline{A} + B} \in \mathbb{A}.$$

(iv) כמו (ii). □

דוגמא 2.1 (1) לכל קבוצה $\Omega \neq \emptyset$ קיימת $\mathbb{A} = \{\emptyset, \Omega\}$ הנקראת σ־אלגברה טריוויאלית. אפשר להעיר כי כל σ־אלגברה מעל Ω מכילה את σ־אלגברה טריוויאלית.
(2) $\mathbb{A}_\Omega$ משפחה של כל התת־קבוצות של Ω היא:

- σ־אלגברה: A תת־קבוצה גורר $\overline{A}$ תת־קבוצה. $\{A_i\}_{i\in I}$ תת־קבוצה גורר $\sum_{i\in I} A_i$ תת־קבוצה.

- כל σ־אלגברה מעל Ω מוכלת ב־ $\mathbb{A}_\Omega$. $\mathbb{A}_\Omega$ נקראת σ־אלגברה אוניברסלית מעל Ω.

השאלה הנשאלת עכשיו איך לבנות σ־אלגברה באופן כללי? בו נשיב על השאלה הזו בצורה הבאה. יהי $E \neq \emptyset$־אסוף כלשהוא של תת־קבוצות של Ω קיימת $\mathbb{A}_E$ σ־אלגברה מינימלית שמכילה את E. נקח אסוף של כל ה σ־אלגבראות שמכילות את E. האוסף הזה לא ריק כי $\mathbb{A}_\Omega$ בו. נגדיר $\mathbb{A}_E$ כחיתוך האוסף. אז נקבל את התכונות הבאות:

- האוסף $\mathbb{A}_E$ אינו קבוצה ריקה מכיוון ש־ $E \subseteq \mathbb{A}_E$.

- $\mathbb{A}_E$ היא σ־אלגברה. רואים את זה על ידי: אם $A \in \mathbb{A}_E$ אז הקבוצה A היא נמצאת בכל σ־אלגברה שבאוסף $\mathbb{A}_E$, אז גם $\overline{A}$ היא נמצאת בכל σ־אלגברה שבאוסף $\mathbb{A}_E$. לכן $\overline{A} \in \mathbb{A}_E$. ואם יש לנו אוסף של קבוצות $\{A_i\}_{i\in I} \subseteq \mathbb{A}_E$ אזי האוסף $\{A_i\}_{i\in I}$ מוכל בכל אחת מ־ σ־אלגבראות שבאוסף $\mathbb{A}_E$, אז גם $\sum_{i\in I} A_i$ נמצא בכל אחת מ־ σ־אלגבראות שבאוסף $\mathbb{A}_E$. לכן $\sum_{i\in I} A_i \in \mathbb{A}_E$.

- $\mathbb{A}_E$ היא σ־אלגברה מינימאלית היחידה שמכילה את E. באמת אם $\mathbb{B}_E$ היא σ־אלגברה מינימאלית שמכילה את E איז $\mathbb{A}_E \subseteq \mathbb{B}_E$ ולפי מינימאליות נקבל שמתקיים $\mathbb{A}_E = \mathbb{B}_E$.

למשל, אם $E = \emptyset$ אז $\mathbb{A}_E = \{\emptyset, \Omega\}$, ואם $E = \{A\}$ אז $\mathbb{A}_E = \{\emptyset, \Omega, A, \overline{A}\}$.

2.3 אלגברת המאורעות

אם זורקים צלחת מזכוכית מהחלון של קומה שלישית היא נופלת למטה זה מה שיקרה תמיד וזאת לפי חוק כוח המשיכה. מאורעות כאלה נקראים וודאיים. מאורעות שאי אפשר שיתקיימו נקראים מאורעות בלתי אפשריים. כגון, אף צלחת זכוכית שנזרקה מהחלון של קומה שלישית לא מרחפת עם הרוח בשמיים. לכן זהו מאורע בלתי אפשרי. תוצאות של הניסוי "זריקת צלחת מהחלון של קומה שלישית" מוגדרות מראש. אך המציאות לא תמיד מוגדרת באופן מוחלט ולא תמיד אפשר לצפות מראש את התוצאות של הניסוי, כי העולם שבו אנו חיים ופועלים הוא עולם של אי־ודאות. באופן קבוע אנו מנסים לצמצם ככל האפשר את אלמנט אי־ הוודאות, תוך רצון לצפות את התוצאות של תופעות ותהליכים אקראיים המתרחשים סביבנו. כיון שאין דרך לקבוע בוודאות את תוצאות של תהליך אקראי, אנו מנסים, לחלופין, להעריך את הסיכויים של התוצאות האפשריות השונות, כדי שנוכל להגיע להחלטות טובות ככל האפשר. נסתכל בשתי השאלות הבאות:

(1) האם משה יעבור את הבחינה בהסתברות?

(2) האם בהטלת קוביה נקבל 6?

כפי שניתן לראות, התשובות לשאלות הללו אי אפשר לצפות מראש. אך יש הבדל גדול בין הניסויים הקשורים למשה לבין הניסויים הקשורים להטלת קוביה, הניסוי הראשון הוא חד פעמי, כי אם משה נכשל בבחינה אז הוא יתחיל להתכונן למועד ב' ואז הסיכוי שלו ישתנה. ואילו אם מטילים קוביה הוגנת הרבה פעמים אז יחס של הטלות שבהם הקוביה נופלת על צלע "6" למלה כלל הטלות יהיה קרוב ל־ $\frac{1}{6}$. במקרה הזה ניתן לומר שיש למאורע יציבות סטטיסטית שנותנת לנו אפשרות לחזור על הניסוי מספר פעמים אשר נרצה.

הגדרה 2.2 מאורע נקרא מקרי אם הוא תוצאה של ניסוי שאפשר לחזור עליו מספר פעמים אשר נרצה וסיכויו לא משתנים מניסוי לניסוי. בהמשך אנו נקרא למאורע מקרי פשוט מאורע.

למשל,

(1) קבלת מספר זוגי בהטלת קוביה.

(2) סוג דם של בן אדם במדגם.

(3) מספר α חלקיקים שנקלטו בקולט בקרינה של אותו חומר במשך זמן מסוים.

הגדרה 2.3 מרחב מדגם Ω של ניסוי הסתברותי הוא קבוצה של כל התוצאות האפשריות של הניסוי. כל תוצאה (ז"א כל איבר בודד ב־ Ω) נקראת מאורע אלמנטרי.

דוגמא 2.2 (1) הטלת קובית משחק: $\Omega=\{1,2,3,4,5,6\}$ ויש 6 מאורעות אלמנטריים. (2) בדיקת סוג דם של בן־ אדם במזגם: $\Omega=\{O,A,B,AB\}$ ויש 4 מאורעות אלמנטריים. (3) מספר חלקיקים α שנקלטו בקולט בקרינה של החומר הנתון: $\Omega=\{1,2,\ldots\}$ ויש אינסוף מאורעות אלמנטריים.

הגדרה 2.4 אלגברת מאורעות $\mathbb{A}$ היא משפחה של תת־ קבוצות של Ω שהיא σ אלגברה. וכל איבר של $\mathbb{A}$ נקרא מאורע.

למה מגדירים $\mathbb{A}$ כ־ σ־אלגברה?

הדרישה שאם A מאורע אז גם $\overline{A}$ מאורע נראית טבעית בהחלט כי $\overline{A}$ אומר רק שלא קיבלנו את A כתוצאה מהניסוי. גם ברור שאיחוד סופי של מאורעות צריך להיות מאורע. אבל למה אנו צריכים שגם איחוד אין־סופיים של מאורעות להיות מאורע? האקסיומה השניה של σ־אלגבראות נראית יותר מלכותית, אבל תהיה הכרחית במקרים כאשר σ־אלגברה מכילה מספר אינסופי של מאורעות אלמנטריים. זה קורה לעיתים קרובות בתורת ההסתברות למשל, אם הניסוי הוא הטלת קוביה הוגנת עד קבלת 6 בפעם הראשונה זאת אומרת $\Omega=\{E_1,E_2,\ldots\}$ כאשר E_i הטלה ראשונה של 6 הייתה בפעם ה־ i. נגדיר $E=\sum_{i=1}^{\infty}E_i$. מאורע זה קרה אחרי מספר סופי של הטלות! כפי שניתן לראות זה הופך לדרישה טבעית.

הגדרה 2.5 תהי Ω־מרחב מדגם ו־ $\mathbb{A}$־אלגברת מאורעות.

1. $A,B\in\mathbb{A}$ נקראים מאורעות זרים אם $AB=\emptyset$.

2. מאורע $\emptyset$ נקרא בלתי אפשרי.

3. מאורע Ω נקרא וודאי.

4. $\{U_i\}_{i\in I}$ נקראים פירוק של Ω אם מתקיימים שני התנאים הבאים:

(א) $\sum_{i\in I} U_i = \Omega$

(ב) $U_iU_j = \emptyset$ לכל $i \neq j \in I$ (I ־סופי או אינסופי).

דוגמא 2.3 נביא שתי דוגמאות. (א) הניסוי הוא הטלת קובית משחק. מרחב המדגם הוא $\Omega = \{1,2,3,4,5,6\}$, ו־ $\mathbb{A}$ היא ה־σ־אלגברה של כל התת־קבוצות של Ω ($|\mathbb{A}| = 2^6$). נגדיר את המאורע הבא: A={קבלת מספר זוגי בהטלה}, B={קבלת 1 או 3 בהטלה}, ו־ C={קבלת 5 בהטלה}. לכן (1) A ו־ B הן קבוצות זרות. (2) $\{A,B,C\}$ פרוק של Ω.

(ב) הניסוי הוא הטלת קוביה עד קבלת 6. מרחב המדגם הוא $\Omega = \{E_1, E_2, \ldots\}$ ו־ $\mathbb{A}$ היא ה־σ־האלגברה של כל התת־קבוצות של Ω ($\aleph = 2^{\aleph_0} = |\mathbb{A}|$). נגדיר A={6 הופיע ב 10 הטלות ראשונות}, B={6 הופיע אחרי הטלה 15}, ו־ C={6 הופיע בין הטלה 11 להטלה 15}. אזי $\{A,B,C\}$ הוא פירוק של Ω כי: $A = \sum_{i=1}^{10} E_i$, $B = \sum_{i=16}^{\infty} E_i$, $C = \sum_{i=11}^{15} E_i$.

2.4 תרגילים

תרגיל 2.1 הוכח את הזהויות הבאות: $(a,b) = \cup_{n=1}^{\infty}(a, b-1/n) = \cup_{n=1}^{\infty}[a+1/n, b)$ ו־ $[a,b] = \cap_{n=1}^{\infty}[a, b+1/n) = \cap_{n=1}\infty(a-1/n, b]$.

תרגיל 2.2 נתון ש־ A ו־ B שתי קבוצות זרות. הוכח גם הקבוצות $A\cap C$ ו־ $B\cap C$ זרות. מה עם הקבוצות $A\cup C$ ו־ $B\cup C$?

תרגיל 2.3 הוכח את הזהות הבאה $A\cap(\cup_i B_i) = \cup_i(A\cap B_i)$.

תרגיל 2.4 תהי $\{A_n\}_{n\geq 1}$ סדרה של קבוצות כלשהי. (א) הוכח או הפרך ש־ $\cap_{i=1}^{\infty} A_i \neq \emptyset$ אם נתון $\cap_{i=1}^{n} A_i \neq \emptyset$ עבור כל $n \geq 1$. (ב) הוכח או הפרך ש־ $\cup_{i=1}^{\infty} A_i \subset B$ אם נתון $\cup_{i=1}^{n} A_i \subset B$ עבור כל $n \geq 1$

תרגיל 2.5 תהי Ω קבוצה, ו־ $X \subseteq \Omega$ תת־קבוצה. תהי $\mathbb{A}$־אלגברה על Ω. הראו שהאוסף

$$\mathbb{A}_X = \{B\cap X | B \in \mathbb{A}\}$$

מהווה σ־אלגברה על X.

פתרון: יש לבדוק את האסיומות: (1) $X \in \mathbb{A}_X$: $X = \Omega\cap X$ ואילו $\Omega \in \mathbb{A}$. (2) אם $A \in \mathbb{A}_X$, למשל $A = B\cap X$ עם $B \in \mathbb{A}$, אז $X - A = (\Omega - B)\cap X'$ ולכן $X - A \in \mathbb{A}_X$, כלומר המשלים של A בתוך X גם ב־ σ־אלגברה. (3) אם $A_i \in \mathbb{A}_X$, $i = 1,2,\ldots$, ו־ $A_i = B_i\cap X$ אז $\cup_i A_i = \cup_i B_i \cap X$ ולכן $\cup_i A_i \in \mathbb{A}_X$. $\square$

תרגיל 2.6 תהי $\mathbb{A}$ σ־אלגברה על קבוצה X. האם יכול להיות ש־ $|\mathbb{A}| = \aleph_0$?

פתרון: לא יכול להיות. קיימת משפחה בת מנייה אינסופית של קבוצות $\{X_1, X_2, \ldots\}$ מ־ $\mathbb{A}$ כך שכל אחת מכילה איבר שאף אחת אחרת אינה מכילה (נראה זאת למטה). נוכל להגדיר פונקציה $2^{\mathbb{N}} \to \mathbb{A}$ על ידי $N_0 \mapsto \cup_{n \in N_0} A_n$, $N_0 \in 2^{\mathbb{N}}$, (כאן לתקינות ההגדרה נזקקים ל־ σ־אדיטיביות). פונקציה זו חח"ע כי אם $N_0 \neq N_1$ אז קיים למשל $n \in N_0$, $n \notin N_1$ ואז לפי התנאי על המשפחה איבר כלשהו ב־ A_n שייך לתמונת N_0 אבל לא לתמונת N_1, כלומר התמונות לא שוות. מהיות ההעתקה חח"ע נסיק $|\mathbb{A}| \geq 2^{\aleph_0} > \aleph_0$.

נראה שקיימת המשפחה שתוארה. נגדיר יחס (שיתברר כיחס שקילות) על Ω כך:

$$a \sim b \Leftrightarrow \forall X \in A (a \in X \to b \in X),$$

(כל X ב־ $\mathbb{A}$ המכיל את a חייב להכיל את b). ברור שהיחס רפלקסיבי וטרנזיטיבי. הוא גם סימטרי: נניח $a \sim b$ ותהי $X \in \mathbb{A}$ כך ש־ $b \in X$. אז אם בסתירה $a \notin X$ יהיה $a \in \overline{X}$ ולכן (מ־ $a \sim b$) $b \in \overline{X}$, סתירה ל־ $b \in X$. לכן היחס הוא יחס שקילות. קל להראות שכל $X \in \mathbb{A}$ ניתן להציג כאיחוד מחלקות שקילות של יחס זה (כל המחלקות שלא זרות לו); לכן מספר מחלקות אלו לא יכול להיות סופי (אחרת מספר הקבוצות ב־ $\mathbb{A}$ היה לכל היותר 2 חזקת מספר זה, סתירה להיות $\mathbb{A}$ אינסופית). אם כך נוכל לקחת $\aleph_0$ מחלקות שקילות שונות $E_1, E_2, \ldots$, ובכל אחת איבר $e_1, e_2, \ldots$. לכל $i \neq j$, $i, j \in \mathbb{N}$ נוכל למצוא $X_{ij} \in \mathbb{A}$ המקיים $e_i \in X_{ij}$, $e_j \notin X_{ij}$ (אחרת באותה מחלקת שקילות). אז $X_i = \cap_{i \neq j} X_{ij}$ יהיו המשפחה הרצויה (e_i נמצא ב־ X_i ולא נמצא בשאר). $\square$

תרגיל 2.7 אם $X \in \{(a,b), (a,b], [a,b), [a,b]\}$ כאשר $a, b, \in [0,1]$ נאמר ש X $segment$. הוכח או הפרך את הטענות הבאות:

א. $\{A_1 \cup \cdots \cup A_n | n \in \mathbb{N}, A_i \ is \ segment\}$ היא σ־אלגברה.

ב. $\{(A_1 \cup \cdots \cup A_n) \cap \mathbb{Q} | n \in \mathbb{N}, A_i \ is \ segment\}$ היא σ־אלגברה.

תרגיל 2.8 יהיו $\mathcal{J}_1, \mathcal{J}_2$ שתי σ־אלגבראות. הוכח או הפרך את הטענות הבאות:

א. $\mathcal{J}_1 \cap \mathcal{J}_2$ היא σ־אלגברה.

ב. $\mathcal{J}_1 \cup \mathcal{J}_2$ היא σ־אלגברה.

ג. $\mathcal{J}_1 \backslash \mathcal{J}_2$ היא σ־אלגברה.

ד. $\mathcal{J}_1 \triangle \mathcal{J}_2$ היא σ־אלגברה.

תרגיל 2.9 יהי $\mathcal{J}$ σ־אלגברה על X ונניח ש־ $|\mathcal{J}| = \aleph_0$. לכל נקודה $a \in X$ נגדיר הקבצוה $X_a = \cap_{a \in A \in \mathcal{J}} A$. הוכח שלכל $a \in X$ מתקיים $X_a \in \mathcal{J}$, וגם שהקבוצה $\{X_a | a \in X\}$ הוא פירוק אינסופי של X.

פרק 3

הסתברות

3.1 מבוא

ז'ירלמו קארדאנו

רשמית תורת ההסתברות נולדה בחצי הראשון של המאה השבע העשרה בהתכתבות בין בלז פקסל ופייר דה פרמה. אבל כבר במאה ה־16 לפני ההתכתבות המפורסמת ,המתמטיקאי הגדול ביותר של התקופה, ז'ירלמו קארדאנו פתר בעיות ראשונות שכולן היו קשורות למשחקי הימורים ופרסם ספר בשם "libre de ludo aleae" ("הספר על משחקי הימורים"). יש לציין שקרדאנו היה מהמר מכור כך שהימורים גזלו הרבה מזמנו, מכספו ומהמוניטין שלו. אבל כנראה קארדאנו הקדים את תקופת ההסתברות וההתעניינות האמיתית והרחבה בתורת ההסתברות אשר התעוררה רק במאה הבאה כאשר גדולי המדענים כמו פרמה, פסקל וגלילאי גלילאו לא רק פתרו הבעיות (שגם היו קשורות למשחקי הימורים) אלא גם ניסחו את היסודות של המדע החדש.

במשחקי הימורים (הוגנים) תמיד יש n תוצאות שונות אפשריות באותה מידה. למשל, בהטלת מטבע הסיכוי לקבלת עץ שווה לסיכוי של פלי, וכמו כן, בהוצאת קלף מחפיסת ברידג' הסיכוי להוציא 6 יהלום שווה לסיכוי להוציא מלך לב וכו'. לכן ההגדרה הראשונה "הקלאסית" של הסתברות מתייחסת רק למרחבי מאורעות "סימטריים" זאת אומרת כאלה שכל המאורעות האלמנטריים במרחב מדגם אפשריים באותה מידה.

הגדרה 3.1 אם במרחב מדגם יש n מאורעות אלמנטריים אפשריים באותה מידה, אז ההסתברות של כל מאורע אלמנטרי היא $\frac{1}{n}$. אם מאורע A כולל m מאורעות אלמנטריים אז ההסתברות שלו היא $\frac{m}{n}$.

סימון 3.1 נסמן ב־$P(A)$ את ההסתברות של מאורע A. לפי ההגדרה רואים כי: "לכל מאורע A מתקיים: $0 \le P(A) \le 1$; הסתברות המאורע הוודאי היא 1, והסתברות המאורע הבלתי אפשרי היא אפס.

דוגמא 3.1 (פרדוקס דה מרה) זוהי הבעיה המפורסמת מימים הראשוניים של תורת ההסתברות. שבליה דה מרה־ המהמר שחיי במאה ה־17 התעניין בתורת ההסתברות השימושית. הוא ערך

ניסיונות כדי לבדוק הסתברויות באופן מעשי וניסח (בהתכתבות עם פסקל) כמה בעיות חשובות. להלן אחת מהן: אחרי חישובים דה־מרה קיבל שסיכויו לקבל הסכום 11 בהטלת 3 קוביות משחק שווה לסיכויו לקבל הסכום 12 וזאת כי שני הסכומים אפשר לקבל ב 6 דרכים:

$$11 : \{(1,4,6),(1,5,5),(2,3,6),(2,4,5),(3,3,5),(3,4,4)\}$$
$$12 : \{(1,5,6),(2,4,6),(2,5,5),(3,3,6),(3,4,5),(4,4,4)\}$$

אבל בבדיקות מעשיות יצא שסיכוי 11 יותר גבוה. פסקל מצא פתרון לפרדוקס דה־מרה: מרחב המאורעות הסימטרי של הניסוי הזה הוא אוסף של כל השלשות האפשריות כאשר מבדלים בין הקוביות. סך הכל המרחב הכולל הוא $6^3 = 216$ מאורעות.

מאורע $A = \{$בקוביה אחת קיבלנו 1, בקוביה אחרת 4 ובקוביה שלישית קיבלנו 6$\}$ כולל 6 מאורעות אלמנטריים $A = \{(1,4,6),(1,6,4),(4,1,6),(4,6,1),(6,1,4),(6,4,1)\}$ כמו כן, כל מאורע של קבלת 3 מספרים נתונים (בין 1 ל־ 6) שונים בהטלת 3 קוביות כולל 6 מאורעות אלמנטריים.

מאורע $B = \{$בקוביה אחת קיבלנו 1 ובשתי קוביות אחרות קיבלנו 5$\}$ כולל 3 מאורעות אלמנטריים $B = \{(1,5,5),(5,1,5),(5,5,1)\}$ כמו כן, כל מאורע של קבלת 2 מספרים נתונים (בין 1 ל־6) שונים בהטלת 3 קוביות כולל 3 מאורעות אלמנטריים.

מאורע $C = \{$קיבלנו 4 בכל שלוש הקוביות$\}$ כולל מאורע אלמנטרי יחיד $C = \{(4,4,4)\}$. כמו כן, כל מאורע של קבלת מספר נתון בין 1 ל־6 בכל השלוש קוביות הוא מאורע אלמנטרי.

לכן מאורע $D = \{$קבלת הסכום 11$\}$ כולל $6+3+6+6+3+3 = 27$ מאורעות אלמנטריים ומאורע $E = \{$קבלת הסכום 12$\}$ כולל $6+6+3+3+6+1 = 25$ מאורעות אלמנטריים. לפי הגדרה קלאסית של הסתברות $P(D) = \frac{27}{216}$ ו־ $P(E) = \frac{25}{216}$ שמסביר פרדוקס דה מרה.

אחרי שתורת ההסתברות יצאה מתחום משחקי ההימורים בלבד היא התחילה להתעסק בשאלות הקשורות למרחבים שמאורעות אלמנטריים שאפשרותן אינה באותה מידה. זה קרה בסוף המאה השמונה־עשרה והיה קשור לתחילת העידן הסטטיסטי. בגלל המורכבות של הבעיות החדשות היה קשה או בלתי אפשרי להגדיר את הסתברויות של מאורעות אלמנטריים משיקולים שונים מראש. לכן הופיע הצורך בהגדרה יותר כללית שתאפשר לפתור בעיות שימושיות בתורת ההסתברות. ומכאן נולדה ההגדרה הסטטיסטית הבאה.

הגדרה 3.2 (סטטיסטית) אם ידוע שלמאורע יש יציבות סטטיסטית ונתונים מראים ששכיחות של מאורע A היא m מתוך n ניסיונות אז מגדירים $P(A) = \frac{m}{n}$.

דוגמא 3.2 1. במשך יותר מ־ 400 שנה באירופה רשמו את כל התינוקות כולל מינם. הנתונים הם אחידים לתקופות ומדינות שונות. לפי הנתונים נולדים 51% בנים ו־ 49% בנות. לפי הגדרה לעיל ההסתברויות הן: $P(B) = 0.51$, $P(G) = 0.49$.

2. לפי נתונים סטטיסטים לקבוצות דם באירופה השכיחויות הן:

קבוצת דם	שכיחות באחוזים
O	33%
A	42%
B	18%
AB	7%

וההסתברויות הן: $P(AB)=0.07$, $P(B)=0.18$, $P(A)=0.42$, ו־ $P(O)=0.33$.

3.2 פונקצית הסתברות

בשנות ה־ 20 של מאה XX היסודות האקיומתית של תורת ההסתברות נבנו. הגדרה כללית אקסיומתית שנשתמש בה היא כדלקמן.

הגדרה 3.3 יהי Ω־מרחב מדגם ו־ $\mathbb{A}$־אלגברה מאורעות מעל Ω. פונקציה $P:\mathbb{A}\to[0,1]$ המקיימת את האקסיומות הבאות:

א. $P(\Omega)=1$.

ב. אם $\{A_i\}_{i\in I}\subseteq\mathbb{A}$ זרים אזי $\sum_{i\in I}P(A_i)=P\left(\sum_{i\in I}A_i\right)$,

נקראת פונקצית הסתברות. השלישייה $(\Omega,\mathbb{A},P)$ נקראת מרחב הסתברות.

אקסיומה (2) נקראת התכונה האדיטיבית של פונקצית ההסתברות. ועכשיו השאלה הנשאלת האם לכל אלגברת מאורעות $\mathbb{A}$ מעל Ω קיימת פונקצית הסתברות?

1. אם Ω ניתן למניה $\Omega=\{x_i\}_{i\in I}$ אזי לכל $\mathbb{A}$ אפשר למצוא P. לכן מספיק להגדיר $P_i=P(x_i)$ כאשר $0\le P_i\le 1$ לכל $i\in I$ כאלה ש־ $\sum_{i\in I}P_i=1$ ומקבלים פונקצית הסתברות ל־ $\mathbb{A}_\Omega$ ומכאן מקבלים פונקצית הסתברות לכל σ־אלגברה מעל Ω.

2. אם Ω לא ניתן למניה אז אי אפשר להגדיר פונקציה הסתברות ל־ $\mathbb{A}_\Omega$. כדוגמא, זורקים נקודה על הישר באופן אחיד. מובן ש־ $P(x)=0$ לכל $x\in\mathbb{R}$ מכאן רואים שיש קושי להגדיר $P:\mathbb{A}_\Omega\to[0,1]$ (לאיחוד סופי או ניתן למניה של נקודות ב Ω הסתברות היא אפס אבל מה לעשות עם למשל קטעים ?)

משפט 3.1 (תכונות של פונקצית ההסתברות) יהי $(\Omega\,\mathbb{A},P)$ מרחב הסתברות, אזי

א. $P(\emptyset)=0$.

ב. לכל $A\in\mathbb{A}$ מתקיים $P(\overline{A})=1-P(A)$.

ג. אם $A,B\in\mathbb{A}$ ו־ $B\subset A$ אזי $P(A-B)=P(A)-P(B)$.

ד. לכל $A,B\in\mathbb{A}$ מתקיים $P(A-B)=P(A)-P(AB)$.

ה. אם $A,B\in\mathbb{A}$ ו־ $B\subset A$ אזי $P(B)\le P(A)$ (מונוטוניות של ההסתברות).

ו. לכל $\{A_i\}_{i\in I}\subseteq\mathbb{A}$ מתקיים $P\left(\sum_{i\in I}A_i\right)\le\sum_{i\in I}P(A_i)$.

ז. לכל $A,B\in\mathbb{A}$ מתקיים $P(A+B)=P(A)+P(B)-P(AB)$.

ח. (א) אם $\{A_i\}_{i\in I}\subseteq\mathbb{A}$ כזאת ש־ $A_i\subset A_{i+1}$ אזי $P\left(\sum_{i=1}^{\infty}A_i\right)=\lim\limits_{i\to\infty}P(A_i)$.

(ב) אם $\{A_i\}_{i\in I}\subseteq\mathbb{A}$ כזאת ש־ $A_{i+1}\subset A_i$ אזי $P\left(\prod_{i=1}^{\infty}A_i\right)=\lim\limits_{i\to\infty}P(A_i)$.

הוכחה:

1. $\Omega, \emptyset$ שתי קבוצות זרות לכן לפי אקסיומה (2) מתקיים
$$P(\Omega) = P(\Omega + \emptyset) = P(\Omega) + P(\emptyset) \Rightarrow P(\emptyset) = P(\Omega) - P(\Omega) = 0.$$

2. $A, \overline{A}$ שתי קבוצות זרות אזי
$$1 = P(\Omega) = P(A + \overline{A}) = P(A) + P(\overline{A}) \Rightarrow P(\overline{A}) = 1 - P(A).$$

3. $B \subset A$ ו־ $A - B, B$ שתי קבוצות זרות, אזי
$$P(A) = P(B + (A - B)) = P(B) + P(A - B) \Rightarrow P(A - B) = P(A) - P(B).$$

4. $A = A \cdot \Omega = A \cdot (B + \overline{B}) = AB + (A - B)$ אזי $P(A) = P(AB) + P(A - B)$, לכן $P(A - B) = P(A) - P(AB)$.

5. אם $B \subset A$ אזי מכיוון ש־ $P(E) \geq 0$ לכל E נקבל ש־ $P(A) = P(B) + P(A - B) \geq P(B)$.

6. קל להראות ש־
$$\sum_{i \in I} A_i = A_1 + (A_2 - A_1) + (A_3 - (A_1 + A_2)) + \ldots = A_1 + \sum_{i \in I,\, i > 1} \left(A_i - \sum_{j=1}^{i-1} A_j \right).$$
היתרון בביטוי האחרון שקיבלנו הוא שכל המאורעות הן קבוצות זרות בזוגות זה לזה. ולכן לפי אקסיומה (2) ותכונה 5 נקבל
$$P\left(\sum_{i \in I} Ai \right) = P(A_1) + \sum_{i \in I,\, i > 1} P\left(A_i - \sum_{j=1}^{i-1} A_j \right) \leq \sum_{i \in I} P(A_i).$$

7. מעובדה ש־ $A + B = A + (B - A)$ ולפי אקסיומה 2 ותכונה 4 מקבלים
$$P(A + B) = P(B + (A - B)) = P(B) + P(A - B) = P(A) + P(B) - P(A \cdot B).$$

8. (א) עבור כל n (סופי) מתקיים $P\left(\sum_{i=1}^{n} A_i\right) = P(A_n)$. לכן, על פי תכונה (5) נקבל $P(A_n)$ סדרה מונוטונית לא יורדת וחסומה על ידי 1, אז היא מתכנסת ונקבל
$$P\left(\sum_{i=1}^{\infty} A_i \right) = \lim_{n \to \infty} P\left(\sum_{i=1}^{n} A_i \right) = \lim_{n \to \infty} P(A_n).$$
(ב) אנו יודעים ש־ $A_{i+1} \subset A_i$ לכל $i \in I$. אזי $\overline{A_i} \subset \overline{A_{i+1}}$ לכל $i \in I$. לכן,
$$\begin{aligned} P\left(\prod_{i=1}^{\infty} A_i\right) &= P\left(\overline{\left(\sum_{i=1}^{\infty} \overline{A_i}\right)}\right) = 1 - P\left(\sum_{i=1}^{\infty} \overline{A_i}\right) \\ &= 1 - \lim_{n \to \infty} P(\overline{A_n}) = \lim_{n \to \infty}(1 - P(\overline{A_n})) = \lim_{n \to \infty} P(A_n). \end{aligned}$$

כנדרש. $\square$

3.3 מרחב מדגם בדיד, ושיטות קומבינטוריות

הגדרה 3.4 אם מרחב מדגם Ω מכיל מספר שניתן להימנות של נקודות (זאת אומרת, $|\Omega| \leq \aleph_0$), אזי מרחב ההסתברות $(\Omega, \mathbb{A}, P)$ נקרא בדיד ופונקציה ההסתברות P נקראת הסתברות בדידה.

הערה 3.1 יהי $\Omega = \{x_i\}_{i\in I}$ מרחב מדגם בדיד, כדי בניית פונקצית הסתברות $P : \Omega \to [0,1]$ כך ש־ $\sum_{i\in I} P(x_i) = 1$, נרחיב את הפונקציה P על ידי $P : \mathbb{A}_\Omega \to [0,1]$ כך שלכל $A = \{x_{i_j}\}_{j\in J} \subseteq \mathbb{A}_\Omega$, $P(A) = \sum_{j\in J} P(x_{i_j})$. אז כל σ־אלגברה $\mathbb{A}$ היא תת־אלגברה של $\mathbb{A}_\Omega$ ולכן ניתן להגדיר $P : \mathbb{A} \to [0,1]$ לכל σ־אלגברה.

הגדרה 3.5 אם מרחב המדגם $\Omega = \{x_i\}_{i=1}^n$, $n \geq 2$, שעבורו מתקיים $p(x_i) = \frac{1}{n}$ לכל $i = 1, 2, \ldots, n$, אזי Ω נקרא מרחב מדגם סימטרי והפונקציה P נקראת הסתברות אחידה. ואומרים גם שהמרחב $(\Omega, \mathbb{A}, P)$ הוא מרחב שווה הסתברות. לעיתים קרובות מנוסחת ההנחה של הסתברות אחידה על ידי משפט הבא "נבחר איבר מתוך קבוצה S באופן מקרי".

דוגמא 3.3 בניין של 8 קומות ובמעלית נמצאים שני אנשים שנכנסו בקומה ראשונה. מהי ההסתברות שהם ייצאו בקומות שונות, אם כל אחד יוצא באופן מקרי (בלי תלות)? נציע שני אופנים לפתרון.

1. מרבח המדגם: כל האפשריות $\Omega = \{(i,j)|2 \leq i,j \leq 8\}$. נתבונן במאורע
$$A = \{(i,j)|2 \leq i,j \leq 8,\, i \neq j\}.$$
אז $|\Omega| = 7 \cdot 7 = 49$ ו־ $|A| = |\Omega| - |\{(i,i)|2 \leq i \leq 8\}| = 49 - 7 = 42$. לכן $P(A) = \frac{42}{49} = \frac{6}{7}$.

2. בן אדם A' יכול לצאת בכל אחת מ 7 הקומות, לבן אדם B' נשאר בכל מקרה 6 מתוך 7 אפשרויות. $P(A) = P(A') \cdot P(B') = 1 \cdot \frac{6}{7} = \frac{6}{7}$.

3.3.1 תזכורת מקומבינטוריקה

נתונה קבוצה של n מספרים, נאמר ש־ $\{1, 2, \ldots n\}$. כל n־יה מסודרת של מספרים האלה נקראת תמורה. ואוסף של כל התמורות מהווה חבורה שנקראת החברוה הסימטרית מסדר n ונסמנה ב־ S_n. ברור שמספר איברי החבורה הסימטרית מסדר $n!$ הוא נתון על ידי $|S_n| = n!$.

כל k־יה מסודרת כאשר $k \leq n$ נקראת חליפה של k מספרים מתוך n. מספר של חליפות כאלה מסומן על ידי

$$A_n^k = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}.$$

כל k־יה לא מסודרת כאשר $k \leq n$, ז"א תת־קבוצה של k מספרים מתוך הקבוצה $\{1, 2, \ldots n\}$, נקראת צירוף ומספר צירופים כאלה הוא $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. למשל, לבחור 5 אנשים מתוך 10 אנשים יש $\binom{10}{5} = \frac{10!}{5! \cdot 5!} = \frac{10!}{5!^2}$. מספר הדרכים לחלק קבוצה בת n איברים $\{1, 2, \ldots n\}$ ל־ k קבוצות לא מסודרות שכל קבוצה בה i_j איברים, $j = 1, 2, \ldots k$, הוא $\binom{n}{i_1, i_2, \ldots, i_k} = \frac{n!}{i_1! i_2! \ldots i_k!}$.

דוגמא 3.4 (בעית ימי־הולדת) בחדר k אנשים, מהי ההסתברות שלפחות לשני אנשים יש אותו יום הולדת באותו תאריך? נציע שני פתרונות כדלקמן:

1. מאורע $A_k = \{$לפחות לשניים מתוך k יש אותו יום הולדת באותו תאריך$\}$ לכן למאורע המשלים $\overline{A_k} = \{$לכולם מתוך k אנשים יש ימי הולדת בתאריכים שונים$\}$. נחשב: $P(\overline{A_1}) = 1$, $P(\overline{A_2}) = \frac{364}{365}$, $P(\overline{A_3}) = \frac{364}{365} \cdot \frac{363}{365}$, ואז קל להראות שמתקיים

$$P(\overline{A_k}) = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365+1-k}{365} = \frac{365!}{(365-k)! \cdot (365^k)},$$

כלומר $P(A_k) = 1 - \frac{365!}{(365-k)! \cdot 365^k}$.

2. יש לנו 365^k אפשרויות לרשימה של ימי הולדת אפשריים ל־ k אנשים, וגם יש A_{365}^k רשימות מסודרות שבהם כל התאריכים שונים. לכן $P(\overline{A_k}) = \frac{A_{365}^k}{365^k} = \frac{365!}{(365-k)!365^k}$.

ועכשיו נתבונן במדגמים מסודרים ולא מסודרים עם ובלי חזרות. מתוך אוכלוסיה של n איברים שונים $1, 2, \ldots n$, מוציאים מדגם של r איברים כאשר הסדר חשוב, כלומר לכתוב וקטור $(i_1, i_2, \ldots i_k)$ כאשר i_j זה מספר שהתקבל במקום j, $1 \leq i_j \leq n$, $1 \leq j \leq r$, ניסוי זה ניתן לתיאור על ידי ניסוי של כדורים ותאים באופן הבא: יש n תאים שונים ו־ r כדורים שונים, אנו מעוניינים לפזר אותם בין התאים כך שיהיה i_j כדורים בתא j. לכן, אם המדגם הוא עם חזרות אז יש n^r אפשרויות שונות, ואם המדגם הוא בלי חזרות אז יש $A_n^r = \frac{n!}{(n-r)!}$ אפשרויות.

דוגמא 3.5 (הכללת בעית ימי הולדת) מהי ההסתברות שבמדגם עם חזרות לבחור מאוכלוסיה של n איברים שונים, r איברים שונים. התשובה ניתנת על ידי הביטוי $\frac{An^r}{n^r}$.
ועכשיו נשאל את השאלה הבאה: מהי ההסתברות שבמדגם מסודר של r איברים יהיה איבר i? עם חזרות יש לנו הסתברות $1 - \frac{(n-1)^r}{n^r}$, ובלי חזרות יש לנו הסתברות $1 - \frac{A_{n-1}^r}{A_n^r} = \frac{r}{n}$.
אם נשאל את השאלה הבאה: מהי ההסתברות שכאשר n כדורים שונים מתפלגים בין n תאים שונים עם חזרות ללא תאים ריקים? ידוע ש־ $n!$ שווה לסדר n כדורים ב־ n תאים ריקים כך שכל תא יכיל בדיוק כדור אחד, וגם n^n שווה לגודל מרחב המדגם. לכן ההסתברות ניתנת על ידי הביטוי $\frac{n!}{n^n}$. למשל, ההסבתרות שב־ 6 הטלות רציפות של קוביה הוגנת הופיעו כל המספרים היא $P = \frac{6!}{6^6}$.

אוכלוסיה של n איברים שונים מוציאים מדגם של r איברים כאשר הסדר לא חשוב על ידי הוקטור $(i_1, i_2, \ldots i_n)$ כאשר i_j מספר הפעמים שהוציאו את האיבר ה־ j כאשר $\sum_{j=1}^n i_j = r$. אפשר לתאר את זה על ידי המודל הבא: r כדורים זהים מתפלגים בין n תאים שונים. אם נתבונן בבעיה זו בלי חזרות אז מספר האפשרויות נתון על ידי $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ (מספר הווקטורים $(i_1, i_2, \ldots i_n)$ עם r אפסים ו־ $n-1$ אחדים. אם נתבונן באותה בעיה עם חזרות אז מספר האפשרויות נתון על ידי $\binom{n+r-1}{r} = \binom{r+n-1}{n-1}$.
איך בונים הוקטור? נדגים את הוקטור $(i_1, i_2, \ldots i_n)$ בעזרת סדרה של אפסים ואחדים

$$\underbrace{0\ldots0}_{i_1}1\underbrace{0\ldots0}_{i_2}1\ldots1\underbrace{0\ldots0}_{i_n}$$

כאשר כל 1 מהווה המחיצות לתאים שמכילים מספר כלשהו של אפסים. למשל, בהינתן הסדרה 0010100010, מכאן נובע שיש לנו 4 תאים, כאשר בתא הראשון מכיל שני אפסים, השני מכיל

אפס אחד, והשלישי מכיל 3 אפסים והרביעי ריק, ז"א הוקטור שלנו הוא $(2,1,3,0)$. ובכך בנינו פונקציה חח"ע ועל בין הוקטורים $(i_1,i_2,\ldots,i_n)$ עם קאורדינטות אי־שליליות לבין סדרות של אפסים ואחדים $0\ldots010\ldots010\ldots010\ldots0$ עם $n-1$ אחדים, ז"א עוצמת שתי הקבוצות שוות, ואני יודעים גם כן שאורך המילה הוא $n-r+1$ ויש לבחור מקומות לאחדים $n-1$ מקומות (מספר המחיצות בין התאים) ומזה נקבל שיש לנו $\binom{n-1+r}{r}=\binom{n-1+r}{n-1}$.

כאשר $r>n$ ובעיה נדונה עם חזרות ואין תאים ריקים אז יש $\binom{n-1+r-k}{n-1}$ אפשרויות, כאשר $n-1$ שווה למספר התאים, $r-k$ שווה למספר הכדורים שצריך לחלק לתאים אחרי שחילקנו כדור אחד בלבד בכל תא.

דוגמא 3.6 1. בבעיית הקלפים. לבחירת 5 קלפים מתוך חפיסת קלפים יש $\binom{52}{5}$ נשאלת השאלה, מהי ההסתברות ש־ 5 הקלפים הללו יהיו מאותו סוג? מאחר ויש לנו 4 סוגים של קלפים ומכל אחד מהסוגים יש 13 קלפים מכאן נובע שמרחב המדגם הוא $\binom{52}{5}$ ולכן ההסתברות היא $\frac{4\binom{13}{5}}{\binom{52}{5}}$.

ההסתברות שכל הקלפים (מאותו סוג כאשר המספרים שונים) יש פנים שונים היא $P=\frac{\binom{13}{5}\cdot 4^5}{\binom{52}{5}}$.

ההסתברות שיהיו 4 קלפים עם אותם פנים היא: $P=\frac{13\cdot\binom{1}{1}^4\binom{48}{1}}{\binom{52}{5}}=\frac{13\cdot 48}{\binom{52}{5}}=0.00025$.

2. r כדורים מתפזרים בין n תאים (עם חזרות),מהי ההסתברות שבתא מסויים יש בדיוק k כדורים?

$$P=\frac{\binom{r}{k}(n-1)^{r-k}}{n^r}=\binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k},$$

כאשר $\binom{r}{k}$ שווה לבחור k כדורים מתוך r כדורים שיש לנו ולשים אותם בתא מסויים, $(n-1)^{r-k}$ שווה לפיזור $(r-k)$ כדורים ב־ $n-1$ התאים שנשארו, ואילו n^r שווה לפיזור r כדורים ב־ n תאים.

3. לפזר r כדורים ב־ n תאים (עם חזרות), מהי ההסתברות שבתא ראשון i_1, בתא שני i_2, ..., בתא n יש i_n כדורים כך ש־ $r=\sum_{j=1}^{n} i_j$:

$$P=\frac{\binom{r}{i_1}\binom{r-i_1}{i_2}\cdots\binom{r-i_1-\cdots-i_{n-1}}{i_n}}{n^r}=\frac{1}{n^r}\frac{r!}{i_1!i_2!\cdots i_n!}.$$

4. כמה נגזרות חלקיות מסדר r יש לפונקציה אנליטית של n משתנים? התשובה לכך היה $\binom{n+r-1}{r}$, שווה למספר מספר הוקטורים מהצורה $(i_1,\ldots,i_n)$ ומקיים $\sum_{j=1}^{n} i_j=r$.

איך לבחור את המודל הנכון? לפעמים זה מובן מתנאי הבעייה, ולפעמים צריך לערוך ניסויים, דוגמא לכך מכניקה סטטיסטית. יש r חלקיקים אלמנטריים מאותו סוג (ז"א בלי חשיבות לסדר). מחלקים מרחב־פזה ל n "תאים" כך שביחידת זמן כל חלקיק יכול להיות בתא אחד.

1. אזי יש תיאורטית n^r אפשריות, לכן למצב $(i_1, \dots i_n)$ יש הסתברות $\frac{1}{n^r} \cdot \frac{r!}{i_1! \dots i_n!}$ התפלגות הזאת נקראת סטטיסטיקה בולצנו־מקסוול, והיא לעולם לא מתקבלת בנסיונות!

2. אם החלקיקים הם פוטונים, נוקלאונים ואטומים עם מספר זוגי של חלקיקים אלמנטריים אז ההתפלגות היא לפי אינשטיין־בוזה יש $\binom{n+r-1}{r}$ אפשריות כולם אפשריים באותה מידה.

3. אם החלקיקים הם אלקטרונים, פרוטונים,נויטרונים ו־ $n > r$ (יש יותר תאים מאשר כדורים) אז הם מתפלגים לפי החוקים הבאים:

   (א) בשום תא אין יותר מחלקיק אחד

   (ב) כל המצבים אפשריים באותה מידה יש $\binom{n}{r}$ מצבים שכולם אפשריים באותה מידה; סטטיסטיקה הזו נקראת דירק־פרמי.

3.3.2 נוסחת סטרילינג

משפט 3.2 (נוסחת סטרילינג)

$$n! \sim \sqrt{2\pi} \cdot n^{n+\frac{1}{2}} \cdot e^{-n} \Leftrightarrow \lim_{n\to\infty} \frac{n!}{\sqrt{2\pi} \cdot n^{n+\frac{1}{2}} \cdot e^{-n}} = 1.$$

הוכחה: נוכיח כי $\lim_{n\to\infty} \frac{n!}{n^{(n+\frac{1}{2})} \cdot e^{-n}} = C$ כאשר $e^{\frac{11}{12}} \le C \le e$. (בהמשך נוכיח כי $C = \sqrt{2\pi}$).
נחלק את ההוכחה לשלבים:

1. במקום לחקור את $n!$ נחקור את $\ln n!$ באופן הבא:

$$\ln(n!) = \ln(1 \cdot 2 \dots n) = \ln 1 + \ln 2 + \dots \ln n = \sum_{i=1}^{n} \ln(i)$$

2. פונקצית $\ln x$ מונטנית עולה לכן לכל k טבעי מקבלים $\int\limits_{k}^{k+1} \ln t dt$ > $\ln k$ > $\int\limits_{k-1}^{k} \ln t dt$ לכל $k \ge 1$ לכן:

$$\sum_{k=1}^{n} \int\limits_{k-1}^{k} \ln t dt < \sum_{k=1}^{n} \ln k < \sum_{k=1}^{n} \int\limits_{k}^{k+1} \ln t dt$$

ששקול ל־

$$\int\limits_{0}^{n} \ln t dt < \sum_{i=1}^{n} \ln i < \int\limits_{1}^{n+1} \ln t dt \qquad (3.1)$$

3. נחשב $\int\limits_a^b \ln t dt$ לפי חלקים:

$$\int\limits_a^b \ln x dx \underbrace{=}_{\substack{u=\ln x\\ dv=dx}} x\ln x|_a^b - \int\limits_a^b \frac{1}{x}\cdot x dx = b\ln(b) - a\ln(a) - (b-a)$$

נציב בשני הצדדים ונקח בחשבון שלפי כלל לופיטל מתקיים: $\lim\limits_{x\to 0^+} x\ln x = 0$ נציב ב־ (3.1) ונקבל

$$n\ln n - \underbrace{\lim_{a\to 0^+} a\ln a}_{0} - n < \sum_{i=1}^{n} \ln i < (n+1)\ln(n+1) - 1\cdot\ln 1 - n.$$

לפי $\lim\limits_{x\to 0^+} x\ln x = \lim\limits_{x\to 0^+} \frac{\ln x}{\frac{1}{x}} \underbrace{=}_{\frac{\infty}{\infty}} \lim\limits_{x\to 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}} = \lim\limits_{x\to 0^+} -x = 0$ נובע ש־

$$n\ln n - n < \ln n! < (n+1)\ln(n+1) - n.$$

4. תהי הסדרה $d_n = \ln n! - (n+\frac{1}{2})\ln n + n$ נראה ש־ d_n סדרה יורדת:

$$\begin{aligned} d_n - d_{n+1} &= \ln n! - (n+\tfrac{1}{2})\ln n + n - \ln((n+1)!) + (n+1+\tfrac{1}{2})\ln(n+1) - (n+1) \\ &= -\ln(n+1) - (n+\tfrac{1}{2})[\ln n - \ln(n+1)] + \ln(n+1) - 1 \\ &= -(n+\tfrac{1}{2})\ln(\tfrac{n}{n+1}) - 1 = (n+\tfrac{1}{2})\ln(\tfrac{n+1}{n}) - 1 \end{aligned}$$

5. נזכיר של $0 < t < 1$ מתקיים $\ln(t+1) = \sum\limits_{i=1}^{\infty} (-1)^{i+1}\frac{t^i}{i}$ ו־ $\ln(1-t) = -\sum\limits_{i=1}^{\infty} \frac{t^i}{i}$, לכן

$$\ln\left(\frac{1+t}{1-t}\right) = \sum_{i\geq 1}((-1)^{i+1}+1)\frac{t^i}{i} = \sum_{j\geq 0}((-1)^{2j+1+1}+1)\frac{t^{2j+1}}{2j+1} = 2\sum_{j\geq 0}\frac{t^{2j+1}}{2j+1}.$$

(3.2)

נשתמש בנוסחה (3.2) לגבי t כך ש־ $\frac{1+t}{1-t} = \frac{n+1}{n}$, ונקבל

$$\frac{n+1}{n} = \frac{2(n+1)}{2n} = \frac{(2n+1)+1}{(2n+1)-1} = \frac{1+\frac{1}{2n+1}}{1-\frac{1}{2n+1}}$$

מכאן נובע

$$\begin{aligned} (n+\tfrac{1}{2})\ln(\tfrac{n+1}{n}) &= (n+\tfrac{1}{2})\cdot 2\sum_{j=0}^{\infty}\frac{(\frac{1}{2n+1})^{2j+1}}{2j+1} = (2n+1)\cdot\frac{1}{2n+1} + (2n+1)\sum_{j=1}^{\infty}\frac{(\frac{1}{2n+1})^{2j+1}}{2j+1} \\ &= 1 + \sum_{j=1}^{\infty}\frac{(\frac{1}{2n+1})^{2j}}{2j+1} \end{aligned}$$

זאת אומרת, $d_n - d_{n+1} = \sum\limits_{j=1}^{\infty}\frac{(\frac{1}{2n+1})^{2j}}{2j+1} > 0$, כלומר קיבלנו ש־ d_n הינה סידרה יורדת.

6. מצד שני מתקיים

$$\begin{aligned}\sum_{j=1}^{\infty}\frac{(\frac{1}{2n+1})^{2j}}{2j+1} &< \frac{1}{3}\sum_{j=1}^{\infty}(\frac{1}{2n+1})^{2j} \\ &= \frac{1}{3}\frac{(\frac{1}{2n+1})^2}{1-(\frac{1}{2n+1})^2} \\ &= \frac{1}{3}\frac{1}{(2n+1)^2-1} = \frac{1}{3}\frac{1}{(2n+2)(2n)} = \\ &= \frac{1}{12}\cdot\frac{1}{(n+1)n} \\ &= \frac{1}{12}\left(\frac{1}{n}-\frac{1}{n+1}\right)\end{aligned},$$

ולכן מתקיים

$$d_n - d_{n+1} < \tfrac{1}{12}\cdot\tfrac{1}{n} - \tfrac{1}{12}\cdot\tfrac{1}{n+1} \Leftrightarrow d_n - \tfrac{1}{12}\cdot\tfrac{1}{n} < d_{n+1} - \tfrac{1}{12}\cdot\tfrac{1}{n+1} \ .$$

נסמן ב־ $c_n = d_n - \frac{1}{12}\cdot\frac{1}{n}$, קל לראות שהסדרה היא עולה ולכן מכאן מקבלים

$$\frac{11}{12} = 1 - \frac{1}{12} = c_1 < c_n < d_n < d_1 = 1.$$

מאחר ו־ d_n הינה סדרה יורדת וחסומה מלמטה נקבל ש־ $\lim_{n\to\infty} d_n$ קיים. ולכן מתקיים

$$d_n = \ln n! - \left(n+\frac{1}{2}\right)\ln n + n = \ln\left(\frac{n!}{n^{n+\frac{1}{2}}\cdot e^{-n}}\right),$$

ומתקיים גם $A = \lim_{n\to\infty}\frac{n!}{n^{n+\frac{1}{2}}\cdot e^{-n}}$ ז"א הגבול קיים כאשר $\frac{11}{12} < \ln A < 1$, ואז נקבל ש־ $e^{\frac{11}{12}} < A < e^1$.

עכשיו נוכיח ש־ $C = \sqrt{2\pi}$. לשם כך נשתמש באינטגרלים הבאים $J_m = \int_0^{\frac{\pi}{2}}\sin^m x dx$.
מחישוב ישיר נקבל $J_1 = \int_0^{\frac{\pi}{2}}\sin x dx = -\cos x|_0^{\frac{\pi}{2}} = 1$, $J_2 = \int_0^{\frac{\pi}{2}}\sin^2 x dx = \int_0^{\frac{\pi}{2}}\frac{1-\cos 2x}{2}dx = \frac{\pi}{4}$
ואילו

$$\begin{aligned}J_m &= \int_0^{\frac{\pi}{2}}\sin^m x dx \underbrace{=}_{\substack{u=\sin^{m-1}x, du=(m-1)\sin^{m-2}x\cos x dx \\ dv=\sin x dx, v=-\cos x}} -\sin^{m-1}x\cos x\Big|_0^{\frac{\pi}{2}} + (m-1)\int_0^{\frac{\pi}{2}}\sin^{m-2}x\cos^2 x dx \\ &= (m-1)\int_0^{\frac{\pi}{2}}\sin^{m-2}x dx - (m-1)\underbrace{\int_0^{\frac{\pi}{2}}\sin^m x dx}_{J_m},\end{aligned}$$

זאת אומרת $J_m = \frac{m-1}{m}J_{m-2}$ עבור $m \geq 3$. נחלק לשני מקרים

1. אם $m=2k$ אזי $J_m=\frac{m-1}{m}\cdot\frac{m-3}{m-2}\cdot\frac{m-5}{m-4}\cdot\cdots\cdot\frac{3}{4}\frac{\pi}{4}=\frac{(m-1)(m-3)\cdots 3\cdot 1}{m(m-2)\cdots 4\cdot 2}\cdot\frac{\pi}{2}=\frac{\pi}{2}\frac{(m-1)!!}{m!!}$

2. אם $m=2k+1$ אזי $J_m=\frac{m-1}{m}\cdot\frac{m-3}{m-2}\cdots\frac{2}{3}\cdot 1=\frac{(m-1)!!}{m!!}$. כלומר קיבלנו ש־

$$J_m=\begin{cases}\frac{\pi}{2}\frac{(m-1)!!}{m!!}; m=2k\\ \frac{(m-1)!!}{m!!}; m=2k+1\end{cases}$$

כעת נותר לחשב את הביטוי $\lim_{n\to\infty}\left(\frac{(2n)!!}{(2n-1)!!}\right)^2\cdot\frac{1}{2n}$, ידוע כי $0<\sin x<1$ בתחום $0<x<\frac{\pi}{2}$ מזה אנו מקבלים

$$0<\sin^{2n+1}x<\sin^{2n}x<\sin^{2n-1}x\Rightarrow J_{2n+1}<J_{2n}<J_{2n-1}$$

ששקול ל־

$$\frac{(2n)!!}{(2n+1)!!}<\frac{(2n-1)!!}{(2n)!!}\cdot\frac{\pi}{2}<\frac{(2n-2)!!}{(2n-1)!!}\Leftrightarrow\frac{((2n)!!)^2}{(2n+1)((2n-1)!!)^2}<\frac{\pi}{2}<\left(\frac{(2n)!!}{(2n-1)!!}\right)^2\frac{1}{2n},$$

וזה גורר ש־

$$\lim_{n\to\infty}\left(\frac{(2n)!!}{(2n-1)!!}\right)^2\frac{1}{2n}=\frac{\pi}{2}.$$

נחשב את $\frac{(2n)!!}{(2n-1)!!}=\frac{((2n)!!)^2}{(2n)!!(2n-1)!!}=\frac{(2n(2n-2)(2n-4)\cdots 2)^2}{(2n)!!(2n-1)!!}=\frac{2^{2n}(n(n-1)\cdots 1)^2}{(2n)!}=\frac{2^{2n}\cdot n!^2}{(2n)!}$, נשתמש בנוסחה $\lim_{n\to\infty}n!/(A\sqrt{n}(\frac{n}{e})^n)=1$, ונקבל

$$\frac{\pi}{2}=\lim_{n\to\infty}\left(\frac{2^{2n}(n!)^2}{(2n)!}\right)^2\cdot\frac{1}{2n}=\lim_{n\to\infty}\frac{2^{4n}\cdot A^4n^2(\frac{n}{e})^{4n}}{A^2\cdot 2n(\frac{2n}{e})^{4n}}\cdot\frac{1}{2n}=\frac{A^2}{4}$$

לכן $A^2=4\cdot\frac{\pi}{2}=2\pi$, כלומר $A=\sqrt{2\pi}$ $(A>0)$. □

לפי נוסחת סטרילינג מתקיים:

n	$n!$	קירוב על פי נוסחת סטרילנג	השגיאה
1	1!	0.9221	0.08
2	2!	1.919	0.04
5	5!	118.019	0.02

עכשיו נביא דוגמא לאחד השימושים של נוסחת סטירלינג. יהי n כדורים ב־ n תאים עם חזרות, מהי ההסתברות שכל הכדורים יהיו בתאים שונים? התשובה לכך נתונה בביטוי הבא

$$p=\frac{n!}{n^n}\approx\frac{\sqrt{2\pi}n^{n+\frac{1}{2}}\cdot e^{-n}}{n^n}=\sqrt{2\pi n}\cdot e^{-n}$$

פרק 4

הסתברות מותנית, הסתברות שלמה ונוסחת בייס

4.1 הגדרת הסתברות מותנית

הגדרה 4.1 נתון מרחב הסתברות $(\Omega, \mathbb{A}, P)$, ויהיו $A, B \in \mathbb{A}$ כאלה ש־ $P(B) > 0$. הסתברות מותנית של מאורע A בתנאי שמאורע B קרה, או בקיצור "הסתברות של A בתנאי B מוגדרת על ידי

$$P(A|B) = \frac{P(A \cdot B)}{P(B)}.$$

טענה 4.1 נתון מרחב הסתברות $(\Omega, \mathbb{A}, P)$ ו־ $B \in \mathbb{A}$ כזה ש־ $P(B) > 0$. נגדיר $(B, \mathbb{A}_B, P_B)$ באופן הבא $\mathbb{A}_B = \{A \cdot B | \, A \in \mathbb{A}\}$ ולכל קבוצה $F \in \mathbb{A}_B$ מתקיים $P_B(F) = \frac{P(F)}{P(B)}$. אז $(B, \mathbb{A}_B, P_B)$ הוא מרחב הסתברות.

הוכחה: צריך להראות ש־ $\mathbb{A}_B$ היא σ־אלגברה מעל B ו־ $P_B : \mathbb{A}_B \to [0, 1]$ היא פונקצית ההסתברות.

1. נוכיח כי $\mathbb{A}_B$ היא σ־אלגברה מעל B.

   (א) $\mathbb{A}_B \neq \emptyset$ כי $B \cdot B = B \in \mathbb{A}_B$

   (ב) תהי $A \in \mathbb{A}$ אזי $\overline{A} \in \mathbb{A}$ ומתקיים: $B = \Omega \cdot B = (A + \overline{A})B = AB + \overline{A}B$ לכן לכל $A \in \mathbb{A}, F = AB \in \mathbb{A}_B$ מתקיים

$$\overline{F}_B = \overline{A}B \in \mathbb{A}_B$$

(ג) לכל $\{F_i\}_{i\in I} \subseteq \mathbb{A}_B$ קיימת $\{A_i\}_{i\in I} \subseteq \mathbb{A}$ כזאת ש־ $F_i = A_i \cdot B$ לכל $i \in I$ לכן $\sum\limits_{i\in I} F_i = \sum\limits_{i\in I} A_i \cdot B = \left(\sum\limits_{i\in I} A_i\right) B \in \mathbb{A}_B$, מכאן נובע ש: $\mathbb{A}_B$ היא σ־אלגברה מעל B.

2. נוכיח כי $P_B : \mathbb{A}_B \to [0,1]$ מקיימת אקסיומות פונקציה ההסתברות.

(א) $0 \le P_B(F) = \frac{P(AB)}{P(B)} \le 1$.

(ב) $P_B(B) = \frac{P(B\cdot B)}{P(B)} = \frac{P(B)}{P(B)} = 1$

(ג) אם $\{F_i\}_{i\in I} \subseteq \mathbb{A}_B$ כמשפחה זרה של קבוצות אזי $(F_i = A_i \cdot B)$ עם $A_i \in \mathbb{A}$, ולכן

$$P_B(\sum F_i) = \frac{P(\sum A_iB)}{P(B)} = \frac{\sum P(A_iB)}{P(B)} = \sum P_B(F_i).$$

□

יש להעיר כי $A|B$ זו לא קבוצה! ולפי טענה 4.1 הסתברות מותנית מקיימת תכונות של פונקציה הסתברות.

טענה 4.2 יהי $(\Omega, \mathbb{A}, P)$ מרחב הסתברות, אזי

1. לכל $A, B \in \mathbb{A}$ כאלה ש־ $P(A) \neq 0$, $P(B) \neq 0$ מתקיים:

$$P(AB) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A).$$

2. אם $\{A_i\}_{i=1}^n \in \mathbb{A}$ קבוצה כזאת ש־ $P(\prod\limits_{i=1}^n A_i) > 0$ אזי

$$P(\prod_{i=1}^n A_i) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cdot A_2) \cdots P(A_n|\prod_{i=1}^{n-1} A_i).$$

הוכחה:

1. לפי הגדרת ההסתברות המותנית נקבל שמתקיים

$$P(A|B) = \frac{P(AB)}{P(B)} \Rightarrow P(AB) = P(A|B) \cdot P(B).$$

2. הטענה השנייה נכונה עבור $n=2$ (הוכחנו ב־ 1), לכן נוכיח את הטענה באינדוקציה על n, לשם כך נניח נכונות הטענה עבור $n-1$ ונראה עבור n. כעת נגדיר הקבוצות $A=A_n$ ו־ $B=A_1A_2\cdots A_{n-1}$ מכאן נובע שמתקיים:

$$\begin{aligned} P(\prod_{i=1}^{n} A_i) &= P(A_n\cdot B)=P(A_n|B)\cdot P(B)= \\ &= P(A_n|B)\cdot P(A_{n-1}|A_1,\ldots,A_{n-2})\ldots P(A_2|A_1)P(A_1). \end{aligned}$$

□

הגדרה 4.2 יהי $(\Omega,\mathbb{A},P)$ מרחב הסתברות. מאורעות $A,B\in\mathbb{A}$ נקראים בלתי תלויים אם מתקיימים $P(AB)=P(A)\cdot P(B)$.

אזי הגדרה אלטרנטיבית היא A,B בלתי תלויים אם: $P(A)=P(A|B)$. שתי ההגדרות האלה מתלקדות כאשר $P(B)>0$ כי אז אם מתקיים $P(AB)=P(A)\cdot P(B)$ אזי

$$P(A|B)=\frac{P(AB)}{P(B)}=P(A).$$

ואם $P(A|B)=P(A)\Rightarrow P(AB)=P(A)\cdot P(B)$. ההגדרה הראשונה טובה כי היא מראה שמוסג "בלתי תלויים" הוא סימטרי. ההגדרה השניה מסבירה את העיקרון של אי־תלות, כלומר זה ש־ B קרה לא משפיע על סיכוי של A לקרות.

הגדרה 4.3 יהי $(\Omega,\mathbb{A},P)$ מרחב הסתברות. מאורעות $\{A_i\}_{i=1}^{n}\subseteq\mathbb{A}$ נקראים בלתי תלויים אם לכל צירוף $\{i_1,\ldots i_k\}$ כאשר $2\le k\le n$ מתקיים: $P(\prod_{j=1}^{k} A_{i_j})=\prod_{j=1}^{k} P(A_{i_j})$.

חשוב להעיר שצריך לבדוק את כל הצירופים האפשריים כי לבדוק רק צירופים מאיזשהו סדר זה לא מספיק!

דוגמא 4.1 הניסוי: הטלת 2 קוביות שונות. יהיו נתונות המאורעות $A=$ {תוצאה על הקוביה הראשונה היא אי זוגית}, $B=$ {תוצאה על הקוביה השניה היא זוגית}, $C=$ {סכום על שתי קוביות הוא אי זוגי} ומתקיים $P(A)=P(B)=P(C)=0.5$ וגם

$$\begin{aligned} &AB=AC=BC=ABC \\ &\Rightarrow P(AB)=P(AC)=P(BC)= \\ &=P(A)\cdot P(B)=P(A)\cdot P(C)=P(B)\cdot P(C) \end{aligned}$$

כאשר $P(AB)=0.5\cdot 0.5=0.25$, ומצד שני מתקיים: $P(ABC)=0.25$ ו־ C תלויה ב־ A ו־ B.

טענה 4.3 (תכונות אלמנטריות של הסתברות מותנית)

1. אם $A_1, A_2 \in \mathbb{A}$ זרים ו־ $B \in \mathbb{A}$ כזה ש־ $P(B) > 0$ אזי

$$P((A_1 + A_2)|B) = P(A_1|B) + P(A_2|B).$$

2. בפרט מתקיים $P(\overline{A}|B) = 1 - P(A|B)$.
3. אם A ו־ B זרים ו־ $P(A+B) > 0$ אזי $P(A|A+B) = \frac{P(A)}{P(A)+P(B)}$.
4. אם A, B בלתי תלויים אזי הזוגות הבאים הם בלתי תלויים: $(A,\overline{B}), (\overline{A},B), (\overline{A},\overline{B})$.
5. אם $0 < P(B) < 1$ אזי B, A בלתי תלויים אס"ם $P(A|B) = P(A|\overline{B})$.

הוכחה: על פי ההנחות מתקיים

1.
$$\begin{aligned} P(A_1 + A_2|B) &= \tfrac{P((A_1+A_2)B)}{P(B)} = \tfrac{P(A_1B+A_2B)}{P(B)} = \tfrac{P(A_1B)+P(A_2B)}{P(B)} \\ &= P(A_1|B) + P(A_2|B). \end{aligned}$$

2. $1 = P(\Omega|B) = P(A+\overline{A}|B) = P(A|B) + P(\overline{A}|B) \Rightarrow P(\overline{A}|B) = 1 - P(A|B).$
3. $P(A|A+B) = \frac{P(A(A+B))}{P(A+B)} = \frac{P(A+\emptyset)}{P(A)+P(B)} = \frac{P(A)}{P(A)+P(B)}.$
4. $P(A\overline{B}) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)(1-P(B)) = P(\overline{B})P(A)$.
5. אם A ו־ B בלתי תלויים, אז $A, \overline{B}$ בלתי תלויים לפי הסעיף הקודם, לכן $P(A|B) = P(A) = P(A|\overline{B})$. להיפך נניח ש־ $P(A|\overline{B}) = P(A|B)$, אז

$$P(A|\overline{B}) = \frac{P(A\overline{B})}{P(\overline{B})} = \frac{P(A) - P(AB)}{1 - P(B)} = \frac{P(AB)}{P(B)} = P(A|B).$$

מכאן נקבל ש־ $P(B)(P(A) - P(AB)) = P(AB)(1 - P(B))$ שגורר ש־ $P(AB) = P(A)P(B)$.

□

4.2 הסתברות שלמה ונוסחת בייס

משפט 4.1 יהי $(\Omega, \mathbb{A}, P)$ מרחב הסתברות. תהי $\{A_i\}_{i\in I}$ פירוק של Ω, ותהי $B \in \mathbb{A}$ אזי:

1. נוסחת **ההסתברות השלמה**

$$P(B) = \sum_{i\in I} P(B|A_i) \cdot P(A_i).$$

2. אם בנוסף נתון ש־ $P(B) > 0$ אזי מתקיימת **נוסחת בייס** הבאה:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum\limits_{i \in I} P(B|A_i) \cdot P(A_i)}.$$

הוכחה: על פי ההנחות מתקיים

1. $P(B) = P(B \cdot \sum\limits_{i\in I} A_i) = P(\sum\limits_{i\in I} B \cdot A_i) = \sum\limits_{i\in I} P(B \cdot A_i) = \sum\limits_{i\in I} P(B|A_i)P(A_i)$

2. $P(A_i|B) = \frac{P(A_i \cdot B)}{P(B)}$ אבל מצד שני מתקיים $P(B|A_i) = \frac{P(B \cdot A_i)}{P(A_i)}$ ומכאן נובע ש־ $P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)}$, ושימוש בסעיף הקודם נשלים את ההוכחה.

□

דוגמא 4.2 נוסחת ההסתברות שלמה מאוד נוחה, כי לעיתים קרובות אנו יודעים משהו רק מותנית. למשל נדון בבעית עיוורון צבעים: עיוורון צבעים קשור לכרומוזמה X. כל השינויים שקשורים לכרומוזומה X ההסתברות שלהם תלויה במין אם הסתברות לגבר היא P אז ההסתברות לאישה היא P^2. נסמן ב־$D = \{$יש עיוורון צבעים$\}$, B=גבר, נתון גם $P(D|B) = \frac{1}{40}$, $P(B) = \frac{1}{2}$, $P(D|\overline{B}) = \frac{1}{40^2}$. מהי ההסתברות שלתינוק שנולד יש עיוורון צבעים?

$$P(D) = P(D|B) \cdot P(B) + P(D|\overline{B}) \cdot P(\overline{B}) = \frac{1}{40} \cdot \frac{1}{2} + \frac{1}{40^2} \cdot \frac{1}{2} = \frac{1}{2}(\frac{1}{40} + \frac{1}{40^2}).$$

דוגמא 4.3 נגדיר תאומים זהים או תואמים חד ביצית, הם תאומים שהתפתחו מביצה אחת, אחרת נאמר תאומים לא זהים (תאומים זהים הם תמיד מאותו מין). נניח שאם תאומים הם לא זהים אזי ההסתברות לכל אחד להיות ממין הנתון הוא 0.5, זאת אומרת בהסתברות 0.5 הם מאותו מין ובהסתברות 0.5 הם ממין שונה. בין התאומרים נולדים 64% של תאומים מאותו מין. כמה מבין התאומים חד־ביצית? נגדיר המאורעות הבאים $A = \{$תאומים מאותו מין$\}$ $B = \{$תאומים חד־ביצית$\}$ וקיים $P(A|\overline{B}) = 0.5, P(A|B) = 1$ ו־ $P(B) = P$ מכאן נובע ש:

$$P(A) = P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B}) == p + 0.5(1-p) = 0.64,$$

זאת אומרת, $0.5p = 0.14$, $P(B) = 0.28$.

בנוסחת בייס יש לנו K הנחות זרות ואנו יודעים ש־ $P(B|A_i)$ שווה להסתברות של המאורע B בהינתן ההנחה A_i, $P(A_i)$ שווה להסתברות של ההנחה A_i, ו־ $P(B)$ שווה להסתברות של המאורע הכללי B. אנו מעוניינים לחשב את ההסתברות של ההנחה בהינתן המאורע B כלומר לחשב $P(A_i|B)$.

דוגמא 4.4 תסמונית דואן היא מחלה תורשתית הקשורה לשגיאת שיכפול הגן. ההסתברות שלה גודל דרסאית עם השנים. ההסתברות של הולדת תינוק עם תסומנת לאישה בגיל 30 היא: $P(D_{30}) = \frac{1}{885}$, ההסתברות של הולדת תינוק עם תסמונת לאישה בגיל 45 הוא: $P(D_{45}) = \frac{1}{32}$. בדיקת חלבון עוברי מראה תוצאות אמת ב־99.5% לכל צד, ז"א

$$\begin{aligned} P(A|D) &= 0.995 \\ P(\overline{A}|\overline{D}) &= 0.995 \\ P(A|\overline{D}) &= 0.005 \end{aligned}$$

כאשר $A = \{$תוצאה חיובית$\}$, $D = \{$איזשהו מקרה$\}$. אישה בגיל 30 מקבלת תוצאה חיובית. מהי ההסתברות שלתינוק יש תסמונת דאון? התשובה להלן:

$$P(D_{30}|A) = \frac{P(A|D_{30}) \cdot P(D_{30})}{P(A)} = \frac{0.995 \cdot \frac{1}{885}}{P(A)}$$

נחשב את $P(A)$ ע"י הנוסחה השלמה ונקבל:

$$P(A|D_{30}) \cdot P(D_{30}) + P(A|\overline{D}_{30}) \cdot P(\overline{D}_{30}).$$

מכאן נובע ש־

$$P(D_{30}|A) = \frac{0.995 \cdot \frac{1}{885}}{0.995 \cdot \frac{1}{885} + 0.005 \cdot \frac{884}{885}} \approx 0.18.$$

דוגמא 4.5 יש שק עם 1000 מטבעות, 999 מטבעות הוגנות ואחת נוטה עם הסתברות $P(\text{עץ}) = 0.9$. מוציאים מטבע מהשק ומתחילים להטיל אותו. כמה פעמים צריך לקבל עץ ברציפות כדי להיות בטוח ב K לפחות שהמטבע הוא נוטה כאשר $K = 50\%$, $K = 90\%$, $K = 99\%$, ו־ $K = 99.9\%$.

כדי לענות על השאלה, נסמן ב־ $A_n = \{n$ פעמים ברציפות התקבל עץ$\}$ וב־$B = \{$ המטבע הוא נוטה$\}$ מכאן נובע ש־ $P(B) = 0.001$, $P(\overline{B}) = 0.999$ הן הסתברויות אפשריוריות כמו כן מתקיים $P(A_n|B) = (0.9)^n$, $P(A_n|\overline{B}) = (0.5)^n$,

$$P(B|A_n) = \frac{P(A_n|B) \cdot P(B)}{P(A_n|B) \cdot P(B) + P(A_n|\overline{B}) \cdot P(\overline{B})}$$

$$P(\overline{B}|A_n) = \frac{P(A_n|\overline{B}) \cdot P(\overline{B})}{P(A_n|B) \cdot P(B) + P(A_n|\overline{B}) \cdot P(\overline{B})}$$

בנוסף,

$$\frac{P(B|A_n)}{P(\overline{B}|A_n)} \geq \frac{k}{1-k} \Leftarrow P(\overline{B}|A_n) \leq 1-k \Leftarrow P(B|A_n) \geq k,$$

אבל

$$\begin{aligned}\frac{P(B|A_n)}{P(\overline{B}|A_n)} &= \frac{(0.9)^n \cdot 0.001}{(0.5)^n \cdot 0.999} \\ &= (\frac{9}{5})^n \cdot \frac{1}{999} \\ &\Leftrightarrow n[\log(9) - \log(5)] - \log(999) \geq \log k - \log(1-k)\end{aligned}$$

ולכן

$$n \geq \frac{\log k - \log(1-k) + \log(999)}{\log 9 - \log 5}.$$

נחלק לכמה מקרים:

$$\begin{aligned}&k = 0.5 \rightarrow n \geq 11.75 \rightarrow n \geq 12 \\ &k = 0.9 \rightarrow n \geq 15.48 \rightarrow n \geq 16 \\ &k = 0.99 \rightarrow n \geq 19.56 \rightarrow n \geq 20 \\ &k = 0.999 \rightarrow n \geq 23.5 \rightarrow n \geq 24\end{aligned}$$

הערה 4.1 משפט בייס זמן רב היה שנוי במחלוקת כי הוא מיועד לתקון של הסתברויות אפריוריות ,כאשר הרעיון המקורי של בביס היה : אם אנחנו לא יודעים כלום על הנחות באו נתן לכל ההנחות אותו סיכוי, זאת אומרת מתרגמים חוסר ידע להתפלגות אחידה ואחרי ניסיונות מתקנים את ההסתברויות. אבל למה לתרגם חוסר ידע דווקא להתפלגות אחידה (כאן חוזרים למקרה נפוץ כאשר אנשים חושבים שאם יש רק שתי אפשריות אז ההסתברות של אפשרות היא חצי). היום יש סטטיסטיקאים שקוראים לעצמם "סטטיסטיקאים בייסאנים" , זאת אומרת שהם משתמשים במשפט בייס במקרה של חוסר ידע, ויש כאלה שמסרבים להשתמש במשפט בייס. מעניין שבייס $(1702-1761)$ לא פרסם את המשפט שלו בדיוק כי הרגיש שתרגום של חוסר ידע להתפלגות אחידה לא נכון ולא ידע איך לפתור את הבעיה של הערכת הסתברויות אפריוריות.

4.3 תרגילים

תרגיל 4.1 בתוכנית הטלויזיה היומית "מי רוצה להיות מילארדר, היום?" ישנם שלושה וילונות. מאחורי אחד הוילונות ישנו צ'ק שמן בשווי מילארד שקלים, ומאחורי שני הנותרים קופסאות סיגרים (גם טוב). המתחרה בוחר וילון. המנחה מרים וילון מהשניים הנותרים בו אין מיליארד שקלים, ושואל את המתחרה האם הוא רוצה להחליף וילון (לשלישי הנותר). איזו טקטיקה כדאית למתחרה, להחליף וילון או להישאר עם הוילון שהוא בחר (בשביל שיהיה לו יותר סיכוי לזכות בכסף)?

פתרון: נניח שהמתחרה נשאר עם הוילון שלו. אז יש לו הסתברות של $1/3$ לזכות. אם המתחרה תמיד מחליף וילון, אז יש לו הסתברות $2/3$ לזכות: הוא יזכה אם בהתחלה הוא בחר בוילון עם סיגרים. לכן עדיף להחליף וילון. □

תרגיל 4.2 בקבוצת לימוד אנגלית לומדים 100 בני מאדים, ו־150 בני נוגה. בוחרים מהקבוצה תת־קבוצה של 10 חיזרים. מה ההסתברות שהיחס בין מספר בני מאדים למספר בני נוגה בתת־הקבוצה יהיה שווה ליחס המתאים בכל קבוצת הלימוד (אין צורך לחשב את התשובה נומרית)?

פתרון: התשובה ניתנת על ידי הביטוי $\frac{\binom{100}{4}\binom{150}{6}}{\binom{250}{10}}$. □

תרגיל 4.3 מתמטיקאי, בזמן חופשתו מההוראה באוניברסיטה, קרא 7 מאמרים בטופולוגיה ו־ 8 מאמרים באלגברה (והוא היה קורא עוד אילולא הסמסטר החדש שהגיע). בוחרים אקראית 5 מאמרים מאלו שהוא קרא. מה הסיכוי שלפחות 4 מהם עוסקים באלגברה?

פתרון: התשובה ניתנת על ידי הביטוי $\frac{\binom{8}{5}\binom{7}{0}}{\binom{15}{5}} + \frac{\binom{8}{4}\binom{7}{1}}{\binom{15}{5}}$. □

תרגיל 4.4 נתאר שני משחקים: בידנו שני כדים; האדום שמכיל n כדורים שחורים ו־ m לבנים והכחול שמכיל $100n$ כדורים שחורים ו־ $100m$ לבנים. במשחק הראשון, אנחנו בוחרים כד, מוציאים ממנו כדור אחד, ואם הכדור לבן, זוכים בחביתה. באיזה כד כדאי יותר לבחור? למה? במשחק השני, אנחנו בוחרים כד, מוציאים ממנו שני כדורים, וזוכים בחביתה אם לפחות אחד מהכדורים לבן. באיזה כד כדאי יותר לבחור? למה?

פתרון: במשחק הראשון לא חשוב איזה כד נבחר כי בכל מקרה יש הסתברות של $\frac{100m}{100n+100m}$ לזכות. במשחק השני עדיף לבחור את הכד האדום: אז ההסתברות לזכות $\frac{m}{n+m} + \frac{m}{n-1+m}$ ואילו כאשר נבחר בכד הכחול ההסתברות לזכות

$$\frac{100m}{100n+100m} + \frac{100m}{100n-1+100m} = \frac{m}{n+m} + \frac{m}{n-1/100+m}.$$

□

תרגיל 4.5 בכד יש 3 כודרים לבנים ו־7 כדורים שחורים. הוציאו באופן אקראי כדור אחד ושמו הצידה. אחרי זה הוציאו עוד כדור, והוא היה לבן. מה ההסתברות שהכדור שהוציאו בפעם הראשונה היה לבן?

פתרון: נסמן W_1־הכדור הראשון היה לבן, W_2־הכדור השני היה לבן. אז

$$P(W_1|W_2) = \frac{P(W_2|W_1)P(W_1)}{P(W_2|W_1)P(W_1) + P(W_2|\overline{W_1})P(\overline{W_1})} = \frac{2/9 \cdot 3/10}{2/9 \cdot 3/10 + 3/9 \cdot 7/10} = \frac{2}{9}.$$

□

תרגיל 4.6 בכל אחד משני כדים יש b כדורים שחורים ו־ w כדורים לבנים. מוציאים כדור מהכד הראשון ושמים אותו בכד השני ואז מוציאים כדור מהכד השני. מה ההסתברות שהכדור השני יהיה לבן?

פתרון: לפי נוסחת ההסתברות השלמה $P = \frac{w}{b+w} \cdot \frac{w+1}{b+w+1} + \frac{b}{b+w} \cdot \frac{w}{b+w+1} = \frac{w}{b+w}$. □

תרגיל 4.7 סביב שולחן עגול עם 12 מקומות מתיישבים 6 זוגות, באופן אקראי לגמרי. מה ההסתברות שכל זוג ישב במקומות צמודים?

פתרון: בכמה אופנים אפשר לסדר את 12 האנשים סביב השולחן? ב־ $12!$ אופנים. בכפה אופנים נקבל סידור "טוב"? ב־ $12\cdot 2\cdot 5!\cdot 2^5$ (ב־12 אופנים נבחר לאן לשים את הגבר בזוג הראשון; ב־2 אופנים נוכל לבחור לאן לשים את אשתו; אז ב־$5!$ אופנים נוכל לבחור איך לסדר את 5 הזוגות הנותרים, וב־2^5 אופנים נבחר את הסידור הפנימי בתוך הזוגות). כלומר התשובה היא $\frac{5!}{11!}2^6 = \frac{12\cdot2\cdot5!\cdot2^5}{12!}$. $\square$

תרגיל 4.8 שני ציידים יורים בצבי בו זמנית ובאופן בלתי־תלוי. ידוע שהצייד הראשון פוגע במטרה בהסתברות 0.8, והשני 0.4. הצבי הרוג, ונמצא בו כדור אחד. איך צריך לחלק את הצבי? (רמז: צריך להשתמש בנוסחת בייס).

פתרון: נסמן מאורעות, A־נמצא כדור אחד בצבי. A_{10}־הציד הראשון פגע והשני לא, A_{01}־הציד השני פגע והראשון לא, A_{11}־שני הצידים פגעו, A_{00}־אף ציד לא פגע. אז לפי נוסחת בייס

$$\begin{aligned} P(A_{10}|A) &= \frac{P(A|A_{10})P(A_{10})}{P(A|A_{00})P(A_{00})+P(A|A_{10})P(A_{10})+P(A|A_{01})P(A_{01})+P(A|A_{11})P(A_{11})} \\ &= \frac{1\cdot(0.8\cdot0.6)}{0\cdot P(A_{00})+1\cdot(0.8\cdot0.6)+1\cdot(0.2\cdot0.4)+0\cdot P(A_{11})} = \frac{6}{7}. \end{aligned}$$

באותו אופן $P(A_{10}|A)=\frac{1}{7}$. ולכן יש להביא לציד הראשון $\frac{6}{7}$ מהצבי ולשני את היתר (תנחומינו לשני, כמו גם לצבי). $\square$

תרגיל 4.9 ישנם 10 מטבעות הוגנים, אבל על 2 מהם ה"עץ" מופיע משני הצדדים. בוחרים באקראי מטבע מ־10 המטבעות הללו ומטילים אותו 3 פעמים. מה ההסתברות שהופיע 3 פעמים "עץ"?

פתרון: לפי נוסחת הסתברות שלמה $\frac{3}{10} = \frac{8}{10}\cdot\frac{1}{8} + \frac{2}{10}\cdot 1 = P$. $\square$

פרק 5

נסיונות ברנולי. התפלגות בדידה קשורות לניסיונות ברנולי.

הגדרה 5.1 ניסוי ש־Ω שלו מכיל רק שתי תוצאות נקרא ניסוי ברנולי. במקרה הזה לתוצאה אחת קוראים "הצלחה" ולתוצאה שניה "כישלון", במילים אחרות $\Omega = \{0,1\}$ כאשר $0 =$ כשלון, $1 =$ הצלחה, $p(1) = p$, ו־ $p(0) = q = 1 - p$.

דוגמא 5.1 (1) בהטלת מטביע הוגן אז מתקיים $p = q = 0.5$. (2) הטלת קוביה $= \{$ "הצלחה"קבלת 5 או 6$\}$. אם הקוביה הוגנת אז מתקיים $q = \frac{2}{3}$ ו־ $p = \frac{1}{3}$.

הגדרה 5.2 סדרת נסיונות נקראת סדרת ברנולי אם:

1. כל נסיון בסדרה הוא נסיון ברנולי.
2. ההסתברות להצלחה לא משתנה מניסוי לניסוי.
3. כל הנסיונות בלתי תלויים.
4. מספר נסיונות n קבוע מראש.

הגדרה 5.3 התפלגות הקשורה לסדרת נסיונות ברנולי העונה על השאלה "מהי ההסתברות שמספר ההצלחה בין n נסיונות יהיה k" נקראת התפלגות בינומית הסמומנת $X \sim B(n,p)$ כאשר $\Omega = \{0,1,2,\cdots,n\}$ ו־ "A_k" $= \{$ בדיוק k הצלחות$\}$ ואז מתקיים $P(A_k) = \binom{n}{k} p^k q^{n-k}$.
מזה נובע ש־ $0 \le P(A_k)$ וגם

$$\sum_{k=0}^{n} P(A_k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1,$$

לכן התפלגות זו נקראת התפלגות בינומית.

5.1 תרגילים

תרגיל 5.1 ההסתברות לפגוע במטרה ביריה בודדת היא 0.2. מה ההסתברות לפגוע לפחות שלוש פעמים במטרה כאשר יורים 20 יריות?

פתרון: לפי נוסחת ברנולי $1-\binom{20}{0}0.2^0 0.8^{20}-\binom{20}{1}0.2^1 0.8^{19}-\binom{20}{2}0.2^2 0.8^{18}$. □

תרגיל 5.2 (1) שני חובבי כדורגל מרמת גן בועטים מקו ה־11 כל אחד שלוש פעמים. לשניהם סיכוי הכנסת גול 0.5. מצא את ההסתברות ששניהם יבקיעו אותו מספר גולים.
(2) שני שחקני כדורסל קולעים לסל כל אחד שלוש פעמים. הסתברות הקליעה לסל של הראשון היא 0.6, ואילו של השני 0.7. מצא את ההסתברות ששניהם יקלעו אותו מספר קליעות.

פתרון: פתרון של (1) הוא

$$P=\sum_{j=0}^{3}\left(\binom{3}{j}\cdot 0.5^j\cdot 0.5^{3-j}\right)^2=\frac{5}{16}.$$

פתרון של (2) הוא

$$P=\sum_{j=0}^{3}\binom{3}{j}^2\cdot 0.6^j\cdot 0.7^j\cdot 0.4^{3-j}\cdot 0.3^{3-j}.$$

□

תרגיל 5.3 מטילים מטבע הוגן n פעמים. נסמן ב־ k_1 את מספר הפעמים שיצא "עץ" וב־ k_2 את מספר הפעמים שיצא "פלי". חשב את ההסתברות ש־ $k_1k_2=n^2-n$.

פתרון: נניח קודם $n=1$. אז ברור ש־ k_1 או k_2 חייבים להיות 0, ולכן שני אגפי השוויון הינם 0 תמיד. נניח $n>1$. אז $(k_1+k_2)^2=n^2$ לכן

$$k_1^2+k_2^2=n^2-2k_1k_2=n^2-2n^2+2n=-n^2+2n>0$$

ששקול ל־ $n<2$ ולכן ההסתברות עבור $n>1$ היא אפסץ □

תרגיל 5.4 במחקר התנהגותי מבצעים את הניסוי הבא: לחולדה נותנים לבחור אחת משלושה דלתות: אם היא בוחרת את הדלת הנכונה אז היא מקבלת מנת מזון ובזה מסתיים הניסוי. אחרת, היא מקבלת מכת חשמל חלשה וחוזרים על הניסוי אך לא יותר משלוש פעמים. 70 אחוז מן החולדות בוחרות כל פעם אחת מהדלתות באקראי ("חולדות טיפשות") ושאר החולדות לא מנסות אותה דלת פעמיים ("חולדות חכמות"). מה תהיה ההסתברות שמתוך 5 חולדות שתיים או שלוש בדיוק ימצאו את הדלת הנכונה?

פתרון: מה ההסתברות שחולדה בודדה שנבחרה באקראי תמצא את הדלת הנכונה? אם זו חולדה חכמה, היא בטוח תצליח. אם זו חולדה טיפשה, היא תצליח בהסתברות $\frac{19}{27} = 1 - \frac{2^3}{3^3}$. לכן לפי נוסחת ההסתברות השלמה $P = 0.7 \cdot \frac{19}{27} + 0.3 \cdot 1 = \frac{107}{135}$. ההסתברות הרצויה היא, לפי נוסחת ברנולי $\binom{5}{2}\left(\frac{107}{135}\right)^2\left(\frac{28}{135}\right)^3 + \binom{5}{3}\left(\frac{107}{135}\right)^3\left(\frac{28}{135}\right)^2 \approx 0.27$. □

תרגיל 5.5 בחינה אמריקאית מכילה 5 שאלות. לכל שאלה יש 4 תשובות שרק אחת מהן נכונה. כדי לעבור את הבחינה צריך לציין לפחות 3 תשובות נכונות. סטודנט אחד שבכלל לא מכיר את החומר חולם בהקיץ על כך שהוא יעבור את המבחן. מה ההסתברות שחלומותיו יתגשמו?

פתרון: ההסתברות שהסטודנט יצליח בשאלה אחת היא $\frac{1}{4}$. ההסתברות שהוא יעבור את המבחן היא ההסתברות שב־5 ניסוי ברנולי עם הסתברות הצלחה $\frac{1}{4}$ כל אחד, יהיו לפחות 3 הצלחות. הסתברות זו, לפי נוסחת ברנולי, היא $\sum_{j=3}^{5}\binom{5}{j}\left(\frac{1}{4}\right)^j\left(\frac{3}{4}\right)^{5-j} = \frac{53}{512}$. □

פרק 6

משתנה מקרי

6.1 תורת המידה על רגל אחת

נתון מרחב מדגם בדיד $\Omega = \{A_i\}_{i\in I}$ כאשר A_i מאורע אלמנטרי. נגדיר $P : \Omega \to \mathbb{R}$ כזאת ש:

1. $0 \le P(A_i) \le 1$ לכל $i \in I$.
2. $\sum_{i\in I} P(A_i) = 1$.

נקח $\mathcal{A}$־משפחה של כל התת־קבוצות של Ω. ראינו ש־ $\mathcal{A}$ היא σ אלגברה. ברור כי כל $A \in \mathcal{A}$ היא מהצורה $A = \{A_{i_j}\}_{i_j\in I}$ תת־אוספים של מאורעות אלמטריים. נרחיב $P : \Omega \to \mathbb{R}$ ל־ $P : \mathcal{A} \to \mathbb{R}$ באופן טבעי על ידי $p(A) := \sum_{i_j} P(A_{i_j})$. ראינו ש־ $(\Omega, \mathcal{A}, P)$ הוא מרחב הסתברות ולהפך אם $(\Omega, \mathcal{A}, P)$־מרבח הסתברות ו־ Ω בדיד אז נקבע

$$\Omega' = \{A_i | \forall A \in \mathcal{A}_{A_i}, A \cdot A_i = \emptyset\}$$

אזי $\Omega' = \coprod_{i\in I} A_i$ $\mathcal{A}$־משפחה של כל התת־קבוצות של Ω' וקיים צימצום $P : \mathcal{A} \to \mathbb{R}$ ל־ $P : \Omega' \to \mathbb{R}$ המקיים

1. $0 \le P(A_i) \le 1$.
2. $\sum_{i\in I} P(A_i) = 1$.

עכשיו נברר מה קורה במקרה ש־ Ω לא בדידה.

הגדרה 6.1 Ω־קבוצה כלשהי. משפחת תת־קבוצות לא ריקה שלה נקראת σ־אלגברה אם היא מקיימת

א. $\mathcal{A} \in A \Rightarrow \overline{A} \in \mathcal{A}$.

ב. $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A} \Rightarrow \sum_{i=1}^{\infty} A_i \in \mathcal{A}$.

הגדרה 6.2 נתונים Ω ו־ $\mathcal{A}$ σ־אלגברה שלה. פונקציה $\mu : \mathcal{A} \to [0, \infty)$ נקראת מידה אם היא מקיימת

א. $\mu(\emptyset) = 0$.

ב. $\forall \{A_i\}_{i \in I}$ זרים $\mu(\sum_{i \in I} A_i) = \sum_{i \in I} \mu(A_i)$.

אם בנוסף לכל $A \in \mathcal{A}$ מתקיים $\mu(A) < \infty$ אזי μ נקראת מידה סופית. נשווה עם הדגרת פונקצית הסתברות ונקבל שפונקצית הסתברות היא מידה סופית מנורמלת, זאת אומרת המקיימת $\mu(\Omega) = 1$. לפעמים לפונקצית הסתברות קוראים מידת הסתברות. נזכיר כי היא מקיימת משפט 1 מפרק 1.3. כאשר דיברנו על σ־אלגבראות אמרנו שדרישה יותר טבעית היא שאם $\{A_i\}_{i=1}^{n}$ במשפחה של תת־קבוצות אזי גם איחוד $\sum_{i=1}^{n} A_i$ במשפחה.

הגדרה 6.3 נקרא למשפחה $\mathcal{H}$ של תת־קבוצות Ω אלגברה אם

א. $\mathcal{H}$ לא ריקה.

ב. אם $A \in \mathcal{H}$ אזי $\overline{A} \in \mathcal{H}$.

ג. אם $\{A_i\}_{i=1}^{n} \subset \mathcal{H}$ אזי $\sum_{i=1}^{n} A_i \in \mathcal{H}$.

$\Omega, \emptyset \in \mathcal{H}$ תמיד, וכל σ־אלגברה היא אלגברה.
השאלה: נתונים $\Omega, \mathcal{H}$־אלגברה ופונקציה $P : \mathcal{H} \to [0, 1]$ המקיימת

1. $P(\Omega) = 1$

2. לכל $\{A_i\}_{i=1}^{n} \subset \mathcal{H}$ וזרים אזי $P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

האם אפשר להרחיב את $\mathcal{H}$ ל־ σ־אלגברה $\mathcal{A}$ זאת אומרת $\mathcal{H} \subset \mathcal{A}$ ופונקציה P לפונקצית הסתברות הבאה $P' : \mathcal{A} \to [0, 1]$ כך שלכל $A \in \mathcal{H}$ מתקיים $P'(A) = P(A)$.

משפט 6.1 (משפט ההרחבה) יהי $\Omega \neq \emptyset$, $\mathcal{H}$־אלגברה מעל Ω ו־ $P : \mathcal{H} \to [0, 1]$ כזאת ש־

א. $P(\Omega) = 1$.

ב. לכל $\{A_i\}_{i=1}^{n}$ זרים מתקיים $P(\sum_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

3. לכל סדרה יורדת $\{A_i\}_{i=1}^{\infty} \subset \mathcal{H}$ המקיימת $A_{i+1} \supset A_i$ ו־ $\prod A_i = \emptyset$ מתקיים

$$\lim_{i\to\infty} P(A_i) = 0,$$

אזי קיימת σ־אלגברה $\mathcal{A}$ מינימלית יחידה שמכילה את $\mathcal{H}$ וקיימת פונקצית הסתברות יחידה : P' $\mathcal{A} \to [0,1]$ כזאת שלכל $A \in \mathcal{H}$ מתקיים $P'(A) = P(A)$.

ראינו בעבר ש־ $\mathcal{A}$ היא חיתוך של כל ה־ σ־אלגבראות שמכילות את $\mathcal{H}$, וזו הדרך שבה אנו בונים את $\mathcal{A}$, יש לציין שאת הקיום של P' תראו בקורס "מבוא לאנליזה פונקציונלית".

הגדרה 6.4 σ־אלגברה $\mathcal{A}$ מעל Ω נקראת שלמה ביחס למידת הסתברות P אם לכל $A \in \mathcal{A}$ כזאת ש־ $P(A) = 0$ מתקיים כל $B \subset A$ היא ב־ $\mathcal{A}$ זאת אומרת $B \in \mathcal{A}$.

משפט 6.2 (משפט השלמה) לכל מרחב הסתברות $(\Omega, \mathcal{A}, P)$ קיימת השלמה $(\Omega, \mathcal{A}', P')$ כך ש־ $\mathcal{A} \subseteq \mathcal{A}'$ ו־ $P'(A) = P(A)$ לכל $A \in \mathcal{A}$.

הגדרה 6.5 יהי (Ω, A, μ) כאשר μ היא מדיה. הקבוצה $A \in \mathcal{A}$ המקיימת $\mu(A) = 0$ נקראת קבוצת מידת אפס.

דוגמא 6.1 יהי $\Omega = [a_0, b_0]$ או $\Omega = \mathbb{R}$. נגדיר $\mathcal{A}$ היא σ־אלגברה הנוצרת על ידי אינטרבלים פתוחים ב־ Ω. נגדיר $\mu((a,b)) = b - a$ כאשר $(a,b) \in \Omega$ אלגברה הנוצרת על ידי אינטרבלים פתוחים מקיימת תנאים של משפט 6.1 (מלבד התנאי $P(\Omega) = 1$ שהוא חשוב רק לעובדה שהמידה היא מידת הסתברות) לכן באמת קיימת מידה על כל התת קבוצות ב־ $\mathcal{A}$.

הגדרה 6.6 אם $\Omega = \mathbb{R}$ או $\Omega = [a_0, b_0]$ אזי כל התת־קבוצות של $\mathcal{A}$ σ־אלגברה הנוצרת על ידי קטעים פתוחים נקראת קבוצת בורל. המידה המוגדרת על ידי $\mu((a,b)) = b - a$ כאשר $(a,b) \in \Omega$. נקראת מידת לבג כאשר $\Omega = \mathbb{R}$, מידת בורל כאשר $\Omega = [a_0, b_0]$.

טענה 6.1 כל אוסף של נקודות ב־ Ω ניתן להמנות הוא קבוצת מידה אפס.

הוכחה: $a \in \Omega$ אזי קיים n_0 כזה ש־ $(a - \frac{1}{n_0}, a + \frac{1}{n_0}) \subset \Omega$ ו־ $\{a\} = \prod_{n=n_0}^{\infty} (a - \frac{1}{n}, a + \frac{1}{n})$. ולכן $\mu(\{a\}) = \lim_{n\to\infty}(a + \frac{1}{n} - a + \frac{1}{n}) = 0$. לכן לכל אוסף של נקודות $\{a_i\}_{i\in I}$ מקבלים $\mu(\sum_{i\in I}\{a_i\}) = \sum_{i\in I}\mu\{a_i\} = 0$. $\square$

יש לציין שיש גם אוספים שלא ניתנים להמנות של מידת אפס. תהי $f : \mathbb{R} \to \mathbb{R}$ פוקנציה רציפה מונוטונית עולה (לא יורדת) נראה כי $\mu : \mathcal{A} \to [0, \infty)$ המסודרת על ידי $\mu(a,b) = f(b) - f(a)$ היא מידה. אבל תחילה נשנה את הקבוצות הבסיסיות של $\mathcal{A}$.

טענה 6.2 יהי $\Omega[a_0, b_0]$ או $\Omega = \mathbb{R}$ נגדיר $\mathcal{A}'$ σ־אלגברה המוגדרת על ידי קטעים חצי פתוחים $[a,b)$ אזי $\mathcal{A} = \mathcal{A}'$ כאשר $\mathcal{A}$ σ־אלגברה של קבוצות בורל.

הוכחה:

1. נראה שלכל $a,b \in \Omega$, $[a,b) \in \mathcal{A}$ מתקיים $(a-\frac{1}{n},b) \prod_{n=n_0}^{\infty} = [a,b)$ לכן $\mathcal{A}' \supset \mathcal{A}$.

2. נראה שלכל $a,b \in \Omega$, $(a,b) \in \mathcal{A}'$ מתקיים $[a+\frac{1}{n},b) \sum_{n=n_0}^{\infty} = (a,b)$. לכן $\mathcal{A} \supset \mathcal{A}'$.
קיבלנו $\mathcal{A} = \mathcal{A}'$.

□

מובן שאלגברה $\mathcal{H}$ של איחודים, חיתוכים והשלמות סופיות μ המוגדרת על ידי $\mu([a,b)) = f(b) - f(a)$ מקיימת תנאי משפט 6.1 נראה זה באופן ישיר.

טענה 6.3 אם $\{E_1, E_2, \ldots E_n\}$ קטעים חצי פתוחים זרים ו־ E_0 קטע כזה ש־ $E_i \subset E_0$ לכל i אזי

$$\sum_{i=1}^{n} \mu(E_i) \leq \mu(E_0).$$

הוכחה: נגדיר $E_0 = [c,d)$, $E_i = [a_i, b_i)$ כאשר הם מסודרים בסדר עולה

$$a_1 < a_2 < \ldots < a_n$$

בגלל שהם זרים נקבל

$$a_1 < b_1 \leq a_2 < \ldots \leq a_n < b_n$$

ומכיוון ש־ $E_i \subset E_0$ נקבל $a_1 \geq c$, $b_n \leq d$, ולכן מתקיים

$$\begin{aligned}\sum_{i=1}^{n} \mu(E_i) = \sum_{i=1}^{n} f(b_i) - f(a_i) \leq \sum_{i=1}^{n} f(b_i) - f(a_i) + \sum_{i=1}^{n-1} f(a_{i+1}) - f(b_i) = \\ = f(b_n) - f(a_1) \leq f(d) - f(c).\end{aligned}$$

□

טענה 6.4 אם קטע סגור $F = [c,d]$ מוכל באיחוד של מספר סופי של קטעים פתוחים

$$\sum_{i=1}^{n} U_i = \sum_{i=1}^{n} (a_i, b_i)$$

אזי

$$\mu(F) \leq \sum_{i=1}^{n} \mu(U_i).$$

הוכחה: $F \subset \sum_{i=1}^{n} U_i$ אזי קיים k_1 כזה ש־ $c \in U_{k_1}$. אם $d \in U_{k_1}$ אז גמרנו כי $\mu(F) = f(d) - f(c) \leq f(b_{k_1}) - f(a_{k_1})$. אחרת, קיים k_2 כזה ש־ $b_{k_1} \in U_{k_2}$. אם $d \in U_{k_2}$ אז גמרנו, אחרת קיים k_3 כזה $b_{k_3} \in U_{k_3}$ וכך הלאה. בסופו של דבר נגיע ל־ U_{k_m} כזה ש־ $d \in U_{k_m}$. נשנה את המספור של U כך ש־ $U_{k_i} = U_i$ ונקבל $a_1 < c < b_1$ ו־ $a_m < d < b_m$ ולכל $i = 1, 2, \ldots, m-1$ נקבל $a_{i+1} < b_i < b_{i+1}$ אז

$$\mu(F) = f(d) - f(c) \leq f(b_m) - f(a_1) = f(b_1) - f(a_1) + \sum_{i=1}^{m-1} f(b_{i+1}) - f(b_i)$$
$$\leq \sum_{i=1}^{m} f(b_i) - f(a_i)$$

□

טענה 6.5 אם $\{E_0, E_1, \ldots\}$ סדרה של קטעים חצי פתוחים כאלה ש־ $E_0 \subset \sum_{i=1}^{\infty} E_i$ אזי

$$\mu(E_0) \leq \sum_{i=1}^{\infty} \mu(E_i).$$

הוכחה: אם $f(c) = f(d)$ אזי $\mu(E_0) = 0$ וגמרנו. אחרת, $f(d) - f(c) = A > 0$ ובגלל ש־ f היא פונקציה רציפה אזי לכל $0 < \varepsilon < A$ קיים δ_ε כזה ש־ $f(d \pm \delta_\varepsilon) > f(d) \pm \varepsilon$.
יהי $\varepsilon > 0$, $f(d) - f(c) > \varepsilon$ ונקח $\delta > 0$ ונגדיר $F_\varepsilon = [c, d - \delta_\varepsilon]$, $U_{i,\delta} = (a_i - \frac{\delta_i}{2^i}, b_i)$ ומקבלים $U_{i,\delta} \supset E_i$, $F_\varepsilon \subset E_0$, לכן $F_\varepsilon \subset \sum_{i=1}^{\infty} U_{i,\delta}$. לפי משפט היינה־בורל קיים n כזה ש־ $F_\varepsilon \subset \sum_{i=1}^{n} U_{i,\delta}$ לכן על פי טענה קודמת מקבלים

$$\mu(F_\varepsilon) = f(d - \delta_\varepsilon) - f(c) \leq \sum_{i=1}^{n} (f(b_i) - f(a_i - \frac{\delta_\varepsilon}{2^i}))$$

מכאן נובע

$$f(d) - f(c) - \varepsilon \leq \sum_{i=1}^{n} \left(f(b_i) - f(a_i) + \tfrac{\delta}{2^i}\right) \leq \sum_{i=1}^{n} \mu(E_i) + \delta \sum_{i=1}^{\infty} \tfrac{1}{2^i} = \sum_{i=1}^{\infty} \mu(E_i) + \delta$$

ובגלל שזה נכון לכל ε, δ מקבלים $\mu(E_0) \leq \sum_{i=1}^{\infty} \mu(E_i)$. □

טענה 6.6 אם $\{E_i\}_{i=1}^{\infty}$־אוסף של קטעים זרים חצי פתוחים כאלה ש־ $E_0 = \sum_{i=1}^{\infty} E_i$־קטע חצי פתוח אזי

$$\mu(E_0) = \sum_{i=1}^{\infty} \mu(E_i).$$

הוכחה: אם $\sum_{i=1}^{\infty} E_i \subseteq E_0$ נקבל ש־ $\mu(E_0) \leq \sum_{i=1}^{\infty} \mu(E_i)$ ומצד שני אם $\sum_{i=1}^{\infty} E_i \supseteq E_0$ נקבל ש־$\mu(E_0) \geq \sum_{i=1}^{\infty} \mu(E_i)$. □

טענה 6.7 על $\mathcal{H}$ בנוי מעל קטעים פתוחים קיימת מידה יחידה סופית $\overline{\mu}$ כזאת ש־ $\overline{\mu}(E) = \mu(E)$ לכל קטע חצי סגור.

הוכחה: יהיו $E \in \mathcal{H}$, $E = \sum_{i=1}^{n} E_i$, $E = \sum_{j=1}^{m} F_j$ ומתקיים $E_i = \sum_{j=1}^{m} E_i F_j$ מכאן מקבלים $\sum_{i=1}^{n}\sum_{j=1}^{m} \mu(E_iF_j) = \sum_{i=1}^{n} \mu(E_i)$ ומצד שני מתקיים $\sum_{j=1}^{m}\sum_{i=1}^{n} \mu(E_iF_j) = \sum_{j=1}^{m} \mu(F_j)$, כלומר הפונקציה $\mu : H \to [0,\infty)$ מוגדרת באופן יחיד. נראה ש־ μ היא σ־אדיטיבית ; יהיו $\{E_i\}$ זרים, ולכל i מתקיים $E_i = \sum_j E_{i_j}$ קטעים חצי פתוחים מתקיים $\mu(E_i) = \sum_j \mu(E_{i_j})$זאת אומרת אם E היא קטע חצי פתוח, אז ראינו שיש σ אדיטיהית של הפונקציה μ. אחרת, $E = \sum F_j$ כאשר F_j־ קטע חצי פתוח ומכאן מקבלים

$$\mu(E) = \sum_{j=1}^{m} \mu(F_j) = \sum_{j=1}^{m}\sum_{i=1}^{n} \mu(E_iF_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mu(E_iF_j) = \sum_{i=1}^{n} \mu(E_i).$$

ולכן μ היא σ אדיטיבית על $\mathcal{H}$. □

כמסקנה מהטענות הקודמות נקבל את המשפט הבא.

משפט 6.3 נקח $f : \mathbb{R} \to \mathbb{R}$ רציפה, מונטונית לא יורדת ונקבל $\mu_f : [a_0, b_0] \to [0,1]$ המוגדרת על ידי $\mu_f(a,b) = \frac{f(b)-f(a)}{f(b_0)-f(a_0)}$ היא מידת הסתברות.

דוגמא 6.2

1. זורקים 2 נקודות על קטע $[0,1]$ מהי ההסתברות ש־$\min(x,y) < \frac{1}{4}$ בתנאי ש־ $\max(x,y) > \frac{2}{3}$? הפתרון הוא כדלקמן המאורעות הם

$$A = \{(x,y) | \max(x,y) > \frac{2}{3}\}, B = \{(x,y) | \min(x,y) < \frac{1}{4}\}.$$

מכאן נובע ש־ $P(A) = \frac{1}{3} + \frac{1}{3} - \frac{1}{9} = \frac{5}{9}$, $P(BA) = (\frac{1}{4} \cdot \frac{1}{3}) \cdot 2 = \frac{1}{6}$ ו־$P(B|A) = \frac{\frac{1}{6}}{\frac{5}{9}} = \frac{9}{30} = 0.3$.

2. אם זורקים באופן מקרי 2 נקודות על קטע $[0,1]$. מהי ההסתברות שמהקטעים $[\max\{x,y\},1]$, $[\min\{x,y\},\max\{x,y\}]$, $[0,\min\{x,y\}]$ אפשר לבנות משולש. כידוע בהינתן a,b,c אורכי קטעים הם יוצרים משולש אם ורק אם מתקיים $(a+b>c) \wedge (a+c>b) \wedge (b+c>a)$.
כדי לפתור את הדוגמא,נטפל בכמה מקרים;

(א) אם $x<y$ אזי נקבל $\begin{cases} y>0.5 \\ x<0.5 \\ y<x+\frac{1}{2} \end{cases} \Leftrightarrow \begin{cases} x+y-x>1-y \\ x+1-y>y-x \\ y-x+1-y>x \end{cases}$

(ב) אם $x>y$ נקבל (לפי סמיטריה) $\begin{cases} x>0.5 \\ y<0.5 \\ x<y+\frac{1}{2} \end{cases}$ הקבוצות שלנו מוגדרות על ריבוע שאם נציירו נקבל ש־$P=\frac{1}{4}$.

התפלגות אחידה על קטע דבר ברור מאוד, אבל לא תמיד "דבר מקרי " מוגדר היטב. למשל **הפרדוקס של ברטרנד:** בוחרים מיתר מקרי במעגל יחידה. מהי ההסתברות שהוא יהיה גדול מהצלע של המשולש המושכלל החסום על ידיד המעגל? (זאות אורמת מ־$\sqrt{3}$). התשובה תלויה בהגדרת מיתר "מיקרי".

1. נבחר נקודה P באופן מיקרי בעיגול ונעביר דרכה מיתר הניצב לרדיוס , אנו יודעים את אורכו של מיתר CD, כי נתון שהוא גדול מהצלע של המשולש של המשוכלל החסום על ידי המעגל, זאת אומרת נתון $|CD|>\sqrt{3}$ אם ורק אם $|OP|<\frac{1}{2}$ מכאן נובע ש־$P=\frac{\Pi\cdot(\frac{1}{2})^2}{\Pi\cdot(1)^2}=\frac{1}{4}$.

2. זורקים 2 נקודות על מעגל. אפשר להניח שנקודה x קבועה ונקודה שניה y מקרית מכאן מקבלים ש־$P=\frac{1}{3}$.

6.2 משתנה מקרי

הגדרה 6.7 יהיו $(\Omega,\mathcal{A})$, $(\Omega',\mathcal{A}')$ פונקציה $f:\Omega\to\Omega'$ נקראת מדידה אם לכל $A'\in\mathcal{A}'$ מתקיים $f^{-1}(A')\in\mathcal{A}$. במקרה שלנו $\Omega=\mathbb{R},\mathbb{R}^n$ ו־ $\mathcal{A}$ היא σ־אלגברת קבוצות בורל, זאת אומרת σ־אלגברה הבנויה על קטעים פתוחים במקרה של $\mathbb{R}$ או תיבות פתוחות במקרה של $\mathbb{R}^n$.

הגדרה 6.8 משתנה מקרי על מרחב הסתברות $(\Omega,\mathcal{A},P)$ זוהי פונקציה $X:\Omega\to\mathbb{R}(\mathbb{R}^n)$ המקיימת את התכונה שלכל קטע פתוח (a,b) מתקיים $X^{-1}(a,b)=C\in\mathcal{A}$ ומגדירים

$$P(a<X<b)=P(C).$$

הערה 6.1 א. אם Ω בדיד אזי $\Omega=\{a_1,a_2,\ldots\}$. במקרה הזה כל פונקציה $X(a_i)=x_i$ כאשר $x_i\in\mathbb{R}(\vee\mathbb{R}^n)$ היא משתנה מקרי ו־ $P(X=x_i)=P(a_i)$, $P(X=x)=0$ לכל $x\notin\{x_1,x_2,\ldots\}$. למשל בהתפלגות ברנולי הכישלון זה q והצלחה זה p וההסתברות של השאר היא תמיד אפס.

ב. לפעמים Ω הוא רציף אבל $X:\Omega\to\mathbb{R}$ בדיד. למשל אם מודדים רמת סוכר בדם מקבלים 4 קטגוריות

(א) רמת סוכר נמוכה מדי ־ 0.

(ב) רמת סוכר בנורמה ־ 1.

(ג) רמת סוכר גבולי ־2.

(ד) רמת סוכר גבוהה ־3.

3. Ω מרחב מדגם רציף ו־ $X:\Omega\to\mathbb{R}$ רציף. למשל התפלגות זמן המתנה לאוטובוס, התפלגות גובה, התפלגות אורך החיים של מכשיר מסויים וכו'. תפקידו של המשתנה המקרי הוא להעיבר את המידה במרחב מדגם לערכים מספריים. אם Ω מספרי בהתחלה, אזי זה אמצעי פורמלי.

הגדרה 6.9 משתנה מקרי X נקרא בדיד אם קיימת קבוצה $\{x_i\}_{i\in I}\in\mathbb{R}(\vee\mathbb{R}^n)$ כזאת שלכל $x\in\mathbb{R}(\vee\mathbb{R}^n)$ אם $x\notin\{x_i\}$ אזי $P(X=x)=0$, $0\le P(X=x_i)\le 1$ ו־ $\sum\limits_{i\in I}P(X=x_i)=1$.

הגדרה 6.10 יהי $(\Omega,\mathcal{A},P)$ מרחב הסתברות. אם משתנה מקרי $X:\Omega\to\mathbb{R}(\vee\mathbb{R}^n)$ מקיים $P(X=x)=0$ לכל $x\in\mathbb{R}(\mathbb{R}^n)$ אזי X נקרא משתנה מקרי רציף.

הגדרה 6.11 פונקצית התפלגות מצטברת של משתנה מקרי X, $X:\Omega\to\mathbb{R}$ היא $F:\mathbb{R}\to[0,1]$ המוגדרת על ידי $F(x)=P(X\le x)$.

דוגמא 6.3

1. נסיון ברנולי של הטלת מטבע הוגן; $P(X=0)=0.5$, $P(X=1)=0.5$ ו־ $P(X=x)=$ 0 לכל $x\notin\{0,1\}$.
$$F(x)=\begin{cases}0, x<0\\ 0.5, 0\le x<1\\ 1, 1\le x\end{cases}$$

2. אם X משתנה מקרי בדיד $\{x_i\}_{i\in I}$ ו־ $x_1<x_2<\cdots$ אזי

$$F(x)=\begin{cases}0, x<x_1\\ P(X=x_1), x_1\le x<x_2\\ P(X=x_1)+P(X=x_2), x_2\le x<x_3\\ \vdots\end{cases}$$

זוהי פונקצית מדריגות. אם X משתנה מקרי אזי צריך לחשב כל ה־ $\{x_i\}$ משמאל לנקודה הנתונה (כולל) ולחשב סכום הסתברויות שלהן.

3. התפלגות אחידה בקטע $[a,b]$, $F(x)=\begin{cases}0, & x\le a\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$

טענה 6.8 תכונות של פונקצית ההתפלגות F

1. F היא פונקציה מונוטונית עולה חלש זאת אאומרת לכל $x_1,x_2\in\mathbb{R}$ כאלה ש־ $x_1<x_2$ מתקיים $F(x_1)\le F(x_2)$.

2. $\lim_{x\to\infty} F(x) = 1$.

3. $\lim_{x\to-\infty} F(x) = 0$.

4. F רציפה מימין $\lim_{x\to a^+} F(x) = F(a)$.

5. $P(X > a) = 1 - F(a)$.

6. $P(a < X \le b) = F(b) - F(a)$.

7. $P(X < a) = \lim_{n\to\infty} F(a - \frac{1}{n}) = F(a^-)$.

8. $P(X \ge a) = 1 - F(a^-)$.

9. $P(X = a) = F(a) - F(a^-)$.

10. אם X משתנה מקרי רציף אזי F היא פונקציה רציפה מונוטונית לא יורדת כך ש־ $F(a) = F(a^-)$ ומתקיים $P(X \ge a) = P(X > a) = 1 - F(a)$.

הוכחה:

1. אם $(-\infty, x_1] \subset (-\infty, x_2]$ אזי $X^{-1}((-\infty, x_1]) \subset X^{-1}((-\infty, x_2])$, לכן אם $x_1 < x_2$ אזי $F(x_1) \le F(x_2)$.

2. תהי $A_n = X^{-1}((-\infty, n])$ כך ש־ $\sum_n A_n = \Omega$ ו־ $\ldots \subseteq A_{n+2} \subseteq A_{n+1} \subseteq A_n$ אזי

$$1 = P(\sum_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} F(n) = \lim_{x\to\infty} F(x).$$

3. $A_n = X^{-1}((-\infty, -n])$ ולכל n מתקיים $A_n \supset A_{n-1}$ כך ש־ $\prod_{n=1}^{\infty} A_n = \emptyset$ אזי

$$\lim_{x\to-\infty} F(x) = \lim_{n\to\infty} P(A_n) = P(\prod_{n=1}^{\infty} A_n) = P(\emptyset) = 0$$

4. $A_n = X^{-1}((-\infty, a + \frac{1}{n}])$ ו־ $\prod A_n = X^{-1}((-\infty, a])$ מכאן מקבלים

$$\lim_{x\to a^+} F(x) = \lim_{n\to\infty} P(A_n) = F(a).$$

5. $F(\infty) = 1 = F(a) + P(X > a)$ ומזה מקבלים $P(X > a) = 1 - F(a)$.

6. $P(a<X\le b)=P(X\le b)-P(X\le a)=F(b)-F(a)$.

7. נסמן $A_n=X^{-1}(-\infty,a-\frac{1}{n}]$, $\sum\limits_{n=1}^{\infty}A_n=X^{-1}((-\infty,a))$ לכל מתקיים $A_n\subset A_{n+1}$

אזי
$$\begin{aligned}P(X<a)&=P(\sum_{n=1}^{\infty}A_n)=\lim_{n\to\infty}P(A_n)=\\&=\lim_{n\to\infty}F(a-\tfrac{1}{n})=F(a^-)\end{aligned}$$

8. $P(X\ge a)=1-P(X<a)=1-F(a^{-1})$.

9. $P(X=a)=P(X\le a)-P(X<a)=F(a)-F(a^-)$.

10. נובע ישירות מהעובדה ש־ $P(X=a)=0$ לכל a כי a הוא משתנה מקרי רציף.

□

6.3 מומנטים של התפלגות בדידה

יהי X משתנה מקרי זאת אומרת $\{x_i\}_{i\in I}\subset\mathbb{R}$ עם $0\le P(X=x_i)\le 1$ ו־ $\sum\limits_{i\in I}P(X=x_i)=1$. לכל $x\notin\{x_i\}_{i\in I}$ מתקיים $P(X=x)=0$. נגדיר $f(x_i)=P(X=x_i)$ אם $\sum\limits_{i\in I}|x_i|f(x_i)$ קיים זאת אומרת טור זה $\sum\limits_{i\in I}x_if(x_i)$ מתכנס בהחלט, אזי $\mu_X=E(X)=\sum\limits_{i\in I}x_if(x_i)$. אחרת אומרים של־$X$ אין תוחלת. אפשר להעיר שההתכנסות בהחלט נדרשת כדי ששינוי בסדר הנקודות לא יביא לשינוי של התוחלת.

דוגמא 6.4 א. אם $X=\{0,1\}$ נגדיר נסיון ברנולי באופן הבא $\begin{cases}f(0)=q\\f(1)=p\end{cases}$ מכאן מקבלים שהתוחלת $E(X)=0\cdot q+1\cdot p=p$.

ב. $X\sim B(n,p)$ המשתנה המקרי מתפלג לפי התפלגות בינומית, כלומר ההסתברות ש־
$$f(k)=P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

לכן
$$\begin{aligned}E(X)&=\sum_{k=0}^{n}k\cdot f(k)=\sum_{k=0}^{n}k\tbinom{n}{k}p^k(1-p)^{n-k}=\sum_{k=1}^{n}\tfrac{n(n-1)!}{(k-1)!(n-k)!}p^k(1-p)^{n-k}=\\&=n\sum_{k=1}^{n}\tbinom{n-1}{k-1}p^k(1-p)^{n-k}\underbrace{=}_{l=k-1}np\sum_{l=0}^{n-1}\tbinom{n-1}{l}p^l(1-p)^{n-1-l}\\&=np(p+(1-p))^{n-1}=np.\end{aligned}$$

למשל הטלת קוביה היא בעלת התוחלת
$$E(X)=1\cdot\tfrac{1}{6}+2\cdot\tfrac{1}{6}+3\cdot\tfrac{1}{6}+4\cdot\tfrac{1}{6}+5\cdot\tfrac{1}{6}+6\cdot\tfrac{1}{6}=\tfrac{1}{6}(1+2+3+4+5+6)=\tfrac{7}{2}.$$

3. $X=\{2^i\}_{i=1}^{\infty}$, הצפיפות בנקודה 2^i היא $f(2^i)=2^{-i}$ לכל $i\geq 1$, מכאן מקבלים ש־ $\sum_{i=1}^{\infty} f(x_i)=\frac{\frac{1}{2}}{1-\frac{1}{2}}=1$, כמו כן מקבלים $E(X)=\sum_{i=1}^{\infty} x_i f(x_i)=\sum_{i=1}^{\infty} 2^i\cdot 2^{-i} \underbrace{=}_{\sum_{i=1}^{\infty} 1} \infty$,
כלומר התוחלת אינה קיימת.

הגדרה 6.12 אם $H:\mathbb{R}\to\mathbb{R}$ פונקציה מדידה (למשל רציפה) אזי $Y=H(X)$ הוא משתנה מקרי $F(y)=P(Y\leq y):=P(X\leq \max H^{-1}(y))$

הגדרה 6.13 מומנט מסדר k של X מוגדר על ידי תוחלת של X^k, כלומר $E(X^k)=\sum_{i\in I} x_i^k f(x_i)$.

הגדרה 6.14 יהי X משתנה מקרי בדיד בעל תוחלת. אם קיים $E((X-E(X))^2)$ אזי הוא נקרא שונות של X, ונסמן ב־ $VAR(X)$.

דוגמא 6.5 בכד יש N כדורים ממוספרים מ־ 1 עד N,מוציאים בלי חזרה n כדורים. מתשנה מקרי Y הוא שווה למספר הגדול ביותר בין המספרים שהוצאו. חשב את $E(Y)$ ואת הקשר בין $E(Y)$ ו־ N. הפתרון לבעיה זו הוא כדלקמן

$$E(Y)=\sum_{k=n}^{N}\frac{k\binom{k-1}{n-1}}{\binom{N}{n}}=\frac{1}{\binom{N}{n}}\cdot\sum_{k=n}^{N}k\binom{k-1}{n-1}=\underbrace{=}_{k\binom{k-1}{n-1}=n\binom{k}{n}}\frac{n}{\binom{N}{n}}\cdot\sum_{k=n}^{N}\binom{k}{n}. \tag{6.1}$$

טענה 6.9

$$\binom{n}{n}+\binom{n+1}{n}+\ldots+\binom{N}{n}=\binom{N+1}{n+1}$$

הוכחה: נוכיח באינדוקציה עבור $N=n$ את הטענה. $\binom{n}{n}=\binom{n+1}{n+1}=1$ נניח עבור M ונוכיח עבור $M+1$

$$\underbrace{\binom{n}{n}+\ldots+\binom{M}{n}}_{\text{הנחת האינדוקציה}}+\binom{M+1}{n}=\binom{M+2}{n+1},$$

לכן מספיק להוכיח על פי הנחת האינדוקציה שמתקיים

$$\binom{M+1}{n+1}+\binom{M+1}{n}=\binom{M+2}{n+1}.$$

טענה זו ניתן להראות על ידי בחירת $n+1$ קופים מגן חיות שמכיל $M+1$ קופים, בשני אופנים, הראשונה היא בחירה ישירה זאת אומרת $\binom{M+1}{n+1}$ או נסמן קוף אחד בעל שתי אפשרויות, או לבחור אותו שזה $\binom{M+1}{n}$ או לא לבחור אותו, ובמקרה זה יש $\binom{M+1}{n+1}$ אפשרויות. לכן מכאן מקבלים שמתקיים $\binom{M+1}{n+1}+\binom{M+1}{n}=\binom{M+2}{n+1}$. □

מ־ (6.1) נקבלת ש־ $E(Y)=\frac{n}{\binom{N}{n}}\cdot\binom{N+1}{n+1}=\frac{(N+1)n}{n+1}$. מכאן $\frac{n+1}{n}E(Y)-1=N$.

דוגמא 6.6 חישוב של תוחלת

1. נסיוי ברנולי כאשר $X=\{0,1\}$ משתנה מקרי ו־ $f(0)=P(X=0)=q$ וגם $f(1)=P(X=1)=p$ אזי
$$E(X)=q\cdot 0+1\cdot p=p.$$
כדי לחשב את השונות נזכיר ש־
$$\begin{aligned} Var(X) &= E((X-E(X))^2)=E(X^2-2XE(X)+E^2(X)) \\ &= E(X^2)-2E(XE(X))+E(E^2(X))=E(X^2)-2E(X)E(X)+E^2(X) \\ &= E(X^2)-E^2(X). \end{aligned}$$
מכאן נובע שהשונות של ניסוי ברנולי היא כדלקמן
$$Var(X)=0^2\cdot q+1^2\cdot p-p^2=p(1-p)=pq.$$

2. התפלגות בינומית, כלומר $X\sim B(n,p)$.
$$\begin{aligned} Var(X) &= \sum_{k=0}^{n} k^2\binom{n}{k}\cdot p^k\cdot q^{n-k}-n^2p^2=n\sum_{k=1}^{n} k\binom{n-1}{k-1}\cdot p^k\cdot q^{n-k}-n^2p^2 \\ &\underbrace{=}_{l=k-1} np\sum_{l=0}^{n-1}(l+1)\cdot\binom{n-1}{l}\cdot p^l\cdot q^{n-1-l}-n^2p^2= \\ &= np\sum_{l=0}^{n-1} l\binom{n-1}{l}\cdot p^l\cdot q^{n-1-l}+np\sum_{l=0}^{n-1}\binom{n-1}{l}\cdot p^l\cdot q^{n-1-l}-n^2p^2= \\ &= np(n-1)p+np(p+q)^{n-1}-n^2p^2= \\ &= n^2p^2-np^2+np-n^2p^2= \\ &= np(1-p)=npq. \end{aligned}$$
אפשר להעיר שאם נתון לנו $E(X)$ ו־$Var(X)$ אז נוכל למצוא את n,p,q.

טענה 6.10 יהי X משתנה מקרי ויהיו a,b קבועים אזי מתקיים

1. $E(aX+b)=aE(X)+b$.
2. $E(X-E(X))=0$.
3. $Var(X)=E(X^2)-E^2(X)$.
4. $Var(aX+b)=a^2Var(X)$.

הוכחה:

1. E היא פונקציה לינארית לכן מתקיים $E(aX+b)=aE(X)+b$:
$$E(aX+b)=\sum_{i\in I}(ax_i+b)f(x_i)=a\sum_{i\in I}x_if(x_i)+b\sum_{i\in I}f(x_i)=aE(X)+b.$$

2. מ־1 נובע כי $E(X-E(X))=1\cdot E(X)-E(X)=0$.

3.
$$\begin{aligned}Var(X)&=E((X-E(X))^2)=E(X^2-2XE(X)+E^2(X))=\\&=E(X^2)-2E(XE(X))+E(E^2(X))=E(X^2)-2E(X)E(X)+E^2(X)=\\&=E(X^2)-E^2(X).\end{aligned}$$

4. לפי 3 נקבל ש־
$$\begin{aligned}Var(aX+b)&=E((aX+b)^2)-E^2(aX+b)=\\&=E(a^2X^2+2abX+b^2)-(aE(X)+b)^2=\\&=a^2E(X^2)+2abE(X)+b^2-a^2E^2(X)-2abE(X)-b^2=\\&=a^2(E(X^2)-E^2(X))=a^2Var(X).\end{aligned}$$

$\square$

משפט 6.4 (אי שיוויון צ'בישיב) לכל $t>0$ מתקיים $P\{|X|\geq t\}\leq t^{-2}E(X^2)$. בפרט אם $E(X)=\mu$ אזי $P\{|X-\mu|\geq t\}\leq t^{-2}Var(X)$.

הוכחה:
$$P\{|X|\geq t\}=\sum_{\underbrace{|x_j|\geq t}_{\frac{x_j^2}{t^2}\geq 1}}f(x_j)\leq t^{-2}\sum_j x_j^2f(x_j)=t^{-2}E(X^2).$$

ובפרט אם $Y=X-\mu$ אזי $P\{|Y|\geq t\}\leq t^{-2}(E(X-\mu)^2)=t^{-2}Var(X)$. $\square$

משפט צ'בישיב יותר מדי כללי ובדרך כלל נותן הערכה מאוד גסה. למשל, בהטלת קוביה
$$E(X)=1\cdot\frac{1}{6}+2\cdot\frac{1}{6}+3\cdot\frac{1}{6}+4\cdot\frac{1}{6}+5\cdot\frac{1}{6}+6\cdot\frac{1}{6}=3.5$$
ו־
$$E(X^2)=1^2\cdot\frac{1}{6}+2^2\cdot\frac{1}{6}+3^2\cdot\frac{1}{6}+4^2\cdot\frac{1}{6}+5^2\cdot\frac{1}{6}+6^2\cdot\frac{1}{6}=15\frac{1}{6}$$
מכאן נקבל ש־ $Var(X)=E(X^2)-(E(X))^2=\frac{35}{12}$ ולכן
$$0=P(|X-3.5|>2.6)\leq 2.6^{-2}\cdot\frac{35}{12}\approx 0.47.$$

6.4 התפלגויות בדידות נפוצות ומומנטים שלהן

1. **התפלגות בינומית** $X \sim B(n,p)$ בעל תוחלת $E(X)=np$ ושונות $Var(X)=npq$.
2. **התפלגות גיאומטרית** $X \sim Geom(P)$ בעל תוחלת

$$E(X)=\sum_{i=0}^{\infty} i\cdot q^i \cdot p = qp\sum_{i=1}^{\infty} i\cdot q^{i-1} = q\cdot p\frac{\partial}{\partial q}(\frac{1}{1-q}) = q\cdot p\cdot \frac{1}{(1-q)^2} = \frac{q}{p}.$$

המומנט השני נתון על ידי

$$E(X^2)=\sum_{i=0}^{\infty} i^2 q^i \cdot p = p\sum_{i=1}^{\infty} i^2 q^i = qp\frac{\partial}{\partial q}(\frac{q}{(1-q)^2}) = qp\frac{(1-q)+2q}{(1-q)^3} = \frac{pq(1+q)}{(1-q)^3} = \frac{q(1+q)}{p^2}\ .$$

מכאן נובע שהשונות היא

$$Var(X)=\frac{q(1+q)}{p^2}-\frac{q^2}{p^2}=\frac{q}{p^2}.$$

ניתן לשים לב שאם יודעים $E(X)$ אזי יודעים ש־ $p=\frac{1}{1+E(X)}$.

טענה 6.11 ל־ $|q|<1$ ולכל $i\geq 1$ מתקיים

$$\sum_{k=0}^{\infty}\binom{i+k-1}{k}\cdot q^k = \frac{1}{(1-q)^i}.$$

הוכחה: זה נכון ל־ $i=1$ כי זהו טור הנדסי. ננית שזה נכון ל־i ונראה ל־$i+1$. לפי הנחת האינדוקציה נקבל

$$\sum_{k=0}^{\infty}\binom{i+k-1}{k}\cdot q^k = \frac{1}{(1-q)^i},$$

לכן

$$\sum_{k=0}^{\infty}\binom{i+k}{k}\cdot q^k = \sum_{k=0}^{\infty}\binom{i+k-1}{k}\cdot q^k + \sum_{k=1}^{\infty}\binom{i+k-1}{k-1}\cdot q^k = \sum_{k=0}^{\infty}\binom{i+k-1}{k}\cdot q^k + q\sum_{k'=0}^{\infty}\binom{i+k'}{k'}\cdot q^{k'},$$

שגורר

$$(1-q)\sum_{k=0}^{\infty}\binom{i+k}{k}\cdot q^k = \sum_{k=0}^{\infty}\binom{i+k-1}{k}\cdot q^k = \frac{1}{(1-q)^i} \Rightarrow \sum_{k=0}^{\infty}\binom{i+k}{k}\cdot q^k = \frac{1}{(1-q)^{i+1}}.$$

□

1. **התפלגות בינומית שלילית** $X \sim f(r,p)$ - מספר כשלונות k, לפני הסגת r הצלחות.

$$P(X=k) = \binom{r+k-1}{r-1} p^r q^k.$$

כמו כן מתקיים $\sum_{k=0}^{\infty} \binom{r+k-1}{r-1} q^k p^r = p^r \frac{1}{(1-q)^r} = 1$, ולכן $f(r,p)$ היא פונקצית צפיפות. נחשב עבורה את התוחלת והשונות באופן הבא:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot P(X=k) = \sum_{K=0}^{\infty} k \cdot \tbinom{r+k-1}{r-1} q^k \cdot p^r = p^r \sum_{k=0}^{\infty} k \tbinom{r+k-1}{r-1} q^k \\ &= p^r \cdot q \cdot \tfrac{\partial}{\partial q} \left(\sum_{k=0}^{\infty} \tbinom{r+k-1}{r-1} \cdot q^k \right) \\ &= p^r \cdot q \cdot \tfrac{\partial}{\partial q} ((1-q)^{-r}) = p^r \cdot q(-r) \cdot (1-q)^{-r-1} \cdot (-1) \\ &= \tfrac{rp^r q}{(1-q)^{r+1}} = \tfrac{rp^r q}{p^{r+1}} = \tfrac{rq}{p} \end{aligned}$$

ואילו

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \cdot \tbinom{r+k-1}{r-1} \cdot q^k \cdot p^r \\ &= p^r \cdot q \cdot \tfrac{\partial}{\partial q} \left(\sum_{K=0}^{\infty} k \cdot \tbinom{r+k-1}{r-1} q^k \right) \\ &= p^r \cdot q \cdot \tfrac{\partial}{\partial q} \left(\tfrac{rq}{(1-q)^{r+1}} \right) = p^r \cdot q \cdot \tfrac{\partial}{\partial q} (rq(1-q)^{-r-1}) \\ &= rp^r q((1-q)^{-r-1} + (r+1)(1-q)^{-r-2} \cdot q) \\ &= rp^r q \tfrac{(1-q)+(r+1)q}{(1-q)^{r+2}} \underbrace{=}_{1-q=p} rp^r q \cdot \tfrac{p+(r+1)q}{p^{r+2}} = \tfrac{rq(p+(r+1)q)}{p^2} \end{aligned}$$

מכאן נובע שהשונות היא

$$Var(X) = E(X^2) - (E(X))^2 = \frac{rq(p+(r+1)q)}{p^2} - \frac{r^2q^2}{p^2} = \frac{rq\overbrace{(p+q)}^{1}}{p^2} = \frac{rq}{p^2}.$$

2. **התפלגות פואסון** $X \sim P(\lambda)$, $\lambda > 0$. $p(X=k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$ ו- $k \in \mathbb{N} \cup \{0\}$, נראה שזו פונקצית צפיפות

$$\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1.$$

נחשב את התוחלת והשונות באופן הבא

$$E(X) = \sum_{k=0}^{\infty} k \cdot \tfrac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \lambda \cdot \tfrac{\partial}{\partial \lambda} \left(\sum_{k=0}^{\infty} \tfrac{\lambda^k}{k!} \right) = e^{-\lambda} \cdot \lambda \cdot \tfrac{\partial}{\partial \lambda} (e^{\lambda}) = \lambda,$$

$$E(X^2)=\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=e^{-\lambda}\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}=e^{-\lambda}\cdot\lambda\cdot\frac{\partial}{\partial\lambda}(\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!})=$$
$$=e^{-\lambda}\cdot\lambda\cdot\frac{\partial}{\partial\lambda}\left(\lambda\underbrace{\frac{\partial}{\partial\lambda}(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!})}_{e^{\lambda}}\right)=e^{-\lambda}\cdot\lambda\cdot\frac{\partial}{\partial\lambda}(\lambda e^{\lambda})=\lambda e^{-\lambda}(e^{\lambda}+\lambda e^{\lambda})=\lambda(1+\lambda)$$

מכאן אנו מקבלים ש־ $Var(X)=\lambda(1+\lambda)-\lambda^2=\lambda$.

דוגמא 6.7 1. פניות למוקדי שירות שונים, 144, מגן דוויד אדום, קרינת חומר ראדיואקטיבי.

2. **קירוב פואסון להתפלגות בינומית** אם (1) $p<<1$, $n>>1$ ($np\leq 5$) ו־ $X\sim B(n,p)$

אזי $E(X)=np$ ו־ $P(X=0)=(1-p)^n=(1-\frac{\overbrace{np}^{\lambda}}{n})^n$. נחשב $\ln(1-\frac{\lambda}{n})^n=$ $n\ln(1-\frac{\lambda}{n})=-\lambda-\frac{\lambda^2}{2n}-\frac{\lambda 3}{3n^2}-\cdots\approx-\lambda$ מכאן מקבלים ש־ $P(X=0)\approx e^{-\lambda}$.

(2) $P(X=k)$ ו־ $k\geq 1$ אזי

$$\begin{aligned}\frac{P(X=k)}{P(X=k-1)}&=\frac{\binom{n}{k}\cdot p^kq^{n-k}}{\binom{n}{k-1}p^{k-1}q^{n+1-k}}=\frac{\frac{n!}{k!(n-k)!}\cdot p^k}{\frac{n!}{(k-1)!(n+1-k)!}p^kp^{-1}q}=\frac{p(n+1-k)}{kq}\\&=\frac{\lambda-(k-1)p}{kq}\approx\frac{\lambda}{k}-\frac{p}{q}(\frac{k-1}{k})\approx\frac{\lambda}{k}\end{aligned}$$

מכאן מקבלים ש־ $P(X=k)\approx\frac{\lambda}{k}P(X=k-1)$ כמו כן,

$$\begin{aligned}&P(X=1)\approx\lambda e^{-\lambda}\\&P(X=2)\approx\frac{\lambda^2}{2}e^{-\lambda}\\&\vdots\\&P(X=k)\approx\frac{\lambda^k}{k!}e^{-\lambda}.\end{aligned}$$

6.5 התפלגויות משותפות

$X_i:\Omega\to\mathbb{R}$ כאשר $1\leq i\leq n$ משתנים מקריים על אותו מרחב מדגם. למשל, מדגם שבודק משקל בתלות בגיל, גובה, מבנה גוף וכו'. בפרט אם Ω־מרחב מדגם בדיד נקבל $X_i:\Omega\to\{X_j^{(i)}\}_{j\in J_i}$, $1\leq i\leq n$ וכל מאורע אלמנטרי $A\in\Omega$ מיוצג על ידי ווקטור $A\leftrightarrow(x_1,x_2,\ldots,x_n)$ ונגדיר $P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)=P(A)$.

הגדרה 6.15 אם $X_i : \Omega \to \mathbb{R}$ בדיד לכל $1 \le i \le n$ אזי משתנה מקרי $(X_1, X_2, \ldots, X_n)$ נקרא משתנה מקרי בדיד ווקטורי n־מימדי. אם $n = 2$ אומרים משתנה מקרי דו־מימד. אם $n = 3$ אומרים משתנה מקרי תלת־מימדי. עבור המקרה $n = 2$ ו־ $X = \{x_i\}_{i=1}^{m_x}$, $Y = \{y_i\}_{j=1}^{m_y}$ כאשר m_x ו־m_y קבועים. נוח לכתוב את ההסתברויות $P(X = x_i, Y = y_j) = f(x_i, y_j)$ בטבלה

$y\backslash x$	x_1	x_2	$\ldots$	x_k	$P(Y = y_i)$
y_1	$f(x_1, y_1)$	$f(x_2, y_1)$	$\ldots$	$f(x_k, y_1)$	$\sum_{i=1}^{k} f(x_i, y_1)$
y_2	$f(x_1, y_2)$	$f(x_2, y_2)$	$\ldots$	$f(x_k, y_2)$	$\sum_{i=1}^{k} f(x_i, y_2)$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
y_m	$f(x_1, y_m)$	$f(x_2, y_m)$	$\ldots$	$f(x_k, y_m)$	$\sum_{i=1}^{k} f(x_i, y_m)$
$p(X = x_i)$	$P(X = x_1)$	$P(X = x_2)$	$\ldots$	$P(X = x_k)$	1

הגדרה 6.16 פונקציה $F(x_1, x_2, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$ נקראת פונקצית התפלגות מצטברת משותפת בדיוק כמו פונקצית התפלגות סקלרית. מההגדרה ניתן להשיג באופן אוטומטי ש־

א. F פונקציה מונטונית עולה באופן חלש כך שעבור $(a_1, a_2, \ldots, a_n)$ ו־ $(b_1, b_2, \ldots, b_n)$ אם מתקיים $a_i \le b_i$ לכל $i = 1, 2, \ldots, n$ אזי $F(a_1, a_2, \ldots, a_n) \le F(b_1, b_2, \ldots, b_n)$.

ב. $\lim\limits_{\forall i: x_i \to \infty} F(x_1, x_2, \ldots x_n) = 1.$

ג. $\lim\limits_{\forall i: x_i \to -\infty} F(x_1, x_2, \ldots x_n) = 0.$

יש להעיר שמעכשיו והלאה נתמקד רק בהתפלגות בדידה, זאת אומרת

$$0 \le P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1, x_2, \ldots x_n) \le 1$$

לכל $(x_1, x_2, \ldots, x_n)$ ויש רק מספר נקודות ניתנות להימנות ב־ $\mathbb{R}^n$ כאלה ש־ $f(x_1, x_2, \ldots, x_n) \neq 0$. נסמן את קבוצה של כל הנקודות כאלה ב־$\mathcal{G}$, לכן $\sum\limits_{(x_1, x_2, \ldots, x_n) \in \mathcal{G}} f(x_1, x_2, \ldots x_n) = 1$.

הגדרה 6.17 נתונה התפלגות משותפת בדידה $(X_1, X_2, \ldots, X_n)$. נגדיר התפלגות שולית על ידי

$$P(X_i = a) = f_i(a) := \sum_{j_1 \in J_1} \sum_{j_2 \in J_2} \cdots \sum_{j_n \in J_n} f(x_{j_1}^{(1)}, \ldots a, \ldots, x_{j_n}^{(n)}),$$

כאשר $\{x_j^{(i)}\}_{j \in J_j}$ ־ זה כל האפשרויות מקבוצה $\mathcal{G}$ של קואורדינטה j. f_i נקראת התפלגות שולית. בהתאם מגדירים התפלגות מצטברת שולית $F_i(x) = P(X_i \le x)$.

הגדרה 6.18 X ו־ Y נקראים בלתי תלויים אם לכל x, y מתקיים $f(x, y) = f_1(x) \cdot f_2(y)$. באופן כללי, $(X_1, X_2, \ldots, X_n)$ נקראים בלתי תלויים אם $f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n)$.

הערה 6.2

$$F(x_1,x_2,\dots,x_n)=P(X_1\le x_1,X_2\le x_2,\dots,X_n\le x_n)=$$
$$=\sum_{x^1_{j_1}\le x_1,\dots,x^n_{j_n}\le x_n} f(x^1_{j_1},x^2_{j_2},\dots,x^n_{j_n})$$

לכן $(X_1,X_2,\dots,X_n)$ בלתי תלויים אם ורק אם לכל x מתקיים $F(x_1,x_2,\dots,x_n)=F_1(x_1)\dots F_n(x_n)$.

הגדרה 6.19 נתונה התפלגות משותפת של X,Y אזי מגדירים התפלגות מותנית באופן הבא

$$P(X=x_i|Y=y_j)=\frac{f(x_i,y_j)}{P(Y=y_j)}.$$

הערה 6.3 אם

$$(X_1,X_2,\dots,X_n)$$

התפלגות משותפת ו־ $(X_1,X_2,\dots,X_n$ בלתי תלויים אזי גם $X_1,X_2,\dots\hat{X}_i,\dots X_n$ זאת אומרת כולם חוץ מ־ X_i בלתי תלויים ומתקיים

$$\begin{aligned}
&P(X_1=x_1,X_2=x_2,\dots X_n=x_n)=\\
&=\sum_{j\in J_j} f(x_1,x_2,\dots x_i^j,\dots)=\\
&=\sum_{j\in J_i} f_1(x_1)f_2(x_2)\dots f_i(x_j^{(i)})\dots f_n(x_n)=\\
&=f_1(x_1)f_2(x_2)\cdots f_n(x_n)\underbrace{\sum_{j\in J_i} f_i(x_j^{(i)})}_{=1}=f_1(x_1)f_2(x_2)\cdots\hat{f}_i\cdots f_n(x_n)
\end{aligned}$$

לכן לא צריך לבדוק כמו במקרה של מאורעות את כל החיתוכים. הסיבה היא ש"אי תלול " משתנים מושג הרבה יותר חזק מאשר אי־תלות של מאורעות ־בודקים לפי כל המאורעות האלמנטריים.

הגדרה 6.20 יהי $(X_1,X_2,\dots,X_n)$ משתנה מקרי וקטורי, ותהי $G:\mathbb{R}^n\to\mathbb{R}$ אזי

$$Z=G(X_1,X_2,\dots,X_n)$$

הוא משתנה מקרי סקלרי שמקיים

$$P(Z=z)=\sum_{G(x_1,x_2,\dots,x_n)=z} f(x_1,x_2,\dots x_n)$$

בפרט אפשר להגדיר $G(X_1,X_2,\dots,X_n)=\sum_{i=1}^n X_i$ וכמו כן אפשר להגדיר $G(X,Y)=XY$.

משפט 6.5 יהי $(X_1,X_2,\dots,X_n)$ משתנה מקרי וקטורי, $Z=X_1+X_2+\dots+X_n$ אזי $E(Z)=\sum_{i=1}^n E(X_i)$.

הוכחה: נראה שזה נכון ל־$n=2$

$$\begin{aligned} E(X+Y) &= \sum_{x_i+y_j=z} z\cdot f(x_i,y_j) = \\ &= \sum_{i\in I}\sum_{j\in J}(x_i+y_j)f(x_i,y_j) = \\ &= \sum_{i\in I}x_i\sum_{j\in J}f(x_i,y_j)+\sum_{j\in J}y_j\sum_{i\in I}f(x_i,y_j) = \\ &= \sum_{i\in I}x_if_X(x_i)+\sum_{j\in J}y_jf_Y(y_j)=E(X)+E(Y) \end{aligned}$$

באופן כללי מתקיים

$$E(X_1+X_2+\dots X_n)=E(X_1+X_2+\dots X_{n-1})+E(X_n)=E(X_1)+E(X_2)+\cdots+E(X_n)$$

$\square$

טענה 6.12 אם X,Y משתנים מקרים בלתי תלויים אזי $E(X\cdot Y)=E(X)E(Y)$.

הוכחה:

$$\begin{aligned} E(XY) &= \sum_{x_i\cdot y_j=z} zf(x_i,y_j)=\sum_{i\in I}\sum_{j\in J}x_iy_jf(x_i,y_j) = \\ &= \sum_{i\in I}\sum_{j\in J}x_iy_jf_X(x_i)f_Y(y_j)=\sum_{i\in I}x_if_X(x_i)\cdot\sum_{j\in J}y_jf_Y(y_j)=E(X)E(Y). \end{aligned}$$

$\square$

הגדרה 6.21 תהי־X,Y התפלגות משותפת כזאת שגם X וגם Y בעלי תוחלת ושונות אזי קו־שונות של X,Y מוגדרת על ידי $Cov(X,Y)=E((X-\mu_X)(Y-\mu_Y))$.

טענה 6.13 תכונות של קו־שונות

1. $Cov(X,Y)=Cov(Y,X)$
2. $Cov(X,X)=Var(X)$
3. $Cov(aX+b,cY+d)=ac\cdot Cov(X,Y)$
4. $Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$
5. $Cov(X,Y)=E(XY)-\mu_X\mu_Y$
6. בפרט אם X,Y בלתי תלויים אזי $Cov(X,Y)=0$

הוכחה:

1. את שתי התכונות הראשונות נוכל להראות ישירות על פי ההגדרה.

2. $$\begin{aligned}&Cov(aX+b,cY+d)=E((aX+b-a\mu_X-b)(cY+d-c\mu_Y-d))=\\&=E(ac(X-\mu_X)(Y-\mu_Y))=\\&=acE((X-\mu_X)(Y-\mu_Y))=\\&=ac\cdot Cov(X,Y)\end{aligned}$$

3. $$\begin{aligned}&Cov(X,Y+Z)=E((X-\mu_X)(Y+Z-\mu_Y-\mu_Z))=\\&=E((X-\mu_X)(Y-\mu_Y)+(X-\mu_X)(Z-\mu_Z))=\\&=E((X-\mu_X)(Y-\mu_Y)+E((X-\mu_X)(Z-\mu_Z))=\\&=Cov(X,Y)+Cov(X,Z)\end{aligned}$$

4. $$\begin{aligned}&Cov(X,Y)=E((X-\mu_X)(Y-\mu_Y))=\\&=E(XY-\mu_YX-\mu_XY+\mu_X\mu_Y)=\\&=E(XY)-\mu_Y\underbrace{E(X)}_{\mu_X}-\mu_X\underbrace{E(Y)}_{\mu_Y}+\mu_X\mu_Y=\\&=E(XY)-\mu_X\mu_Y\end{aligned}$$

5. נובע מיידית מטענה ראשונה ותכונה מספר 5.

□

משפט 6.6 אם $X_1,X_2,\ldots,X_n$ התפלגות משותפת כזאת שכל X_i בעלת תוחלת μ_i ושונות σ_i^2 אזי

$$\mathrm{Var}(X_1+X_2+\ldots+X_n)=\sum_{i=1}^{n}\sigma_i^2+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}Cov(X_i,X_j)$$

בפרט אם $X_1,X_2,\ldots,X_n$ בלתי תלויים אזי $\mathrm{Var}(X_1+X_2+\ldots+X_n)=\sum_{i=1}^{n}\sigma_i^2$

הוכחה: נוכיח עבור $n=2$

$$\begin{aligned}&Var(X_1+X_2)=E((X_1+X_2-\mu_1-\mu_2)^2)=\\&=E(((X_1-\mu_1)+(X_2-\mu_2))^2)=\\&=E((X_1-\mu_1)^2+2(X_1-\mu_1)(X_2-\mu_2)+(X_2-\mu_2)^2)=\\&=\sigma_1^2+2Cov(X_1,X_2)+\sigma_2^2=\\&=\sum_{i=1}^{2}\sigma_i^2+2\sum_{i=1}^{1}\sum_{j=2}^{2}Cov(X_i,X_j)\end{aligned}$$

נניח עבור $n-1$ ונוכיח עבור n.

$$\begin{aligned}
&Var(X_1+X_2+\cdots X_n)=E((X_1+X_2+\cdots X_n-\mu_1-\mu_2-\cdots\mu_n)^2)=\\
&=E(((X_1+X_2+\cdots X_{n-1}-\mu_1-\mu_2-\cdots\mu_{n-1})+(X_n-\mu_n))^2)=\\
&=Var(X_1+X_2+\cdots X_{n-1})+\sigma_n^2+2Cov(X_1+X_2+\ldots+X_{n-1},X_n)=\\
&=\sum_{i=1}^{n-1}\sigma_i^2+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n-1}Cov(X_i,X_j)+\sigma_n^2+2\sum_{i=1}^{n-1}Cov(X_i,X_n)=\\
&=\sum_{i=1}^{n}\sigma_i^2+2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}Cov(X_i,X_j)
\end{aligned}$$

בפרט אם $X_1,X_2,\cdots X_n$ בלתי תלויים אזי $Cov(X_i,X_j)=0$ לכל i,j לכן

$$Var(X_1+X_2+\cdots X_n)=\sum_{i=1}^{n}\sigma_i^2.$$

□

הגדרה 6.22 נתון משתנה מקרי X בעל תוחלת μ_x ושנות σ_x^2 אזי $Z_x=\frac{X-\mu_x}{\sigma_x}$ נקרא נורמליזציה של X.

טענה 6.14 לכל X בעל תוחלת ושונות מתקיים $Var(Z_x)=1$ ו־ $E(Z_x)=0$

הוכחה:

$$E(Z_x)=E(\frac{X-\mu_x}{\sigma_x})=\frac{E(X)-\mu_x}{\sigma_x}\underbrace{=}_{\mu_x=E(X)}\frac{E(X)-E(X)}{\sigma_x}=0$$

ואילו

$$\begin{aligned}
&Var(Z_x)=Cov(Z_x,Z_x)=Cov(\tfrac{1}{\sigma_x}X-\tfrac{\mu_x}{\sigma_x},\tfrac{1}{\sigma_x}X-\tfrac{\mu_x}{\sigma_x})=\\
&=\tfrac{1}{\sigma_x^2}Cov(X,X)=\tfrac{1}{\sigma_x^2}Var(X)=1
\end{aligned}$$

□

הגדרה 6.23 נתונה התפלגות משותפת X,Y כאשר XY־בעלי תוחלת ושונות, אזי $\rho(X,Y)=Cov(Z_x,Z_y)$ נקרא מקדם קורלציה של X,Y.

טענה 6.15
1. $\rho(X,Y)=\frac{Cov(X,Y)}{\sigma_x\sigma_y}$
2. $-1\le\rho(X,Y)\le 1$
3. $\rho(X,Y)=\pm 1$ אם ורק אם קיימים $a,b\in\mathbb{R}$ כאלה ש־ $Y=aX+b$

הוכחה:

1. $\rho(X,Y) = Cov(\frac{X-\mu_x}{\sigma_x}, \frac{Y-\mu_y}{\sigma_y}) = \frac{1}{\sigma_x\sigma_y}Cov(X,Y)$

2.
$$0 \leq Var(Z_x \pm Z_y) = Var(Z_x) + Var(Z_y) \pm 2Cov(Z_x, Z_y) =$$
$$= Var(Z_x) + Var(Z_y) \pm 2\rho(X,Y) = 2 \pm 2\rho(X,Y)$$
$$\Rightarrow 0 \leq 2 \pm 2\rho(X,Y) \Rightarrow -1 \leq \rho(X,Y) \leq 1$$

3. נניח $\rho(X,Y) = 1$ אם ורק אם

$$Var(Z_x - Z_y) = Var(Z_x) + Var(Z_y) - 2Cov(Z_x, Z_y) = 1 + 1 - 2 \cdot 1 = 0$$
$$\Leftrightarrow \exists a, b \in \mathbb{R}, Y = aX + b$$

נניח $\rho(X,Y) = -1$ אם ורק אם $Var(Z_x + Z_y) = 0 \Leftrightarrow \exists a, b \in \mathbb{R}, Y = aX + b$

$\square$

דוגמא 6.8 יהי X בעל שלוש נקודות שונות

X	-1	0	1
	0.25	0.5	0.25

ויהי $Y = X^2$. אז $E(X) = 0$, $E(Y) = 0.5$, ומכאן נובע ש־ $Cov(X,Y) = E(XY) = -1 \cdot 0.25 + 0 \cdot 0.5 + 1 \cdot 0.25 = 0$

הערה 6.4 $Cov(X,Y)$ מודד את התוחלת הלינארית בין X לבין Y. אם $Cov(X,Y) = 0$ אזי בין X לבין Y אין תוחלת לינארית , זאת אומרת אין פונקציה לינארית של X שמקרבת את Y.

שימושים

1. התפלגות ניסיון ברנולי $\begin{array}{c|c|c} X_i & 0 & 1 \\ \hline & q & p \end{array}$ כאשר $E(X_i) = p$, $Var(X_i) = pq$ והתפלגות בינומית $X \sim B(n,p)$ היא סכום $X = \sum_{i=1}^{n} X_i$ ומתקיים $E(X) = np$ כמו כן, $Var(X) = npq$

2. התפלגות גיאומטרית מתקיים $E(X_i) = \frac{q}{p}$ ו־ $Var(X_i) = \frac{q}{p^2}$

3. התפלגות בינומית שלילית $X \sim f(k,p)$ הוא סכום של משתנים מקריים מתפלגים גיאומטרית, כלומר עבור $X = \sum_{i=1}^{k} X_i$ מתקיים $E(X) = \frac{kq}{p}$ כמו כות $Var(X) = \frac{kq}{p^2}$.

דוגמא 6.9 בכד יש כדורים ממוספרים מ־ 1 עד n. מוצאים אותם (כולם) בלי החזרה. X הוא מספר כדורים שהמספר שלהם שווה למקום במזגם. $X = \{0, 1, 2, \ldots, n-1, n\}$ מכאן מקבלים ש־ $P(X = n) = \frac{1}{n!}$, $P(X = n-1) = 0$, $P(X = n-2) = \frac{\binom{n}{2}}{n!} = \frac{1}{2(n-2)!}$, ועבור כל k נקבל $P(X = k) = \frac{1}{k!}(\frac{1}{2!} - \frac{1}{3!} + \cdots \pm \frac{1}{(n-k)!})$. נגדיר X_i המסמן כדור עם מספר i באופן הבא

X_i	0	1
	$\frac{n-1}{n}$	$\frac{1}{n}$

מזה נובע ש־ $E(X_i) = 0 \cdot \frac{n-1}{n} + 1 \cdot \frac{1}{n} = \frac{1}{n}$, $Var(X_i) = (0^2 \cdot \frac{n-1}{n} + 1^2 \cdot \frac{1}{n}) - \frac{1}{n^2} = \frac{n-1}{n^2}$, יהי $X = \sum_{i=1}^{n} X_i$ המשתנה המקרי הדרוש, אזי

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{1}{n} = 1,$$

$$E(X_i \cdot X_j) = \frac{1}{n} \cdot \frac{1}{n-1} = \frac{1}{n(n-1)},$$

$$Cov(X_i, X_j) = \frac{1}{n(n-1)} - \frac{1}{n^2} = \frac{1}{n^2(n-1)},$$

כמו כן מתקיים

$$\begin{aligned}
Var(X) &= \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} Cov(x_i, x_j) = \\
&= n \cdot \tfrac{n-1}{n^2} + \tfrac{2}{n^2(n-1)} \cdot \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 1 = \\
&= \tfrac{n-1}{n} + \tfrac{2}{n^2(n-1)} \cdot \sum_{i=1}^{n-1} (n-i) = \\
&= \tfrac{n-1}{n} + \tfrac{2}{n^2(n-1)} \cdot \sum_{j=1}^{n-1} j = \\
&= \tfrac{n-1}{n} + \tfrac{2}{n^2(n-1)} \cdot \tfrac{(n-1)n}{2} = \tfrac{n-1}{n} + \tfrac{1}{n} = 1
\end{aligned}$$

פרק 7

משתנה מקרי רציף ומומנטים שלו

7.1 הגדרות

יהיו $X:\Omega\to\mathbb{R}$, $F(x)=P(X\le x)$, ותהי $F:\mathbb{R}\to[0,1]$ ־ פונקצית הסתברות מצטברת, אזי $X:\Omega\to\mathbb{R}$ נקרא משתנה מקרי רציף אם לכל $x\in\mathbb{R}$ מתקיים $P(X=x)=0$ או שקול אם $F(x)$ פונקציה רציפה.

הערה 7.1 ראינו שהפונקציה $F(x)$ מונטונית עולה חלש רציפה מימין ומקיימת $\lim\limits_{x\to-\infty}F(x)=0$ ו־ $\lim\limits_{x\to\infty}F(x)=1$. מוסיפים רציפות מצד שמאל ומקבלים משתנה מקרי רציף שלכל $[c,d]\in\mathbb{R}$ מתקיים $P(c\le X\le d)=F(d)-F(c)$.

דוגמא 7.1 התפלגות אחידה בקטע $[a,b]$,

$$F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a}, & a\le x\le b\\ 1, & x>b\end{cases}$$

ומתקיים $P(c\le X\le d)=F(d)-F(c)$.

הגדרה 7.1 יהי X משתנה מקרי רציף. קבוצה C ב־ $\mathbb{R}$ נקראת קבוצת מידה אפס אם מתקיים $P(X\in C)=0$.

טענה 7.1 אם X משתנה מקרי רציף אזי כל אוסף של נקודות נתן להמנות (סופי או בן מניה) הוא קבוצת מידה אפס.

הוכחה: לפי ההגדרה של משתנה מקרי רציף $P(X=x_i)=0$ לכל $x_i\in\{x_j\}_{j\in I}$ ולכן $P(X\in\{x_i\}_{i\in I})=\sum\limits_{i\in I}P(X=x_i)=0$. $\square$

הערה 7.2 יש גם קבוצות שאינן בנות מניה בעלות מידה אפס.

הגדרה 7.2 אם טענה נכונה לכל $x \in \mathbb{R}$ חוץ מקבוצת מידת אפס אומרים שהטענה נכונה כמעט בכל מקום.

דוגמא 7.2 $g(x) = i$ אם $i < x \leq i+1$ לכל $i \in \mathbb{Z}$ היא פונקציה רציפה כמעט בכל מקום.

הגדרה 7.3 יהי X משתנה מקרי רציף. אם $F(x)$ גזירה כמעט בכל מקום, אזי $F'(x)$ נקראת פונקצית צפיפות של X ונסמנה $F'(x) = f(x)$. בנקודות של קבוצת מידה אפס $f(x)$ לא מוגדרת.

דוגמא 7.3 התפלגות אחידה בקטע $[a,b]$:

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$$

אזי

$$f(x) = \begin{cases} 0, & x < a \\ \frac{1}{b-a}, & a < x < b \\ 0, & x > b \end{cases}$$

הפונקציה רציפה אך לא גזירה, מכיוון בנקודות a ו־b הנגזרת אינה מוגדרת.

הערה 7.3 בקורס זה נדון רק בהתפלגויות רציפות בעלות פונקצית צפיפות. אם נתונה $F(x)$ פונקצית התפלגות מצטברת של מתשנה מקרי רציף גזירה כמעט בכל מקום אזי $F'(x) = f(x)$ מקיימת

1. $f(x) \geq 0$ כמעט בכל מקום.
2. $\lim\limits_{x\to\infty} F(x) = \int\limits_{-\infty}^{\infty} f(x)dx = 1$
3. $\lim\limits_{x\to-\infty} f(x) = \lim\limits_{x\to-\infty} F(x) = 0$

מצד שני אם $f(x)$ מוגדרת כמעט בכל מקום ומקיימת את שתי התכונות הראשונות אזי היא משמשת כפונקצית צפיפות של איזשהו משתנה מקרי רציף X ואז פונקצית התפלגות מצטברת $F(x)$ מוגדרת על ידי $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ כאשר $\int\limits_{-\infty}^{x} f(t)dt$ שווה לשטח מתחת גרף פונקצית הצפיפות בקטע $[-\infty, x]$.

הגדרה 7.4 יהי X משתנה מקרי רציף בעל פונקצית צפיפות $f(x)$. אם $\int\limits_{-\infty}^{\infty} |x|f(x)dx$ קיים אזי $\mu_X := E(X) = \int\limits_{-\infty}^{\infty} xf(x)dx$ נקראת התוחלת של X.

דוגמא 7.4 1. התפלגות אחידה בקטע $[a,b]$:

$$E(X)=\int_{-\infty}^{\infty}x\cdot f(x)dx=\int_a^b\frac{x}{b-a}dx=\left.\frac{x^2}{2(b-a)}\right|_a^b=\frac{b^2-a^2}{2(b-a)}=\frac{b+a}{2}.$$

2. התפלגות מערכית $X\sim Exp(\lambda)$, $\lambda>0$, עם פונקצית צפיפות

$$f(x)=\begin{cases}0, x<0\\ \lambda e^{-\lambda x}, x>0.\end{cases}$$

אזי

$$\begin{aligned}E(X)&=\int_0^{\infty}\lambda xe^{-\lambda x}dx\underbrace{=}_{v=\lambda x,\,du=e^{-\lambda x}dx}\left.\frac{\lambda xe^{-\lambda x}}{-\lambda}\right|_0^{\infty}-\int_0^{\infty}\frac{\lambda e^{-\lambda x}}{-\lambda}dx\\&=0+\int_0^{\infty}e^{-\lambda x}dx\\&=\left.\frac{e^{-\lambda x}}{-\lambda}\right|_0^{\infty}\\&=\frac{1}{\lambda}.\end{aligned}$$

הגדרה 7.5 יהי X משתנה מקרי רציף ותהי $G:\mathbb{R}\to\mathbb{R}$ פונקציה אזי $Y=G(X)$ הוא משתנה מקרי חדש שפונקצית ההתפלגות המצטברת שלו מוגדרת על ידי $F_Y(y)=P(G(x)\le y)$.

הערה 7.4 1. אם G פונקציה עולה אזי $G(x)\le y\Leftrightarrow x\le G^{-1}(y)$, לכן

$$F_Y(y)=F_X(G^{-1}(y)).$$

2. אם G פונקציה יורדת אזי $G(x)\le y\Leftrightarrow x\ge G^{-1}(y)$, לכן

$$F_Y(y)=1-F_X(G^{-1}(y)).$$

3. נניח ש־ X בעלת פונקצית צפיפות $f_X(x)$ השאלה איך מגדירים $f_Y(y)$ במקרה ש־ $G(x)$ הנה פונקציה גזירה כמעט בכל מקום, והתשובה היא כדלקמן

$$f_Y(y)=\begin{cases}(F_X(G^{-1}(y)))', G\text{ עולה}\\(1-F_X(G^{-1}(y)))', G\text{ יורדת}\end{cases}=\begin{cases}\frac{f_X(G^{-1}(y))}{(G'(y))}\\-\frac{f_X(G^{-1}(y))}{(G'(y))}\end{cases}=\frac{f_X(G^{-1}(y))}{|G'(y)|}.$$

הגדרה 7.6 $G(x)$ רציפה נקראת פונקציה בעלת ווריאציה חסומה אם יש לה לא יותר ממספר ניתן להמנות של מינימומים מקומיים ומקסימומים מקומיים.

משפט 7.1 כל פונקציה רציפה בעלת ווריאציה חסומה היא סכום של שתי פונקציות רציפות שאחת מהן יורדת ושניה עולה.

משפט 7.2 אם X משתנה מקרי רציף בעל פונקצית צפיפות $f_X(x)$ ו־$Y = G(X)$ כאשר $G(x)$ הנה פונקציה רציפה, גזירה כמעט בכל מקום, בעלת ווריאציה חסומה ומקיימת

$$\int_{-\infty}^{\infty} |G(x)|\cdot f_X(x)dx < \infty,$$

אזי $E(Y) = \int_{-\infty}^{\infty} G(t)\cdot f_X(t)dt.$

הוכחה:

1. נניח ש־ $G(x)$ מונטונית, אז $E(Y) = \int_{-\infty}^{\infty} y\cdot f_Y(y)dy = \int_{-\infty}^{\infty} y\frac{f_X(G^{-1}(y))}{|G'(y)|}dy$. נסמן

$$x = G^{-1}(y) \Rightarrow y = G(x)$$

ו־ $dx = \frac{1}{G'(y)}dy$ מזה מקבלים

$$E(Y) = \int_{-\infty}^{\infty} G(x)\cdot f_X(x)dx < \infty$$

ובכך הוכחנו נכונות המשפט לגבי פונקציות מונוטוניות.

2. אם $G(x)$ היא בעלת ווריאציה חסומה אזי קיימות שתי פונקציות G_1, G_2 כך שמתקיים $G(x) = G_1(x) + G_2(x)$ כאשר $G_1(x)$ עולה ו־ $G_2(x)$ יורדת. כפי שנראה בהמשך אם X_1, X_2 התפלגות משותפת רציפה אזי $E(X_1 + X_2) = E(X_1) + E(X_2)$ בדיוק כמו במקרה הבדיד. אזי $Y = Y_1 + Y_2$ כאשר $Y_1 = G_1(X)$, $Y_2 = G_2(X)$ ולכן

$$\begin{aligned}
&E(Y) = E(Y_1) + E(Y_2) = E(G_1(X)) + E(G_2(X)) = \\
&= \int_{-\infty}^{\infty} G_1(x)f_X(x)dx + \int_{-\infty}^{\infty} G_2(x)f_X(x)dx = \\
&= \int_{-\infty}^{\infty} (G_1(x) + G_2(x))f_X(x)dx = \\
&= \int_{-\infty}^{\infty} G(x)f_X(x)dx
\end{aligned}$$

□

7.2 מומנט שני

הגדרה 7.7 יהי X משתנה מקרי רציף בעל פונקצית צפיפות ותוחלת μ_X, אם קיים $E(X^2)$ אזי $Var(X) = E((X-\mu_x)^2)$ נקרא שונות של X.

הערה 7.5 הגדרת שונות במקרה של בדיד ורציף אינה משתנה כי משתמשים בפונקצית תוחלת.

טענה 7.2 תכונות של תוחלת ושונות של משתנה מקרי רציף יהיX משתנה מקרי רציף בעל פונקצית צפיפות, תוחלת ושונות אזי

1. $E(aX+b) = aE(X)+b$ לכל $a,b \in \mathbb{R}$.
2. בפרט $E(X-\mu_X) = 0$.
3. לכל $a,b \in \mathbb{R}$ מתקיים $Var(aX+b) = a^2 \cdot Var(X)$.
4. $Var(X) = E(X^2) - \mu_X^2$.
5. אם $Y = G(X)$ כאשר $G(x)$ רציפה בעלת ווריאציה חסומה, אזי $E(Y) = \int\limits_{-\infty}^{\infty} G(x)f(x)dx$ ו־ $Var(Y) = \int\limits_{-\infty}^{\infty} G^2(x)f(x)dx - \mu_X^2$
6. בפרט אם $Z_X = \frac{X-\mu_X}{\sigma_X}$ אזי $E(Z_X) = 0$ ו־$Var(Z_X) = 1$.
7. אם $f(x)$ סימטרית סביב נקודה a זאת אומרת לכל x מתקיים $f(a+x) = f(a-x)$, אזי $E(X) = a$

הוכחה: ניתן לראות שארבע התכונות הראשונות בנוסף לתכונה השישית ההוכחה שלהם היא בדיוק כמו במקרה של משתנה מקרי בדיד, כמו כן התכונה החמישית נובעת ממשפט הקודם יחד עם התכונה הרביעית ולכן אין צורך בלהוכיח טענות אלו מחדש. אם כך נותר להוכיח רק את התכונה האחרונה, להלהן ההוכחה

$$E(X) = \int\limits_{-\infty}^{\infty} x \cdot f(x)dx = \underbrace{\int_{-\infty}^{a} x \cdot f(x)dx}_{\substack{y=-x+a \\ dy=-dx}} + \underbrace{\int_{a}^{\infty} x \cdot f(x)dx}_{\substack{y=x-a \\ dy=dx}} =$$
$$= -\int\limits_{\infty}^{0} (a-y) \cdot f(a-y)dy + \int\limits_{0}^{\infty} (y+a) \cdot f(y+a)dy =$$
$$= \int\limits_{0}^{\infty} (a-y) \cdot f(a-y)dy + \int\limits_{0}^{\infty} (y+a) \cdot f(y+a)dy =$$
$$= \int\limits_{0}^{\infty} (a-y+y+a) \cdot f(a+y)dy = 2a \underbrace{\int_{0}^{\infty} \cdot f(y+a)dy}_{\frac{1}{2}} = a$$

□

(1) $X \sim Un[a,b]$ אזי $E(X)=\frac{a+b}{2}$ כמו כן,

$$Var(X)=\int_a^b \frac{x^2}{b-a}dx-(\frac{a+b}{2})^2=\frac{1}{3}\cdot\frac{b^3-a^3}{b-a}-\frac{(a+b)^2}{4}=\frac{(b-a)^2}{12}.$$

(2) $X \sim Exp(\lambda)$, עם $\lambda>0$ ופונקצית צפיפות $f(x)=\begin{cases} 0; x<0 \\ \lambda e^{-\lambda x}; x>0 \end{cases}$ אזי

$$E(X)=\int_0^\infty \lambda x e^{-\lambda x}dx=-xe^{-\lambda x}\Big|_0^\infty+\int_0^\infty e^{-\lambda x}dx=\frac{1}{\lambda}$$

וגם

$$\begin{aligned} VAR(X) &= \int_0^\infty \lambda x^2 e^{-\lambda x}dx-(\frac{1}{\lambda})^2 \\ &= -x^2e^{-\lambda x}\Big|_0^\infty+\frac{2}{\lambda}\underbrace{\int_0^\infty \lambda x e^{-\lambda x}dx}_{\frac{1}{\lambda}}-\frac{1}{\lambda^2}=\frac{1}{\lambda^2}. \end{aligned}$$

(3) התפלגות גמא מוגדרת באופן הבא. תחילה נגדיר פונקצית גאמא כדלקמן

$$\Gamma(t)=\int_0^\infty x^{t-1}e^{-x}dx.$$

יש לציין שהיאנטגרל מתכנס לכל $t>0$.

טענה 7.3 לכל $t>1$ מתקיים $\Gamma(t)=(t-1)\Gamma(t-1)$

הוכחה:

$$\begin{aligned} \Gamma(t) &= \int_0^\infty x^{t-1}e^{-x}dx \\ &\underbrace{=}_{u=x^{t-1},\ v=-e^{-x},\ du=(t-1)x^{t-2}, dv=e^{-x},} \frac{x^{t-1}e^{-x}}{-1}\Big|_0^\infty+\int_0^\infty (t-1)x^{t-2}e^{-x}dx \\ &= (t-1)\Gamma(t-1). \end{aligned}$$

□

טענה 7.4 $\Gamma(t)=\frac{1}{t}\prod\limits_{n=1}^{\infty}\frac{(1+\frac{1}{n})^t}{1+\frac{t}{n}}$.

נסמן $X\sim\Gamma(\nu,\alpha)$ הכללה של התפלגות מערכית עם פונקצית צפיפות

$$f_{\alpha,\nu}(x)=\begin{cases}0; & x<0\\ \frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}; & x>0\end{cases}$$

נראה שזוהי התפלגות באופן הבא

$$\int\limits_0^{\infty}\frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}dx\underbrace{=}_{\substack{y=\alpha x\\ dy=\alpha dx}}$$
$$=\frac{1}{\Gamma(\nu)}\int\limits_0^{\infty}y^{\nu-1}\cdot e^{-y}dy=\frac{1}{\Gamma(\nu)}\cdot\Gamma(\nu)=1$$

נחשב את השונות והתוחלת של X,

$$E(X)=\frac{1}{\Gamma(\nu)}\int\limits_0^{\infty}x\alpha^{\nu}x^{\nu-1}e^{-\alpha x}dx=\frac{1}{\alpha\Gamma(\nu)}\Gamma(\nu+1)=\frac{\nu}{\alpha}$$

ואילו

$$Var(X)=\int\limits_0^{\infty}x^2\frac{1}{\Gamma(\nu)}\alpha^{\nu}x^{\nu-1}e^{-\alpha x}dx-(\tfrac{\nu}{\alpha})^2=\frac{1}{\Gamma(\nu)}\int\limits_0^{\infty}\alpha^{\nu}x^{\nu+1}e^{-\alpha x}dx-\frac{\nu^2}{\alpha^2}=$$
$$=\frac{\Gamma(\nu+2)}{\Gamma(\nu)\alpha^2}-\frac{\nu^2}{\alpha^2}=\frac{\nu(\nu+1)}{\alpha^2}-\frac{\nu^2}{\alpha^2}=\frac{\nu}{\alpha^2}$$

התפלגות גמה מופיעה בזמן המתנה בזרם פואסוני,ועונה על השאלה כמה זמן עובר בין מאורעות פואסונים.

(4) התפלגות נורמלית $X\sim N(\mu,\sigma)$ פונקצית ההסבתרות של התפלגות נורמלית היא

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

נראה תחילה שזוהי התפלגות ל־ $X\sim N(0,1)$

$$(\int\limits_{-\infty}^{\infty}f(x)dx)^2=\int\limits_{-\infty}^{\infty}f(x)dx\cdot\int\limits_{-\infty}^{\infty}f(y)dy=$$
$$=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}f(x)f(y)dxdy\underbrace{=}_{\substack{\sigma=1\\ \mu=0}}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\frac{1}{2\pi}\cdot e^{-\frac{x^2}{2}-\frac{y^2}{2}}dxdy=$$
$$=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}e^{-\frac{1}{2}\cdot(x^2+y^2)}dxdy\underbrace{=}_{\substack{x=r\cos\phi\\ y=r\sin\phi\\ J=r}}\frac{1}{2\pi}\int\limits_{r=0}^{\infty}\int\limits_{\phi=0}^{2\pi}re^{-\frac{1}{2}r^2}drd\phi=$$
$$=\frac{1}{2\pi}\int\limits_{r=0}^{\infty}2\pi r\cdot e^{-\frac{1}{2}r^2}dr=(-e^{-\frac{1}{2}r^2})\Big|_0^{\infty}=0-(-1)=1$$

שגורר ש־ $\int_{-\infty}^{\infty} f(x)dx = 1$, זאת אומרת $f(x)$ פונצקית צפיפות.
עעת נראה שזוהי התפלגות עבור המקרה $N(\mu, \sigma)$:

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \underbrace{=}_{\substack{y=\frac{x-\mu}{\sigma}\\ dy=\frac{dx}{\sigma}}} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}dy = 1.$$

כעת נחשב תוחלת ושנות של התפלגות נורמלית:

1. פונקציה $f(x)$ סימטרית סביב $x = \mu$ לכן $E(X) = \mu$

2.

$$Var(X) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx \underbrace{=}_{y=x-\mu} \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2\sigma^2}}dy =$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\underbrace{y}_{u}\cdot\underbrace{y\cdot e^{-\frac{y^2}{2\sigma^2}}dy}_{dv} = \underbrace{\frac{1}{\sigma\sqrt{2\pi}}y\sigma^2 e^{-\frac{y^2}{2\sigma^2}}\Bigg|_{-\infty}^{\infty}}_{=0} - \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} -\sigma^2 e^{-\frac{y^2}{2\sigma^2}}dy =$$

$$= \frac{\sigma^2}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma^2}}dy \underbrace{=}_{\substack{z=\frac{y}{\sigma}\\ \sigma dz=dy}} \frac{\sigma^2}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz\cdot\sigma =$$

$$= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz = \sigma^2\cdot(\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz}_{=1}) = \sigma^2$$

נורמליזציה של התפחגות נורמלית נעשתה על ידי $Z_X = \frac{X-\mu}{\sigma}$ ואז נקבל התפלגות נורמלית סטנדרטית $Z_X \sim N(0,1)$.
לחישוב $F_{N(0,1)}(x) = \phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}dy$ נעזרים בטבלה. $\phi(-x) = 1-\phi(x)$ לכן לכל התפלגות נורמלית $X \sim N(\mu,\sigma)$ כדי לחשב $F(X)$ יש לחשב $F(x) = \phi(\frac{x-\mu}{\sigma})$.
כפי שנראה בעתיד התפלגות נורמלית היא התפלגות מאוד נפוצה ־ לפי משפט הגבול המרכזי. כל הערכים שתלויים בהרבה מרכיבים בלתי תלויים למשל גובה או משקל של קבוצה הומוגנית של אנשים, או ציוני $I.Q$ או אורך החיים של מכשירים אלקטרוניים מתפלג לפי התפלגות נורמלית.

טענה 7.5 אי שיוויון צ'בישב
יהי X משתנה מקרי רציף בעל פונקצית צפיפות $f(x)$,תוחלת μ_x ושונות σ_x^2 אזי לכל $\epsilon > 0$ מתקיים $P\{|X| \geq \epsilon\} \leq \epsilon^{-2}E(X^2)$, בפרט $P\{|X-\mu_x| \geq \epsilon\} \leq \frac{\sigma_x^2}{\epsilon^2}$.

הוכחה:

$$P\{|X|\geq\epsilon\}=\int\limits_{|x|\geq\epsilon}f(x)dx=\frac{1}{\epsilon^2}\int\limits_{|x|\geq\epsilon}\epsilon^2\cdot f(x)dx\leq\frac{1}{\epsilon^2}\int\limits_{|x|\geq\epsilon}x^2\cdot f(x)dx\qquad(x^2\geq\epsilon^2)$$
$$\leq\frac{1}{\epsilon^2}\int\limits_{-\infty}^{\infty}x^2\cdot f(x)dx=\frac{E(X^2)}{\epsilon^2}$$

נקח $Y=X-\mu_x$ אזי $P(|X-\mu_x|\geq\epsilon)=P(|Y|\geq\epsilon)\leq\frac{E(Y^2)}{\epsilon^2}=\frac{\sigma_x^2}{\epsilon^2}$, $\square$

הערה 7.6 אי שיוויון צ'בישב נכון גם למשתנה מקרי בדיד, וגם למשתנה מקרי רציף. בעיקרון אי שיוויון צ'בישב נכון לכל משתנה מקרי.

7.3 התפלגות משותפת של משתנים מקריים רציפים. קונבולוציות

הגדרה 7.8 $X_i:\Omega\to\mathbb{R}$, X_i משתנה מקרי רציף לכל i המקיים $1\leq i\leq n$ אזי $X=(X_1,X_2,\dots X_n)$, $X:\Omega\to\mathbb{R}^n$ נקרא משתנה מקרי וקטורי רציף, או התפלגות רציפה משותפת של $(X_1,X_2,\dots X_n)$.

הגדרה 7.9 $F(x_1,x_2,\dots x_n)=P(X_1\leq x_1,X_2\leq x_2,\dots X_n\leq x_n)$ נקראת פונקצית התפלגות מצטברת משותפת. X משתנה מקרי רציף אם ורק אם $F(x_1,x_2,\dots x_n)$ רציפה.

טענה 7.6 א. $\lim\limits_{\forall i\ x_i\to-\infty}F(x_1,x_2,\dots,x_n)=0$

ב. $$\lim_{\substack{x_1\to\infty\\ x_2\to\infty\\ \vdots\\ x_n\to\infty}}F(x_1,x_2,\dots,x_n)=F(\infty,\infty,\dots,\infty)=1$$

הגדרה 7.10 נדרוש בנוסף ש־ $F(x_1,x_2,\dots,x_n)$ תהיה גזירה n פעמים לפי מתשנים שונים ושהנגזרת המעורבת תהיה רציפה כמעט בכל מקום אזי $f(x_1,x_2,\dots,x_n)=\frac{d^nF(x_1,x_2,\dots,x_n)}{dx_1dx_2\cdots dx_n}$ נקראת פונקצית צפיפות משותפת של $X_1,X_2,\dots,X_n$. ברור כי $F(x_1,x_2,\dots,x_n)=\int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{x_2}\cdots\int\limits_{-\infty}^{x_n}f(t_1,t_2,\dots,t_n)dt_ndt_{n-1}\cdots dt_1.$ אנחנו חוקרים רק התפלגויות משותפות בעלות פונקצית צפיפות.

הגדרה 7.11 $F_i(x_i)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\cdots\int\limits_{-\infty}^{x_i}\cdots\int\limits_{-\infty}^{\infty}f(t_1,t_2,\dots,t_n)dt_ndt_{n-1}\cdots dt_1$ נקראת פונקצית התפלגות מצטברת שולית ו־ $f_i(x_i)=F_i'(x_i)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\cdots\cdots\int\limits_{-\infty}^{\infty}f(t_1,t_2,\dots,t_{i-1},x_i,t_{i+1}\dots,t_n)dt_ndt_{n-1}\cdots dt_1$ נקראת פונקצית צפיפות שולית.

הגדרה 7.12 תהיינה X,Y התפלגות משותפת רציפה, ו־ $F_1(x),F_2(x)$ פונקציות התפלגות מצטברת שוליות. אם לכל $(x,y)\in\mathbb{R}^2$ מתקיים $F(x,y)=F_1(x)\cdot F_2(y)$ אזי X,Y נקראים בלתי תלויים.

טענה 7.7 X,Y בלתי תלויים אם ורק אם $f(x,y)=f_1(x)\cdot f_2(y)$.

הוכחה:

1. נניח כי $F(x,y)=F_1(x)\cdot F_2(y)$ אזי

$$f(x,y)=\frac{d^2F(x,y)}{dxdy}=\frac{d}{dx}F_1(x)\cdot\frac{d}{dy}F_2(y)=f_1(x)\cdot f_2(y)$$

2. נניח כי $f(x,y)=f_1(x)\cdot f_2(y)$ אזי

$$\begin{aligned}F(x,y)&=\int_{-\infty}^{x}\int_{-\infty}^{y}f(t_1,t_2)dt_2dt_1=\int_{-\infty}^{x}\int_{-\infty}^{y}f_1(t_1)f_2(t_2)dt_2dt_1\\&=\int_{-\infty}^{x}f_1(t_1)dt_1\cdot\int_{-\infty}^{y}f_2(t_2)dt_2=F_1(x)\cdot F_2(y)\end{aligned}$$

□

הגדרה 7.13 תהי X,Y התפלגות משותפת רציפה עם פונקצית צפיפות $f(x,y)$. יהי $Z=X+Y$ משתנה מקרי רציף המקיים $F_Z(z)=F_{XY}(x,z-x)$.
נחשב $F(z)$ ו־ $f(z)$ באופן הבא

$$F(z)=\int_{-\infty}^{\infty}\int_{-\infty}^{z-t_1}f(t_1,t_2)dt_2dt_1\underbrace{=}_{\substack{t=t_1+t_2\\t_2=t-t_1\\dt_2=dt}}\int_{-\infty}^{\infty}\int_{-\infty}^{Z}f(t_1,t-t_1)dtdt_1.$$

$f(z)=F'(z)=\int_{-\infty}^{\infty}f(t_1,z-t_1)dt_1\underbrace{=}\int_{-\infty}^{\infty}f(z-t_1,t_1)dt_1$. אם X,Y בלתי תלויים אזי

$$f(z)=\int_{-\infty}^{\infty}f_1(t)\cdot f_2(z-t)dt=\int_{-\infty}^{\infty}f_1(z-t)f_2(t)dt=f_1*f_2(z)=f(z)=f_2*f_1(z)$$

הפונקציה הזו נקראת קונבולוציה של f_1,f_2.

הערה 7.7 אם $f_1(x)=0$ לכל $x\le 0$ וגם $f_2(y)=0$ לכל $y\le 0$ אזי

$$f_1*f_2(z)=\int_{0}^{z}f_1(t)f_2(z-t)dt=\int_{0}^{z}f_1(z-t)f_2(t)dt.$$

משפט 7.3 תהי X,Y התפלגות משותפת רציפה כזאת ש־ $E(X)$ ו־ $E(Y)$ קיימות אזי

$$E(X+Y)=E(X)+E(Y).$$

הוכחה:

$$E(X+Y)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}zf(t,z-t)dtdz\underbrace{=}_{v=z-t}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(t+v)f(t,v)dtdv=$$
$$=\int\limits_{-\infty}^{\infty}t(\int\limits_{-\infty}^{\infty}f(t,v)dv)dt+\int\limits_{-\infty}^{\infty}v(\int\limits_{-\infty}^{\infty}f(t,v)dt)dv=$$
$$=\int\limits_{-\infty}^{\infty}tf_1(t)dt+\int\limits_{-\infty}^{\infty}vf_2(v)dv=E(X)+E(Y)$$

□

מכאן אפשר להסיק שניתן להוכיח להתפלגות רציפה משותפת באינדוקציה

$$E(X_1+X_2+\cdots+X_n)=E(X_1)+E(X_2)+\cdots+E(X_n).$$

משפט 7.4 אם X,Y רציפים בלתי תלויים אזי $E(XY)=E(X)E(Y)$.

הוכחה: יהי $Z=XY$ אזי

$$F(z)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\frac{z}{x}}f(s,t)dtds=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\frac{z}{x}}f_1(s)f_2(t)dtds\underbrace{=}_{t'=ts}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{z}\frac{1}{s}f_1(s)f_2(\frac{t'}{s})dt'ds,$$

כלומר

$$f(z)=\int\limits_{-\infty}^{\infty}\frac{1}{s}f_1(s)f_2(\frac{z}{s})ds.$$

$$E(Z)=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\tfrac{z}{s}f_1(s)f_2(\tfrac{z}{s})dzds\underbrace{=}_{y=\frac{z}{s}}\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}y\cdot s\cdot f_1(s)f_2(y)dyds=$$
$$=\int\limits_{-\infty}^{\infty}sf_1(s)ds\int\limits_{-\infty}^{\infty}y\cdot f_2(y)dy=E(X)E(Y)$$

□

טענה 7.8 אם X,Y בלתי תלויים אזי מתקיים $Var(X+Y)=Var(X)+Var(Y).$

הוכחה:

$$Var(X+Y)=E((X+Y)-(\mu_x+\mu_y))^2)=E((X-\mu_x)^2+(Y-\mu_y)^2+2(X-\mu_x)(Y-\mu_y))$$

$$=Var(X)+Var(Y)+2E(X-\mu_x)(Y-\mu_y))$$

$$=Var(X)+Var(Y)+2E(XY-X\mu_y-Y\mu_x+\mu_x\mu_y)=Var(X)+Var(Y)$$

□

דוגמא 7.5 1. יהיו X,Y בלתי תלויים ומתקיים $X\sim\Gamma(\alpha,\nu)$, $Y\sim\Gamma(\alpha,\mu)$ אזי לפי משפט קודם מקבלים ש־ $E(X+Y)=E(X)+E(Y)=\frac{\nu}{\alpha}+\frac{\mu}{\alpha}=\frac{\nu+\mu}{\alpha}$ ו־ $Var(X+Y)=\frac{\nu+\mu}{\alpha^2}$
נחשב את הביטוי $f_1*f_2(z)$ באופן הבא

$$f_1*f_2(z)=\int\limits_0^z f_1(x)f_2(z-x)dx$$

$$=\int\limits_0^z\left(\frac{1}{\Gamma(\nu)}\alpha^\nu x^{\nu-1}e^{-\alpha x}\right)\cdot\left(\frac{1}{\Gamma(\mu)}\alpha^\mu(z-x)^{\mu-1}e^{-\alpha(z-x)}\right)dx\underbrace{=}_{\substack{dx=zdt\\x=tz}}$$

$$=\frac{\alpha^{\nu+\mu}e^{-\alpha z}}{\Gamma(\nu)\Gamma(\mu)}\int\limits_0^z x^{\nu-1}(z-x)^{\mu-1}dx$$

$$=\frac{\alpha^{\nu+\mu}e^{-\alpha z}}{\Gamma(\nu)\Gamma(\mu)}\int\limits_0^1(zt)^{\nu-1}(z-zt)^{\mu-1}zdt=\frac{\alpha^{\nu+\mu}e^{-\alpha z}}{\Gamma(\nu+\mu)}z^{\nu+\mu-1}\cdot\frac{\Gamma(\nu+\mu)}{\Gamma(\nu)\Gamma(\mu)}\underbrace{\int\limits_0^1 t^{\nu-1}(1-t)^{\mu-1}dt}_{=\beta_{\mu,\nu}=\frac{\Gamma(\nu)\Gamma(\mu)}{\Gamma(\nu+\mu)}}$$

$$=\frac{\alpha^{\nu+\mu}e^{-\alpha z}}{\Gamma(\nu+\mu)}z^{\nu+\mu-1}=f_{\nu+\mu}(z)$$

2. יהיו X,Y מתשנים מקריים בלתי תלויים כך ש־ $X\sim N(\mu,\sigma)$, $Y\sim N(\nu,\tau)$ מעוניינים לחשב את $f_1*f_2(z)$

$$f_1*f_2(z)=\int\limits_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\cdot\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi}\cdot\tau}e^{-\frac{(z-x-\nu)^2}{2\tau^2}}dx$$

$$= \frac{1}{2\pi\cdot\sigma\cdot\tau}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}-\frac{(z-x-\nu)^2}{2\tau^2}}dx = \frac{1}{2\pi\sigma\tau}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}-\frac{(z-x-\nu)^2}{2\tau^2}}dx \underbrace{=}_{\substack{u=x-\mu\\ z-x-\nu=v-u\\ v=z-x-\nu+x-\mu=z-\nu-\mu}}$$

$$\frac{1}{2\pi\sigma\tau}\int_{-\infty}^{\infty} e^{-(\frac{u^2}{2\sigma^2}+\frac{(v-u)^2}{2\tau^2})}du = \frac{1}{2\pi\sigma\tau}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2\tau^2}(\tau^2u^2+\sigma^2(v-u)^2)}du = (*)$$

נחשב את הביטוי

$$\tau^2u^2+\sigma^2(v-u)^2 = (\tau^2+\sigma^2)u^2+\sigma^2v^2-2\sigma^2vu \underbrace{=}_{\text{לריבוע השלמה}}$$

$$(\tau^2+\sigma^2)u^2-2\sigma^2vu+\sigma^2v^2 =$$

$$= (\sqrt{\tau^2+\sigma^2}u - \underbrace{\frac{\sigma^2}{\sqrt{\tau^2+\sigma^2}}v}_{b})^2 + \frac{\sigma^2\tau^2}{\tau^2+\sigma^2}v^2$$

נציב בחזרה ונקבל

$$(*) = \frac{e^{-\frac{v^2}{2(\sigma^2+\tau^2)}}}{2\pi\cdot\sigma\cdot\tau}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2\tau^2}(\sqrt{\sigma^2+\tau^2}u-b)^2}du \underbrace{=}_{t=\sqrt{\sigma^2+\tau^2}u} \frac{e^{-\frac{v^2}{2(\sigma^2+\tau^2)}}}{2\pi\cdot\sigma\cdot\tau}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2\tau^2}(t-b)^2}\frac{dt}{\sqrt{\sigma^2+\tau^2}}$$

$$= \frac{e^{-\frac{(z-(\mu+\nu))^2}{2(\sigma^2+\tau^2)}}}{\sqrt{\sigma^2+\tau^2}\sqrt{2\pi}} \cdot \underbrace{\frac{1}{\sigma\tau\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2\tau^2}(t-b)^2}dt}_{1}$$

$$= \frac{e^{-\frac{(z-(\mu+\nu))^2}{2(\sigma^2+\tau^2)}}}{\sqrt{\sigma^2+\tau^2}\sqrt{2\pi}},$$

זאת אומרת קיבלנו ש־ $Z = X+Y \sim N(\mu+\nu, \sqrt{\sigma^2+\tau^2})$

פרק 8

חוקי המספרים הגדולים ומשפט הגבול המרכזי

בגדול חוקי המספרים הגדולים אומרים שאם עורכים מספר גדול של נסיונות בלתי תלויים אז הממוצע לא תלוי בפילוג של כל אחד מהתוצאות, הוא תלוי רק בממוצע של כל אחד מהנסיונות. חוקי המספרים הגדולים אומרים שלמרות שאין לנו שום אפשרות לצפות תוצאה של ניסוי בודד אנחנו יכולים בהסתברות גבוהה לצפות תוצאות של נסיונות רבים ונותנים בסיס להגדרת הסתברות כשכיחות יחסית בסידרה גדולה של נסיונות בלתי תלויים. משפט הגבול המרכזי טוען שאם משתנים מקריים בלתי תלויים עם תוחלת ושונות שמקיימים תנאים מוסיימים אז התפלגות של סכום או ממוצע לא תלוי בהתפלגויות של כל אחד מהמשתנים המקריים ומתפלג לפי התפלגות נורמלית. ננסח אותו כאן נוכיח אותו בהמשך.

משפט 8.1 משפט הגבול המרכזי יהיו $X_1, \dots X_n$ משתנים מקריים בלתי תלויים בעלי פונקצית צפיפות f_i, תוחלת μ_i ושונות σ_i^2. נסמן $S_n = \sum\limits_{i=1}^{n} X_i$ מכאן נקבל $Var(S_n) = b_n = \sum\limits_{i=1}^{n} \sigma_i^2$. אם עבור $\tau > 0$ מתקיים $\lim\limits_{n\to\infty} \frac{1}{b_n} \sum\limits_{k=1}^{n} \int\limits_{|x-\mu_k|>\tau b_n} (x-\mu_k)^2 f_k(x)dx = 0$, אזי

$$P(S_n < x) \underset{n\to\infty}{\rightarrow} \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{\frac{1}{b_n}(x-\sum\limits_{k=1}^{n}\mu_k)} e^{-\frac{z^2}{2}} dz.$$

במילים אחרות $S_n \sim N(\sum\limits_{i=1}^{n} \mu_i, \sqrt{b_n})$ כאשר $n \to \infty$.

בפרט אם $S_n \sim B(n,p)$ אזי $S_n \underset{n\to\infty}{\rightarrow} N(np, \sqrt{npq})$, משפט לפלס־דה מאובר.

8.1 חוקים חלשים של מספרים גדולים

טענה 8.1 אי שיוויון צ'בישב. יהי X משתנה מקרי בעל תוחלת ושונות (רציף או בדיד) אזי מתקיים $P\{|X| \geq \epsilon\} \leq \epsilon^{-2} E(X^2)$ כמו כן מתקיים $P\{|X - \mu_x| \geq \epsilon\} \leq \frac{\sigma_x^2}{\epsilon^2}$

משפט 8.2 חוק המספרים הגדולים של צ'בישב אם $X_1, X_2, \ldots, X_n$ סדרה של משתנים מקריים בלתי תלויים בעלי תוחלת μ_i ושונות σ_i^2 המקיימים $\sigma_i^2 \le c$ לכל i אזי לכל $\epsilon > 0$ מתקיים $\lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \frac{1}{n}\sum\limits_{i=1}^{n}\mu_i| < \epsilon\} = 1$

הוכחה: נסמן ב־ $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ מכאן מקבלים

$$\begin{aligned} Var(\overline{X_n}) &= \frac{1}{n^2}Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2 \le \frac{1}{n^2}\sum_{i=1}^{n} c \\ &= \frac{c}{n} \end{aligned}$$

בנוסף $\mu_{\overline{X_n}} = \frac{1}{n}E\left(\sum\limits_{i=1}^{n} X_i\right) = \frac{1}{n}\sum\limits_{i=1}^{n}\mu_i$ לכן לפי אי שיוויון צ'בשיב מתקיים

$$P\{|\overline{X_n} - \mu_{\overline{X_n}}| \ge \epsilon\} \le \frac{\sigma^2_{\overline{X_n}}}{\epsilon^2} \Rightarrow P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i| \ge \epsilon\} \le \frac{\frac{c}{n}}{\epsilon^2}$$

$$\Rightarrow P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i| < \epsilon\} \ge 1 - \frac{\frac{c}{n}}{\epsilon^2} \underset{n\to\infty}{\to} 1.$$

$\square$

כמסקנה מהמשפט הקודם נקבל את הטענות הבאות

טענה 8.2 חוק המספרים הגדולים של ברנולי. יהי $X \sim B(n,p)$, ויהי m מספר הצלחות ב־n נסיונות אזי $\lim\limits_{n\to\infty} P(|\frac{m}{n} - p| < \epsilon) = 1$ זאת אומרת שכיחות יחסית של הצלחות שואפת להסתברות P.

הוכחה: יהי $\begin{array}{c|cc} X_i & 0 & 1 \\ \hline & q & p \end{array}$ ונסמן ב־ $\overline{X_n} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ מזה מקבלים ש־ $\mu_{\overline{X_n}} = \frac{1}{n}\sum\limits_{i=1}^{n} P = P$

כמו כן $Var(\overline{X_n}) = \frac{1}{n^2}\sum\limits_{i=1}^{n} pq = \frac{pq}{n} < 1$, אזי על פי משפט צ'בישב למספרים הגדולים מתקיים

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i| < \epsilon\} = 1 \Leftrightarrow \lim_{n\to\infty} P\{|\frac{m}{n} - p| < \epsilon\} = 1. \qquad \square$$

טענה 8.3 משפט פואסון יש סדרה של נסיונות בלתי תלויים והסתברות בניסיון k היא P_k, אזי $P\{|\frac{m}{n} - \frac{\sum p_i}{n}| < \epsilon\} = 1.$ ההוכחה היא בדיוק כמו במקרה של ברנולי.

משוואות של משפט פואסון: יש נסיונות שאי אפשר לחזור עליהם באותם תנאים הרבה פעמים. למשל לכל בית בעיר יש הסתברות שונה P_i לשריפה, כי אם הסתברות הייתה שווה אזי מספר פניות

לתחנת כיבוי אש היה $m = pn.$ אבל זה ממשיך להיות נכון גם כאשר הסתברויות שונות כל עוד מספר הפניות ממשיך להיות קבוע m.

טענה 8.4 אם X_i בלתי תלויים וכולם מתפלגים לפי אותה התפלגות בעלת תוחלת μ ושונות σ^2 אזי $\lim_{n\to\infty} P(|\overline{X_n} - \mu| < \epsilon) = 1$, זאת אומרת ממוצע מתקרב לתוחלת.

השאלה הנשאלת האם אפשר לקבל משפטים חלשים של מספרים גדולים עם תנאים יותר חלשים על שונות?

משפט 8.3 מרקוב תהי X_i סדרה של משתנים מקריים בלתי תלויים המקיימים

$$\frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) \to 0.$$

זאת אומרת שונות יכולה לגדול אבל לא מהר מדי, אזי $\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu_i| < \epsilon\} = 1.$

הוכחה: נסמן $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ומקבלים $Var(\overline{X_n}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$ ו־ $\mu_{\overline{X_n}} = \frac{1}{n}\sum_{i=1}^{n}\mu_{X_i}$ אזי על פי אי שיוויון צ'בשיב למספרים גדולים מקבלים

$$P(|\overline{X_n} - \mu_{\overline{X_n}}| < \epsilon) \geq 1 - \frac{Var(\overline{X_n})}{\epsilon^2} = 1 - \frac{\frac{1}{n^2}\cdot\sum_{i=1}^{n}\sigma_i^2}{\epsilon^2} \underset{n\to\infty}{\longrightarrow} 1$$

. □

משפט 8.4 חינצ'ין יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים המתפלגים לפי אותה התפלגות עם $E(X_i) = \mu$ אזי $\lim_{n\to\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \epsilon) = 1$

הוכחה: נקח $\delta > 0$ ונגדיר $Y_k = \begin{cases} X_k; |X_k| < \delta n \\ 0; |X_k| \geq \delta n \end{cases}$ ו־ $Z_k = \begin{cases} 0; |X_k| < \delta n \\ X_k; |X_k| \geq \delta n \end{cases}$ ומתקיים $X_k = Y_k + Z_k$ מכאן מקבלים ש־ $E(Y_k) = \int_{-\delta n}^{\delta n} xf(x)dx = \mu_n$ וגם

$$Var(Y_k) = \int_{-\delta n}^{\delta n} (x^2 f(x)dx) - \mu_n^2 \leq \delta n \underbrace{\int_{-\delta n}^{\delta n} |x| f(x)dx}_{b} = \delta n b.$$

מכאן נובע $P\{|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu_n| \geq \epsilon\} \leq \frac{\delta nb}{n\epsilon^2}$. באותו אופן מקבלים ביטוי דומהל־ Z_k,

$$P\{|\frac{1}{n}\sum_{i=1}^{n} Z_i - \nu_n| \geq \epsilon\} \leq \frac{1}{n} - \frac{\delta b}{\epsilon^2}.$$

מזה נובע שמתקיים

$$P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon\} \leq \frac{1}{n} \underset{n\to\infty}{\to} 0.$$

□

8.2 חוקים חזקים של מספרים גדולים

הגדרה 8.1 בהינתן מרחב הסתברות כללי $(\Omega, \mathbb{B}, P)$ וסדרת משתנים מקריים $\{X_n\}_{n=1}^{\infty}$ על Ω:

1. אומרים שהסדרה X_n מתכנסת למשתנה מקרי Y כמעט בוודאות *(almost surely)* אם
$$P(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = Y(\omega)\}) = 1.$$

2. אומרים שהסדרה X_n מתכנסת למשתנה מקרי Y בהסתברות אם לכל $\epsilon > 0$ מתקיים
$$\lim_{n\to\infty} P(\{\omega \in \Omega \mid |X_n(\omega) - Y(\omega)| \geq \epsilon\}) = 0,$$
במילים אחרות אם לכל $\epsilon, \delta > 0$ קיים N כזה שלכל $n \geq N$ מתקיים
$$P(|X_n - Y| \geq \epsilon) < \delta.$$

טענה 8.5 אם X_n מתכנסת ל־ Y כמעט בוודאות אזי מתכנסת ל־ Y בהסתברות.

הוכחה: נסמן $E = \{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = Y(\omega)\}$, ונניח ש־ $P(E) = 1$, כלומר ב־ E יש התכנסות נקודתית של הסדרה. בהינתן $\epsilon > 0$, נסמן E_N את קבוצת הנקודות "הרעות" ב־ E, כך
$$E_N = \{\omega \in E \mid \exists n \geq N, \, |X_n(\omega) - Y(\omega)| \geq \epsilon\}.$$
אז זו סדרה יורדת בהכלה: $E_{N+1} \subseteq E_N$, ולפי הגדרת E, $\prod_{N\geq 1} E_N = \emptyset$. אם נסמן $A_0 = E \backslash E_1$ ו־ $A_n = E_n \backslash E_{n+1}$ לכל $n \geq 1$, אז A_n הינם מאורעות זרים ומתקיים $\sum_{n\geq 0} A_n = E$. לכן
$$1 = P(E) = \sum_{n\geq 0} P(A_n),$$

שגורר ש־ $\lim_{n\to\infty}\sum_{i\geq n}P(A_i)=0$. מצד שני $\cup_{n\geq 0}A_n=E$ ולכן $P(E_n)=\sum_{i\geq n}P(A_i)$, ומצאנו שקיים N כך ש־ $P(E_N)<\epsilon$ לכל $n\geq N$. אולם

$$E_N=\cup_{n\geq N}\{\omega\in E \mid |X_n(\omega)-Y(\omega)|\geq\epsilon\},$$

ולכן בוודאי לכל $n\geq N$ מתקיים $P(\{\omega\in E \mid |X_n(\omega)-Y(\omega)|>\epsilon)<\delta$, כנדרש. $\square$

ההפך לא נכון, להלן דוגמא נגדית. נגדיר משתנה מקרי על $[0,1]$ (עם הסתברות האחידה) כך ש־ $X_n=\chi_{[\frac{i}{2^k},\frac{i+1}{2^k})}$ כאשר $2^{k+1}>n\geq 2^k$ וגם $n=2^k+i$. אז $X_n\to 0$ בהסתברות, אך X_n אינו מתכנס ל־ 0 לאף נקודה ב־ $[0,1]$!

טענה 8.6 אם לכל $\epsilon>0$ מתקיים $\sum_{n=1}^{\infty}P\{|X_n-Y|\geq\epsilon\}<\infty$ אזי X_n מתכנס ל־ Y כמעט בוודאות.

הוכחה: נסמן ב־ $E_{n,\epsilon}$ מאורע $|X_n-Y|\geq\epsilon$ ונגדיר $S_{n,\epsilon}=\sum_{k=1}^{\infty}E_{n+k,\epsilon}$, מכאן מקבלים ש־ $P(S_{n,\epsilon})\leq\sum_{k=1}^{\infty}P(E_{n+k,\epsilon})=\sum_{l=n+1}^{\infty}P(|X_l-Y|\geq\epsilon)$, כאשר $S_{n,\epsilon}$ הנה סדרה מונוטונית יורדת כלומר $\cdots\supset S_{2,\epsilon}\supset S_{1,\epsilon}$ ומתקיים $\lim_{n\to\infty}P(S_{n,\epsilon})=0$. נגדיר $S_\epsilon=\prod_{n=1}^{\infty}S_{n,\epsilon}$ כאשר $P(S_\epsilon)=0$. נסמן $S=S_1+S_{\frac{1}{2}}+S_{\frac{1}{4}}\cdots$, אז

$$S=\{|X_{n+k}-Y|\geq\epsilon \text{ אחד } k(n) \text{ לפחות ל־ } n \text{ לכל }\}$$

מכאן נובע ש־

$$P(S)\leq\sum_{i=1}^{\infty}P(S_{(\frac{1}{2})^i})=0\Leftrightarrow P(\overline{S})=P(\lim_{n\to\infty}X_n=Y\}=1.$$

$\square$

הערה 8.1 כל המשפטים של הפרק הקודם נותנים התכנסות בהסתברות. כאן אנו בודקים מה אפשר לומר על התכנסות כמעט בוודאות. מזה נובע חוקים חזקים של מספרים גדולים. מובן שכדי לקבל חוק חזק צריך או תנאים יותר חזקים או עבודה יותר קשה (או שניהם).

משפט 8.5 חוק חזק של המספרים הגדולים של בורל תהי $\{X_n\}$ סדרה של n נסיונות ברנולי, ויהי m_n מספר ההצלחות ב־$S_n=X_1+X_2+\cdots+X_n$, ותהי p ההסתברות להצלחה בודדת, אזי $P(\lim_{n\to\infty}\frac{m_n}{n}=p)=1$

הוכחה: אם באי שיוויון צ'בישב נשנה את ϵ^2 ל- ϵ^4 נקבל $P(|X-\mu|\geq\epsilon^2)\leq\frac{E((X-\mu)^4)}{\epsilon^4}$ אם X מתפלג באופן הבא $\begin{array}{cc} 0 & 1 \\ \hline q & p \end{array}$ אזי $E(X^i)=q\cdot 0^i+p\cdot 1^i=p$ נחשב את המומנט הרבעי של משתנה ברנולי

$$\begin{aligned}
E((\frac{1}{n}\sum_{i=1}^{n}X_i-p)^4)=&\frac{1}{n^4}\sum_i\sum_j\sum_k\sum_l E((X_i-p)(X_j-p)(X_k-p)(X_l-p)) \\
&=\frac{1}{n^4}(6\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}E((X_i-p)^2(X_j-p)^2)+\sum_{i=1}^{n}E(X_i-p)^4) \\
&=\frac{1}{n^4}(6(pq)^2\cdot\frac{n(n-1)}{2}+npq(p^3+q^3))\underbrace{=}_{pq\leq\frac{1}{4}}\frac{pqn}{n^4}\underbrace{(3pq(n-1)+p^3+q^3)}_{\leq\frac{3}{4}n+\frac{1}{4}<n}<\frac{n}{4n^3} \\
&=\frac{1}{4n^2}
\end{aligned}$$

כלומר קיבלנו שמתקיים $P\{|\frac{m_n}{n}-p|\geq\epsilon^2\}\leq\frac{1}{\epsilon^4}\cdot\frac{1}{4n^2}$, זאת אומרת הטור $\frac{1}{4\epsilon^4}\sum_{n=1}^{\infty}\frac{1}{n^2}$ מתכנס ולכן גם טור $\sum_{n=1}^{\infty}P\{|\frac{mn}{n}-p|\geq\epsilon^2\}$ מתכנס לפי הטענות הקודמות ולכן גם הטור $\sum_{n=1}^{\infty}P\{|\frac{mn}{n}-p|\geq\epsilon\}$ מתכנס. לבסוף מקבלים

$$P(\lim_{n\to\infty}\frac{m_n}{n}=p)=1.$$

$\square$

הערה 8.2 אם $\{X_n\}$ סדרה של משתנים מקריים בלתי תלויים כאלה ש-

$$Var(X_i)<c,\quad E((X_i-\mu_i)^4)<c$$

לכל i אזי לפי אותה הוכחה בדיוק מקבלים

$$P(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mu_i)=1.$$

קיימים משפטים של חוקים חזקים שדורשים הרבה פחות, למשל שני משפטי קולמוגרוב הבאים

משפט 8.6 קולמוגרוב אם $\{X_n\}$ בלתי תלויים בעלי תוחלת μ_n ושונות σ_n^2 המקיימות $\sum_{i=1}^{\infty}\frac{\sigma_i^2}{i^2}<\infty$ אזי $P(\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mu_i)=1$.

משפט 8.7 קיום תוחלת הוא תנאי הכרחי ומספיק לחוק חזק של מספרים גדולים ל־ $\{X_n\}$ בלתי תלויים שכולם מתפלגים לפי אותה התפלגות ובעלי תוחלת μ אזי $P(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu) = 1$

למה 8.1 הלמה של בורל קנטלי (Borel-Cantelli) תהי $\{E_i\}$ סדרת מאורעות במרחב הסתברות $(\Omega, \mathbb{B}, P)$ ונגדיר $E = \limsup_{n\to\infty} E_n := \prod_{n=1}^{\infty}\sum_{k=n}^{\infty} E_k$ זאת אומרת חיתוך של "הזנבות".

1. אם $\sum_{k=1}^{\infty} P(E_k) < \infty$ אזי $P(E) = 0$. זאת אומרת לא יותר ממספר סופי של מאורעות מהסדרה קורים.

2. אם כל המאורעות $E_1, E_2, \ldots$ בלתי תלויים ו־ $\sum_{k=1}^{\infty} P(E_k) = \infty$ אזי $P(E) = 1$, זאת אומרת מספר אינסופי של מאורעות קורים.

הוכחה:

1. $E \subset \sum_{k=n}^{\infty} E_k$ לכל n, לכן $P(E) \leq P(\sum_{k=n}^{\infty} E_k) \leq \sum_{k=n}^{\infty} P(E_k)$ לכל n. טור $\sum_{k=1}^{\infty} P(E_k)$ מתכנס אזי לכל $\epsilon > 0$ קיים N כזה שלכל $n > N$ מתקיים $\sum_{k=n}^{\infty} P(E_k) < \epsilon$, זאת אומרת $P(E) < \epsilon$ לכל $\epsilon > 0$ נקבל $P(E) = 0$.

2. $\sum_{k=1}^{\infty} P(E_k) = \infty$, נגדיר $A_n = \overline{E_n + E_{n+1} + E_{n+2} + \cdots} = \prod_{k=n}^{\infty} \overline{E_k}$, לפי חוקי דה־מורגן מתקיים $\overline{E} = \sum A_n$ וגם $A_1 \subset A_2 \subset \cdots$ מכאן מקבלים $P(\overline{E}) = \lim_{n\to\infty} P(A_n)$, אבל $\prod_{k=n}^{N} (1 - P(E_k)) = \prod_{k=n}^{N} P(\overline{E_k}) = P(\prod_{k=n}^{N} \overline{E_k}) \leq P(\prod_{k=n}^{\infty} \overline{E_k}) = P(A_n)$. לכל x בקטע $(0,1)$ מתקיים $1 - x \leq e^{-x}$ (למה?) ולכן $0 \leq P(A_n) \leq \prod_{k=n}^{N} e^{-P(E_k)} = e^{-\sum_{k=n}^{N} P(E_k)} \underset{n\to\infty}{\to} 0$. לכן לכל $\epsilon > 0$ קיים N כך שלכל $n > N$ מתקיים $P(A_n) < \epsilon$ זאת אומרת $P(\overline{E}) = 0 \Rightarrow P(E) = 1$.

$\square$

משפט 8.8 (החוק החזק של המספרים הגדולים) יהיו $\{X_n\}_{n\geq 1}$ משתנים מקריים בלתי תלויים עם אותה התפלגות על מרחב הסתברות $(\Omega, \mathbb{B}, P)$ עם תוחלת μ, שונות σ^2, ונניח ש־ $E(X_n^4) < \infty$.
אז

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1.$$

במלים אחרות, הממוצעים $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ שואפים כמעט בוודאות למשתנה מקרי הקבוע μ.

הוכחה: ראשית, נשם לב שאם $\{Y_i\}_{i\geq 1}$ משתנים מקריים אי־שליליים ו־ $\sum_{i\geq 1} Y_i < \infty$, אז $Y_n \to 0$ כמעט בוודאות. זאת משום שאם נסמן $Z_n = Y_1 + \cdots + Y_n$ אז ה־ Z_n מהווים סדרה עולה חלש, ולכן הם מתכנסים לגבול $Z = \lim_{n\to\infty} Z_n$ (אולי סופי). לפי משפט ההתכנסות המונוטונית של תורת המידה, מתקיים

$$E(Z) = \lim_{n\to\infty} E(Z_n) = \lim_{n\to} \sum_{k=1}^{n} E(Y_k) = \sum_{n\geq 1} E(Y_n) < \infty.$$

לכן Z משתנה מקרי סופי (מלבד אולי לקבוצת A בהסתברות 0). אולם $Z(\omega) = \sum_{n\geq 1} Y_n(\omega)$ לכל $\omega \notin A$ ולכן $Y_n(\omega) \to 0$ לכל $\omega \notin A$ וזה בדיוק אומר ש־ $Y_n \to 0$ כמעט בוודאות.

עתה ניישם את האמור עבור המתשנה המקרי $(S_n - \mu)^4$. כאן נשתמש באי־התלות ונמצא ש־

$$E\left(\left(\sum_{i=1}^{n} X_i - \mu\right)^4\right) = nE((X_1-\mu)^4) + 6\binom{n}{2}\sigma^4 + 4\binom{n}{2}E((X_1-\mu)^3) < Cn^2,$$

עבור קבוע C. לכן $E\left(\frac{1}{n}S_n - \mu\right)^4 \leq C/n^2$ ואז על פי לעיל $Y_n \to 0$ מתכנסת כמעט בוודאות.
$\square$

הערה 1: משפט זה בעצם הוא נכון גם בלי להניח שהשונות או $E(X_n^4)$ סופיים - והוא גם נכון למקרה $\mu = \infty$.

הערה 2: במקום להשתמש במשפט ההתכנסות החסומה, נוכל להשתמש בלמה של בורל־קנטאלי כדלקמאן: בהוכחת אי־שוויון צ'בישב למעשה הוכחנו ש־ $P(|X| \geq t) \leq \frac{E(|X|^k}{t^k}$ לכל $k, t \geq 0$ - ובפרט במקרה שלנו לכל $\epsilon > 0$ מתקיים

$$P\left(\left|\sum_{k=1}^{n} X_k - \mu\right| \geq n\epsilon\right) \leq \frac{Cn^2}{n^4\epsilon^4},$$

או $P(|Y_n| \geq \epsilon) \leq \frac{C}{\epsilon^4 n^2}$. עתה לכל $\epsilon > 0$, נגדיר

$$A_n = \{\omega \in \Omega \mid |Y_n(\omega)| \geq \epsilon\}.$$

אז $P(A_n) \leq \frac{K}{n^2}$, ולכן $\sum_{n\geq 1} P(A_n) < \infty$. לפי הלמה של בורל־קנטאלי (i) המאורע $B_\epsilon = \prod_{k\geq 1} \sum_{n\geq k} A_n$ יש הסתברות אפס - כלומר

$$P(B_\epsilon) = P(\{\omega \in \Omega \mid \forall k \, \exists n \geq k, \, |Y_n(\omega)| \geq \epsilon\}) = 0,$$

ולכן המאורע המשלים $\overline{B_\epsilon}$ הוא מאורע וודאי (בהסתברות 1). אך חיתוך בן מנייה של מאורעות וודאיים הוא וודאי, ולכן $P\left(\prod_{i\geq 1} \overline{B_{1/i}}\right) = 1$. זאת אומרת, בדיוק בהסתברות 1 לכל $\epsilon > 0$ קיים N כך שלכל $n \geq N$, $|Y_n(\omega)| < \epsilon$. כלומר, $\lim_{n\to\infty} |Y_n(\omega)| = 0$. ומכאן ההוכחה של המשפט ממשיכה כפי שראינו.

הערה 3: בהוכחת המשפט השתמשנו בעובדה הבאה.

למה 8.2 אם $\{X_i\}_{i=1}^n$ משתנים מקריים עם אותה התפלגות, תוחלת μ ושונות σ^2, אז

(א) $Var\left(\sum_{i=1}^n X_i - \mu\right) = E\left(\left(\sum_{i=1}^n X_i - \mu\right)^2\right) = \sum_{i=1}^n Var(X_i) = n\sigma^2$.

(ב) $E\left(\left(\sum_{i=1}^n X_i - \mu\right)^3\right) = nE\left((X_1 - \mu)^4\right)$.

(ג) $E\left(\left(\sum_{i=1}^n X_i - \mu\right)^4\right) = nE\left((X_i - \mu)^4\right) + 6\binom{n}{2}\sigma^4$.

הוכחה: (א) נזכור שכבר ראינו ש־ $Var\left(\sum_{i=1}^n X_i\right) = n\sigma^2$.

(ב) נסמן $Y_i = X_i - \mu$. אז $E(Y_i) = 0$ ו־ $E(Y_i^2) = Var(Y_i) = Var(X_i) = \sigma^2$. אז

$$\begin{aligned} E((Y_1+Y_2)^3) &= E(Y_1^3 + 3Y_1^2Y_2 + 3Y_1Y_2^2 + Y_2^3) \\ &= E(Y_1^3) + 3E(Y_1^2)E(Y_2) + 3E(Y_1)E(Y_2^2) + E(Y_2^3) \\ &= E(Y_1^3) + E(Y_2^3) = 2E(Y_1^3). \end{aligned}$$

באינדוקציה מתקיים

$$\begin{aligned} E\left(\left(\sum_{i=1}^n Y_i\right)^3\right) &= E\left((Z+Y_n)^3\right) \\ &= E(Z^3) + 3E(Z^2)E(Y_n) + 3E(Z)E(Y_n^2) + E(Y_n^3) \\ &= (n-1)E(Y_1^3) + E(Y_n^3) \\ &= nE(Y_1^3). \end{aligned}$$

(ג) שוב

$$\begin{aligned} E((Y_1+Y_2)^4) &= E(Y_1^4 + 4Y_1^3Y_2 + 6Y_1^2Y_2^2 + 4Y_1Y_2^3 + Y_2^4) \\ &= 2E(Y_1^4) + 6E(Y_1^2)E(Y_2^2) \\ &= 2E(Y_1^4) + 6\sigma^4. \end{aligned}$$

באינדוקציה מתקיים

$$\begin{aligned} E\left(\left(\sum_{i=1}^n Y_i\right)^4\right) &= E\left((Z+Y_n)^4\right) \\ &= E(Z^4) + 4E(Z^4)E(Y_n) + 6E(Z^2)E(Y_n^2) + 4E(Z)E(Y_n^3) + E(Y_n^4) \\ &= (n-1)E(Y_1^4) + 6\binom{n-1}{2}\sigma^2 + 6(n-1)\sigma^4 + E(Y_1^4) \\ &= nE(Y_1^4) + 6\binom{n}{2}\sigma^4. \end{aligned}$$

□

שימוש למשפט ויירשטראס: תהי $f:[0,1]\to\mathbb{R}$ פונקציה חסומה ונניח ש־ $\{X_i\}_{i=1}^n$ משתנים מקריים ברנולי בלתי תלויים עם אותה התפלגות p להצלחה וגם $S_n = X_1 + \cdots + X_n$. פולינום ברנשטיין המתאים הוא

$$E(F(S_n/n)) = \sum_{k=0}^n f(k/n)\binom{n}{k}p^kq^{n-k} = B_n(p).$$

זהו פולינום ב־ p שהמקדמים שלו תלויים בפונצקיה f מדרגה n, ועבור n גדול רב המשקל מרוכז באותם מקומות כך ש־ k/n קרוב ל־ p.

משפט 8.9 אם $f:[0,1]\to\mathbb{R}$ פונקציה רציפה, אז

$$\lim_{n\to\infty}\max_{0\le p\le 1}|B_n(p)-f(p)|=0,$$

כלומר לכל $\epsilon>0$ קיים n כך ש־ B_n קרוב ל־ f עד כדי ϵ בנורמת הסיפרימום.

הוכחה: ברור ש־ f חסומה בערכה המוחלט (על ידי M) ורציפה במידה שווה ־ לכן לכל $\epsilon>0$ קיים $\delta>0$ כך שלכל $x,y\in[0,1]$ מתקיים $|x-y|<\delta$ גורר ש־ $|f(x)-f(y)|<\epsilon$. לכן

$$\begin{aligned}&|B_n(p)-f(p)|\\ &=\left|\sum_{k=0}^{n}(f(k/n)-f(p))\binom{n}{k}p^kq^{n-k}\right|\\ &\le\sum_{k=0}^{n}|f(k/n)-f(p)|\binom{n}{k}p^kq^{n-k}\\ &=\sum_{|k/n-p|<\delta}|f(k/n)-f(p)|\binom{n}{k}p^kq^{n-k}+\sum_{|k/n-p|\ge\delta}|f(k/n)-f(p)|\binom{n}{k}p^kq^{n-k}\\ &\le\sum_{|k/n-p|<\delta}\epsilon\binom{n}{k}p^kq^{n-k}+\sum_{|k/n-p|\ge\delta}|f(k/n)-f(p)|\binom{n}{k}p^kq^{n-k}\\ &\le\sum_{k=0}^{n}\epsilon\binom{n}{k}p^kq^{n-k}+2M\sum_{|k/n-p|\ge\delta}\binom{n}{k}p^kq^{n-k}\\ &=\epsilon+2MP(|\tfrac{1}{n}S_n-p|\ge\delta),\end{aligned}$$

ולכן על פי אי־שוויון צ'יבשיב נקבל

$$\begin{aligned}|B_n(p)-f(p)|&\le\epsilon+2MP(|\tfrac{1}{n}S_n-p|\ge\delta)\\ &\le\epsilon+2M\tfrac{pq}{n\epsilon^2}\to_{n\to\infty}\epsilon.\end{aligned}$$

□

8.3 משפט הגבול המרכזי

משפט 8.10 דה־מאוור $B(n,0.5)\to N(0.5n,0.5\sqrt{n})$

הוכחה: נתבונן רק במקרה ש־ $n=2m$ ונסמן

$$a_k:=P(X=m+k)=\binom{2m}{m+k}0.5^{2m}\underbrace{=}_{\binom{n}{k}=\binom{n}{n-k}}\binom{2m}{m-k}0.5^{2m}=a_{-k},$$

זאת אומרת מתקיים $a_k=a_{-k}$ לכל k.
נחשב עבור $k=1$ ונקבל $a_0=\binom{2m}{m}0.5^{2m}$, ואילו $a_1=\binom{2m}{m+1}0.5^{2m}$, כלומר

$$\frac{a_1}{a_0}=\frac{m}{m+1}\Rightarrow a_1=\frac{m}{m+1}\cdot a_0.$$

נותר לחשב את

$$\frac{a_k}{a_{k-1}}=\frac{(2m)!(0.5)^{2m}}{(m+k)!(m-k)!}\cdot\frac{(m+k-1)!(m-k+1)!}{(2m)!(0.5)^{2m}}=\frac{m-k+1}{m+k},$$

כלומר לכל $k \geq 1$ מתקיים

$$\begin{aligned} a_k &= \frac{m-k+1}{m+k} \cdot a_{k-1} = \frac{m-k+1}{m+k} \cdot \frac{m-k+2}{m+k-1} \cdot a_{k-2} \\ &= \underbrace{\cdots}_{\text{באינדוקציה}} = \frac{m-k+1}{m+k} \cdot \frac{m-k+2}{m+k-1} \cdots \frac{m}{m+1} \cdot a_0, \end{aligned}$$

כאשר $a_0 = \binom{2m}{m} 0.5^{2m}$.

לפי טור טיילור סביב האפס לפונקציה e^x מתקיים $e^{\frac{d}{m}} = 1 + \frac{d}{m} + O((\frac{d}{m})^2)$, לכן

$$a_k \approx \frac{e^{\frac{-1}{m}} e^{\frac{-2}{m}} e^{\frac{-3}{m}} \cdots e^{-\frac{k-1}{m}}}{e^{\frac{1}{m}} e^{\frac{2}{m}} \cdots e^{\frac{k}{m}}} \cdot a_0 = \frac{e^{\frac{-1}{m}(1+2+3+\cdots+(k-1))}}{e^{\frac{1}{m}(1+2+3+\cdots+k)}} \cdot a_0 = e^{\frac{-1}{m}(\frac{k(k-1)}{2} + \frac{k(k+1)}{2})} \cdot a_0 = a_0 e^{\frac{-k^2}{m}}$$

מכאן מקבלים $a_k = (e^{\frac{-k^2}{m}} + O(\frac{k^3}{m^2})) \cdot a_0$. תוך שימוש בנוסחת סטרילינג נקבל ש-

$$\begin{aligned} a_0 &= \binom{2m}{m} 0.5^{2m} \\ &= \frac{(2m)!}{(m!)^2} \cdot 0.5^{2m} \approx \frac{\sqrt{2m \cdot 2\pi} \cdot \left(\frac{2m}{e}\right)^{2m}}{\left(\sqrt{m \cdot 2\pi} \cdot \left(\frac{m}{e}\right)^m\right)^2} \cdot 0.5^{2m} \\ &= \frac{2\sqrt{m\pi} \cdot \left(\frac{m}{e}\right)^{2m}}{m \cdot 2\pi \cdot \left(\frac{m}{e}\right)^{2m}} = \frac{\sqrt{m\pi}}{m\pi} \\ &= \frac{1}{\sqrt{m\pi}} \end{aligned}$$

וגם

$$a_k = \frac{1}{\sqrt{m\pi}} \cdot \left(e^{-\frac{k^2}{m}} + O(\frac{k^3}{m^2}) \right).$$

יהי $a_k = P(x = m+k)$, נחשב $X \sim N(m, 0.5\sqrt{2m})$ באופן הבא: ראשית כל פונקצית הצפיפות היא $f(x) = \frac{1}{\sqrt{\pi m}} \cdot e^{-\frac{1}{2} \cdot \frac{(x-m)^2}{0.5m}} = \frac{1}{\sqrt{\pi m}} \cdot e^{-\frac{(x-m)^2}{m}}$, מכאן מקבלים גם

$$f(m+k) = \frac{1}{\sqrt{\pi m}} \cdot e^{-\frac{k^2}{m}},$$

כלומר $\mathrm{B(n, 0.5)} \underset{\mathrm{n}\to\infty}{\rightarrow} N(0.5n, o.5\sqrt{n})$. $\square$

8.4 פונקציות אופייניות והוכחת משפט הגבול המרכזי

נתון משתנה מקרי X נגדיר פונקציה אופינית או יוצרת שלו באופן הבאה $g_X(t) = E(e^{itX})$.

1. אם X משתנה מקרי בדיד אזי $g_X(t)=\sum\limits_{j\in I} e^{itx_j}P(X=x_j)$.

2. אם X משתנה מקרי רציף אזי $g_X(t)=\int\limits_{-\infty}^{\infty} e^{itx}f(x)dx$, זה נקרא טרנספורם פוריה של פונקציה $f(x)$.

דוגמא 8.1 1. אם X משתנה מקרי ברנולי, כלומר

X	0	1
	q	p

אזי

$$g_X(t)=e^{it\cdot 0}\cdot q+e^{it\cdot 1}\cdot p=q+e^{it}\cdot p.$$

2. $X\sim N(0,1)$ אזי

$$g_X(t)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{itx-\frac{x^2}{2}}dx \underbrace{=}_{z=x-it,\, x=z+it,\, dz=dx} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty} e^{-\frac{z^2}{2}}dz=e^{-\frac{t^2}{2}}.$$

3. X מתפלג מערכי, פואסוני, לפי פרמטר λ, כלומר $X\sim P(\lambda)$ ו- $P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ אזי

$$g_X(t)=\sum_{k=0}^{\infty} e^{itk}\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{1}{k!}(\lambda e^{it})^k \underbrace{=}_{e^x=\sum\limits_{k=0}^{\infty}\frac{1}{k!}x^k} e^{-\lambda}(e^{\lambda e^{it}})=e^{\lambda e^{it}-\lambda}=e^{\lambda(e^{it}-1)}.$$

4. X מתפלג אחיד כלומר $X\sim Uni[-a,a]$ אזי

$$\begin{aligned} g_X(t)&=\int\limits_{-a}^{a} e^{itx}\cdot\frac{1}{2a}dx\\ &=\frac{1}{2a}\int\limits_{-a}^{a} e^{itx}dx\\ &=\frac{1}{2a}\cdot\frac{e^{itx}}{it}\bigg|_{-a}^{a}\\ &=\frac{e^{ita}}{2iat}-\frac{e^{-iat}}{2ait}=\frac{\sin(at)}{at}. \end{aligned}$$

משפט 8.11 תכונות של פונקציות יוצרות

1. $g_X(t)$ קיימת לכל t.

2. $g_X(t)$ רציפה במידה שווה ומקיימת. $g_X(0)=1$ (1). $|g_X(t)|\le 1$ (2).

הוכחה:

.1

$$|g_X(t)| = |\int_{-\infty}^{\infty} e^{itx} f(x)dx| \leq \int_{-\infty}^{\infty} |e^{itx}| \cdot f(x)dx = \int_{-\infty}^{\infty} f(x)dx = 1$$

לכן תמיד אינטגרל זה חסום על ידי 1 וקיים לכל t. מזה נובע שמתקיים $g_x(0) = 1$ כמו כן $|g_X(t)| \leq 1$.

2. נותר להוכיח ש־ $g_X(t)$ רציפה במידה שווה, כלומר להוכיח שלכל h מתקיים

$$g_X(h+t) - g_X(t) = \int_{-\infty}^{\infty} e^{i(h+t)x} f(x)dx - \int_{-\infty}^{\infty} e^{itx} f(x)dx = \int_{-\infty}^{\infty} e^{itx}(e^{ihx} - 1) f(x) \cdot dx$$

, כלומר

$$|g_X(h+t) - g_X(t)| \leq \int_{-\infty}^{\infty} |e^{ihx} - 1| f(x)dx,$$

לכן לכל $\epsilon > 0$ קיים A כזה ש־ $\int_{-\infty}^{-A} f(x)dx + \int_{A}^{\infty} f(x)dx < \frac{\epsilon}{4}$. נמצא h כזה שלכל $|x| < A$ מתקיים $|e^{ihx} - 1| < \frac{\epsilon}{2}$ אזי

$$|g_X(h+t) - g_X(t)| \leq \int_{-A}^{A} \frac{\epsilon}{2} \cdot f(x)dx + 2\frac{\epsilon}{4} \leq \frac{\epsilon}{2} \cdot 1 + 2 \cdot \frac{\epsilon}{4} = \epsilon.$$

□

משפט 8.12 אם X, Y משתנים מקריים כך ש־ $Y = aX + b$, a, b קבועים אזי $g_Y(t) = g_X(at) \cdot e^{ibt}$.

הוכחה:

$$g_Y(t) = E(e^{itY}) = E(e^{it(aX+b)}) = E(e^{itaX} \cdot e^{itb}) = e^{itb} E(e^{itaX}) = e^{itb} g_X(at).$$

□

משפט 8.13 אם X_1, X_2 בלתי תלויים אזי $g_{X_1+X_2}(t) = g_{X_1}(t) g_{X_2}(t)$.

הוכחה:

$$g_{X_1+X_2}(t) = E(e^{it(X_1+X_2)}) = E(e^{itX_1} \cdot e^{itX_2}) = E(e^{itX_1})E(e^{itX_2}) = g_{X_1}(t)g_{X_2}(t)$$

□

הערה 8.3 אם $\{X_i\}$ בלתי תלויים אזי לכל n מתקיים $g_{\sum\limits_{j=1}^{n} X_j}(t) = \prod\limits_{j=1}^{n} g_{X_j}(t)$.

משפט 8.14 אם $E(X^k)$ קיים לכל $1 \le k \le n$, זאת אומרת מתקיים $\int\limits_{-\infty}^{\infty} |x^k| \cdot f(x)dx < \infty$, אזי ל- $g_X(t)$ יש n נגזרות ומתקיים $g_X^{(k)}(0) = i^k E(X^k)$.

הוכחה: תהי $g_X(t) = \int\limits_{-\infty}^{\infty} e^{itx} f(x)dx$, אזי

$$g_X^{(k)}(t) = \frac{\partial^k}{\partial t^k} g_X(t) = \int\limits_{-\infty}^{\infty} \frac{\partial^k}{\partial t^k}(e^{itx}) f(x)dx = \int\limits_{-\infty}^{\infty} (ix)^k \cdot e^{itx} \cdot f(x)dx = i^k \int\limits_{-\infty}^{\infty} e^{itx} \cdot x^k \cdot f(x)dx$$

מכאן נובע $|g_X^{(k)}(t)| \le \int\limits_{-\infty}^{\infty} |x|^k f(x)dx$ ואז $g_X^{(k)}(0) = i^k \cdot E(X^k)$. □

הגדרה 8.2 $\rho_X(t) = \ln(g_X(t))$.

נחשב

$$\rho'_X(0) = \left.\frac{g'_X(t)}{g_X(t)}\right|_{t=0} = \frac{g'_X(0)}{g_X(0)} = i \cdot E(X).$$

$$\begin{aligned}\rho''_X(0) &= \left.\tfrac{g''_X(t) \cdot g_X(t) - (g'_X(t))^2}{(g_X(t))^2}\right|_{t=0} = \tfrac{-E(X^2) - (i \cdot E(X))^2}{1^2} \\ &= -E(X^2) + (E(x))^2 = (E(X))^2 - E(X^2) = -Var(X).\end{aligned}$$

משפט 8.15 יהי X משתנה מקרי רציף בעל פונקצית צפיפות $f(x)$ ופונקצית התפלגות מצטברת $F(x)$, אזי

1. לכל $x_1, x_2 \in \mathbb{R}$ מתקיים

$$F(x_2) - F(x_1) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} \frac{e^{-itx_1} - e^{-itx_2}}{it} g_x(t)dt.$$

2. טרנספורמציה הפוכה לטרנספורם פוריה

$$f(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx} g_x(t)dt.$$

הוכחה:

.1

$$J_c=\frac{1}{2\pi}\int_{-c}^{c}\frac{e^{-itx_1}-e^{-itx_2}}{it}g_x(t)dt=\frac{1}{2\pi}\int_{-c}^{c}\int_{-\infty}^{\infty}\frac{e^{-itx_1}-e^{-itx_2}}{it}\cdot e^{itz}f(z)dzdt$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}f(z)\int_{-c}^{c}\frac{e^{it(z-x_1)}-e^{it(z-x_2)}}{it}dtdz \quad \underbrace{=}_{\int_{-c}^{0}\frac{e^{ixA}}{ix}dx\underbrace{=}_{y=-x}-\int_{0}^{c}\frac{e^{-iyA}}{iy}dy}$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}f(z)\underbrace{\int_{0}^{c}\frac{e^{it(z-x_1)}-e^{-it(z-x_1)}}{it}}_{=\frac{2\sin t(z-x_1)}{t}}-\underbrace{\frac{e^{it(z-x_2)}-e^{-it(z-x_2)}}{it}}_{=\frac{2\sin t(z-x_2)}{t}}dtdz$$

$$=\frac{1}{\pi}\int_{-\infty}^{\infty}f(z)\int_{0}^{c}\frac{\sin t(z-x_1)}{t}-\frac{\sin t(z-x_2)}{t}dtdz.$$

לכל $\delta>0$ ולכל $|\alpha|>\delta$ מתקיים

$$\frac{1}{\pi}\int_{0}^{\infty}\frac{\sin\alpha x}{x}dx=\begin{cases}\frac{1}{2};\alpha>0\\-\frac{1}{2};\alpha<0\end{cases}$$

לכן ההתכנסות היא במידה שווה. אם $\delta<|\alpha|$ אזי לכל c מתקיים $|\frac{1}{\pi}\int_{0}^{c}\frac{\sin\alpha t}{t}dt|<1$, לכן

$$J_c=\int_{-\infty}^{x_1-\delta}\cdots+\int_{x_1-\delta}^{x_1+\delta}\cdots+\int_{x_1+\delta}^{x_2-\delta}\cdots+\int_{x_2-\delta}^{x_2+\delta}\cdots+\int_{x_2+\delta}^{\infty}\cdots$$

ל־ δ כזה ש־ $x_1+\delta<x_2-\delta$ אזי $\begin{cases}z-x_2<-\delta\\x_1-z<-\delta\end{cases}$ ולכן מתקיים $\int_{-\infty}^{x_1-\delta}\cdots\underset{\delta\to0}{\to}0$. באותו אופן $\int_{x_2+\delta}^{\infty}\cdots\underset{\delta\to0}{\to}0$.

עבור $z-x_1>\delta, z-x_2>\delta$ מתקיים

$$\int_{x_1+\delta}^{x_2-\delta}\cdots\underset{\delta\to0}{\to}F(x_2-\delta)-F(x_1+\delta).$$

עבור $z - x_1 > \delta, z - x_2 < \delta$ מתקיים בצורה דומה ש־

$$|\int_{x_1-\delta}^{x_1+\delta} \cdots| < 2\int_{x_1-\delta}^{x_1+\delta} f(z)dz = 2(F(x_1+\delta) - F(x_1-\delta))$$

בנוסף גם

$$|\int_{x_2-\delta}^{x_2+\delta} \cdots| < 2\int_{x_2-\delta}^{x_2+\delta} f(z)dz = 2(F(x_2+\delta) - F(x_2-\delta)),$$

וזה נכון לכל $\delta > 0$ ולכן $J_c \underset{c\to\infty}{\longrightarrow} F(x_2) - F(x_1)$

.2

$$f(x) = F'(x) = \lim_{h\to 0}\frac{F(x+h) - F(x)}{h} = \lim_{h\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx} - e^{-it(x+h)}}{ith}g_x(t)dt$$

$$= \lim_{h\to 0}\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}(1-e^{-ith})}{ith}g_x(t)dt \underbrace{=}_{\text{שווה במידה התכנסות}} \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}g_x(t)\underbrace{(\lim_{h\to 0}\frac{1-e^{-ith}}{ith})}_{=1}dt$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}g_x(t)dt.$$

□

מסקנה 8.1 $f(x)$ מוגדרת באופן יחיד על ידי $g_X(t)$, זאת אומרת אם $g_X(t) = g_Y(t)$ לכל t אזי ל־ X ול־ Y יש אותה התפלגות.

משפט 8.16 תהי $\{X_j\}$ סדרה של משתנים מקריים בעלי פונקצית צפיפות $f_j(x)$ ופונקציות יוצרות $g_j(t)$ בהתאם, ידוע לנו גם ש־ $g_j(t) \to g(t)$ אזי $f_j(x) \to f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}g(t)dt.$

משפט 8.17 הגבול המרכזי יהי $\{X_j\}_{j=1}^{\infty}$ משתנים מקריים בלתי תלויים, בעלי אותה התפלגות עם תוחלת μ וסטיית תקן σ. נסמן $S_n = \sum_{j=1}^{n} X_j$ אזי $S_n \to N(n\mu, \sigma\sqrt{n}).$

הוכחה: נניח כי $\mu = 0$ אזי

$$g_X(t) = \underbrace{g_X(0)}_{=1} + g_X'(0)t + \frac{g_X''(0)}{2} \cdot t^2 + \alpha(t) \cdot t^2.$$

מאחר ו־ $E(X^2) - \underbrace{(E(X))^2}_{=0} = \sigma^2$ אזי מקבלים ש־ $E(X^2) = \sigma^2 \Rightarrow \frac{g_X''(0)}{2} = -\frac{\sigma^2}{2}$ מקבלים ש־ $g_X(t) = 1 - (\frac{\sigma^2}{2} - \alpha(t))t^2$.

מצד שני $S_n = \sum_{j=1}^{n} X_j$ ו־ $Z_n = \frac{S_n}{\sigma\sqrt{n}}$ ולכן $g_{Z_n}(t) = g_{S_n}(\frac{t}{\sigma\sqrt{n}}) = (1-(\frac{\sigma^2}{2}-\alpha(\frac{t}{\sigma\sqrt{n}}))\frac{t^2}{n\sigma^2})^n$

כמו כן $\ln g_{Z_n}(t) = n\ln((1 - \underbrace{(\frac{\sigma^2}{2} + \alpha(\frac{t}{\sigma\sqrt{n}}))\frac{t^2}{n\sigma^2}}_{\underbrace{\rightarrow}_{n\to\infty} 0})$. מזה ש־ $\ln(1-x) = -x - x^2/2 - \cdots$ נובע

$$\lim_{n\to\infty} \ln g_{Z_n}(t) = \lim_{n\to\infty} -n(\frac{\sigma^2}{2})\frac{t^2}{n\sigma^2} = \frac{-t^2}{2},$$

כלומר

$$\lim_{n\to\infty} g_{Z_n}(t) = e^{\frac{-t^2}{2}} = g_Y(t),$$

זאת אומרת קיבלנו ש־ $Y \sim N(0,1)$ ו־ $Z_n \sim N(0,1)$, לכן

$$S_n \sim N(0, \sigma\sqrt{n}).$$

עבור μ כללי, נסמן $X_i = Y_i + \mu$ ונקבל

$$S_n = \sum Y_i + n\mu \Rightarrow S_n \to N(n\mu, \sigma\sqrt{n}).$$

□