

סיכום הקורס מבוא להסתברות וסטטיסטיקה למדעי המחשב א', שנת הלימודים תשע"ב

מאת סשה גולדשטיין, sashag@cs

הסיכום להלן מבוסס על הרצאותיו של ד"ר גל אלידן ושיעורי התרגול של מר לירון רבנר. זהו אינו סיכום רשמי של סגל הקורס, ואינני מתחייב על דיוק הנאמר בו. לא כל הדוגמאות שהובאו בשיעור נמצאות בסיכום. אם אתם מוצאים טעות, אי-דיוק, השמטה – אשמח לקבל הערות למייל שלי, sashag@cs.

הסתברות בסיסית

מטרת הקורס היא לבנות שפה פורמלית לטיפול באי-וודאות. נעסוק במונחים ניסוי, מרחב מדגם, הסתברות.

מרחב מדגם ופונקציית ההסתברות

הגדרה: מרחב המדגם הוא אוסף כל התוצאות האפשריות בניסוי (מאורעות), ונסמנו Ω .

הגדרה: פונקציית ההסתברות $P: \Omega \rightarrow \mathbb{R}$ ממפה מאורעות למספרים. פונקציית ההסתברות מקיימת את האקסיומות הבאות (אקסיומות קולמוגורוב):

$$P(\Omega) = 1 \quad (1)$$

$$P(A) \geq 0, A \text{ מאורע} \quad (2)$$

$$P(A \cup B) = P(A) + P(B), A \cap B = \emptyset \text{ זוג מאורעות זרים} \quad (3)$$

הגדרה: עבור מרחב מדגם בדיד וסופי, פונקציית ההסתברות המקיימת $P(A) = \frac{|A|}{|\Omega|}$ לכל מאורע A נקראת פונקציית ההסתברות אחידה.

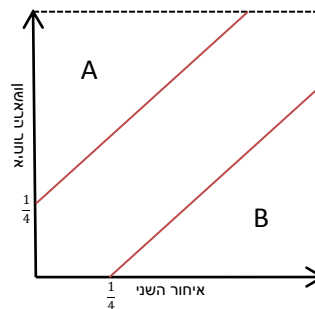
קל לראות שפונקציית ההסתברות אחידה אכן מקיימת את האקסיומות הדרושות מפונקציית ההסתברות.

הגדרה: נחלק את מרחב המדגם למאורעות $A_1, \dots, A_n$ כך שלכל $i \neq j$ מתקיים $A_i \cap A_j = \emptyset$ וגם $\bigcup_{i=1}^n A_i = \Omega$. למאורעות $A_1, \dots, A_n$ נקרא מאורעות אלמנטריים.

דוגמאות:

- הטלת מטבע לא מוטת. מרחב המדגם הוא $\{H, T\}$ וכאן $P(T) = P(H) = \frac{1}{2}$.
- בהטלת קובייה 5 פעמים, ניתן להסתכל על מספר מרחבי מדגם:
 - תוצאה ספציפית – יש 6^5 תוצאות אפשריות, למשל $\{1, 2, 4, 6, 4\}$
 - מקסימום בכל 5 ההטלות – יש 6 תוצאות אפשריות
 - כמה פעמים היה מספר זוגי בקובייה – יש 6 תוצאות אפשריות
- מרחב מדגם של הטלת קובייה לא מוטת פעמיים – 36 תוצאות אפשריות, והוא סימטרי (או אחיד) – ההסתברות של כל מאורע אלמנטרי היא אותו הדבר
- מהי ההסתברות של-12 אנשים יש ימי הולדת ב-12 חודשים שונים?
 - במרחב המדגם 12^{12} מאורעות שונים, ומתוכם $12!$ מאורעות העונים על הדרישה. בהנחה שמרחב המדגם הוא סימטרי, התוצאה היא $\frac{12!}{12^{12}} \sim 0.00005$

- נגדיר פונקציית הסתברות כך ש- $P(A) = 1$ עבור כל A כך שקיים מאורע אלמנטרי ω שמקיים $\omega \in A$, ועבור כל A אחר, $P(A) = 0$. זוהי אכן פונקציית הסתברות: שתי האקסיומות הראשונות מתקיימות באופן טריוויאלי, וגם השלישית מתקיימת מבדיקה.
- ניתן להגדיר גם פונקציית הסתברות על מרחבי מדגם רציפים (אינסופיים):
 - למשל, מסובבים גלגל מזל שהיקפו 1 כך שהגלגל יכול להיעצר על כל מספר בין 0 ל-1. כאן, למשל, $P(0.52) = 0$ וכך גם לכל תוצאה ספציפית אחרת.
 - כדי להגיע להסתברויות חיוביות, נסתכל על קטעים ונגדיר $P([a, b] \subseteq [0, 1]) = b - a$, ועבור איחוד של קטעים זרים $P([a, b] \cup [c, d]) = (b - a) + (d - c)$.
 - קל לראות שהאקסיומות מתקיימות עבור בנייה מהסוג הזה.
- זורקים מטבע בקוטר 3 ס"מ על שולחן אינסופי של משבצות 4-על-4 ס"מ. מה ההסתברות שהמטבע נופל בתוך משבצת?
 - במרכז כל משבצת 4-על-4 יש ריבוע בגודל 1-על-1 שאם מרכז המטבע נופל בתוכו – מקבלים את המאורע המבוקש. בהנחה שההסתברות אחידה, ההסתברות לכך היא $\frac{1}{16}$.
- שני חברים שקבעו להיפגש מאחרים לפגישה בין 0 ל-60 דקות באופן אחיד; כל אחד מחכה לשני רבע שעה, ואם חברו לא מגיע – עוזב את המקום. מה ההסתברות שייפגשו?
 - ניתן להסתכל גם על הבעיה הזאת כבעיית שטח – השטח בין שני הישרים האדומים בגרף הבא:



○ קל לראות ששטח זה הוא: $1 - 2 \left(\frac{3}{4} \cdot \frac{3}{4} \cdot \frac{1}{2} \right) = \frac{7}{16}$, וזהו הסיכוי שייפגשו.

טענה: $P(\emptyset) = 0$ לכל פונקציית הסתברות.

הוכחה: $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$ ומהעברת אגפים, $P(\emptyset) = 0$. ■

טענה: $A \subseteq B \Rightarrow P(B) \geq P(A)$

הוכחה: $P(B) = P(B \cup (B \setminus A)) = P(A) + P(B \setminus A) \geq P(A)$. ■

טענה: $P(\bar{A}) = 1 - P(A)$

הוכחה: $1 = P(\Omega) = P(A \cup \bar{A}) = P(A) + P(\bar{A})$. ■

טענה: לכל שני מאורעות A, B מתקיים $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

הוכחה: ראשית, $P(A) = P(A \cap B) + P(A \cap \bar{B})$. באופן דומה, $P(B) = P(A \cap B) + P(\bar{A} \cap B)$. אם מחברים את שתי התוצאות מקבלים $P(A) + P(B) = 2P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B)$, כלומר:

$$P(A) + P(B) - P(A \cap B) = P(A \cap B) + P(A \cap \bar{B}) + P(\bar{A} \cap B) = P(A \cup B)$$

כפי שרצינו. ■

ניתן גם להכליל טענה זו למספר כלשהו של מאורעות, באמצעות נוסחת ההכללה וההדחה.

$$P(\cup_{i=1}^n A_i) = P(A_1) + P(\overline{A_1} \cap A_2) + \dots + P(\overline{A_1} \cap \dots \cap \overline{A_{n-1}} \cap A_n) \quad \text{טענה:}$$

הוכחה: אינדוקציה על n . המקרה $n = 3$, למשל, נובע מפירוק $A_1 \cup A_2 \cup A_3$ לאיחוד של שלושה מאורעות זרים:
 $P(A_1 \cup A_2 \cup A_3) = P(A_1 \cup (\overline{A_1} \cap A_2) \cup (\overline{A_1} \cap \overline{A_2} \cap A_3))$ המקרה הכללי:

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= P\left(\left(\bigcup_{i=1}^{n-1} A_i\right) \cup A_n\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P\left(\overline{\bigcup_{i=1}^{n-1} A_i} \cap A_n\right) = \\ &= P(A_1) + P(\overline{A_1} \cap A_2) + \dots + P(\overline{A_1} \cap \dots \cap \overline{A_{n-2}} \cap A_{n-1}) + P(\overline{A_1} \cap \dots \cap \overline{A_{n-1}} \cap A_n) \end{aligned}$$

כאשר המעבר הראשון הוא מהנחת האינדוקציה על $n = 2$ והמעבר השני הוא מהנחת האינדוקציה וחוק דה-מורגן המוכלל על הגורם $\overline{\bigcup_{i=1}^{n-1} A_i}$. ■

הסתברות מותנית

רוצים לשאול מה ההסתברות למאורע מסוים בהינתן שמאורע אחר כבר התרחש, כלומר בתנאי שמהו כבר נתון לנו. ההתנייה משנה את העולם ההסתברותי – את פונקציית ההסתברות.

הגדרה: בהינתן מאורעות A, B נסמן ב- $P(B|A)$ את ההסתברות של המאורע B בהינתן המאורע A . זוהי הסתברות מותנית. נגדיר: $P(B|A) = \frac{P(A \cap B)}{P(A)}$

טענה: הפונקציה $P(A|B) = \frac{P(A \cap B)}{P(B)}$ היא אכן פונקציית הסתברות.

הוכחה: ברור ששתי האקסיומות הראשונות מתקיימות. האקסיומה השלישית מתקיימת גם כן:

$$\begin{aligned} P(A_1 \cup A_2 | B) &= \frac{P((A_1 \cup A_2) \cap B)}{P(B)} = \frac{P((A_1 \cap B) \cup (A_2 \cap B))}{P(B)} = \\ &= \frac{P(A_1 \cap B) + P(A_2 \cap B)}{P(B)} = P(A_1 | B) + P(A_2 | B) \end{aligned}$$

■ כנדרש.

טענה (חוק הכפל או חוק השרשרת): $P(A \cap B) = P(A)P(B|A)$ וכן $P(\cap_{i=1}^n A_i) = P(A_1)P(A_2|A_1) \dots P(A_n|A_1 \cap \dots \cap A_{n-1})$

הוכחה: החלק הראשון בהעברת אגפים בהגדרת הסתברות מותנית, החלק השני באינדוקציה על n . ■

טענה (נוסחת ההסתברות השלמה): עבור זוג מאורעות A, B מתקיים $P(B) = P(A)P(B|A) + P(\overline{A})P(B|\overline{A})$

הוכחה: $P(B) = P((A \cap B) \cup (\overline{A} \cap B)) = P(A \cap B) + P(\overline{A} \cap B)$ ■

ניתן להסתכל על הנוסחא כפירוק המאורע B לשני עולמות – עולם שבו A קרה, ועולם שבו A לא קרה. ניתן גם להכליל את הנוסחא ל- n מאורעות זרים בזוגות $A_1, \dots, A_n$ שאיחודם Ω , ואז מתקבלת הנוסחא הכללית יותר, לפיה $P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(A_i)P(B|A_i)$

דוגמאות:

- מה ההסתברות שיצא 3 בקוביה, בהינתן שיצא מספר אי-זוגי?
 - אפשר לספור מאורעות אלמנטריים – מדובר במאורע 1 מתוך 3, ולכן התוצאה היא $\frac{1}{3}$. אפשר לקבל אותה תוצאה גם בעזרת ההגדרה לעיל על חיתוך מאורעות, כמובן.
- שתי קבוצות קידוד ברפא"ל התבקשו לכתוב את תוכנת ה"חץ" בתוך שנה. הסיכוי של A להצליח הוא $\frac{2}{3}$ והסיכוי של B להצליח הוא $\frac{1}{2}$. הסיכוי שלפחות אחת תצליח הוא $\frac{3}{4}$. בהינתן שקבוצה אחת בלבד הצליחה, מה ההסתברות שזו B?
 - המאורעות האפשריים הם $\Omega = \{(1,1), (1,0), (0,1), (0,0)\}$ המייצגים הצלחה/כישלון של כל אחת מהקבוצות. לפי הנתונים, $P(\{(1,1), (1,0)\}) = P(\{(1,1)\}) + P(\{(1,0)\}) = \frac{2}{3}$, וכך אפשר לבטא גם את שאר המידע. פותרים שלוש משוואות בשלושה נעלמים ומקבלים $P(\{(1,0)\}) = \frac{1}{4}$, $P(\{(1,1)\}) = \frac{5}{12}$, $P(\{(0,1)\}) = \frac{1}{12}$, $P(\{(0,0)\}) = \frac{1}{4}$.
 - כעת נשתמש בהסתברות מותנית ונקבל את התוצאה המבוקשת:

$$P(\{(0,1), (1,1)\} | \{(1,0), (0,1)\}) = \frac{P(\{(0,1)\})}{P(\{(0,1), (1,0)\})} = \frac{\frac{1}{12}}{\frac{1}{4} + \frac{1}{12}} = \frac{1}{4}$$
- תוכנת אבטחה מתריעה כשיש חדירה של וירוס בסיכוי של 0.99, וכשאין וירוס בסיכוי של 0.1. הסיכוי לוירוס הוא 0.05. מה הסיכוי שיהיה במערכת וירוס והמערכת לא תתריע על כך?
 - נסמן ב-A את המאורע שוירוס חדר למערכת, וב-B את המאורע שהייתה התרעה. אם כן,

$$P(A \cap \bar{B}) = P(A)P(\bar{B}|A) = P(A)(1 - P(B|A)) = 0.05 \times 0.01 = 0.0005$$
- משחק הדלתות: במשחק שלוש דלתות. בשלב הראשון שמים פרס מאחורי דלת ודחליל מאחורי האחרות. המתחרה בוחר דלת באקראי. בשלב השני המנחה פותח דלת שמאחוריה דחליל שאינה זו שהמתחרה בחר. בשלב השלישי נותנים למתחרה אפשרות להחליף את הדלת – מה כדאי לו לעשות?
 - ברור שההסתברות להצליח בהינתן שלא משנים את הדלת היא $\frac{1}{3}$. השאלה מה ההסתברות להצליח בהינתן שמשנים את הדלת. נניח שהפרס הוא מאחורי דלת מס' 1, ונכתוב את כל האפשרויות לפי נוסחת ההסתברות השלמה:

$$P(\$) = P(\$|1)P(1) + P(\$|2)P(2) + P(\$|3)P(3) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = \frac{2}{3}$$
 - באופן דומה מתקבלת תוצאה של $\frac{2}{3}$ כאשר הפרס הוא מאחורי כל דלת אחרת, ולכן ההסתברות להצליח אם מחליפים דלת היא $\frac{2}{3}$ – וזה הדבר שכדאי לעשות.
- פרדוקס ימי ההולדת (Birthday Paradox): בכיתה של 40 תלמידים, מה הסיכוי שלשניים או יותר יום הולדת באותו יום (תחת ההנחה שהסיכוי להיוולד ביום מסוים שווה לכל יום אחר)?
 - נסדר את התלמידים שרירותית ונסמן A_1 – המאורע שיום ההולדת של התלמיד הראשון לא חופף עם קודמיו (הסתברות מאורע זה היא 1), נסמן A_2 – המאורע שיום ההולדת של השני לא חופף עם קודמיו (הסתברות מאורע זה היא $\frac{364}{365}$), וכו'. אנו מחפשים את ההסתברות של המאורע $A_1 \cap \dots \cap A_{40}$.
 - לפי חוק הכפל מקבלים:

$$P(\overline{A_1 \cap \dots \cap A_{40}}) = 1 - P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \dots P(A_{40}|A_1 \cap \dots \cap A_{39}) = 1 - \frac{364}{365} \cdot \frac{363}{365} \dots \frac{326}{365} \approx 0.89$$

הגדרה: נאמר ש-B נותן עדות חיוביות ל-A אם $P(A|B) \geq P(A)$.

טענה: A נותן עדות חיובית ל-B אם B נותן עדות חיובית ל-A.

הוכחה: $P(A|B) \geq P(A) \Rightarrow \frac{P(A \cap B)}{P(B)} \geq P(A) \Rightarrow \frac{P(A \cap B)}{P(A)} \geq P(B) \Rightarrow P(B|A) \geq P(B)$ ■

טענה (חוק בייס – Bayes): $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

בניסוח אחר, באמצעות נוסחת ההסתברות השלמה: $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}$

הוכחה: מיידית מההגדרה, $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{P(A \cap B)}{P(A)} P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$ ■

ניתן להכליל את החוק עבור כל אוסף מאורעות $A_1, \dots, A_n$ המהווים חלוקה של Ω (זרים בזוגות שאיחודם Ω) ולקבל את הניסוח הבא:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

דוגמאות:

- מבחן למחלה נדירה מחזיר תשובה חיובית (רעה) ב-95% מהמקרים אם יש את המחלה, ותשובה שלילית בסיכוי 95% אם אין את המחלה. שכיחות המחלה באוכלוסיה היא 0.1%. מה איכות המבחן?
 ○ נסמן כעת B – המאורע התשובה חיובית, A – המאורע שיש מחלה, ונמצא את ההסתברות שיש מחלה בהינתן שהתשובה הייתה חיובית:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})} = \frac{0.95 \times 0.001}{0.95 \times 0.001 + 0.05 \times 0.999} \cong 0.0187$$

- וזו תוצאה גרועה ממש – המבחן מדייק בזיהוי המחלה בפחות מ-2% מהמקרים!
- **פרדוקס סימפסון:** במחקר להשוואת שני טיפולים בכליות נמצא שטיפול א' מצליח בסיכוי של 78% וטיפול ב' מצליח בסיכוי של 83%.
 ○ מידע נוסף – בטיפול א' הייתה הצלחה של 93% בטיפול באבנים קטנות והצלחה של 73% בטיפול באבנים גדולות, ואילו בטיפול ב' הייתה הצלחה של 87% בטיפול באבנים קטנות והצלחה של 69% בטיפול באבנים גדולות.
 ○ התוצאות נראות כאילו טיפול א' דווקא עדיף! – אבל העניין הוא שבטיפול א' היו הרבה יותר אנשים עם אבנים גדולות, מה שגרם לטיפול ב' להיראות מוצלח יותר באופן אבסולוטי. נסתכל על טיפול א' (הצלחה) H , אבנים קטנות S , אבנים גדולות B :

$$0.78 = P(H) = P(H|S)P(S) + P(H|B)P(B) = 0.93 \cdot p + 0.73 \cdot (1 - p) \Rightarrow p = 0.25$$

 ○ כלומר, בטיפול א' היו ל-75% אבנים גדולות, ובאופן דומה בטיפול ב' היו 22% בלבד עם אבנים גדולות. זו הסיבה להטיית התוצאות האבסולוטיות.

הגדרה: אם מתקיים $P(A|B) = P(B)$ אז נאמר שהמאורע A בלתי-תלוי במאורע B . בניסוח אחר, נאמר $P(A \cap B) = P(A)P(B)$. לפעמים גם נסמן $A \perp B$.

אינטואיטיבית, התרחשות המאורע B לא משפיע על ההסתברות המאורע A . אי-תלות היא סימטרית – אם A בלתי-תלוי ב- B , אזי גם B בלתי-תלוי ב- A (ישירות מהגדרת הסתברות מותנית), ולכן פשוט נאמר שהמאורעות בלתי-תלויים.

טענה: $A \perp B \Rightarrow A \perp \bar{B}$

הוכחה: $P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A)P(B) + P(A \cap \bar{B})$ וכעת מעבירים אגפים ומקבלים $P(A \cap \bar{B}) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$ ■ כנדרש.

הגדרה: נאמר שהמאורעות A, B הם בלתי-תלויים בהינתן C אם $P(A \cap B|C) = P(A|C)P(B|C)$.

דוגמאות:

- מאורעות זרים אינם בלתי-תלויים (כמעט תמיד): אם אחד מהם קרה, אז השני בוודאי לא קרה, ולכן הם משפיעים זה על זה. האפשרות היחידה שבה הם בלתי-תלויים היא המקרה המנוון $P(A) = P(B) = 0$.
- בכד 10 כדורים ממוספרים, ומוציאים מתוכו 3 כדורים ללא החזרה. A המאורע שהכדור הראשון הוא 1, ו- B המאורע שהכדור השני הוא 3. המאורעות A, B אינם בלתי-תלויים כיוון שההוצאה היא ללא החזרה.
- מסובבים סביבון הוגן פעמיים. ההסתברות לכל אפשרות היא $\frac{1}{16}$. נגדיר A_i – המאורע שבסיבוב הראשון יצא i , ו- B_j – המאורע שבסיבוב השני יצא j . האם A_i ו- B_j בלתי-תלויים לכל i, j ?
 ○ מהנתון, $P(A_i \cap B_j) = \frac{1}{16}$. מספירה, $P(A_i) = P((A_i \cap B_1) \cup \dots \cup (A_i \cap B_4)) = \frac{1}{4}$ וכך גם לגבי B_j . לכן מתקיים $P(A_i \cap B_j) = \frac{1}{16} = \frac{1}{4} \cdot \frac{1}{4} = P(A_i)P(B_j)$ ואכן המאורעות הם ב"ת.
- מסובבים סביבון הוגן פעמיים. A – המקסימום בשני הסיבובים הוא 2, B – המינימום בשני הסיבובים הוא 2. האם המאורעות בלתי-תלויים?

- כאן $P(A \cap B) = \frac{1}{16}$, ואילו $P(A) = \frac{3}{16}$. אפילו לא צריך לחשב את $P(B)$ כעת, כי זה לא יכול לצאת $\frac{1}{3}$ על פני 16 מאורעות אלמנטריים שווים הסתברות. מספירה, $P(A)P(B) = \frac{15}{16^2} \neq \frac{1}{16}$.
- מטילים פעמיים מטבע, שתי ההטלות בלתי-תלויות. נסמן H_1 – יצא ראש בהטלה הראשונה, כנ"ל H_2 , ו- D – בשתי ההטלות יצא משהו שונה. מהנתון, $H_1 \perp H_2$ – אך האם $H_1 \perp H_2 | D$?
 - נסתכל על ההסתברויות: $P(H_1 | D) = \frac{1}{2} = P(H_1)$, כלומר $H_1 \perp D$ ובאופן דומה גם $H_2 \perp D$. אלא ש- $P(H_1 \cap H_2 | D) = 0$ כי לא ייתכן שהיו שתי הטלות שיצא בכל אחת משהו אחר ובשתייהן יצא ראש, ולכן אין כאן אי-תלות מותנית.
- נתונים שני מטבעות, אחד עם הסתברות 0.99 לראש והשני עם הסתברות 0.01 לראש. בוחרים בהסתברות שווה מטבע (א' או ב') ואז מטילים שתי הטלות בלתי-תלויות. יהי A המאורע שנבחר מטבע א', והמאורעות H_1, H_2 כמו בסעיף הקודם.
 - מהנתון, $H_1 \perp H_2 | A$.
 - האם $H_1 \perp H_2$? ובכן, $P(H_1) = P(H_1 | A)P(A) + P(H_1 | \bar{A})P(\bar{A}) = 0.5$, מצד שני, מקבלים $P(H_2) = 0.5$. מכאן $P(H_1 \cap H_2) = P(H_1 \cap H_2 | A)P(A) + P(H_1 \cap H_2 | \bar{A})P(\bar{A}) = 0.5 \times 0.99^2 + 0.5 \times 0.01^2 \approx 0.5$. כלומר כאן לא מתקיים $P(H_1 \cap H_2) = P(H_1)P(H_2)$ ולכן המאורעות האלה אינם בלתי-תלויים.

הגדרה: נאמר שהמאורעות $A_1, \dots, A_n$ בלתי-תלויים אם לכל $S \subseteq \{1, \dots, n\}$ מתקיים $P(\cap_{i \in S} A_i) = \prod_{i \in S} P(A_i)$.

הסיבה שאנו דורשים את זה לכל תת-קבוצה ולא משהו חלש יותר היא שאנו לא רוצים "לפספס" תלות. למשל, אם היינו דורשים רק אי-תלות בזוגות, ייתכן שצירוף של שלושה מאורעות היה יוצר תלות. לחלופין, אם היינו דורשים רק אי-תלות של כל הקבוצה, ייתכן שזוג מאורעות היה יוצר תלות שהייתה נעלמת עם מאורעות אחרים בקבוצה.

הגדרה: ניסוי ברנוולי (Bernoulli) הוא ניסוי עם שתי תוצאות אפשריות. נקרא לאחת מהן "הצלחה" ונסמן את ההסתברות להצלחה ב- p .

דוגמאות:

- ספק אינטרנט מתקין C מודמים המיועדים לשרת N אנשים, כאשר $N > C$. בכל זמן נתון לקוח מתחבר בהסתברות p באופן ב"ת מהאחרים. מה ההסתברות שיותר מ- C לקוחות יתחברו (נסמן מאורע זה ב- A)?
 - נחשב ישירות כאשר אנו משתמשים בהגדרת אי-תלות על קבוצת המאורעות של כל הלקוחות:

$$P(A) = \sum_{k=C+1}^N P(k) = \sum_{k=C+1}^N \binom{N}{k} p^k (1-p)^{N-k}$$
 למשל, עבור $N = 100, C = 15, p = 0.1$ מקבלים $P(A) \approx 0.0399$.
- נניח שהשמש זורחת בכל יום בהסתברות p לא ידועה. ההנחה היא שכל יום הוא ניסוי ברנוולי, כאשר p נבחר אקראית מהאוסף $\{0, \frac{1}{m}, \dots, \frac{m-1}{m}, \frac{m}{m}\}$. רוצים להשתמש בניסיון של 5,000 שנים כדי לקבוע את p .
 - נסמן ב- R_N את המאורע שהשמש זרחה N פעמים, וב- T את המאורע שהשמש תזרח מחר (בפעם ה- $N+1$). רוצים למצוא את $P(T | R_N)$, ולמעשה מספיק למצוא את R_N כי $P(T | R_N) = \frac{P(T \cap R_N)}{P(R_N)} = \frac{P(R_{N+1})}{P(R_N)}$.
 - נשתמש בנוסחת ההסתברות השלמה על כל ערכי p האפשריים:

$$P(R_N) = \sum_{i=0}^m P\left(R_N \mid p = \frac{i}{m}\right) P\left(p = \frac{i}{m}\right) = \sum_{i=0}^m \frac{1}{m+1} \left(\frac{i}{m}\right)^N = \frac{1}{m+1} \frac{1}{m^N} \sum_{i=0}^m i^N \approx \frac{1}{m+1} \frac{1}{m^N} \int_0^m i^N di = \frac{1}{m+1} \frac{1}{m^N} \frac{m^{N+1}}{N+1} = \frac{m}{m+1} \frac{1}{N+1}$$
 - כעת כאשר $m \rightarrow \infty$ מקבלים $P(R_N) = \frac{1}{N+1}$ ו- $P(T | R_N) = \frac{N+1}{N+2}$. לסיכום, ככל שהשמש זורחת יותר ימים ברציפות, אנו נעשים יותר ויותר משוכנעים שהיא תזרח מחר.
- לאבי ולביתיה $2n+1$ מטבעות. אבי מטיל n מטבעות ובתיה מטילה $n+1$ מטבעות. המטבעות הוגנים, וההטלות הן בלתי-תלויות. מה ההסתברות שלביתיה יהיה יותר "עץ" מאשר לאבי?
 - נסמן B – לבתיה יש יותר "עץ" מלאבי. נסתכל על n ההטלות הראשונות שלהם (ולביתיה נותרה עוד הטלה אחת) ונסמן: A_1 – לבתיה יותר "עץ", A_2 – לאבי יותר "עץ", A_3 – לשניהם אותו מספר הטלות "עץ". נחשב את ההסתברויות של המאורעות:

$$P(A_1) = P(A_2), \quad P(A_3) = 1 - P(A_1) - P(A_2) = 1 - 2P(A_1)$$

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) = 1 \cdot P(A_1) + 0 \cdot P(A_2) + \frac{1}{2} \cdot (1 - 2P(A_1)) = \frac{1}{2}$$

משתנים מקריים בדידים

הגדרה: משתנה מקרי הוא פונקציה $\mathbb{R} \rightarrow \Omega$. תומך זהו אוסף הערכים שהמ"מ יכול לקבל (תמונת הפונקציה).

נשתמש במשתנים מקריים כדי לסנן מתוך מרחב המדגם (שיכול להיות עצום ומפחיד) את המספרים שמעניינים אותנו שנוכל לבצע עליהם חישובים בקלות רבה יותר. בהינתן משתנה מקרי X , האמירה " $X = 1$ " מגדירה מאורע – זהו אוסף המאורעות $\mathcal{F}$ כך ש- $X(A) = 1$. $\forall A \in \mathcal{F}, X(A) = 1$ נשים לב גם שלכל $x \neq y$, המאורעות $X = x, X = y$ הם מאורעות זרים.

דוגמאות:

- בניסוי מטילים שני מטבעות, $\Omega = \{TT, TH, HT, HH\}$. נגדיר משתנה מקרי הסופר את מספר הפעמים שיצא T בשתי הטלות: $X: TT \mapsto 2, TH \mapsto 1, HT \mapsto 1, HH \mapsto 0$.
- מסובבים שני סביבונים. נגדיר X – מספר התוצאות הזוגיות. התומך הוא $\{0, 1, 2\}$.
- מבצעים N הטלות מטבע, ורואים רק את מספר ה- H שיצאו: $\Omega = \{0, 1, \dots, N\}$. נגדיר X מ"מ המקבל 1 אם מספר ה- H זוגי, ו-0 אחרת. כעת X זהו ניסוי ברנולי.
- מסובבים שני סביבונים. נגדיר X – המינימום שיצא בשני הסיבובים. התומך: $\{1, 2, 3, 4\}$.
- זורקים חץ על הישר $[0, 1]$. התוצאות האפשריות: $\Omega = [0, 1]$. נגדיר מ"מ X המקבל 1 אם פגענו בין 0 ל-0.5, ו-0 אחרת.

הגדרה: פונקציית התפלגות (Probability Mass Function או Probability Distribution) של מ"מ X זו הפונקציה

$$P_X(x) := P(X = x) = \sum_{\omega \in \Omega, X(\omega)=x} P(\omega)$$

כמובן, תמיד מתקיים $\sum_x P(X = x) = 1$, כיוון ש- P היא פונקציית הסתברות ו- $\cup_x (X = x) = \Omega$.

משתנים מקריים טיפוסיים

הגדרה: משתנה מקרי אחיד מסומן $X \sim U(a, b)$ עם תומך $\{a, a + 1, \dots, b - 1, b\}$ אם פונקציית ההתפלגות שלו

$$\text{מקיימת } P_X(x) = \frac{1}{b-a+1} \text{ לכל } x \in \{a, a + 1, \dots, b - 1, b\}$$

דוגמאות:

- המספר שיוצא בקוביה הוגנת.
- המספר שיוצא בכדור הראשון בלוטו ($a = 1, b = 46$).
- לוטו עם 10 מספרים בכל כדור, X_1 – המספר שיצא בכדור הראשון ($X_1 \sim U(1, 10)$).

הגדרה: משתנה מקרי ברנולי מסומן $X \sim \text{Ber}(p)$ עם תומך $\{0, 1\}$ אם פונקציית ההתפלגות שלו מקיימת

$$P(X = 1) = p, P(X = 0) = 1 - p$$

בכתיב אחר (שנשתמש בו בהמשך), $P(X = x) = p^x (1 - p)^{1-x}$.

הגדרה: משתנה מקרי בינומי המסומן $X \sim \text{Bin}(p, n)$ הוא מספר ההצלחות ב- n ניסויי ברנולי ב"ת בעלי הסתברות

זוהי p לכל אחד. התומך הוא $\{0, 1, \dots, n\}$ ופונקציית ההתפלגות היא $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$.

פורמלית, יש להראות שזוהי אכן פונקציית הסתברות. כרגיל, החלק המאתגר הוא להראות שההסתברויות מסתכמות ל-1 על פני כל ערכי x , וזה נובע טריוויאלית מנוסחת הבינום (כי $p + (1 - p) = 1$ ולכן זהו למעשה פיתוח של 1^n).

הגדרה: משתנה מקרי היפר-גיאומטרי מסומן $X \sim HG(k, n, m)$ הוא בעל תומך $\{0, 1, \dots, m\}$ ופונקציית התפלגות

$$P(X = x) = \frac{\binom{m}{x} \binom{n}{k-x}}{\binom{n+m}{k}}$$

זהו מ"מ המייצג את המצב שבו יש שתי קבוצות פריטים (בגודל n ובגודל m) ואנו מבצעים k ניסויים שבהם בוחרים פריטים מהקבוצות, ורוצים לדעת את ההסתברות שייבחרו x פריטים מהקבוצה בגודל m .

הגדרה: משתנה מקרי גיאומטרי המסומן $X \sim Geo(p)$ הוא מספר ניסויי ברנולי ב"ת בעלי הסתברות זהה p עד להצלחה, כולל. התומך הוא $\{1, 2, \dots\}$ ופונקציית ההתפלגות היא $P(X = x) = (1 - p)^{x-1} p$.

פורמלית, יש להראות שזוהי אכן פונקציית הסתברות. אכן, מסכום סדרה הנדסית אינסופית מתכנסת מקבלים:

$$\sum_{x=1}^{\infty} p(1-p)^{x-1} = p \cdot \frac{1}{1-(1-p)} = 1$$

הגדרה: ההסתברות לשרוד (Survival Function) על מ"מ $X \sim Geo(p)$ זוהי ההסתברות $P(X > s)$, ומתקיים:

$$P(X > s) = 1 - P(X \leq s) = 1 - \sum_{x=1}^s P(X = x) = 1 - \sum_{x=1}^s p(1-p)^{x-1} = (1-p)^s$$

טענה (תכונת חוסר הזיכרון של מ"מ גיאומטרי): יהי $Y \sim Geo(p)$, אזי $P(Y > s + t | Y \geq t) = P(Y > s)$.

הוכחה:

$$\begin{aligned} P(Y > s + t | Y \geq t) &= \frac{P((Y > s + t) \cap (Y \geq t))}{P(Y \geq t)} = \frac{P(Y > s + t)}{P(Y \geq t)} = \frac{P(Y > s + t)}{P(Y > t - 1)} = \\ &= \frac{(1-p)^{s+t+1}}{(1-p)^{t-1+1}} = (1-p)^{s+1} = P(Y > s) \end{aligned}$$

■ כנדרש.

הגדרה: משתנה מקרי פואסוני (Poisson) המסומן $X \sim Pois(\lambda)$ הוא מספר מאורעות ביחידת זמן המתרחשים בקצב היסטורי של $\lambda > 0$. התומך הוא $\{0, 1, 2, \dots\}$ ופונקציית ההתפלגות היא $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$.

זהו משתנה מקרי המשמש לעתים קרובות לתכנון וניתוח קצב או קצב הגעה (למשל, מספר פצצות בקמ"ר של לונדון במלחמת העולם השנייה, או מספר טלפונים למרכזייה בשעה).

נראה שזו אכן פונקציית התפלגות, שוב האתגר הוא להראות הסתכמות ל-1, ונשתמש לשם כך בתוצאה מאינפי:

$$\sum_x P(X = x) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \left[1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \dots \right] = e^{-\lambda} e^{\lambda} = 1$$

טענה: יהי $X \sim Bin(p, n)$. הגבול $\lim_{n \rightarrow \infty} P(X = x)$ זהה להסתברות $P(Y = x)$ עבור $Y \sim Pois(np)$.

הוכחה: נסמן $\lambda = np$. נתחיל מ- $X = 0$, אכן מתקיים: $P(X = 0) = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\lambda}$. כעת נסתכל על המקרה הכללי:

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{n(n-1) \dots (n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} = \end{aligned}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{n^k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}$$

כעת נתבונן בכל גורם בנפרד כאשר $n \rightarrow \infty$: $\frac{\lambda^k}{k!}$ זהו קבוע, ואינו מושפע מכך. הגורם $\frac{n(n-1)\cdots(n-k+1)}{n^k}$ שואף ל-1 בגלל כל נכפל בנפרד. לבסוף, $\left(1 - \frac{\lambda}{n}\right)^{-k}$ שואף ל-1 ונותרנו עם $\left(1 - \frac{\lambda}{n}\right)^n$ שכמו קודם, שואף ל- $e^{-\lambda}$. מכאן מקבלים: $\lim_{n \rightarrow \infty} P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ כפי שרצינו. ■

טענה זו מראה את הקשר בין התפלגות בינומית להתפלגות פואסונית – במובן מסוים, התפלגות פואסונית היא מודל טוב להתפלגות בינומית שבה מספר רב מאוד של ניסויים והסתברות נמוכה להצלחה בכל ניסוי.

תוחלת של משתנה מקרי

הגדרה: יהי X מ"מ. התוחלת של X היא: $E[X] = \sum_x P(X = x)x$.

נתייחס רק לתוחלות קיימות, כלומר כאלה שעבורן הטור מתכנס בהחלט. זה לא בהכרח המצב. כעת נחשב תוחלות של משתנים מקריים מסוגים שכבר ראינו, שלכולם התוחלת קיימת.

דוגמאות:

- במשחק מזל, מקבלים 5.5 ש"ח בהסתברות 0.1, 3 ש"ח בהסתברות 0.5 וכלום בהסתברות 0.4. עלות ההשתתפות היא 2.5 ש"ח. מהי תוחלת הרווח? – חישוב ממוצע משוקלל.
- התוחלת של ערך ההטלה בקוביה היא 3.5.
- התוחלת של ערך ההטלה בקוביה מוטה עם $P(6) = 0.99$ היא מאוד קרובה ל-6.

טענה: יהי $X \sim \text{Pois}(\lambda)$. אזי $E[X] = \lambda$.

הוכחה:

$$\begin{aligned} E[X] &= \sum_{x=0}^{\infty} P(X = x)x = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} x = \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^x}{(x-1)!} = \\ &= \lambda \sum_{x=1}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} = \lambda \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^{x-1}}{(x-1)!} = \lambda \end{aligned}$$

כאשר השתמשנו בכך שכבר ראינו שפונקציית ההתפלגות של מ"מ פואסוני מסתכמת ל-1 על כל ערכי התומך. ■

טענה: יהי X מ"מ ונגדיר $Y = g(X)$ כאשר $Y: \mathbb{R} \rightarrow \mathbb{R}$ פונקציה כלשהי. מתקיים: $E[Y] = \sum_x g(x)P(X = x)$.

הוכחה: נתייחס ל- Y כמשתנה מקרי ונחשב את התוחלת שלו:

$$\begin{aligned} E[Y] &= \sum_y yP(Y = y) = \sum_y yP(X = x, g(x) = y) = \sum_y y \sum_{x: g(x)=y} P(X = x) = \\ &= \sum_y \sum_{x: g(x)=y} g(x)P(X = x) = \sum_x g(x)P(X = x) \end{aligned}$$

■ **כנדרש.**

דוגמאות:

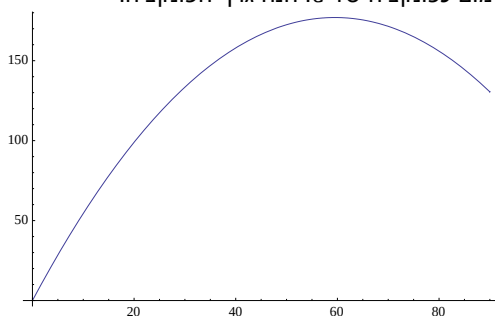
- מוכר פלאפל מכין מנות בתחילת היום. מספר הלקוחות הוא $X \sim U(1, 90)$. כל מנה עולה 3 ש"ח ונמכרת ב-9 ש"ח. כמה מנות על המוכר להכין כדי למקסם את תוחלת הרווח?

נסמן Y – הרווח, C מספר המנות שנכין, וכאן $Y = 9 \min\{X, C\} - 3C$.

כעת ניתן לחשב את התוחלת של Y ולמצוא את מספר המנות שיש להכין:

$$E[Y] = \sum_{x=1}^{90} \frac{(9 \min\{x, C\} - 3C)}{90} = \frac{1}{90} \left[9 \sum_{x=1}^{90} \min\{x, C\} - 270C \right] = \frac{1}{90} \left[9 \sum_{x=1}^C x + 9 \sum_{x>C}^90 C - 270C \right]$$

לביטוי שהתקבל אפשר למצוא מינימום כפונקציה של C . הנה גרף הפונקציה:



- חנות אלקטרוניקה מקבלת בכל שבוע משלוח של 40 מכשירים מהמפעל. ההסתברות לקבל מוצר פגום היא 7%, באופן ב"ת מהמוצרים האחרים. מהי ההסתברות שבמשלוח יהיו לפחות 4 מוצרים פגומים?

נגדיר מ"מ $X \sim \text{Bin}(40, 0.07)$ המייצג את מספר הפריטים הפגומים, ועלינו למצוא את ההסתברות הבאה:

$$P(X \geq 4) = 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \approx 0.306$$

את ההסתברות הזאת אנו יודעים לחשב לפי פונקציית ההתפלגות של מ"מ בינומי.

כעת נניח שמגיע משלוח, בחנות בוחרים 10 מוצרים ללא החזרה, ואם נמצא מכשיר לא תקין – מחזירים את המשלוח.

נשלחו 10 משלוחים שבכל אחד 3 מוצרים פגומים, מהי ההסתברות שבדיוק חצי יוחזרו?

נגדיר כאן מ"מ $Y \sim \text{HG}(10, 37, 3)$ מספר המוצרים הפגומים בבדיקה. ההסתברות שמשלוח אחד יוחזר מתקבלת לפי

$$P(Y > 0) = 1 - P(Y = 0) \approx 0.59$$

$$P(T = 5) = \binom{10}{5} 0.59^5 0.41^{10-5} \approx 0.21$$

- מבצעים n הטלות ב"ת של מטבע עם הסתברות p . מהי ההסתברות שמספר ה"עץ" יהיה זוגי?

נגדיר מ"מ $X \sim \text{Bin}(n, p)$ הסופר את מספר ה"עצים" ואנו מחפשים את ההסתברות שערך יהיה זוגי. כעת נכתוב נוסחת

נסיגה להסתברות הרצויה. נסמן P_n ההסתברות למס' זוגי של "עץ" לאחר n הטלות. מקרה הבסיס הוא 0 הטלות, ושם

$P_0 = 1$. נוסחת הנסיגה היא: $P_n = P_{n-1}(1-p) + (1-P_{n-1})p = p + (1-2p)P_{n-1}$. ניתן להוכיח באינדוקציה ש-

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

דרך אחרת – נפתח ישירות:

$$P_n = \frac{1}{2} \left[\sum_{i=0}^n \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=0}^n (-1)^i \binom{n}{i} p^i (1-p)^{n-i} \right] = \frac{1}{2} \left[1 + \sum_{i=0}^n \binom{n}{i} (-p)^i (1-p)^{n-i} \right] = \frac{1}{2} [1 + (-p + (1-p))^n] = \frac{1}{2} (1 + (1-2p)^n)$$

$$E[X] = \frac{b+a}{2} \text{ טענה: יהי } X \sim U(a, b) \text{ אזי}$$

הוכחה: נחשב ישירות:

$$E[X] = \sum_{x=a}^b \frac{x}{b-a+1} = \frac{1}{b-a+1} \sum_{x=a}^b x = \frac{1}{b-a+1} \frac{1}{2} (b-a+1)(b+a) = \frac{b+a}{2}$$

■ כנדרש.

טענה: יהי $X \sim \text{Bin}(n, p)$, אזי $E[X] = np$.

הוכחה: נחשב ישירות:

$$\begin{aligned} E[X] &= \sum_{x=0}^n x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} = \\ &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} (1-p)^{n-x} = np \sum_{x=0}^n \frac{(n-1)!}{x!(n-x-1)!} p^x (1-p)^{n-x-1} = \\ &= np \sum_{x=0}^n \frac{(n-1)!}{x!((n-1)-x)!} p^x (1-p)^{(n-1)-x} = np \end{aligned}$$

■ כנדרש.

טענה: יהי X מ"מ אי-שלילי, אזי $E[X] = \sum_{x=0}^{\infty} P(X > x)$.

הוכחה:

$$\begin{aligned} \sum_{x=0}^{\infty} P(X > x) &= \sum_{x=1}^{\infty} P(X \geq x) = [P(X=1) + P(X=2) + \dots] + [P(X=2) + P(X=3) + \dots] + \dots = \\ &= \sum_{x=1}^{\infty} xP(X=x) = E[X] \end{aligned}$$

■ כנדרש.

טענה (ליניאריות התוחלת): יהי X משתנה מקרי, אזי $E[aX + b] = aE[X] + b$.

הוכחה: נחשב ישירות:

$$E[aX + b] = \sum_x (ax + b)P(X=x) = a \sum_x xP(X=x) + b \sum_x P(X=x) = aE[X] + b$$

■ כנדרש.

לא נכון באופן כללי ש- $E[g(X)] = g(E[X])$ עבור פונקציה $g: \mathbb{R} \rightarrow \mathbb{R}$ כלשהי.

דוגמאות:

- אם מזג האוויר הוא טוב (בהסתברות 0.6), אז שולה הולכת 2 ק"מ לכיתה במהירות 5 קמ"ש; אחרת, שולה משתמשת באופנוע (באותו מסלול) במהירות 30 קמ"ש. מהו זמן ההגעה הממוצע של שולה?
 - נסמן אותו ב- T , ונשים לב ש- $P\left(T = \frac{2}{5}\right) = 0.6$ ו- $P\left(T = \frac{2}{30}\right) = 0.4$. התוחלת מתקבלת משני הערכים האפשריים,

$$E[T] = \frac{2}{5} \cdot 0.6 + \frac{2}{30} \cdot 0.4 = \frac{4}{15}$$
 - נסמן כעת ב- V את המהירות של שולה, ומתקיים $E[V] = 0.6 \cdot 5 + 0.4 \cdot 30 = 15$.
 - אמנם, הדרך = זמן כפול המהירות, אך עבור התוחלת זה לא נכון!
- בחידון שתי שאלות. הסיכוי לענות על שאלה 1 נכון הוא 0.8 והפרס הוא \$100, הסיכוי לענות על שאלה 2 הוא 0.5 והפרס הוא \$200. אחרי שנכשלים בשאלה, אי אפשר להמשיך. על איזו שאלה כדאי לענות קודם?

- נסתכל על המקרה הכללי כאשר p_1, p_2 ההסתברויות ו- v_1, v_2 ערכי הזכייה. נסתכל על תוחלת הרווח בכל אסטרטגיה:

$$E[1] = p_1(1 - p_2)v_1 + p_1p_2(v_1 + v_2)$$

$$E[2] = p_2(1 - p_1)v_2 + p_1p_2(v_1 + v_2)$$
- ברור שההבדל נקבע ע"י הגורם הראשון. במקרה הספציפי של נתוני השאלה, עדיף לענות קודם על השאלה הקלה יותר.
- **פרדוקס סנט-פטרסבורג:** במשחק מזל מטילים מטבע הוגן עד שיוצא T . מקבלים 2^n דולרים כאשר n הוא מספר ההטלות (הכולל). כמה אתם מוכנים לשלם עבור ההשתפות במשחק?
 - נחשב את תוחלת הרווח במשחק:

$$E[\$] = 2^1 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} + \dots = 1 + 1 + \dots$$
 - הטור לא מתכנס – כביכול, ב"ממוצע" מקבלים החזר של ∞ , אז כדאי להמר בכל סכום. ה"פרדוקס" הוא ההבדל בין התוחלת לבין מה שסביר – התוחלת מייצג את הרווח על פני הרבה משחקים, בעוד שאם נהמר \$1024, הסיכוי שנחזיר את מה שהימרנו במשחק אחד הוא בסך הכל $\frac{1}{1024}$.

שוונות של משתנים מקריים

הגדרה: השוונות של מ"מ X מסומנת ומוגדרת $Var(X) := E[(X - E[X])^2]$. סטיית התקן של X מסומנת ומוגדרת $Std(X) = \sqrt{Var(X)}$.

השוונות היא מדד לפיזור של המשתנה המקרי סביב התוחלת שלו. מכאן הרצון לחסר מהמשתנה המקרי את התוחלת שלו, כאשר הריבוע נועד להבטיח שלכל ערך של המ"מ שרחוק מהתוחלת, אנו מוסיפים לסטייה – אילו היינו לוקחים $E[X - E[X]]$, הסטיות היו מבטלות זו את זו והיינו מקבלים תמיד שוונות 0.

הגדרה: המומנט ה- k של מ"מ X הוא $E[X^k]$.

טענה: $Var(X) = E[X^2] - E[X]^2$

הוכחה: נחשב ישירות:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] = E[X^2 - 2E[X]X + E[X]^2] = \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 = E[X^2] - E[X]^2 \end{aligned}$$

■ כנדרש.

טענה: יהי X משתנה מקרי, אזי $Var(aX + b) = a^2 Var(X)$.

הוכחה: נחשב ישירות:

$$Var(aX + b) = E[(aX + b) - E[aX + b]]^2 = E[a^2(X - E[X])^2] = a^2 Var(X)$$

■ כנדרש.

בפרט, קיבלנו שהזזה בקבוע של מ"מ לא משנה את השוונות שלו – וזה אכן הגיוני, כי הזזה בקבוע משנה את התוחלת אבל לא יכולה לשנות את הפיזור שהיה קיים במילא סביב התוחלת.

טענה: יהי $X \sim \text{Bin}(n, p)$, אזי $Var(X) = np(1 - p)$.

הוכחה: **הרבה אלגברה מגעילה.** ■

טענה: יהי $X \sim U(a, b)$, אזי $Var(X) = \frac{(b-a+1)^2 - 1}{12}$.

הוכחה: ראשית נחשב את השונות של $Y \sim U(1, b)$ ואז נסתכל עליו כהזזה של X :

$$\begin{aligned} \text{Var}(Y) &= E[Y^2] - E[Y]^2 = \sum_{x=1}^b \frac{x^2}{b} - \left(\frac{b+1}{2}\right)^2 = \frac{1}{b}(1^2 + 2^2 + \dots + b^2) - \left(\frac{b+1}{2}\right)^2 = \\ &= \frac{1}{6} \frac{1}{b} b(b+1)(2b+1) - \left(\frac{b+1}{2}\right)^2 = \frac{1}{6}(b+1)(2b+1) - \frac{1}{4}(b+1)^2 = \\ &= (b+1) \left(\frac{1}{6}(2b+1) - \frac{1}{4}(b+1) \right) = (b+1) \left(\frac{b}{3} + \frac{1}{6} - \frac{b}{4} - \frac{1}{4} \right) = \frac{b^2 - 1}{12} \end{aligned}$$

כעת, בהינתן $X \sim U(a, b)$ נסמן $Y = X - a + 1$, ו- $Y \sim U(1, b - a + 1)$ והשונות שלו זהה לשונות של המ"מ המקורי לפי הטענה שראינו קודם. לכן, $\text{Var}(X) = \text{Var}(Y) = \frac{(b-a+1)^2 - 1}{12}$ כנדרש. ■

טענה: יהי $X \sim \text{Geo}(p)$, אזי $E[X] = \frac{1}{p}$ ו- $\text{Var}(X) = \frac{1-p}{p^2}$.

הוכחה: נתחיל מהתוחלת:

$$E[X] = \sum_{x=1}^{\infty} xp(1-p)^{x-1} = \frac{p}{(1 - (1-p))^2} = \frac{1}{p}$$

נסתכל על שתי הנגזרות הראשונות של הטור הגיאומטרי שיעזרו לנו בהמשך (משתמשים באותו טור כמו בשוויון הקודם):

$$\begin{aligned} \left(\sum_{x=0}^{\infty} ar^x \right)' &= \sum_{x=0}^{\infty} axr^{x-1} = \sum_{x=1}^{\infty} axr^{x-1} = \frac{a}{(1-r)^2} \\ \left(\sum_{x=0}^{\infty} ar^x \right)'' &= \sum_{x=2}^{\infty} ax(x-1)r^{x-2} = \frac{2a}{(1-r)^3} \\ &= \sum_{x=1}^{\infty} ax(x+1)r^{x-1} = \sum_{x=1}^{\infty} ar^{x-1}(x^2 + x) \end{aligned}$$

וכעת ניגש למומנט השני שבאמצעותו נחשב את השונות:

$$\begin{aligned} E[X^2] &= \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} = \sum_{x=1}^{\infty} (x^2 + x)p(1-p)^{x-1} - \sum_{x=1}^{\infty} xp(1-p)^{x-1} = \frac{2}{p^2} - \frac{1}{p} \\ \text{Var}(X) &= E[X^2] - E[X]^2 = \frac{2}{p^2} - \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2} \end{aligned}$$

כנדרש. ■

טענה: יהי $X \sim \text{Pois}(\lambda)$ אזי $\text{Var}(X) = \lambda$.

הוכחה: בתרגיל 6. ■

טענה: יהי $X \sim \text{Bin}(n, p)$. השכיחות היחסית היא מ"מ $\frac{X}{n}$, ומתקיים $E\left[\frac{X}{n}\right] = p$ ו- $\text{Var}\left(\frac{X}{n}\right) = \frac{p(1-p)}{n}$.

הוכחה: מליניאריות התוחלת והתכונה שהוכחנו לגבי שונות. ■

מהתוצאה הזאת נובע ש- $\lim_{n \rightarrow \infty} \text{Var}\left(\frac{X}{n}\right) = 0$, כלומר ככל שעושים יותר ניסויים, השונות של השכיחות היחסית הולכת ומתקרבת ל-0 (כלומר, השכיחות היחסית מתקרבת לתוחלת שלה).

דוגמא:

- ההסתברות לעץ במטבע 1 היא p_1 וההסתברות לעץ במטבע 2 היא p_2 . נגדיר מ"מ X – סופר את מספר העצים כאשר בוחרים באקראי מטבע ומטילים אותו פעמיים, ומ"מ Y – סופר את מספר העצים כאשר לפני כל הטלה בוחרים מטבע באקראי (פעמיים). בשני המקרים ההטלות ב"ת. נחשב את התוחלת והשונות של המשתנים.
 - עבור X , מקבלים:

$$P(X=1) = \frac{1}{2}p_1(1-p_1) \cdot 2 + \frac{1}{2}p_2(1-p_2) \cdot 2$$

$$P(X=2) = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2$$

$$E[X] = 0 \cdot P(X=0) + 1 \cdot (p_1(1-p_1) + p_2(1-p_2)) + 2 \cdot \left(\frac{1}{2}p_1^2 + \frac{1}{2}p_2^2\right) = p_1 + p_2$$

$$E[X^2] = 0^2 \cdot P(X=0) + 1^2(p_1(1-p_1) + p_2(1-p_2)) + 4\left(\frac{1}{2}p_1^2 + \frac{1}{2}p_2^2\right) = p_1 + p_2 + p_1^2 + p_2^2$$

$$\text{Var}(X) = p_1 + p_2 + p_1^2 + p_2^2 - (p_1 + p_2)^2 = p_1 + p_2 - 2p_1p_2$$

- עבור Y , ההסתברות של עץ בהטלה הראשונה היא $\frac{1}{2}p_1 + \frac{1}{2}p_2$ וכנ"ל גם בהטלה השנייה. לכן אפשר לומר שההתפלגות

$$\text{היא } Y \sim \text{Bin}\left(2, \frac{1}{2}(p_1 + p_2)\right) \text{ ו- } E[Y] = p_1 + p_2 \text{ ו- } \text{Var}(Y) = 2 \cdot \frac{1}{2}(p_1 + p_2) \left(1 - \frac{1}{2}(p_1 + p_2)\right)$$

- כלומר, התוחלת בשני המקרים זהה, אבל השונות של X גדולה יותר למעט המקרה ש- $p_1 = p_2$. (זה גם מתבקש אינטואיטיבית)

התפלגות משותפת והתפלגות שולית

הגדרה: יהיו X, Y מ"מ. ההתפלגות המשותפת: $P_{X,Y}(x, y) = P(X=x, Y=y) = P(\{X=x\} \cap \{Y=y\})$.

צריך להראות ש- $P_{X,Y}(x, y)$ כזאת היא אכן פונקציית הסתברות חוקית. כרגיל, הקושי הוא להראות הסתכמות ל-1 על כל הערכים, אבל כאן זה נובע מנוסחת ההסתברות השלמה:

$$\sum_x \sum_y P_{X,Y}(x, y) = \sum_x \sum_y P(\{X=x\} \cap \{Y=y\}) = \sum_x P(X=x) = 1$$

שכן הקבוצה $\{\{Y=y\}\}_y$ מהווה חלוקה של Ω .

הגדרה: יהיו X, Y מ"מ. ההתפלגות השולית של X מוגדרת להיות $P_X(x) = P(X=x) = \sum_y P_{X,Y}(x, y)$.

דוגמאות:

- מבצעים 3 הטלות מטבע הוגן, נסמן ב- X את אורך הרצף הארוך ביותר, וב- Y את מספר הראשים שהתקבלו.
 - נצייר את זה בטבלה:

	$X=1$	$X=2$	$X=3$
$Y=0$	0	0	$\frac{1}{8}$
$Y=1$	$\frac{1}{8}$	$\frac{1}{4}$	0

$Y = 2$	$\frac{1}{8}$	$\frac{1}{4}$	0
$Y = 3$	0	0	$\frac{1}{8}$

- מסובבים שני סביבונים, נסמן ב- X התוצאה המקסימלית וב- Y התוצאה המינימלית. נצייר את ההתפלגות המשותפת בטבלה:

	$X = 1$	$X = 2$	$X = 3$	$X = 4$	$P_Y(y)$
$Y = 1$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{7}{16}$
$Y = 2$	0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{5}{16}$
$Y = 3$	0	0	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$
$Y = 4$	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
$P_X(x)$	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	

- כעת קל לחשב את התוחלת של כל משתנה באמצעות ההתפלגויות השוליות.

נסתכל על פונקציה של משתנים מקריים, למשל $Z = X + Y$ כאשר X, Y מ"מ. גם למ"מ החדש ניתן לבנות פונקציית התפלגות, $P_Z(z) = \sum_{(x,y): g(x,y)=z} P_{X,Y}(x,y)$. כמו כן, בדומה למה שראינו עבור תוחלת של פונקציה של משתנה מקרי יחיד, גם כאן מתקיים $E[Z] = \sum_z P_Z(z) = \sum_x \sum_y g(x,y) P_{X,Y}(x,y)$; כמובן, תופעה זו ניתנת להכללה למספר סופי כלשהו של משתנים מקריים.

טענה (ליניאריות התוחלת): יהיו X, Y משתנים מקריים, אזי $E[X + Y] = E[X] + E[Y]$.

הוכחה: נחשב ישירות:

$$\begin{aligned} E[X + Y] &= \sum_x \sum_y (x + y) P_{X,Y}(x, y) = \sum_x \sum_y x P_{X,Y}(x, y) + \sum_x \sum_y y P_{X,Y}(x, y) = \\ &= \sum_x x \sum_y P_{X,Y}(x, y) + \sum_y y \sum_x P_{X,Y}(x, y) = \sum_x x P_X(x) + \sum_y y P_Y(y) = E[X] + E[Y] \end{aligned}$$

■ **כנדרש.**

את הטענה הזאת אפשר להכליל באינדוקציה על מספר סופי כלשהו של משתנים מקריים.

דוגמאות:

- n אנשים זורקים לאוויר את כובעיהם וכל אחד תופס כובע באופן אקראי. מה תוחלת מספר האנשים שיקבלו את הכובע שלהם בחזרה?
- נסמן ב- X את מספר התופסים את הכובע שלהם. נפרק את המ"מ המורכב הזה לסכום של משתנים מקריים מציניים X_i - מקבל 1 אם האיש i -י תפס את כובעו, ו-0 אחרת. כעת:

$$E[X] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \left(1 \cdot \frac{1}{n} + 0 \cdot \frac{n-1}{n}\right) = \sum_{i=1}^n \frac{1}{n} = 1$$

- נשתמש בליניאריות התוחלת כדי לחשב את התוחלת של מ"מ בינומי, $X \sim \text{Bin}(n, p)$. נפרק את המ"מ המורכב למ"מ מציניים עבור כל אחד מ- n ניסויי ברנולי, ומליניאריות התוחלת מקבלים:

$$E[X] = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n (1 \cdot p + 0 \cdot (1 - p)) = np$$

- חברת קורנפלקס יוצאת במבצע שבו בכל חבילה שמים תמונה של אחד מ-100 הספורטאים הגדולים בכל הזמנים באופן אחיד. כמה קופסאות בממוצע נצטרך לקנות כדי להשלים את סדרת 100 הספורטאים?

- נסמן ב- X את המ"מ הזה ורוצים לחשב את תוחלתו. נפרק אותו למ"מ X_i – מספר הקופסאות הנוספות שהיה עלינו לקנות עד שקיבלנו את התמונה ה- i , כולל הקופסא האחרונה. מתקיים: $\sum_{i=1}^{100} X_i = X$.
- ניתן לראות בבירור ש- $\{X_i\}$ מתפלגים גיאומטרית, בפרט: $X_i \sim \text{Geo}\left(\frac{100-i+1}{100}\right)$. לכן נוכל לחשב את התוחלת כולה באמצעות התוחלת של כל אחד מהמ"מ הגיאומטריים:

$$E[X] = \sum_{i=1}^{100} E[X_i] = \sum_{i=1}^{100} \frac{100}{100-i+1} = 100 \cdot \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{100}\right) \cong 518.73$$

נניח שמטילים קוביה עם k פאות, כשלפאה ה- i הסתברות p_i . נסמן ב- X_i את מספר הפעמים שיצאה הפאה ה- i ב- N הטלות. רוצים להסתכל על ההתפלגות המשותפת. $P_{X_1, \dots, X_k}(n_1, \dots, n_k)$. ברור שיש תלות ביניהם. ה- $k-1$ ערכים הראשונים קובעים את הערך האחרון. זה נקרא התפלגות מולטינומית ומתקיים:

$$P_{X_1, \dots, X_k}(n_1, \dots, n_k) = \binom{N}{n_1, n_2, \dots, n_k} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k} = \frac{N!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

יתר על כן, אפשר עכשיו לכתוב את ההתפלגות השולית של משתנה מסוים:

$$P_{X_i}(n_i) = \sum_{n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_k} P_{X_1, \dots, X_k}(n_1, \dots, n_k)$$

ההתפלגות הזאת היא התפלגות בינומית – על כל משתנה אפשר לחשוב כהצלחה/כישלון בבחירת i . כלומר, X_i זהו ניסוי שבו או שיוצא i או שלא, ולכן זו התפלגות בינומית. $X_i \sim \text{Bin}(N, p_i)$.

התפלגות מותנית של משתנים מקריים

הגדרה: יהי X מ"מ ו- A מאורע, נגדיר $P_{X|A}(x) = P(X=x|A) := \frac{P(\{X=x\} \cap A)}{P(A)}$ ההתפלגות המותנית של X בהינתן המאורע A .

כצפוי, זוהי אכן פונקציית התפלגות חוקית (יש לבדוק זאת – נובע מנוסחת ההסתברות השלמה שכן $\{\{X=x\}\}_x$ זוהי חלוקה של Ω).

הגדרה: יהיו X, Y מ"מ, נגדיר $P_{X|Y}(x, y) = P(X=x|Y=y) = \frac{P(\{X=x\} \cap \{Y=y\})}{P(Y=y)} = \frac{P_{X,Y}(x, y)}{P(Y=y)}$ ההתפלגות המותנית של X בהינתן Y .

בניסוח אחר. $P(X=x, Y=y) = P(Y=y)P(X=x|Y=y)$, שזהו חוק הכפל (או חוק השרשרת) עבור משתנים מקריים. באופן דומה אפשר לקבל גם את נוסחת ההסתברות השלמה:

$$P(X=x) = \sum_y P(X=x, Y=y) = \sum_y P(Y=y)P(X=x|Y=y)$$

דוגמאות:

- סטודנט ניגש למבחן לכל היותר n פעמים, עם הסתברות הצלחה p בכל ניסיון. מה התפלגות מספר הניסיונות בהינתן שעבר בסופו של דבר?
 - נסמן ב- X את מספר הניסיונות עד להצלחה, $X \sim \text{Geo}(p)$.
 - בעצם אנחנו שואלים מה התפלגות מספר הניסיונות בהינתן שלא נבחן יותר מ- n פעמים. נסמן ב- A את המאורע ש- $X \leq n$, וקעת ניתן למצוא את ההתפלגות המותנית המבוקשת:

$$P(X = x|A) = \begin{cases} x \leq n: \frac{P(\{X = x\} \cap A)}{P(A)} = \frac{P(X = x)}{P(A)} = \frac{(1-p)^{(x-1)}p}{\sum_{i=1}^n (1-p)^{i-1}p} \\ x > n: 0 \end{cases}$$

- מרצה מפוזר טועה בתשובה לשאלות תלמידים בסיכוי 0.25 ללא תלות בשאלות קודמות. בכיתה הוא נשאל 0, 1, או 2 שאלות בסיכוי אחיד. X – מספר השאלות שנשאלו, Y – מספר השאלות שנענו לא נכון. רוצים למצוא את ההסתברות לטעות אחת לפחות.

○ נכתוב התפלגות משותפת בטבלה:

	$X = 0$	$X = 1$	$X = 2$
$Y = 0$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{3} \left(\frac{3}{4}\right)^2$
$Y = 1$			
$Y = 2$			

○ ואנו יכולים כבר אחרי השורה הראשונה להשיב ש- $P(Y \geq 1) = 1 - P(Y = 0)$.

- משדר שולח 1 בהסתברות p ו-0 בהסתברות $1-p$, באופן ב"ת בשידורים קודמים. נגדיר Z – מספר השידורים ליחידת זמן, וידוע ש- $Z \sim \text{Pois}(\lambda)$. מה ההתפלגות של ה-1-ים ביחידת זמן?

○ נסמן ב- X את מספר ה-1-ים, Y מספר ה-0-ים ונחשב את ההתפלגות המשותפת:

$$\begin{aligned} P(X = n, Y = m) &= P(X = n, Y = m, Z = n + m) = P(X = n, Y = m | Z = n + m)P(Z = n + m) = \\ &= e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} \binom{n+m}{n} p^n (1-p)^m \\ P(X = n) &= \sum_{m=0}^{\infty} P(X = n, Y = m) = \sum_{m=0}^{\infty} e^{-\lambda} \frac{\lambda^{n+m}}{(n+m)!} \binom{n+m}{n} p^n (1-p)^m = \sum_{m=0}^{\infty} e^{-\lambda} \frac{\lambda^{n+m}}{n! m!} p^n (1-p)^m = \\ &= \frac{(\lambda p)^n e^{-\lambda p}}{n!} \sum_{m=0}^{\infty} \frac{(1-p)^m e^{-\lambda(1-p)} \lambda^m}{m!} = \frac{(\lambda p)^n e^{-\lambda p}}{n!} e^{-\lambda(1-p)} \sum_{m=0}^{\infty} \frac{(\lambda(1-p))^m}{m!} = \frac{(\lambda p)^n e^{-\lambda p}}{n!} \end{aligned}$$

- כאשר כאן המעבר האחרון היה לפי טור טיילור של e^x (במקרה הזה, $x = \lambda(1-p)$). לבסוף קיבלנו ש- $X \sim \text{Pois}(\lambda p)$, מה שדי סביר אינטואיטיבית.

הגדרה: יהיו X, Y מ"מ ו- A מאורע. התוחלת המותנית של X בהינתן Y או A מוגדרת להיות:

$$\begin{aligned} E[X|Y = y] &= \sum_x x P(X = x | Y = y) \\ E[X|A] &= \sum_x x P(X = x | A) \end{aligned}$$

טענה (נוסחת התוחלת השלמה עבור מאורעות): תהי $A_1, \dots, A_n$ חלוקה של Ω . אזי $E_i[E_X[X|A_i]] = E[X]$.

הוכחה: נפתח ישירות:

$$\begin{aligned} E[X] &= \sum_x x P(X = x) = \sum_x x \sum_{i=1}^n P(X = x | A_i) P(A_i) = \\ &= \sum_{i=1}^n P(A_i) \sum_x x P(X = x | A_i) = \sum_{i=1}^n P(A_i) E[X | A_i] = E_i[E_X[X|A_i]] \end{aligned}$$

■ כנדרש.

טענה (נוסחת התוחלת השלמה עבור משתנים מקריים): יהיו X, Y מ"מ. אזי $E_Y[E_X[X|Y]] = E[X]$.

הוכחה: שוב נפתח ישירות:

$$E[X] = \sum_x P(X=x) = \sum_x x \sum_y P(X=x|Y=y)P(Y=y) = \sum_y P(Y=y) \sum_x xP(X=x|Y=y) = \sum_y P(Y=y)E[X|Y=y] = E_Y[E_X[X|Y]]$$

■ כנדרש.

הסימון המעט-מבלבל $E[E[X|Y]]$ בסה"כ אומר שאנו מגדירים פונקציה $g(Y) = E[X|Y]$, שהיא בעצמה מהווה משתנה מקרי. הפונקציה עבור y מסוים נותנת מספר. $g(y) = E[X|Y=y]$, שזו תוחלת מותנית של M . כיוון ש- g בעצמה משתנה מקרי, ניתן לחשב את התוחלת שלה (ע"פ ערכי Y).

הגדרה: מ"מ X, Y ייקראו בלתי-תלויים אם לכל x, y מתקיים $P(X=x, Y=y) = P(X=x)P(Y=y)$.

הגדרה: יהיו X, Y, Z מ"מ. המ"מ X, Y ייקראו בלתי-תלויים בהינתן Z אם לכל x, y, z מתקיים $P(X=x, Y=y|Z=z) = P(X=x|Z=z)P(Y=y|Z=z)$.

דוגמאות:

- כותבים תרגיל תכנות שוב ושוב שבכל פעם הוא עובד בהסתברות p (באופן ב"ת בניסיונות הקודמים). מהו מספר הפעמים הממוצע שנעבוד על התרגיל עד שנצליח?

○ ראינו כבר שמדובר בתוחלת של מ"מ גיאומטרי, $X \sim Geo(p)$, שהיא $\frac{1}{p}$. נפתח את זה בדרך אחרת:

$$E[X] = P(X=1)E[X|X=1] + P(X>1)E[X|X>1] = p \cdot 1 + (1-p) \cdot (1+E[X]) = p + 1 + E[X] - p - pE[X]$$

$$E[X] = \frac{1}{p}$$

○ בדרך השתמשנו בתכונת חוסר הזיכרון של מ"מ גיאומטרי, שבגללה $E[X|X>1] = 1 + E[X]$.

- יהי Y מ"מ שהוא ההסתברות שהשמש זורחת מחר (אחיד, עם תומך $\{0, \frac{1}{n}, \dots, \frac{n-1}{n}, 1\}$ ו- X מ"מ הסופר את מספר הזריחות ב-100 הימים הקרובים אז:

○ ומתקיים $X \sim Bin(100, Y)$, $E[X|Y] = 100Y$ (תוחלת של מ"מ בינומי).

○ כעת אפשר להשתמש בנוסחת התוחלת השלמה ולקבל:

$$E[X] = E[E[X|Y]] = \sum_{i=0}^n P\left(Y = \frac{i}{n}\right) E\left[X \mid Y = \frac{i}{n}\right] = \frac{1}{n+1} \sum_{i=0}^n \frac{100i}{n}$$

- בכד 3 כדורים ממוספרים. מוציאים 2 כדורים ללא החזרה, X – הכדור הראשון, Y – הכדור השני. ניתן להראות שמתקיים $P(X = \frac{1}{3}, Y = y) = \frac{1}{3}$.

○ נכתוב טבלת התפלגות משותפת כדי לראות האם המ"מ בלתי-תלויים:

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

○ ...ורואים מכאן שהמ"מ אינם ב"ת, כי קביעת ערך של X משפיעה על ההסתברויות של Y .

- מתעניינים בקשר בין נוכחות וירוס ($V, \bar{V}$) למין הנבדק (M, F).

○ נסתכל על שתי קבוצות נבדקים:

ב'	V	$\bar{V}$	
F	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$
M	$\frac{1}{12}$	$\frac{1}{4}$	$\frac{1}{3}$

א'	V	$\bar{V}$	
F	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$
M	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$

	$\frac{1}{4}$	$\frac{3}{4}$			$\frac{1}{3}$	$\frac{2}{3}$	
--	---------------	---------------	--	--	---------------	---------------	--

- בקבוצה א' המשתנים אינם ב"ת; בקבוצה ב' המשתנים אכן ב"ת.
- מטילים מטבע עם הסתברות p לעץ עד לקבלת עץ בפעם הראשונה. לאחר מכן מטילים שוב את אותו מספר הטלות וסופרים כמה עץ יצא. מה ההסתברות שלא יצא עץ בכלל בסיבוב השני?
 - נסמן X – מספר ההטלות בסיבוב הראשון, $X \sim Geo(p)$. נסמן גם Y – מספר ההטלות בסיבוב השני, וכאן יש תלות בין המשתנים: $Y|X = x \sim Bin(x, p)$, כלומר $Y \sim Bin(X, p)$.
 - כעת אנו מחשבים את $P(Y = 0)$:

$$P(Y = 0) = \sum_{x=1}^{\infty} P(Y = 0, X = x) = \sum_{x=1}^{\infty} P(Y = 0|X = x)P(X = x) = \sum_{x=1}^{\infty} (1-p)^x p (1-p)^{x-1} = \frac{1-p}{2-p}$$

- אם התקבלו 0 עצים בסיבוב השני, מהי ההסתברות שמספר הניסיונות היה 2? כאן עלינו להשתמש בכלל בייס על מאורעות:

$$P(X = 2|Y = 0) = \frac{P(Y = 0|X = 2)P(X = 2)}{P(Y = 0)} = \frac{(1-p)^2 p (1-p)}{\frac{1-p}{2-p}}$$

- יהיו $X_1 \sim Geo(p_1), X_2 \sim Geo(p_2)$. אזי $\min\{X_1, X_2\} \sim Geo(1 - (1-p_1)(1-p_2))$. הוכחה: בתרגיל 5.
- בלהקה יש N ציפורים. בכל יחידת זמן כל ציפור הולכת לאכול בהסתברות p או לשמור בהסתברות $1-p$ באופן בלתי תלוי בציפורים האחרות. בכל יחידת זמן מופיע נץ ואם אין אף ציפור במשמר, אוכל ציפור אחת. מה תוחלת אורך החיים של הלהקה?
 - נסמן Y – מספר יחידות הזמן עד להיכחדות הלהקה. (כיוון שלא בוקעים גוזלים חדשים, הזמן הזה סופי).
 - נסמן גם Y_k – מספר יחידות הזמן עד שהאוכלוסיה עוברת מ- $k+1$ ציפורים ל- k ציפורים.
 - ההסתברות שציפור מסוימת תיאלץ ביחידת זמן מסוימת כשיש N ציפורים היא p^N , וזה חוזר על עצמו באופן ב"ת כל יום, ולכן $Y_k \sim Geo(p^{k+1})$. כמו כן, $Y = \sum_{k=0}^{N-1} Y_k$ ולכן ניתן לחשב את התוחלת ישירות:

$$E[Y] = \sum_{k=0}^{N-1} E[Y_k] = \sum_{k=0}^{N-1} \frac{1}{p^{k+1}} = \sum_{k=0}^{N-1} \left(\frac{1}{p}\right)^{k+1} = \frac{1-p^N}{p(1-p)}$$

הגדרה: יהיו $X_1, \dots, X_n$ מ"מ. הם ייקראו בלתי-תלויים אם $P_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n P_{X_i}(x_i)$.

נשים לב שעבור מ"מ זה מספיק, בניגוד למאורעות – שם היינו צריכים לדרוש אי-תלות על כל תת-קבוצה של המאורעות. הסיבה לכך תתבהר מהמקרה הפרטי הבא, שאפשר להכלילו באמצעות אינדוקציה:

טענה: אי-תלות בשלוש גוררת אי-תלות בזוגות.

הוכחה: נתון ש- $P(X = x, Y = y, Z = z) = P(X = x)P(Y = y)P(Z = z)$. נתבונן בהסתברות השולית של המ"מ X, Y על פני ערכי Z :

$$\begin{aligned} P(X = x, Y = y) &= \sum_z P(X = x, Y = y, Z = z) = \sum_z P(X = x)P(Y = y)P(Z = z) = \\ &= P(X = x)P(Y = y) \sum_z P(Z = z) = P(X = x)P(Y = y) \end{aligned}$$

באופן דומה מקבלים את התוצאה עבור כל זוג אחר של מ"מ. ■

דוגמאות:

- X, Y מ"מ גיאומטריים בלתי-תלויים עם פרמטר p . מהי ההתפלגות $P(X = i|X + Y = n)$?
 - נחשב אותה:
$$\begin{aligned} P(X = i|X + Y = n) &= \frac{P(\{X = i\} \cap \{X + Y = n\})}{P(X + Y = n)} = \frac{P(\{X = i\} \cap \{Y = n - i\})}{P(X + Y = n)} = \\ &= \frac{P(X = i)P(Y = n - i)}{P(X + Y = n)} = \frac{(1-p)^{i-1}p(1-p)^{n-i-1}p}{P(X + Y = n)} = \end{aligned}$$

$$= \frac{p^2(1-p)^{n-2}}{P(X+Y=n)}$$

○ כעת נשים לב ש-

$$1 = \sum_{i=1}^{n-1} P(X=i|X+Y=n) = \sum_{i=1}^{n-1} \frac{p^2(1-p)^{n-2}}{P(X+Y=n)}$$

○ וכאן הגורם בתוך הסכום הוא קבוע, ויש $n-1$ גורמים כאלה, ולכן ההסתברות שחיפשונו שווה בדיוק ל- $\frac{1}{n-1}$ לכל i .

- נתון $n \geq 3$ מספר טבעי. בשלב 1 מגרילים K מתוך $\{1, 2, \dots, n\}$ בהסתברות לא דווקא אחידה, אבל יודעים את $\mu = E[K]$. בשלב 2, דוגמים מתוך $\{1, 2, \dots, n\}$ איברים ללא החזרה בהסתברות אחידה, ונסמנם S . נסמן גם X – מספר האיברים ב- S שהם ב- $\{1, 2, 3\}$.
○ נמצא את התוחלת של X : בהינתן K , נסמן X_i (עבור $i \in \{1, \dots, K\}$) מ"מ המקבל 1 אם המספר ה- i שנבחר הוא ב- $\{1, 2, 3\}$ ו-0 אחרת. מתקיים $X|K = \sum_{i=1}^K X_i|K$ וכן $E[X|K] = E[X_i] = \frac{3}{n}$ (משתנה בינומי עם פרמטר $\frac{3}{n}$).
○ כעת נשתמש בנוסחת התוחלת השלמה:

$$E[X] = E[E[X|K]] = E\left[E\left[\sum_{i=1}^K X_i|K\right]\right] = E[KE[X_1]] = E\left[K \frac{3}{n}\right] = \frac{3}{n}E[K] = \frac{3\mu}{n}$$

- כעת בהינתן K , נמצא את ההתפלגות של X . זוהי התפלגות היפר-גיאומטרית: יש שני סוגים של איברים $\{1, 2, 3\}$ וכל השאר), מוציאים K איברים ושואלים מה ההסתברות ש- X מתוכם הם מהסוג הראשון. כלומר, $X \sim HG(K, 3, n-3)$.
• נבצע שתי הטלות של מטבע הוגן באופן ב"ת, X – ההטלה הראשונה, X – השנייה. נגדיר $Z = X \oplus Y$ (xor).
○ ברור ש- X, Y ב"ת מהנתון, אבל אפשר לראות גם ש- $Z \perp X, Z \perp Y$, כלומר מתקיימת אי-תלות בזוגות.
○ עם זאת, בשלשה המ"מ אינם ב"ת, וזה ברור - X, Y קובעים את ערכו של Z .
• מטילים מטבע שהסתברותו לעץ היא p , עד שמתקבל רצף של $r > 0$ עצים. מהי תוחלת מס' ההטלות?
○ נסמן X את מספר ההטלות, Y מספר ההטלות עד שהתקבל פלי (לראשונה). כמובן, $Y \sim Geo(1-p)$. כעת נסתכל על התוחלת באמצעות נוסחת התוחלת השלמה:

$$E[X] = E[E[X|Y]] = \sum_{y=1}^{\infty} P(Y=y)E[X|Y=y] = (*)$$

○ על התוחלת המותנית אנו יכולים לומר את הדבר הבא:

$$E[X|Y=y] = \begin{cases} r & y > r \\ y + E[X] & y \leq r \end{cases}$$

- זאת משום שאם $y \leq r$, אז y ניסיונות כבר קרו, ועכשיו שוב ממשיכים עם X כאילו שחזרנו להתחלה. לכן אפשר לפרק את מה שקיבלנו לשני סכומים:

$$(*) = \sum_{y=1}^r (y + E[X])(1-p)p^{y-1} + \sum_{y=r+1}^{\infty} r(1-p)p^{y-1} = \frac{1-p^r}{1-p} + (1-p^r)E[X] \Rightarrow E[X] = \frac{1-p^r}{p^r(1-p)}$$

- כך נוכל לחשב גם את השונות, שכן $E[X^2|Y=y] = \begin{cases} r^2 & y > r \\ E[(y+X)^2] & y \leq r \end{cases}$.

טענה: יהיו X, Y מ"מ ב"ת. אזי $E[XY] = E[X]E[Y]$.

הוכחה: נוכיח ישירות:

$$\begin{aligned} E[XY] &= \sum_x \sum_y xyP(X=x, Y=y) = \sum_x x \sum_y yP(X=x, Y=y) = \sum_x x \sum_y yP(X=x)P(Y=y) = \\ &= \sum_x xP(X=x) \sum_y yP(Y=y) = \sum_x xP(X=x)E[Y] = E[Y]E[X] \end{aligned}$$

■ כנדרש.

כמובן, ניתן להכליל את הטענה לגבי הפעלת פונקציות כלשהן על המ"מ, כלומר על $g(X), h(Y)$ (הפעלת פונקציה על מ"מ לא יכולה לגרום לו להיות תלוי במ"מ אחר).

טענה: יהיו X, Y מ"מ ב"ת. אזי $Var(X + Y) = Var(X) + Var(Y)$.

הוכחה: באמצעות הטענה הקודמת:

$$\begin{aligned} Var(X + Y) &= E \left[((X + Y) - (E[X] + E[Y]))^2 \right] = E \left[((X - E[X]) + (Y - E[Y]))^2 \right] = \\ &= E[(X - E[X])^2 + 2(X - E[X])(Y - E[Y]) + (Y - E[Y])^2] = \\ &= E[(X - E[X])^2] + 2E[(X - E[X])(Y - E[Y])] + E[(Y - E[Y])^2] = \\ &= Var(X) + Var(Y) + 2E[X - E[X]]E[Y - E[Y]] = Var(X) + Var(Y) \end{aligned}$$

כאשר בדרך השתמשנו באי-תלות כדי לפתוח את התוחלת של המכפלה. ■

כמובן, גם טענה זו ניתנת להכללה (באמצעות אינדוקציה) למספר סופי כלשהו של מ"מ ב"ת.

דוגמאות:

- יהי $X \sim Bin(n, p)$. נחשב את $Var(X)$ באמצעות הטענה הקודמת:
 - נסמן $X = \sum_{i=1}^n X_i$ כאשר $X_i \sim Ber(p)$ בלתי-תלויים כולם.
 - כעת אפשר מיידית להגיע לתוצאה – שהיינו צריכים נגזרות של טורים כדי לקבל, בהוכחה המקורית:

$$Var(X) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) = \sum_{i=1}^n (E[X_i^2] - E[X_i]^2) = \sum_{i=1}^n (1^2 p + 0^2 (1-p) - p^2) = np(1-p)$$
- נניח שיש n מחוזות בחירה ואחוז התמיכה במועמד מסוים במחוז ה- i הוא p_i . אנו מבצעים סקר ובו שואלים נסקר אחד מכל מחוז. נסמן X_i מ"מ מציינ המקבל 1 אם הנסקר תמך במועמד 0-אחרת. נסמן גם $X = \sum_{i=1}^n X_i$. המטרה היא להקטין את השונות של X ע"י תכנון המחוזות – האם כדאי שכל ה- p_i יהיו קרובים זה לזה, או שיהיו שונים זה מזה?
 - נסמן $\mu := E[X] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p_i$ (אנו מניחים שהמחוזות ב"ת).
 - כעת אפשר לחשב את השונות:

$$Var(X) = \sum_{i=1}^n Var(X_i) = \sum_{i=1}^n p_i(1-p_i) = \sum_{i=1}^n p_i - \sum_{i=1}^n p_i^2 = \mu - \sum_{i=1}^n p_i^2$$
 - החלק שאפשר להשפיע עליו הוא הגורם השני – ככל שהוא קטן יותר, זה יקטין את השונות. ובכן:

$$\sum_{i=1}^n p_i^2 = \sum_{i=1}^n \left(p_i - \frac{\mu}{n}\right)^2 + \frac{2\mu}{n} \sum_{i=1}^n p_i - n \left(\frac{\mu}{n}\right)^2 = \sum_{i=1}^n \left(p_i - \frac{\mu}{n}\right)^2 + C$$
 - הביטוי השמאלי כאן מגיע למינימום כאשר $p_i = \frac{\mu}{n}$ לכל i , ואז השונות היא מקסימלית! כלומר, כדי לקבל שונות מינימלית צריך מחוזות הטרופניים.
- בעקבות התחרות החדשה (רמי לוי), n סניפי אורנג' רוצים להתאחד. ידוע כי בכל סניף i מספר הלקוחות היומי X_i מתפלג פואסון עם פרמטר λ_i באופן ב"ת בסניפים האחרים. נתעניין במספר הלקוחות בסניף המאוחד, שנסמנו X :
 - ראשית נמצא את התוחלת: $E[X] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \lambda_i$.
 - נפתח את פונקציית ההתפלגות לסניף המאוחד – לצורך הנוחות נסתכל כרגע על $n = 2$ ואפשר להכליל אחר כך לערכים גדולים יותר באמצעות אינדוקציה:

$$\begin{aligned} P(X = x) &= P(X_1 + X_2 = x) = P(\{X_1 = 0\} \cap \{X_2 = x\}) \cup \dots \cup (\{X_1 = x\} \cap \{X_2 = 0\}) = \\ &= \sum_{m=0}^x P(X_1 = m, X_2 = x - m) = \sum_{m=0}^x P(X_1 = m)P(X_2 = x - m) = \\ &= \sum_{m=0}^x e^{-\lambda_1} \frac{\lambda_1^m}{m!} e^{-\lambda_2} \frac{\lambda_2^{x-m}}{(x-m)!} = e^{-(\lambda_1 + \lambda_2)} \sum_{m=0}^x \frac{\lambda_1^m \lambda_2^{x-m}}{m! (x-m)!} = \frac{1}{x!} e^{-(\lambda_1 + \lambda_2)} \sum_{m=0}^x \frac{x!}{m! (x-m)!} \lambda_1^m \lambda_2^{x-m} = \\ &= \frac{1}{x!} e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^x \end{aligned}$$

- כלומר, קיבלנו ש- $X \sim Pois(\lambda_1 + \lambda_2)$, מה שיהיה נכון גם עבור n סניפים.
- כעת, בהינתן $X = s$, מהי התפלגות X_i בה"כ $i = 1$ ואז:

$$P(X_1 = x | X = s) = \frac{P(X_1 = x, X = s)}{P(X = s)} = \frac{P(X_1 = x, X_2 + \dots + X_n = s - x)}{P(X = s)} = \frac{P(X_1 = x)P(X_2 + \dots + X_n = s - x)}{P(X = s)} =$$

$$\begin{aligned}
&= \frac{e^{-\lambda_1} \frac{\lambda_1^x}{x!} e^{-(\lambda_2 + \dots + \lambda_n)} (\lambda_2 + \dots + \lambda_n)^{s-x}}{e^{-(\lambda_1 + \dots + \lambda_n)} (\lambda_1 + \dots + \lambda_n)^s} = \frac{\frac{s!}{x!(s-x)!} \lambda_1^x (\lambda_2 + \dots + \lambda_n)^{s-x}}{(\lambda_1 + \dots + \lambda_n)^x (\lambda_1 + \dots + \lambda_n)^{s-x}} = \\
&= \binom{s}{x} \left(\frac{\lambda_1}{\sum_{i=1}^n \lambda_i} \right)^x \left(\frac{\sum_{i=2}^n \lambda_i}{\sum_{i=1}^n \lambda_i} \right)^{s-x}
\end{aligned}$$

- וכעת אם נסמן $p = \frac{\lambda_1}{\sum_{i=1}^n \lambda_i}$, אז מקבלים ש- $X_1 | X = s \sim \text{Bin}(s, p)$. (במובן מסוים, זה אינטואיטיבי. יש לנו s לקוחות ואנו רוצים לדעת כמה מהם באו לסניף הראשון. ההגעה של לקוח מסוים לסניף הראשון היא ניסוי ברנולי והפרמטר שקיבלנו, p , מייצג את היחס בין "משקל" הסניף הראשון למשקל כל סניפים, כלומר חלקו היחסי של הסניף הראשון בעומס הלקוחות).

משתנים מקריים רציפים

הגדרה: מ"מ רציף הוא פונקציה $X: \Omega \rightarrow \mathbb{R}$, כך שקיימת פונקציה $f_X(x)$ המקיימת $P(X \in B) = \int_B f_X(x) dx$, הנקראת פונקציית צפיפות הסתברות (באנגלית: *PDF, Probability Density Function*), והמקיימת את האקסיומות הבאות:

$$(1) f_X(x) \geq 0 \text{ לכל } x$$

$$(2) \int_{-\infty}^{\infty} f_X(x) dx = 1 \text{ מתקיים}$$

בדרך כלל הקבוצה B תהיה קטע או איחוד של מספר קטעים. נשים לב שכאשר גודל הקטע $\delta = |B|$ קטן, ניתן לקרב את האינטגרל (ולכן את ההסתברות) באמצעות $\int_c^{c+\delta} f_X(x) dx \cong f_X(x) \delta$.

כמו כן, כיוון שאינטגרל של פונקציית הצפיפות נותן את פונקציית ההסתברות, מתבקש להסתכל על הנגזרת של פונקציית ההסתברות מסוימת – שתיתן פונקציית צפיפות.

לבסוף, מתקיים $P(a \leq X \leq b) = P(a < X < b)$, כלומר אין יותר צורך לדון במה שקורה בקצוות הקטע – האינטגרל יקבל אותו ערך על קטעים פתוחים או סגורים.

הגדרה: מ"מ רציף אחיד על קטע $[a, b]$ מסומן $X \sim U(a, b)$ והוא בעל פ' צפיפות $f_X(x) = \frac{1}{b-a}$.

דוגמאות:

- תהי $f_X(x) = \begin{cases} c & x \in [0, 1] \\ 0 & x \notin [0, 1] \end{cases}$. מהתכונה השנייה מתקיים $1 = \int_{-\infty}^{\infty} f_X(x) dx = c$, כלומר הקבוע נכפה עלינו. המצב יהיה דומה עבור כל קטע אחר.
- זמן נסיעתו של משה לעבודה מתפלג אחיד בין 15 ל-20 דקות ביום שמש ובין 20 ל-25 דקות ביום חורף (ההסתברות ליום חורף היא $\frac{1}{3}$). מה פ' הצפיפות של זמן הנסיעה?

$$f_X(x) = \begin{cases} c_1 & 15 \leq x \leq 20 \\ c_2 & 20 \leq x \leq 25 \\ 0 & \text{otherwise} \end{cases}$$

- ברור מהנתונים שהפונקציה נראית כך: $f_X(x) = \begin{cases} c_1 & 15 \leq x \leq 20 \\ c_2 & 20 \leq x \leq 25 \\ 0 & \text{otherwise} \end{cases}$.
- כדי למצוא את הקבועים, ניזכר בכך שהאינטגרל של f על הקטע הראשון (יום שמש) הוא $\frac{2}{3}$ ועל הקטע השני הוא $\frac{1}{3}$.

תוחלת של משתנה מקרי רציף

הגדרה: אם X מ"מ רציף ו- f_X פונקציית הצפיפות שלו, התוחלת שלו היא: $E[X] := \int_{-\infty}^{\infty} x f_X(x) dx$.

באופן דומה, לכל פונקציה של מ"מ $g: \mathbb{R} \rightarrow \mathbb{R}$, $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$.

הגדרה: אם X מ"מ רציף ו- f_X פונקציית הצפיפות שלו, השונות שלו היא: $\text{Var}(X) := \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$.

כמו עבור מ"מ בדידים, גם כאן נכון ש- $Var(X) = E[X^2] - E[X]^2$. למעשה, רוב התכונות של תוחלת ושונות (למשל, ליניאריות) מתקיימות גם במקרה הרציף ולא נזכיר אותן או נזכיר אותן בנפרד.

דוגמאות:

- יהי $X \sim U(a, b)$, נחשב את התוחלת והשונות שלו:

○ התוחלת:

$$E[X] = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \frac{b^2 + a^2}{2} = \frac{b+a}{2}$$

○ השונות:

$$E[X^2] = \frac{1}{b-a} \int_a^b x^2 dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{a^2 + b^2 + ab}{3}$$

$$Var(X) = \frac{a^2 + b^2 + ab}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{(b-a)^2}{12}$$

- נשים לב שהתוחלת לא תמיד קיימת. למשל, יהי $X \sim Cauchy$, כלומר בעל פונקציית צפיפות: $f_X(x) = \frac{1}{\pi(1+x^2)}$. זוהי אכן פונקציית צפיפות חוקית, האינטגרל על כל הישר הוא 1. לעומת זאת, התוחלת לא קיימת, כלומר האינטגרל הלא-אמיתי $\lim_{N \rightarrow \infty} \lim_{M \rightarrow \infty} \int_{-N}^M \frac{x}{1+x^2} dx$ לא מתכנס, ואפילו לא ניתן לומר אם הוא $+\infty$ או $-\infty$, כי זה מושפע מסדר האינטגרציה. לעומת זאת, $E[|X|]$ זהו אינטגרל שעדיין לא מתכנס, אבל בביור שואף ל- $+\infty$, ולכן אפשר לומר שהתוחלת קיימת והיא אינסוף.
- באופן כללי נוכל לומר ש- $E[g(X)]$ קיימת וסופית אם $E[|g(X)|] < \infty$.

הגדרה: עבור קבוע $\lambda > 0$, מ"מ אקספוננציאלי (מעריכי) מסומן $X \sim \exp(\lambda)$ והוא בעל פ' צפיפות $f_X(x) = \lambda e^{-\lambda x}$ עבור $x \geq 0$ ו-0 אחרת.

נראה שזו פ' צפיפות – האי-שליליות ברורה, וההסתכמות ל-1: $\int_0^\infty \lambda e^{-\lambda x} dx = [-e^{-\lambda x}]_0^\infty = 0 - (-1) = 1$. המ"מ הזה, כפי שתיתקף נראה, מזכיר מ"מ גיאומטרי מבחינת היחס של התוחלת שלו לפרמטר.

טענה: יהי $X \sim \exp(\lambda)$, אזי $E[X] = \frac{1}{\lambda}$ ו- $Var(X) = \frac{1}{\lambda^2}$.

הוכחה: נחשב ישירות באמצעות אינטגרציה בחלקים:

$$E[X] = \int_{-\infty}^\infty x f_X(x) dx = \int_0^\infty (-x)(-\lambda e^{-\lambda x}) dx = [-x e^{-\lambda x}]_0^\infty - \int_0^\infty -e^{-\lambda x} dx = 0 - \left[\frac{e^{-\lambda x}}{\lambda} \right]_0^\infty = \frac{1}{\lambda}$$

$$E[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \int_0^\infty (-x^2)(-\lambda e^{-\lambda x}) dx = [-x^2 e^{-\lambda x}]_0^\infty - \int_0^\infty (-2x) e^{-\lambda x} dx$$

$$= 2 \int_0^\infty x e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$Var(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$$

■ כנדרש.

טענה (תכונת חוסר הזיכרון): יהי $X \sim \exp(\lambda)$, אזי $P(X > t + s | X > s) = P(X > t)$.

הוכחה: בפתרון תרגיל 6. ■

פונקציית התפלגות מצטברת

הגדרה: יהי X מ"מ עם פ' צפיפות $f_X(x)$, לפונקציה F_X הבאה נקרא פונקציית ההתפלגות המצטברת של X :
 (באנגלית: $F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt$ (CDF, Cumulative Distribution Function))

מתקיים, כמובן, הקשר בין אינטגרל לנגזרת. $f_X(x) = \frac{dF_X(x)}{dx}$. כמו כן, מתקיימות עוד מספר תכונות חשובות.

$$P(a \leq X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx \quad (1)$$

$$F_X(x) \text{ לא יורדת} \quad (2)$$

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \text{ מתקיים} \quad (3)$$

$$\lim_{x \rightarrow \infty} F_X(x) = 1 \text{ מתקיים} \quad (4)$$

$$F_X(x) \text{ רציפה מימין, כלומר } \lim_{\varepsilon \rightarrow 0^+} F_X(x + \varepsilon) = F_X(x) \quad (5)$$

דוגמאות:

- פונקציית הצפיפות המצטברת של מ"מ אחיד $X \sim U(a, b)$ היא:

$$F_X(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$
- מותר לגשת למבחן 3 פעמים והציון יהיה המקסימום מה-3. יהיו $\{X_i\}_{i=1}^3$ הציונים, ונסמן $X = \max\{X_1, X_2, X_3\}$, וכן נניח שהציונים מתפלגים אחיד (בדיד) בין 1 ל-10 ובלתי-תלויים ביניהם. השאלה: מה ה-CDF של X ?
 - נכתוב במפורש:

$$F_X(x) = P(X \leq x) = P(X_1 \leq x, X_2 \leq x, X_3 \leq x) = P(X_i \leq x)^3 = \left(\frac{x}{10}\right)^3$$

$$P(X = x) = F_X(x) - F_X(x-1) = \left(\frac{x}{10}\right)^3 - \left(\frac{x-1}{10}\right)^3$$
 - מסקנה כללית: לכל קבוצת מ"מ ב"ת $\{X_i\}_{i=1}^n$ קבוצת מ"מ ב"ת $X = \max\{X_1, \dots, X_n\}$ מתקיים: $F_X(x) = F_{X_1}(x) \cdots F_{X_n}(x)$
- פונקציית התפלגות מצטברת מאפשרת לערבב משתנים מקריים בדידים ורציפים, כי היא מאפיינת את כל המקרים. למשל, יהיו שני מ"מ $X \sim U(0,1)$ ו- $Z \sim \text{Ber}(p)$ ב"ת, ונגדיר $Y = XZ$.
 - נמצא את ה-CDF שלו:

$$F_Y(u) = P(Y \leq u) = \begin{cases} 0 & u < 0 \\ ? & u = 0 \\ ? & 0 < u < 1 \\ 1 & u \geq 1 \end{cases}$$
 - כעת נתבונן על האפשרויות האמצעיות:

$$P(Y \leq 0) = P(Y = 0) = P(XZ = 0) = P(XZ = 0, Z = 1) + P(XZ = 0, Z = 0) = P(XZ = 0, Z = 1) + P(Z = 0) =$$

$$= P(X = 0, Z = 1) + P(Z = 0) = P(X = 0)P(Z = 1) + P(Z = 0) = 0 \cdot p + (1 - p) = 1 - p$$

$$P(Y \leq u) = P(XZ \leq u) = P(XZ \leq u, Z = 1) + P(XZ \leq u, Z = 0) = P(X \leq u, Z = 1) + P(Z = 0) =$$

$$= P(X \leq u)P(Z = 1) + (1 - p) = u \cdot p + (1 - p)$$
 - המ"מ שקיבלנו (Y) אינו רציף ואינו בדיד.
- יהיו $X_1, \dots, X_n$ מ"מ רציפים בלתי-תלויים שווי התפלגות עם פ' צפיפות $f_X(x)$.
 - אם $Y = \max\{X_1, \dots, X_n\}$ אז $F_Y(y) = (F_X(y))^n$
 - אם $Y = \min\{X_1, \dots, X_n\}$ אז $F_Y(y) = 1 - (1 - F_X(y))^n$
 - הוכחת שתי הטענות – בפתרון תרגיל 7.
- יהיו $X_i \sim \exp(\lambda_i)$ מ"מ ב"ת, אזי $\min\{X_1, \dots, X_n\} \sim \exp(\sum_{i=1}^n \lambda_i)$. הוכחה – בפתרון תרגיל 7.

טענה: יהי X מ"מ רציף אי-שלילי, אזי $E[X] = \int_0^\infty P(X > x)dx$

הוכחה: תמיד מתקיים ש- $F_X(x) = 1 - \int_x^\infty f_X(t)dt$. לכן נסתכל על האינטגרל הכפול שלפנינו:

$$\int_0^\infty P(X > x) dx = \int_0^\infty \int_x^\infty f_X(t) dt dx = \int_0^\infty \int_0^t f_X(t) dx dt = \int_0^\infty t f_X(t) dt = \int_{-\infty}^\infty t f_X(t) dt = E[X]$$

כאשר בדרך נדרשנו להחליף את סדר האינטגרציה. ■

הגדרה: X נקרא משתנה מקרי נורמלי $X \sim N(\mu, \sigma^2)$ אם פונקציית הצפיפות שלו היא $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

יש להראות שזו פונקציית צפיפות חוקית, כלומר שהאינטגרל שלה על הישר הוא 1. זה לא אינטגרל פשוט, אבל נשתמש בעובדה זו בהמשך כדי לפשט חישובים אחרים. כמו כן נשים לב שהפונקציה הזאת סימטרית, כלומר $f_X(x) = f_X(-x)$ לכל $x \in \mathbb{R}$.

טענה: יהי $X \sim N(\mu, \sigma^2)$. אזי $E[X] = \mu$.

הוכחה: נחשב את האינטגרל:

$$\begin{aligned} E[X] &= \int_{-\infty}^\infty \frac{x}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^\infty \frac{\sigma y + \mu}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2}} \sigma dy = \\ &= \sigma \int_{-\infty}^\infty \frac{y}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy + \mu \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 0 + \mu = \mu \end{aligned}$$

כאשר במעבר האחרון השתמשנו בכך שהפונקציה $y f_X(y)$ היא אי-זוגית, ולכן האינטגרל שלה על הישר הוא 0, ובכך שבגורם הימני יש לנו אינטגרל של פ' צפיפות של $N(0,1)$. ■

טענה: יהי $X \sim N(\mu, \sigma^2)$. אזי $Var(X) = \sigma^2$.

הוכחה: אינטגרציה בחלקים. ■

הגדרה: מ"מ $Z \sim N(0,1)$ נקרא משנתה נורמלי סטנדרטי, וה-CDF שלו מסומנת $\Phi(z)$. עבור $0 \leq \alpha \leq 1$ נסמן גם ב- z_α את הערך המקיים $P(Z \leq z_\alpha) = \alpha$.

טענה: אם $Z \sim N(0,1)$ אז $aZ + b \sim N(b, a^2)$.

הוכחה: בפתרון תרגיל 8. ■

דוגמאות:

- רוצים לשדר מסר $x \in \{0,1\}$ דרך ערוץ עם רעש $Z \sim N(0,1)$, כלומר המקלט רואה $Y = x + Z$.
 - טרם הוכחנו, אבל אם $x = 1$ אז $Y \sim N(1,1)$.
 - המקלט מחליט על האות שנשלח כך: $g(Y) = \begin{cases} 1 & y > \frac{1}{2} \\ 0 & y < \frac{1}{2} \end{cases}$. מה הסתברות השגיאה כאשר שולחים את האות "1"?
 - המספר שהתקבל הוא ערך של ה-CDF של $Z \sim N(0,1)$, כלומר $\Phi(-\frac{1}{2})$.

מ"מ רציפים במשותף

הגדרה: מ"מ X, Y נקראים רציפים במשותף אם לכל $D \subseteq \mathbb{R}^2$ מתקיים: $P((x, y) \in D) = \iint_D f_{X,Y}(u, v) du dv$, כאשר $f_{X,Y}$ נקראת פונקציית הצפיפות המשותפת שלהם.

פונקציית הצפיפות המשותפת (Joint Probability Density Function) מקיימת את התכונות הבאות:

$$(1) \text{ לכל } u, v \text{ מתקיים } f_{X,Y}(u, v) \geq 0$$

$$(2) \text{ מתקיים } \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(u, v) du dv = 1$$

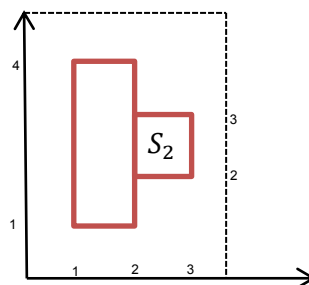
הגדרה: פונקציית הצפיפות השולית של X אם X, Y רציפים במשותף היא $f_X(u) = \int_{-\infty}^{\infty} f_{X,Y}(u, v) dv$.

הגדרה: יהיו (X, Y) מ"מ. הפונקציה $F_{X,Y}(u, v) = P(X \leq u, Y \leq v)$ נקראת פונקציית הצטברות משותפת (Joint Cumulative Distribution Function).

כאשר (X, Y) רציפים במשותף, מתקיים $F_{X,Y}(u, v) = \int_{-\infty}^u \int_{-\infty}^v f_{X,Y}(s, t) ds dt$. כאשר המ"מ בדידים, מתקיים $F_{X,Y}(u, v) = \sum_{x \leq u} \sum_{y \leq v} P_{X,Y}(x, y)$. אפשר לדבר על JCDF גם על וקטורים מקריים שאינם רציפים במשותף ואינם בדידים במשותף.

דוגמאות:

- פונקציית הצפיפות המשותפת של X, Y היא $f_{X,Y}(x, y) = \begin{cases} c & (x, y) \in S \\ 0 & \text{otherwise} \end{cases}$ כאשר הקבוצה S היא שני המלבנים האדומים בציר הבא:



- ניתן לחשב מכאן את ערכו של c :

$$1 = \iint_S f_{X,Y}(u, v) du dv = (4-1)(2-1)c + (3-2)(3-2)c = 4c \Rightarrow c = \frac{1}{4}$$

- כעת ניתן לחשב הסתברות של תת-קבוצה מסוימת:

$$\iint_{S_2} f_{X,Y}(u, v) du dv = \frac{1}{4}$$

- וניתן גם למצוא את ההסתברות השולית של X :

$$f_X(u) = \int_{-\infty}^{\infty} f_{X,Y}(u, v) dv = \begin{cases} 0 & u \notin [1, 3] \\ \frac{3}{4} & u \in [1, 2] \\ \frac{1}{4} & u \in [2, 3] \end{cases}$$

- המחט של בופון – דוגמה שניתנה בכיתה, אבל אפשר למצוא הסבר מלא בויקיפדיה.
- יהי $X \sim U(0, 1)$ ונגדיר $Y = X$. האם (X, Y) רציפים במשותף?

- נניח שקיימת פונקציית צפיפות משותפת $f_{X,Y}$ כנדרש מההגדרה. כעת נבחר $D = \{(u, v) : u = v\}$. נשים לב שמשקולי הסתברות, $P((X, Y) \in D) = 1$ כי $Y = X$. מאידך, האינטגרל על D הוא 0 כי זהו נפח של קו בטול שטח. זוהי סתירה לכך שפונקציה כזו בכלל קיימת.

- בכל מקרה, קיימת פונקציית הצטברות משותפת:

$$F_{X,Y}(u, v) = P(X \leq u, Y \leq v) = P(X \leq u, X \leq v) = P(X \leq \min\{u, v\}) = F_X(\min\{u, v\})$$

הגדרה: יהיו X, Y מ"מ רציפים במשותף ו- $g: \mathbb{R}^2 \rightarrow \mathbb{R}$, אז התוחלת המשותפת על $g(X, Y)$ מוגדרת להיות

$$E[g(X, Y)] = \iint_{\mathbb{R}^2} g(u, v) f_{X,Y}(u, v) du dv$$

ניתן להראות בקלות את כל העובדות שכבר ראינו על תוחלת, למשל ליניאריות בכמה משתנים, ועוד.

אפשר גם להכליל את הדיון על וקטור של מ"מ רציפים במשותף, $(X_1, \dots, X_n)$, שאז קיימת פונקציה $f_{X_1, \dots, X_n}$ שעבורה לכל $D \subseteq \mathbb{R}^n$ מתקיים

$$P((X_1, \dots, X_n) \in D) = \int \dots \int_D f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \dots dx_n$$

התניה על מ"מ רציפים

הגדרה: יהי X מ"מ רציף ו- A מאורע. פונקציית התפלגות מצטברת מותנית של X בהינתן המאורע A מוגדרת להיות $F_{X|A}(u) = P(X \leq u | A)$, ופונקציית הצפיפות המותנית מוגדרת להיות $f_{X|A}(u) = \frac{d}{du} F_{X|A}(u)$.

ההגדרה נבנתה כך שכעת מתקיים $P(X \in \Gamma | A) = \int_{\Gamma} f_{X|A}(u) du$ לכל $\Gamma \subseteq \mathbb{R}$.

טענה (נוסחת ההסתברות השלמה): אם $A_1, \dots, A_n$ חלוקה של Ω , אזי $f_X(x) = \sum_{i=1}^n P(A_i) f_{X|A_i}(x)$

הוכחה: ישירות מההגדרה:

$$\begin{aligned} f_X(x) &= \frac{d}{dx} F_X(x) = \frac{d}{dx} P(X \leq x) = \frac{d}{dx} \left\{ \sum_{i=1}^n P(A_i) P(X \leq x | A_i) \right\} = \\ &= \frac{d}{dx} \left\{ \sum_{i=1}^n P(A_i) F_{X|A_i}(x) \right\} = \sum_{i=1}^n P(A_i) f_{X|A_i}(x) \end{aligned}$$

■ כנדרש.

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X \in \Gamma)} & x \in \Gamma \\ 0 & x \notin \Gamma \end{cases} \quad \text{טענה: אם } A = \{X \in \Gamma\} \text{ עבור } \Gamma \text{ קטע כלשהו, אזי}$$

הוכחה: נסמן $\Gamma = [a, b]$ ונפתח מההגדרה:

$$f_{X|A}(x) = \frac{d}{dx} F_{X|A}(x) = \frac{d}{dx} P(X \leq x | A) = \frac{d}{dx} P(X \leq x | X \in \Gamma) = \frac{d}{dx} \frac{P(X \leq x, X \in \Gamma)}{P(X \in \Gamma)}$$

כעת לפנינו שלושה מקרים. אם $x < a$, המנה היא 0 ולכן גם נגזרתה 0. אם $x > b$, המנה היא 1 ולכן נגזרתה שוב 0. בחלק Γ יש לנו משהו מעניין:

$$x \in \Gamma \Rightarrow \frac{d}{dx} \frac{P(X \leq x, X \in [a, b])}{P(X \in \Gamma)} = \frac{d}{dx} \frac{P(X \in [a, x])}{P(X \in \Gamma)} = \frac{d}{dx} \frac{F_X(x) - F_X(a)}{P(X \in \Gamma)} = \frac{f_X(x)}{P(X \in \Gamma)}$$

■ שזה מה שרצינו.

ניתן להרחיב את הטענה ל- Γ שמהווה איחוד סופי של קטעים.

דוגמאות:

- מגיעים לתחנת הרכבת בין 7:10 ל-7:30. הרכבת מגיעה בדיוק בכל רבע שעה (7:00, 7:15 וכו'). נסמן את זמן ההגעה מ-7:10 ב- X דקות וניח $X \sim U(0,20)$. רוצים למצוא את התפלגות הזמן שנצטרך לחכות לרכבת, שנסמנו Y .
 - מתקיים: $Y = \begin{cases} 5-X & 0 \leq X \leq 5 \\ 20-X & 5 < X \leq 20 \end{cases}$. מהנתון, $f_X(x) = \frac{1}{20}$ עבור $x \in [0,20]$ ו-0 אחרת.
 - כעת נוכל לכתוב את פונקציית הצפיפות המותנית:

$$f_{X|X \leq 5}(x) = \begin{cases} \frac{f_X(x)}{P(X \in \Gamma)} & x \leq 5 \\ 0 & x > 5 \end{cases} = \begin{cases} \frac{1}{5} & x \leq 5 \\ 0 & x > 5 \end{cases}$$

$$f_{X|X \leq 5}(x) = \begin{cases} \frac{1}{15} & 5 < x \leq 20 \\ 0 & x \leq 5 \end{cases}$$

- עכשיו אפשר לעבור לטפל במ"מ המקורי Y – נסמן $A = \{X \leq 5\}$ וננסה להשתמש במשפט שלמה מתקיים כאן:
- $$f_Y(y) = f_{Y|A}(y)P(A) + f_{Y|\bar{A}}(y)P(\bar{A})$$
- נפתח את שני החלקים:

$$\begin{aligned} f_{Y|A}(y) &= \frac{d}{dy} F_{Y|A}(y) = \frac{d}{dy} P(Y \leq y|A) = \frac{d}{dy} P(5 - X \leq y|A) = \frac{d}{dy} P(X \geq 5 - y|A) = \\ &= \frac{d}{dy} \{1 - F_{X|A}(5 - y)\} = -(-f_{X|A}(5 - y)) = f_{X|A}(5 - y) \end{aligned}$$

$$f_{Y|\bar{A}}(y) = f_{X|\bar{A}}(5 - y)$$

- זה נותן לנו אפיון מלא של הפונקצייה, כיוון שאנו יודעים את f_X המותנית.

כעת אנו רוצים להרחיב את ההגדרה עבור מ"מ ולא מאורעות, למשל $P(X \in [2,3] | Y = 8)$. הבעיה היא שבהגדרה יש לנו $P(\{Y = 8\})$ במקנה, וזהו 0. לכן נזדקק להגדרה קצת עדינה יותר, שיהיה צריך להראות שתמיד עובדת:

הגדרה: יהיו X, Y מ"מ רציפים, ההסתברות של X בהינתן Y מוגדרת להיות:

$$.P(X \in \Gamma|Y = v) = \lim_{\varepsilon \rightarrow 0^+} \frac{P(X \in \Gamma, |Y - v| \leq \varepsilon)}{P(|Y - v| \leq \varepsilon)}$$

נראה שההגדרה כשרה, כלומר שהגבול תמיד קיים:

$$\begin{aligned} P(X \in \Gamma | Y = v) &= \lim_{\varepsilon \rightarrow 0^+} \frac{P(X \in \Gamma, |Y - v| \leq \varepsilon)}{P(|Y - v| \leq \varepsilon)} = \lim_{\varepsilon \rightarrow 0^+} \frac{\int_{\Gamma} \int_{v-\varepsilon}^{v+\varepsilon} f_{X,Y}(s, t) dt ds}{\int_{v-\varepsilon}^{v+\varepsilon} f_Y(s, t) ds} = \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{2\varepsilon \int_{\Gamma} f_{X,Y}(s, v) ds + o(\varepsilon)}{2\varepsilon f_Y(v) + o(\varepsilon)} = \int_{\Gamma} \frac{f_{X,Y}(s, v)}{f_Y(v)} ds := \int_{\Gamma} f_{X|Y}(s, v) ds \end{aligned}$$

כאשר כאן ε הוא מספר חיובי קטן, $\varepsilon \rightarrow 0^+$, האינטגרל על קטע באורך 2ε של פונקציה רציפה מתקרב כדי $O(\varepsilon)$ ל- 0 . כדי להיפטר מ- $O(\varepsilon)$, ניתן לחלק את המונה והמכנה ב- ε , ואז לפי ההגדרה, $\lim_{\varepsilon \rightarrow 0^+} \frac{O(\varepsilon)}{\varepsilon} = 0$.

דוגמאות:

- במשחק קליעה למטרה נגדיר (X, Y) מ"מ המייצגים את קואורדינטות הפגיעה של החץ, ונניח שהם מפולגים יוניפורמית על המעגל D סביב הראשית ברדיוס r .

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi r^2} & x^2 + y^2 \leq r^2 \\ 0 & x^2 + y^2 > r^2 \end{cases}, \text{ כלומר } \circ$$

- נמצא את הצפיפות המותנית $f_{X|Y}(u, v) = \frac{f_{X,Y}(u,v)}{f_Y(v)}$, ולשם כך רק צריך למצוא את הצפיפות השולית של Y .

$$f_Y(v) = \int_{-\sqrt{r^2-v^2}}^{\sqrt{r^2-v^2}} f_{X,Y}(s,v) ds = \frac{2}{\pi r^2} \sqrt{r^2-v^2}$$

- שוברים מקל באורך L באופן מקרי ואחיד בנקודה מסוימת, נסמן Y אורך המקל שנותר. כעת שוברים את המקל שנותר שוב באופן מקרי ואחיד, ונסמן X אורך המקל הסופי. רוצים למצוא את הצפיפות המשותפת $f_{X,Y}$.
 - ראשית נשים לב ש- $Y \sim U(0, L)$, כלומר $f_Y(y) = \frac{1}{L}$ לכן $F_Y(y) = \frac{y}{L}$ וגם $E[Y] = \frac{L}{2}$.
 - כמו כן, $X|Y \sim U(0, Y)$. כעת מהגדרת צפיפות מותנית אנו יכולים למצוא את הצפיפות המשותפת:

$$f_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \Rightarrow f_{X,Y}(x,y) = f_{X|Y}(x,y)f_Y(y) = \begin{cases} \frac{1}{L} \frac{1}{y} & 0 \leq x \leq y \leq L \\ 0 & \text{otherwise} \end{cases}$$
 - כעת אפשר למצוא את ההסתברות השולית של X :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy = \int_x^L \frac{1}{Ly} dy = \frac{1}{L} [\log(y)]_x^L = \begin{cases} \frac{1}{L} (\log L - \log x) & 0 \leq x \leq L \\ 0 & \text{otherwise} \end{cases}$$
 - ונעת אפשר לחשב את התוחלת של X , נשתמש לשם כך שתוחלת מותנית:

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x,y) dx = \frac{y}{2} \Rightarrow E[X|Y] = \frac{Y}{2} \Rightarrow E[E[X|Y]] = E\left[\frac{Y}{2}\right] = \frac{1}{2} E[Y] = \frac{L}{4}$$
 - באופן דומה אפשר לחשב את $E\left[\frac{Y^2}{2}\right]$ וגם את השונות המותנית.

$$\text{טענה (חוק בייס הרציף): יהיו } X, Y \text{ מ"מ רציפים, אזי } f_{X|Y}(u,v) = \frac{f_{Y|X}(v,u)f_X(u)}{f_Y(v)}$$

$$\text{בגרסה אחרת, אפשר לכתוב במפורש את ההתפלגות השולית של } Y \text{ ולקבל: } f_{X|Y}(u,v) = \frac{f_{Y|X}(v,u)f_X(u)}{\int_{-\infty}^{\infty} f_{Y|X}(v,t)f_X(t)dt}$$

דוגמאות:

- לנורה יש אורך חיים אקספוננציאלי עם פרמטר λ , כאשר $\Lambda \sim U(1, 1.5)$. בוחנים את המנורה ומודדים אורך חיים y ; מה ניתן לומר עכשיו על Λ ?
 - נתחיל במציאת הצפיפות המותנית:

$$f_{\Lambda|Y}(\lambda, y) = \frac{f_{Y|\Lambda}(y, \lambda)f_{\Lambda}(\lambda)}{f_Y(y)} = \frac{2\lambda e^{-\lambda y}}{f_Y(y)} = \frac{\lambda}{e^{\lambda y}} C$$

- כעת אפשר למצוא את המקסימום של הפונקציה כדי לראות מה ה- λ הסביר ביותר בהינתן y שקיבלנו מהמידה:

$$\frac{d}{d\lambda} f_{\Lambda|Y}(\lambda, y) = c[e^{-\lambda y} + \lambda e^{-\lambda y}(-y)] = 0 \Leftrightarrow ce^{-\lambda y}[1 + \lambda y] = 0 \Leftrightarrow \lambda = \frac{1}{y}$$

$$\text{טענה (כלל בייס המעורב): יהיו } X \text{ מ"מ בדיד, } Y \text{ מ"מ רציף, אז מתקיים: } P_{X|Y}(x,y) = \frac{f_{Y|X}(y,x)P(X=x)}{f_Y(y)}$$

הוכחה: באופן דומה למה שעשינו עבור שני מ"מ רציפים:

$$\begin{aligned} P_{X|Y}(x,y) &= \lim_{\varepsilon \rightarrow 0^+} \frac{P(X=x, |Y-y| \leq \varepsilon)}{P(|Y-y| \leq \varepsilon)} = \lim_{\varepsilon \rightarrow 0^+} \frac{P(X=x, y-\varepsilon \leq Y \leq y+\varepsilon)}{P(y-\varepsilon \leq Y \leq y+\varepsilon)} = \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{\int_{y-\varepsilon}^{y+\varepsilon} f_{Y|X}(t,x)P(X=x)dt}{\int_{y-\varepsilon}^{y+\varepsilon} f_Y(t)dt} = \lim_{\varepsilon \rightarrow 0^+} \frac{2\varepsilon f_{Y|X}(t,y)P(X=x) + o(\varepsilon)}{2\varepsilon f_Y(y) + o(\varepsilon)} = \\ &= \frac{f_{Y|X}(y,x)P(X=x)}{f_Y(y)} \end{aligned}$$

כשגם כאן השתמשנו באותו רעיון של אינטגרל על קטע מצטמצם סביב נקודה. ■

הערה: אפשר לכתוב את המכנה בעזרת נוסחת הסתברות שלמה: $f_Y(y) = \sum_k f_{Y|X}(y, k)P(X = k)$ כדי להימנע מהצורך לחשב ישירות את הצפיפות השולית.

$$f_{Y|X}(y, x) = \frac{P_{X|Y}(x, y)f_Y(y)}{\int_{-\infty}^{\infty} P_{X|Y}(x, t)f_Y(t)dt} \quad \text{טענה (כלל בייס המעורב, 2): יהיו } X \text{ מ"מ בדיד, } Y \text{ מ"מ רציף, אז מתקיים:}$$

הוכחה: בפתרון תרגיל 8. במכנה מופיעה, למעשה, ההסתברות השולית $P_X(x)$. ■

דוגמאות:

- יהי $X \sim \text{Ber}(p)$ ו- $Y|X \sim \exp(X+1)$, ומחפשים את $P_{X|Y}(x, y)$:

$$P_{X|Y}(x, y) = \begin{cases} \frac{f_{Y|X}(y, 0)P(X=0)}{f_{Y|X}(y, 0)P(X=0) + f_{Y|X}(y, 1)P(X=1)} & x=0 \\ \frac{f_{Y|X}(y, 1)P(X=1)}{f_{Y|X}(y, 0)P(X=0) + f_{Y|X}(y, 1)P(X=1)} & x=1 \end{cases} = \begin{cases} \frac{e^{-y}(1-p)}{e^{-y}(1-p) + 2e^{-2y}p} & x=0 \\ \frac{2e^{-2y}p}{e^{-y}(1-p) + 2e^{-2y}p} & x=1 \end{cases}$$
- יהי $Y \sim \exp(1)$ ו- $X|Y \sim \text{Geo}(1 - e^{-y})$. מחפשים את פונקציית הצפיפות $f_{Y|X}(y, x)$:

$$f_{Y|X}(y, x) = \frac{d}{dy} F_{Y|X}(y, x) = \frac{d}{dy} P(Y \leq y | X = x) = \frac{d}{dy} \frac{P(Y \leq y, X = x)}{P(X = x)} = \frac{1}{P(X = x)} \frac{d}{dy} P(Y \leq y, X = x) = \dots = \frac{P_{X|Y}(x, y)f_Y(y)}{\int_{-\infty}^{\infty} P_{X|Y}(x, t)f_Y(t)dt} = \frac{(1 - e^{-y})(e^{-y})^{x-1}e^{-y}}{\int_0^{\infty} (1 - e^{-t})(e^{-t})^{x-1}e^{-t}dt} = x(x+1)(1 - e^{-y})e^{-yx}$$

○ כאשר כאן המעבר האחרון הוא שלב חישובי לא מעניין, והמעבר המסומן נובע מזהות (חוק בייס המעורב – מהצד השני) שראינו בתרגיל.

הגדרה: תוחלת מותנית של מ"מ רציפים X, Y מוגדרת ע"י $E[X|Y = y] = \int_{-\infty}^{\infty} xf_{X|Y}(x, y)dx$.

הגדרה: מ"מ רציפים X, Y נקראים בלתי-תלויים אם $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

מהגדרה זו אפשר לקבל גם הסתברות משותפת:

$$P(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x, y)dy dx = \int_A \int_B f_X(x)f_Y(y)dy dx = \int_A f_X(x) \int_B f_Y(y)dy dx = \left(\int_A f_X(x)dx \right) \left(\int_B f_Y(y)dy \right) = P(X \in A)P(Y \in B)$$

ולכן האי-תלות שקולה בהגדרה גם לגבי התפלגות מצטברת משותפת: $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

טענה: מ"מ X, Y הם בלתי-תלויים אםם לכל זוג פונקציות φ, ψ מתקיים $E[\varphi(X)\psi(Y)] = E[\varphi(X)]E[\psi(Y)]$.

הוכחה: בפתרון תרגיל 9. ■

דוגמאות:

- בהתקפת קניות שולה מבקרת במספר מקרי N של חנויות, ובחנות ה- i מבזבזת סכום מקרי X_i . נניח ש- N הוא שלם חיובי עם פונקציית התפלגות ידועה ו- X_i כולם בעלי אותה תוחלת μ ושונות σ^2 , וכל המ"מ בשאלה ב"ת. מהי תוחלת עלות מסע הקניות?
 - נסמן ב- T את סך הכסף ששולה הוציאה, $T = \sum_{i=1}^N X_i$. כאן מתקיים:

$$E[T|N = n] = E\left[\sum_{i=1}^n X_i \mid N = n\right] = \sum_{i=1}^n E[X_i|N = n] = \sum_{i=1}^n E[X_i] = n\mu$$

$$E[T] = E[E[T|N]] = E[N\mu] = \mu E[N]$$

$$E[T^2] = \sum_{n=1}^{\infty} P(N=n)E[T^2|N=n] = \sum_{n=1}^{\infty} P(N=n)E[(X_1 + \dots + X_n)^2|N=n] =$$

$$= \sum_{i=1}^{\infty} P(N=n)E[nX_1^2 + n(n-1)X_1X_2]$$

○ המעבר הזה נכון כיוון שהמ"מ כולם אותו דבר, ויש לנו רק שני סוגים שונים של גורמים – ריבועים של אותו מ"מ, ומכפלות של שני מ"מ שונים. בשני המקרים תוחלות שונות כי X_i, X_j כאן ב"ת אדם $i \neq j$. אפשר להמשיך מכאן:

$$= \sum_{n=1}^{\infty} P(N=n) \left[E[nX_1^2|N=n] + E[n(n-1)X_1X_2|N=n] \right] = \sum_{i=1}^{\infty} P(N=n) (n\sigma^2 + n(n-1)\mu^2)$$

• יהיו שני מ"מ $X_1 \sim \exp(\lambda_1), X_2 \sim \exp(\lambda_2)$ ב"ת, ונסמן $Y = \min\{X_1, X_2\}$.

○ נמצא את ההתפלגות שלו:

$$F_Y(y) = P(Y \leq y) = P(\min\{X_1, X_2\} \leq y) = 1 - P(\min\{X_1, X_2\} > y) = 1 - P(X_1 > y, X_2 > y) =$$

$$= 1 - P(X_1 > y)P(X_2 > y) = 1 - (1 - F_{X_1}(y))(1 - F_{X_2}(y))$$

○ אפשר גם לבדוק מי המינימום (גם כאן משתמשים באי-תלות):

$$P(X_1 \leq X_2) = \int_0^{\infty} \int_0^{x_2} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = \int_0^{\infty} \int_0^{x_2} f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = \int_0^{\infty} \lambda_2 e^{-\lambda_2 x_2} (1 - e^{-\lambda_1 x_2}) dx_2 =$$

$$= 1 - \int_0^{\infty} \lambda_2 e^{-(\lambda_1 + \lambda_2)x_2} dx_2 = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} \int_0^{\infty} (\lambda_1 + \lambda_2) e^{-(\lambda_1 + \lambda_2)x_2} dx_2 = 1 - \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

טענה (נוסחת השונות השלמה): יהיו X, Y מ"מ. אזי $Var(Y) = E[Var(Y|X)] + Var(E[Y|X])$.

הוכחה: בפתרון תרגיל 8. ■

דוגמאות:

• יהיו $X \sim N(0,1)$ ו- $Y|X \sim U(0, X^2)$.

○ נחשב את התוחלת: $E[Y] = E[E[Y|X]] = E\left[\frac{X^2}{2}\right] = \frac{1}{2}E[X^2] = \frac{1}{2}(Var(X) + E[X]^2) = \frac{1}{2}$

○ נחשב את השונות המותנית:

$$Var(Y|X) = E[(Y - E[Y|X])^2|X] = E\left[\left(Y - \frac{X^2}{2}\right)^2 \middle| X\right] = E\left[Y^2 - \frac{2YX^2}{2} + \frac{X^4}{4} \middle| X\right] = E[Y^2|X] + X^2 E[Y|X] + \frac{X^4}{4} =$$

$$= E[Y^2|X] - X^2 \frac{X^2}{2} + \frac{X^4}{4} = E[Y^2|X] - \frac{X^4}{4}$$

○ נחשב את התוחלת המותנית של המומנט השני:

$$E[Y^2|X=x] = \int_0^{x^2} \frac{1}{x^2} y^2 dy = \frac{1}{x^2} \left[\frac{y^3}{3} \right]_0^{x^2} = \frac{1}{x^2} \frac{x^6}{3} = \frac{x^4}{3}$$

$$E[Y^2|X] = \frac{X^4}{3} \Rightarrow Var(Y|X) = \frac{X^4}{3} - \frac{X^4}{4} = \frac{X^4}{12}$$

○ לבסוף, נציב בנוסחת השונות השלמה:

$$Var(Y) = E[Var(Y|X)] + Var(E[Y|X]) = E\left[\frac{X^4}{12}\right] + Var\left(\frac{X^2}{2}\right)$$

○ מציאת הערכים האלה כרוכה בחישוב אינטגרלים קצת לא סימפטיים, ובכל מקרה בסוף מקבלים מספר.

טרנספורמציות על משתנים מקריים

יהיו X מ"מ, $Y = g(X)$, ונתונה $f_X(x)$. רוצים למצוא את $f_Y(y)$ וכל מה שנובע מזה. כדי לעשות זאת, נעבור דרך ההתפלגות המצטברת: $F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{x: g(x) \leq y} f_X(x) dx$. במקרים מסוימים, זה יהיה מספיק פשוט כדי שנוכל לחשבו, לגזור את התוצאה ולהגיע ל- $f_Y(y)$.

דוגמאות:

- יהיו $X \sim U(0,1)$ ו- $Y = \sqrt{X}$. נמצא את $f_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} P(\sqrt{X} \leq y) = \frac{d}{dy} P(X \leq y^2) = \frac{d}{dy} F_X(y^2) = \frac{d}{dy} y^2 = 2y$$
- משה נוסע מרחק של 180 ק"מ במהירות קבועה $V \sim U(30,60)$. מה פונקציית הצפיפות של משך נסיעתו של משה, שנסמנה כ- T ?

○ כמו קודם, נעבור דרך פונקציית ההתפלגות המצטברת:

$$f_T(t) = \frac{d}{dt} F_T(t) = \frac{d}{dt} P(T \leq t) = \frac{d}{dt} P\left(\frac{180}{V} \leq t\right) = \frac{d}{dt} P\left(V \geq \frac{180}{t}\right) = \frac{d}{dt} \left\{1 - P\left(V \leq \frac{180}{t}\right)\right\} =$$

$$= \frac{d}{dt} \left\{1 - F_V\left(\frac{180}{t}\right)\right\} = \frac{d}{dt} \left\{1 - \frac{\frac{180}{t} - 30}{30}\right\} = \frac{d}{dt} \begin{cases} 0 & t \leq 3 \\ 2 - \frac{6}{t} & 3 \leq t \leq 6 \\ 1 & t \geq 6 \end{cases}$$

○ כלומר, עבור הערכים המעניינים של t , מקבלים $f_T(t) = \frac{6}{t^2}$.

- יהי $Y = g(X) = X^2$ ונתונה $f_X(x)$. כאן מתקיים, באופן כללי:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} P(X^2 \leq y) = \frac{d}{dy} P(-\sqrt{y} \leq X \leq \sqrt{y}) =$$

$$= \frac{d}{dy} \{F_X(\sqrt{y}) - F_X(-\sqrt{y})\} = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y}))$$
- יהיו $X \sim U(-1,1)$ ו- $Y = \sqrt{|X|}$. הטרנספורמציה הזאת אינה מונוטונית ולכן נצטרך למצוא את מבוקשו ידנית:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(Y \leq y) = \frac{d}{dy} P(\sqrt{|X|} \leq y) = \frac{d}{dy} P(|X| \leq y^2) = \frac{d}{dy} P(-y^2 \leq X \leq y^2) = \frac{d}{dy} \frac{2y^2}{2} = 2y$$

במקרה של טרנספורמציה מונוטונית ממש, יש ל- g הופכית, ומכך אפשר לקבל נוסחא כללית:

טענה: יהי X מ"מ, $Y = g(X)$ כאשר g מונוטונית ממש. אזי $f_Y(y) = f_X(g^{-1}(y)) \left \frac{d}{dy} g^{-1}(y) \right$

הוכחה: נניח ש- g יורדת (הכיוון השני סימטרי):

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} P(g(X) \leq y) = \frac{d}{dy} P(X \geq g^{-1}(y)) = \frac{d}{dy} \{1 - F_X(g^{-1}(y))\} =$$

$$= -f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

■ במעבר האחרון השתמשנו בכך ש- g יורדת ולכן גם g^{-1} יורדת ולכן הנגזרת שלה שלילית.

דוגמאות:

- רומיאו ויוליה קובעים להיפגש. כל אחד מאחר באופן ב"ת בזמן אקספוננציאלי עם פרמטר λ . מה הצפיפות של הפרש הזמנים בהם הם מגיעים?

○ נסמן X, Y זמני ההגעה של האוהבים ו- $Z = X - Y$ ההפרש המבוקש. נתחיל מהמקרה שההפרש חיובי:

$$F_Z(z) = P(X - Y \leq z) = 1 - P(X - Y \geq z) = 1 - \int_0^\infty \int_{z+y}^\infty f_{X,Y}(x,y) dx dy = 1 - \int_0^\infty f_Y(y) \int_{z+y}^\infty f_X(x) dx dy =$$

$$= 1 - \int_0^\infty \lambda e^{-\lambda y} \int_{z+y}^\infty \lambda e^{-\lambda x} dx dy = 1 - \int_0^\infty \lambda e^{-\lambda y} [-e^{-\lambda x}]_{z+y}^\infty dy = 1 - \int_0^\infty \lambda e^{-\lambda y} e^{-\lambda(z+y)} dy =$$

$$= 1 - e^{-\lambda z} \int_0^\infty \lambda e^{-2\lambda y} dy = 1 - e^{-\lambda z} \cdot \frac{1}{2}$$

$$f_Z(z) = \frac{d}{dz} \left\{1 - e^{-\lambda z} \cdot \frac{1}{2}\right\} = \frac{\lambda}{2} e^{-\lambda z}$$

○ במקרה שההפרש שלילי, אפשר לראות משיקולי סימטריה שמקבלים:

$$F_Z(z) = P(Z \leq z) = P(-Z \geq -z) = P(Z \geq -z) = 1 - F_Z(-z) = 1 - \left(1 - \frac{e^{\lambda z}}{2}\right) = \frac{e^{\lambda z}}{2}$$

$$f_Z(z) = \frac{d}{dz} \frac{e^{\lambda z}}{2} = \frac{\lambda}{2} e^{\lambda z}$$

○ זאת משום של- $(X - Y)$ ול- $(Y - X)$ צריכה להיות אותה התפלגות – סימטריה מוחלטת בין המ"מ.

שונוות משותפת

המטרה היא לנסח מדד לפיזור משותף של משתנים מקריים (התנהגות משותפת) המאפשר באופן מספרי להבין האם יש התאמה מסוימת בין אופן השינוי של המשתנים.

הגדרה: השונוות המשותפת (Covariance) של מ"מ X, Y היא: $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$.

הגדרה: X, Y נקראים בלתי-מתואמים אם $Cov(X, Y) = 0$.

נשים לב שמתקיים: $Cov(X, X) = E[(X - E[X])(X - E[X])] = Var(X)$.

טענה: יהיו X, Y מ"מ ב"ת. אזי $Cov(X, Y) = 0$.

הוכחה: נחשב ישירות ונשתמש באי-תלות כדי לפרק את תוחלת המכפלה:

$$Cov(X, Y) = E[XY - E[X]Y - E[Y]X + E[X]E[Y]] = E[XY] - E[X]E[Y] = 0$$

■ **כנדרש.**

ההיפך לא נכון: ייתכנו מ"מ בלתי-מתואמים שאינם בלתי-תלויים.

טענה: יהיו X, Y, Z מ"מ. אזי $Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$.

הוכחה: ישירות מההגדרה:

$$\begin{aligned} Cov(X + Y, Z) &= E[((X + Y) - E[X + Y])(Z - E[Z])] = \\ &= E[(X - E[X])(Z - E[Z]) + (Y - E[Y])(Z - E[Z])] = \\ &= E[(X - E[X])(Z - E[Z])] + E[(Y - E[Y])(Z - E[Z])] = Cov(X, Z) + Cov(Y, Z) \end{aligned}$$

■ **כנדרש.**

טענה: יהיו X, Y מ"מ. אזי $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$.

הוכחה: ישירות מההגדרה:

$$\begin{aligned} Var(X + Y) &= E[(X + Y - E[X + Y])^2] = E[(X - E[X] + Y - E[Y])^2] = \\ &= E[(X - E[X])^2 + (Y - E[Y])^2 + 2E[(X - E[X])(Y - E[Y])]] = \\ &= Var(X) + Var(Y) + 2Cov(X, Y) \end{aligned}$$

■ **כנדרש.**

הגדרה: יהיו X, Y מ"מ. מקדם המתאם שלהם מוגדר להיות $\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$.

מקדם המתאם מקיים כמה תכונות טובות:

(1) $-1 \leq \rho(X, Y) \leq 1$, כלומר הוא תמיד נותן חיווי כמותי על "מידת ההתאמה"

(2) $X = aY + b \Leftrightarrow |\rho(X, Y)| = 1$

(3) הוא חסר יחידות (היחידות של המונה והמכנה מבטלות אלה את אלה)

דוגמאות:

- בכד 3 כדורים ממוספרים, ומוציאים 2 כדורים. X – מספר הכדור הראשון, Y – מספר הכדור השני. נמצא את השונות המשותפת:

$$E[XY] = \frac{1}{6} \cdot 1 \cdot 2 + \frac{1}{6} \cdot 1 \cdot 3 + \frac{1}{6} \cdot 2 \cdot 3 + \frac{1}{6} \cdot 2 \cdot 1 + \frac{1}{6} \cdot 3 \cdot 1 + \frac{1}{6} \cdot 3 \cdot 2 = \frac{11}{3}$$

$$Cov(X, Y) = E[XY] - E[X]E[Y] = \frac{11}{3} - 2 \cdot 2 = -\frac{1}{3}$$

- ברשותנו n כדורים ממוספרים שבוחרים כמה מתוכם, ו- X_i הכדור ה- i שהוצאנו. ניתן להוכיח שכל המ"מ שווי התפלגות ללא קשר למספר הכדורים שמוציאים.

- נתחיל בחישוב התוחלת: X_i מ"מ אחיד, ולכן $E[X_i] = \frac{n+1}{2}$, $Var(X_i) = \frac{n^2-1}{12}$
- נשתמש בהכללת הקשר שהוכחנו קודם ונקבל:

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) = \frac{n(n^2-1)}{12} + n(n-1)Cov(X_1, X_2)$$

- כעת נשתמש בכך ש- $Var(\sum_{i=1}^n X_i) = 0$, כי הסכום הזה הוא קבוע (הוצאנו את כל הכדורים!), ולכן ניתן לבדוד את השונות המשותפת ולקבל $Cov(X_1, X_2) = -\frac{n+1}{12}$.
- התוצאה הזאת לא משמחת כל כך, כי ככל שיש יותר כדורים, נראה שהשונות המשותפת גדלה – וזה לא אינטואיטיבי, כיוון שככל שיש יותר כדורים, השפעת הוצאת כדור אחד "אמורה" להשפיע פחות על הוצאת כדור שני. אם נסתכל על מקדם המתאם, $\rho(X_1, X_2) = -\frac{\frac{n+1}{12}}{\frac{n^2-1}{12}} = -\frac{1}{n-1}$, מה שיותר נוח אינטואיטיבית.

- מבצעים n הטלות מטבע עם סיכוי p לראש, כאשר X – מספר הראשים, Y – מספר הזנבות. כמה הם מתואמים?
 $X + Y = n, E[X + Y] = n \Rightarrow (X - E[X]) + (Y - E[Y]) = 0 \Rightarrow Cov(X, Y) = E[-(X - E[X])^2] = -Var(X)$
 $\rho(X, Y) = -\frac{Var(X)}{\sqrt{Var(X)Var(Y)}} = -1$

- n אנשים זורקים באוויר את כובעיהם ותופסים כובע באקראי. נגדיר X – מספר האנשים שתפסו את כובעם, ראינו כבר שהתוחלת של מ"מ זה היא 1. כעת נטפל גם בשונות – נגדיר X_i – מ"מ ברנולי המקבל 1 כאשר האדם ה- i קיבל את כובעו, ונפתח לפי מה שראינו (ונזכור שכל המ"מ סימטריים):

$$Var(X) = Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j) = n \cdot \frac{1}{n} \left(1 - \frac{1}{n}\right) + n(n-1)(E[X_i X_j] - E[X_i]E[X_j]) =$$

$$= \left(1 - \frac{1}{n}\right) + n(n-1) \left(1 \cdot P(X_i = 1, X_j = 1) - \frac{1}{n^2}\right) = \left(1 - \frac{1}{n}\right) + n(n-1) \left(\frac{1}{n(n-1)} - \frac{1}{n^2}\right) = 1 - \frac{1}{n} + 1 - \frac{n-1}{n} = 1$$

- נתונים שני מ"מ, X, Y . "מנחשים" ש- Y הוא בערך פונקציה ליניארית של X . נבנה מודל עם $\hat{Y} = aX + b$. המטרה היא שהשונות תהיה מינימלית, כלומר למזער את $Var(Y - \hat{Y})$.

- נתבונן בביטוי הזה:

$$Var(Y - (aX + b)) = Var(Y - aX) = E\left[\left((Y - aX) - E[Y - aX]\right)^2\right] = E\left[\left((Y - E[Y]) - a(X - E[X])\right)^2\right] =$$

$$= Var(Y) - 2aCov(X, Y) + a^2Var(X)$$

- כעת נגזור את הפונקציה שקיבלנו כדי למצוא מינימום:

$$\frac{d}{da}\{Y - (aX + b)\} = -2Cov(X, Y) + 2aVar(X)$$

$$\hat{a} = \frac{Cov(X, Y)}{Var(X)}$$

- יהיו $T = Y - X$ נגדיר $X \sim U(0, 1), Y \sim U(0, 1)$. נמצא את הצפיפות של T . לשם כך נצטרך לחשב אינטגרל: $P(T \leq t) = P(Y - X \leq t) = \iint f_{X,Y}(x, y) dy dx$ על קבוצת הנקודות שעבורן $y \leq t + x$. את זה נצטרך לעשות בשני אזורים שונים:
 ■ עבור $-1 \leq t \leq 0$

$$\int_{-t}^0 \int_0^{t+x} f_{X,Y}(x,y) dy dx = \int_{-t}^0 \int_0^{t+x} f_X(x) f_Y(y) dy dx = \int_{-t}^0 \int_0^{t+x} 1 dy dx = \frac{t^2}{2} + t$$

■ עבור $0 \leq t \leq 1$ נעדיף לחשב את השטח של המשולש מעל האזור שמעניין אותנו:

$$1 - P(Y - X \geq t) = 1 - \int_0^{1-t} \int_{t+x}^1 dy dx = t - \frac{t^2}{2} + \frac{1}{2}$$

$$f_T(t) = \begin{cases} 1+t & -1 \leq t \leq 0 \\ 1-t & 0 \leq t \leq 1 \end{cases}$$

○ את מה שקיבלנו אפשר עכשיו לגזור: עכשיו נמצא את מקדם המתאם של X, T - נתחיל מהשונויות המשותפת:

$$Cov(X, T) = E[XT] - E[X]E[T] = E[X(Y - X)] - E[X] \cdot 0 = E[XY] - E[X^2] = E[X]E[Y] - E[X^2] = \frac{1}{2} \cdot \frac{1}{2} - \frac{1}{3} = -\frac{1}{12}$$

$$\rho(X, T) = \frac{-\frac{1}{12}}{\sqrt{\frac{1}{126}}} = -\frac{1}{\sqrt{2}} \quad \text{○ נותר לחשב את } Var(T) = Var(X - Y) = Var(X) + Var(Y) = \frac{1}{6}$$

• מטילים קובייה הוגנת n פעמים. נסמן X_1 - מספר תוצאות "1", X_2 - מספר תוצאות "2". רוצים למצוא את מקדם המתאם בין שני המשתנים.

$$E[X_1] = E[X_2] = \frac{n}{6}, Var(X_1) = Var(X_2) = \frac{5n}{36} \quad \text{○ ראשית נשים לב ש-} X_1, X_2 \sim Bin\left(n, \frac{1}{6}\right) \text{ ולכן}$$

$$\text{○ עוד נשים לב שאם נסמן } Y = X_1 + X_2 \text{ אז } Y \sim Bin\left(n, \frac{1}{3}\right) \text{ ומצד שני } Y = X_1 + X_2 \text{ ולכן אנו יודעים איך}$$

מתפלג הסכום ויכולים לבדוד את השונויות המשותפת:

$$\frac{8n}{36} = Var(Y) = Var(X_1 + X_2) = Var(X_1) + Var(X_2) + Cov(X_1, X_2) \Rightarrow Cov(X_1, X_2) = -\frac{n}{36}, \rho(X_1, X_2) = -\frac{1}{5}$$

פונקציה יוצרת מומנטים

הגדרה: יהי X מ"מ. פונקציה יוצרת מומנטים שלו מוגדרת להיות $M(s) = E[e^{sX}]$.

דוגמאות:

• $X \sim Pois(\lambda)$ והפונקציה יוצרת מומנטים שלו היא:

$$M(s) = \sum_{x=0}^{\infty} e^{sx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^s)^x}{x!} = e^{-\lambda} e^{\lambda e^s} = e^{\lambda(e^s - 1)}$$

• $X \sim \exp(\lambda)$ והפונקציה יוצרת מומנטים שלו היא:

$$M(s) = \int_0^{\infty} e^{sx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{x(s-\lambda)} dx = \lambda \left[\frac{e^{s-\lambda}}{s-\lambda} \right]_0^{\infty} = \frac{\lambda}{\lambda - s}$$

נוכל להשתמש בפונקציה יוצרת מומנטים כדי לחשב את המומנטים של מ"מ יחסית בקלות. ובכן:

טענה: יהי X מ"מ, $M(s)$ פונקציה יוצרת מומנטים שלו. אזי $E[X^k] = M^{(k)}(0)$.

הוכחה: נתחיל מהמומנט הראשון (התוחלת):

$$\frac{d}{ds} M(s) = \frac{d}{ds} \int_{-\infty}^{\infty} e^{sx} f_X(x) dx = \int_{-\infty}^{\infty} x e^{sx} f_X(x) dx \Rightarrow \frac{d}{ds} M(0) = E[X]$$

באופן דומה, אפשר להסתכל על הנגזרת השנייה:

$$M''(s) = \int_{-\infty}^{\infty} x^2 e^{sx} f_X(x) dx \Rightarrow M''(0) = E[X^2]$$

■ באופן כללי ניתן להראות באינדוקציה שמתקיים $M^{(k)}(0) = E[X^k]$ כנדרש.

משפט: (ללא הוכחה) תחת תנאי חלקות מסוימים, המומנטים (מכל סדר) של מ"מ X קובעים את f_X ביחידות.

כלומר, אם אנו יודעים את הפונקציה יוצרת מומנטים של מ"מ מסוים, אנו בעצם יודעים מה ההתפלגות שלו.

דוגמאות:

- יהיו $X \sim \text{Pois}(\lambda)$, $Y \sim \text{Pois}(\mu)$ ב"ת ו- $Z = X + Y$. נסתכל על הפונקציה יוצרת מומנטים שלו:

$$M_Z(s) = E[e^{sZ}] = E[e^{s(X+Y)}] = E[e^{sX}e^{sY}] = E[e^{sX}]E[e^{sY}] = M_X(s)M_Y(s) = e^{\lambda(e^s-1)}e^{\mu(e^s-1)} = e^{(\lambda+\mu)(e^s-1)}$$
 - המסקנה: $Z \sim \text{Pois}(\lambda + \mu)$ (נקבע ע"י הפונקציה יוצרת מומנטים שלו).
 - ניתן, כמובן, להרחיב תופעה זו באינדוקציה לכל מספר סופי של מ"מ פואסון ב"ת.
- באופן כללי, אם X, Y מ"מ ב"ת, הראינו כרגע ש- $M_{X+Y}(s) = M_X(s)M_Y(s)$.
- יהי $b = aX + Y$, אזי: $M_Y(s) = E[e^{s(aX+b)}] = e^{sb}E[e^{s aX}] = e^{sb}M_X(sa)$.
- זבובים נכנסים לחדר, ובחדר מלכודת. הזבוב ה- i שנכנס ייתפס במלכודת בהסתברות p , כלומר $X_i \sim \text{Ber}(p)$ וכן כל ה- $\{X_i\}$ ב"ת וש"ה. נסמן N – מספר הזבובים שנכנסו לחדר, ונתון $N \sim \text{Pois}(\lambda)$. נגדיר Y – מספר הזבובים שנתפסו ורוצים למצוא את ההתפלגות שלו.

○ ניזכר ש- $M_{X_i}(s) = (1 - p + pe^s)$ ו- $M_N(s) = e^{\lambda(e^s-1)}$ מכאן נפתח:

$$M_Y(s) = E[e^{sY}] = E\left[e^{s \sum_{i=1}^N X_i}\right] = E\left[E\left[e^{s \sum_{i=1}^N X_i} \middle| N\right]\right]$$

○ כדי להמשיך נסתכל על $N = n$ ספציפי:

$$E\left[e^{s \sum_{i=1}^N X_i} \middle| N = n\right] = E\left[e^{s \sum_{i=1}^n X_i}\right] = \prod_{i=1}^n (1 - p + pe^s) = (1 - p + pe^s)^n$$

○ ומכאן אפשר לחזור לתוחלת:

$$\begin{aligned} E[(1 - p + pe^s)^N] &= \sum_{n=0}^{\infty} P(N = n)(1 - p + pe^s)^n = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} (1 - p + pe^s)^n = \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda(1 - p + pe^s))^n}{n!} = \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda(1 - p + pe^s))^n}{n!} = e^{-\lambda} e^{\lambda(1 - p + pe^s)} = e^{\lambda p(e^s - 1)} \end{aligned}$$

○ וכעת קיבלנו את הפונקציה יוצרת מומנטים של מ"מ פואסון: $Y \sim \text{Pois}(\lambda p)$.

טענה: יהי $Y \sim N(\mu, \sigma^2)$ הפונקציה יוצרת מומנטים שלו היא $M_Y(s) = e^{\frac{s^2 \sigma^2}{2} + s\mu}$.

הוכחה: נתחיל במ"מ נורמלי סטנדרטי, $Z \sim N(0, 1)$. עבורו:

$$\begin{aligned} M_Z(s) &= E[e^{sX}] = \int_{-\infty}^{\infty} \frac{e^{sx}}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} - sx} dx = e^{\frac{s^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2} + sx - \frac{s^2}{2}} dx = \\ &= e^{\frac{s^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-s)^2}{2}} dx = e^{\frac{s^2}{2}} \end{aligned}$$

המעבר האחרון נובע מכך שקיבלנו אינטגרל על פונקציית הצפיפות של מ"מ נורמלי $N(s, 1)$. כעת נתבונן על המ"מ המקורי, $Y = \sigma Z + \mu$ (בתרגיל ראינו ש- $Y \sim N(\mu, \sigma^2)$). עבורו:

$$M_Y(s) = e^{s\mu} M_Z(s\sigma) = e^{s\mu} e^{\frac{s^2 \sigma^2}{2}} = e^{\frac{s^2 \sigma^2}{2} + s\mu}$$

כפי שרצינו. ■

מסקנה: יהיו $X \sim N(\mu_X, \sigma_X^2)$, $Y \sim N(\mu_Y, \sigma_Y^2)$ ב"ת, ו- $Z = X + Y$. אזי $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

הוכחה: מפתוח פ' יוצרת מומנטים:

$$M_Z(s) = M_X(s)M_Y(s) = e^{\frac{s^2 \sigma_X^2}{2} + \mu_X s} e^{\frac{s^2 \sigma_Y^2}{2} + \mu_Y s} = e^{\frac{s^2}{2}(\sigma_X^2 + \sigma_Y^2) + s(\mu_X + \mu_Y)}$$

התכנסות בהסתברות

הגדרה: תהי $(X_n)_{n=1}^{\infty}$ סדרת מ"מ. נאמר שהסדרה (X_n) מתכנסת בהסתברות ל- X אם לכל $\varepsilon > 0$, $\lim_{n \rightarrow \infty} P(|X_n - X| \geq \varepsilon) = 0$.

טענה (אי-שוויון מרקוב): יהי X מ"מ אי-שלילי. אזי $P(X \geq c) \leq \frac{E[X]}{c}$.

הוכחה: מההגדרה:

$$E[X] = \int_0^{\infty} x f_X(x) dx = \int_0^c x f_X(x) dx + \int_c^{\infty} x f_X(x) dx \geq \int_0^c x f_X(x) dx + \int_c^{\infty} c f_X(x) dx \geq c P(X \geq c)$$

ומהעברת אגפים מתקבל הדרוש. ■

טענה (אי-שוויון צ'בישב): יהי X מ"מ עם תוחלת μ ושונות σ^2 . אזי $P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}$.

הוכחה: נגדיר מ"מ חדש $Y = (X - \mu)^2$ ונציב אותו בא"ש מרקוב:

$$P(|X - \mu| \geq c) = P(Y \geq c^2) \leq \frac{E[(X - \mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

כפי שרצינו. ■

מסקנה: $P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}$

זה מאפשר לראות את ההסתברות שהמשתנה המקרי מתרחק מהתוחלת במכפלות של סטיית התקן. למשל, ההסתברות להתרחק 3 סטיות תקן מהתוחלת היא פחות מ- $\frac{1}{9}$.

משפט (החוק החלש של המספרים הגדולים): תהי (X_n) סדרה של מ"מ ב"ת שווי תפלגות עם תוחלת μ . אזי הסדרה $\left(\frac{X_1 + \dots + X_n}{n}\right)$ מתכנסת בהסתברות ל- μ .

הוכחה: ראינו כבר ש- $\mu = E\left[\frac{X_1 + \dots + X_n}{n}\right]$ ונציב את זה בא"ש צ'בישב:

$$P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2} \frac{1}{n}$$

ועכשיו, עבור $\varepsilon > 0$ נתון, הגבול הנ"ל הוא 0 (כאשר $n \rightarrow \infty$) ומכאן המסקנה. ■

הגדרה: סדרה (X_n) של מ"מ מתכנסת בהתפלגות למ"מ X אם $\lim_{n \rightarrow \infty} F_{X_n} = F_X$ (שוויון פונקציות).

דוגמאות:

- מתקיימות בחירות, p ההסתברות לתמיכה במפלגה מסוימת. נסתכל על מדגם $X_1, \dots, X_n$ שכולם ב"ת שווי התפלגות $(X_i \sim \text{Ber}(p))$, והממוצע שלהם M_n .

- אנו יודעים מא"ש צ'בישב ש- $P(|M_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$ ומכאן $P(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2}$, שזו עובדה מועילה אפילו אם לא יודעים את p .

- במכון סקרים רוצים 95% ביטחון עבור ε דיוק ורוצים לדעת מהו ה- n שמישיג את זה. ובכן:

$$P(|M_n - p| \geq \varepsilon) \leq \frac{1}{4n\varepsilon^2} \leq 0.05 \Rightarrow n \geq \frac{1}{4\varepsilon^2 \cdot 0.05}$$
 - למשל, עבור $\varepsilon = 0.001$ מקבלים $n \geq 50,000$
- יהיו $X_i \sim U(0,1)$ ב"ת ש"ה, ו- $Y_n = \min\{X_1, \dots, X_n\}$.
 - נראה ש- $Y_n \xrightarrow{p} 0$ (מתכנסת בהסתברות ל-0).

$$P(|Y_n - 0| \geq \varepsilon) = P(\min\{X_1, \dots, X_n\} \geq \varepsilon) = P(X_1 \geq \varepsilon, \dots, X_n \geq \varepsilon) = \prod_{i=1}^n P(X_i \geq \varepsilon) = (1 - \varepsilon)^n \xrightarrow{n \rightarrow \infty} 0$$
 - נגדיר $W_n = nY_n$. סדרה זו לא מתכנסת בהסתברות ל-0:

$$1 - F_{W_n}(\varepsilon) = P(W_n \geq \varepsilon) = P(nY_n \geq \varepsilon) = P(\min\{X_1, \dots, X_n\} \geq \frac{\varepsilon}{n}) = \left(1 - F_{X_i}\left(\frac{\varepsilon}{n}\right)\right)^n = \left(1 - \frac{\varepsilon}{n}\right)^n \xrightarrow{n \rightarrow \infty} e^{-\varepsilon}$$
 - בעצם, מה שקיבלנו כאן זה ש- $1 - e^{-\varepsilon} \xrightarrow{n \rightarrow \infty} 1$ לכל ε , וזו פונקציית ההתפלגות המצטברת של מ"מ אקספוננציאלי.
 - לכן אנו מסיקים ש- $W_n \xrightarrow{D} \exp(1)$ (מתכנסת בהתפלגות ל- $\exp(1)$).
 - יהיו $X_1, \dots, X_n$ מ"מ ב"ת כך ש- $X_i \sim \exp(\lambda)$ ונגדיר $Y_n = \min\{X_1, \dots, X_n\}$. נראה ש- Y_n מתכנסת בהסתברות ל-0:

$$P(|Y_n| \geq \varepsilon) = P(Y_n \geq \varepsilon) = P(\min\{X_1, \dots, X_n\} \geq \varepsilon) = P(X_1 \geq \varepsilon, \dots, X_n \geq \varepsilon) = P(X_1 \geq \varepsilon) \cdots P(X_n \geq \varepsilon) =$$

$$= P(X_i \geq \varepsilon)^n = \left(1 - F_{X_i}(\varepsilon)\right)^n = \left(1 - (1 - e^{-\lambda\varepsilon})\right)^n = e^{-\lambda\varepsilon n} \xrightarrow{n \rightarrow \infty} 0$$

משפט הגבול המרכזי

משפט (משפט הגבול המרכזי – Central Limit Theorem): יהיו $X_1, \dots, X_n$ מ"מ בלתי-תלויים שווי התפלגות ונגדיר $S_n = X_1 + \dots + X_n, M_n = \frac{S_n}{n}$. נניח גם תוחלת ושונות סופיים לכל X_i . נסמן $E[X_i] = \mu, Var(X_i) = \sigma^2$.
נגדיר מ"מ חדש $Z_n = \frac{S_n - \mu n}{\sigma\sqrt{n}}$. אזי מתקיים $\lim_{n \rightarrow \infty} P(Z_n \leq z) = \Phi(z)$, כלומר Z_n מתכנסת בהתפלגות למ"מ נורמלי סטנדרטי, $N(0,1)$.

הוכחה: תהי $M_{X_i}(s)$ פונקציה יוצרת מומנטים של X_i , סופית בתחום $-d \leq s \leq d$. נניח גם בה"כ שהתוחלת של המ"מ היא 0. כעת:

$$M_{Z_n}(s) = E[e^{sZ_n}] = E\left[e^{s \frac{X_1 + \dots + X_n}{\sigma\sqrt{n}}}\right] = E\left[\prod_{i=1}^n e^{s \frac{X_i}{\sigma\sqrt{n}}}\right] = \prod_{i=1}^n E\left[e^{s \frac{X_i}{\sigma\sqrt{n}}}\right] =$$

$$= \prod_{i=1}^n M_{X_i}\left(\frac{s}{\sigma\sqrt{n}}\right) = \left(M_{X_i}\left(\frac{s}{\sigma\sqrt{n}}\right)\right)^n$$

כעת נשתמש בפיתוח טיילור של $M_{X_i}(s)$ סביב $s = 0$: $s = 0$ סביב $s = 0$ $M_{X_i}(s) = \sum_{k=0}^{\infty} M_{X_i}^{(k)}(0) \frac{s^k}{k!}$, וניקח שלושה איברים מהטור:

$$M_{X_i}(s) = M_{X_i}(0) + M_{X_i}'(0)s + M_{X_i}''(0)\frac{s^2}{2} + o(s^2) = 1 + E[X_i]s + E[X_i^2]\frac{s^2}{2} + o(s^2)$$

תחת ההנחות שלנו, $E[X_i] = 0$ ולכן $Var(X_i) = E[X_i^2]$ ולכן $1 + Var(X_i)\frac{s^2}{2} + o(s^2)$, כלומר נשארו עם $1 + Var(X_i)\frac{s^2}{2} + o(s^2)$. נציב את זה בביטוי שהיה לנו קודם:

$$M_{Z_n}(s) = \left(M_{X_i} \left(\frac{s}{\sigma\sqrt{n}} \right) \right)^n = \left(1 + \frac{\sigma^2 \frac{s^2}{\sigma^2 n}}{2} + o \left(\frac{s^2}{\sigma^2 n} \right) \right)^n = \left(1 + \frac{s^2}{2n} + o \left(\frac{s^2}{\sigma^2 n} \right) \right)^n$$

כאשר $n \rightarrow \infty$, הביטוי שקיבלנו שואף ל- $e^{\frac{s^2}{2}}$ (יש להצדיק את השמטת $o \left(\frac{s^2}{\sigma^2 n} \right)$ ולא נעשה זאת). לסיכום, קיבלנו שהפונקציה יוצרת מומנטים של Z_n מתכנסת ל- $e^{\frac{s^2}{2}}$, שהיא הפונקציה יוצרת מומנטים של מ"מ נורמלי סטנדרטי. לכן (לפי משפט שלא למדנו שלפיו התכנסות נקודתית של פונקציה יוצרת מומנטים גוררת התכנסות בהסתברות של מ"מ – [ויקידפיה](#)), $Z_n \xrightarrow{D} N(0,1)$ כנדרש. ■

באמצעות המשפט ניתן לבצע קירוב נורמלי – אנו יודעים שהמוצע של קבוצת משתנים מקריים ב"ת ש"ה בעלי תוחלת μ ושונות σ^2 ניתן לקירוב ע"י $N \left(\mu, \frac{\sigma^2}{n} \right)$, והסכום שלהם ניתן לקירוב ע"י $N(n\mu, \sigma^2 n)$.

דוגמאות:

- מעלים למטוס 100 חבילות שמשקל כל אחת מתפלג $U(5,50)$ והן ב"ת זו בזו. מה ההסתברות שסך המשקל יהיה גדול מ-3000 ק"ג? – נוכל לעשות קירוב נורמלי:
 - נגדיר X_i משקל החבילה ה- i , ומתקיים $E[X_i] = 27.5$ ולכן $\mu n = 27.5 \times 100$, $Var(X_i) = \frac{(50-5)^2}{12} = 168.75$, והשונות של הסכום היא לפיכך $Var(S_n) = 168.75 \times 100$. מכאן:

$$P(S_n \geq 3000) = 1 - P(S_n \leq 3000) = 1 - P(S_n - 100 \times 27.5 \leq 3000 - 100 \times 27.5) =$$

$$= 1 - P \left(\frac{S_n - 100 \times 27.5}{\sqrt{168.75 \times 100}} \leq \frac{3000 - 100 \times 27.5}{\sqrt{168.75 \times 100}} \right) \approx 1 - \Phi \left(\frac{3000 - 100 \times 27.5}{\sqrt{168.75 \times 100}} \right) \approx 0.0274$$
- ההסתברות שמערכת תקרוס ביום מסוים היא 5%. בשנה יש 300 ימי עבודה – מהי ההסתברות שהמערכת תקרוס בלא יותר מ-5 ימים בשנה?
 - נסמן את המ"מ של הקריסה בכל יום ב- $X_i \sim Ber(0.05)$ וכעת אנו מסתכלים על $Y = \sum_{i=1}^{300} X_i$ ורוצים לדעת מה ההסתברות שערכו יהיה קטן או שווה ל-5. במקום החישוב ע"י $Y \sim Bin(300, 0.05)$ נבצע קירוב נורמלי, שהרי $E[Y] = 300 \times 0.05 = 15$, $Var(Y) = 300 \times 0.05 \times 0.95 = 14.25$ ו- $E[X_i] = 0.05$, $Var(X_i) = 0.05 \times 0.95$

$$P(Y \leq 5) = P \left(\frac{Y - 15}{\sqrt{14.25}} \leq \frac{5 - 15}{\sqrt{14.25}} \right) \approx \Phi(-2.65) \approx 0.004$$

סטטיסטיקה

באופן כללי בפרק זה נניח שה"טבע" מייצר דוגמאות, וננסה להבין מהדוגמאות האלה מה המאפיינים של האקראיות בטבע – בדרך כלל, פרמטר מסוים של התפלגות מסוים. למשל, נניח שהטבע מתנהג לפי התפלגות אקספוננציאלית, ונרצה לאמוד את הפרמטר שלה מדוגמאות שקיבלנו בניסוי.

אמידה נקודתית

כאן אנו מניחים צורה הסתברותית מסוימת של הטבע, $X_i \sim P_{X_i}(x)$, שיש לה פרמטר $\theta^* \in \Theta$ כלשהו, והדגימות שאנו מקבלים תמיד בלתי-תלויות ושוות התפלגות. כדי לאמוד את הפרמטרים של הטבע, מגדירים פונקציה:

הגדרה: אומד (או סטטיסטי) הוא פונקציה $\hat{\theta}: (X_1, \dots, X_n) \rightarrow \Theta$, המחזירה עבור מדגם אקראי (לעתים נסמנו פשוט X) **אומד** – את הערך של הפרמטר θ שמתאים למדגם שהתקבל: $\hat{\theta}(x_1, \dots, x_n) = \theta$.

הגדרה: הטייה של אומד $\hat{\theta}$ מסומנת ומוגדרת $Bias(\hat{\theta}) = E[\hat{\theta}(X)] - \theta^*$, כאשר θ^* הערך ה"אמיתי" של הפרמטר. אומד שההטייה שלו היא 0 נקרא אומד **חסר-הטייה**.

הגדרה: אומד ייקרא קונסיסטנטי (עקבי) אם $\hat{\theta}(X) \xrightarrow{p} \theta^*$.

תכונות שנרצה שיהיו לאומד:

- (1) אם נחזור על התהליך, נקבל תוצאות (אומדנים) דומים – כלומר, $Var(\hat{\theta}(X))$ נמוך
- (2) כשיש הרבה דגימות, נקבל את הדבר ה"נכון": $\hat{\theta}(X) \xrightarrow{p} \theta^*$ (קונסיסטנטיות)
- (3) "טעות קטנה" – לכל n נרצה שהאומדן $\hat{\theta}_n$ יהיה קרוב לאמת: $E[\hat{\theta}_n(X) - \theta^*]$ יהיה קטן, כלומר שההטייה של האומד תהיה קטנה

דוגמאות:

- מטילים מטבע. מניחים ש- $X_i \sim Ber(p)$. נגדיר אומד: $\hat{g}(X) = \frac{1}{n} \sum_{i=1}^n X_i$. כלומר הממוצע של כל ההטלות.
 - אומד זה הוא חסר-הטייה: $E[\hat{g}(X)] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} n E[X_i] = \theta^*$
 - השונות מקיימת: $Var(\hat{g}(X)) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} n Var(X_i) = \frac{\theta^*(1-\theta^*)}{n} \leq \frac{1}{4n}$ כאשר n גדל.
 - משני הסעיפים הקודמים, תוחלת האומד היא θ^* והשונות שלו שואפת ל-0 ככל שמספר הדגימות גדל, ולכן קיבלנו התכנסות בהסתברות ל- θ^* (אומד קונסיסטנטי).

כדי לבדוק את איכותם של אומדים, מגדירים פונקציות נוספות הדומות להטייה. באופן כללי, אפשר להשתמש בכל פונקציית מרחק כדי למדוד כמה רחוק האומדן (בתוחלת) מהערך האמיתי של הפרמטר:

הגדרה: פונקציית הפסד (או טעות) של אומד $\hat{\theta}$ מוגדרת $L(\hat{\theta}(X), \theta^*)$ וההפסד (או סיכון) תחת פונקציית הפסד כזאת מוגדר $E[L(\hat{\theta}(X), \theta^*)]$.

למשל, הטייה היא פונקציית הפסד, שבה $L(\hat{\theta}(X), \theta^*) = \hat{\theta}(X) - \theta^*$ (פונקציית ההפרש).

הגדרה: Mean Squared Error היא פונקציית הפסד המוגדרת $L(\hat{\theta}(X), \theta^*) = (\hat{\theta}(X) - \theta^*)^2$, וההפסד לפי פונקציה זו מסומן $R := E[(\hat{\theta}(X) - \theta^*)^2]$.

מתקיים: $R = Var(\hat{\theta}(X) - \theta^*) + E[\hat{\theta}(X) - \theta^*]^2 = Var(\hat{\theta}(X)) + Bias(\hat{\theta})^2$, כלומר זהו מדד התלוי גם בשונות וגם בהטייה – ואנו, בעיקרון, שואפים להקטין את שניהם. לעתים יהיה trade-off ביניהם שלא יאפשר להקטין את שניהם.

אומד ניראות מירבית

הגדרה: יהיו $X_i \sim f_X(x; \theta)$ (או $P_X(x; \theta)$). אומד הניראות המירבית (Maximum Likelihood Estimator) מוגדר להיות $\hat{\theta}(X) = \operatorname{argmax}_{\theta} f_X(x_1, \dots, x_n; \theta)$.

הרעיון: לבחור את ה- θ שבדיעבד אילו הוא היה הפרמטר של הטבע, המדגם שקיבלנו היה הכי סביר.

דוגמאות:

- יהיו $X_i \sim Ber(\theta^*)$. נמצא את אומד הניראות המירבית:
 - נתחיל מלכתוב את ההסתברות:

$$P(X_1, \dots, X_n; \theta) = \prod_{i=1}^n P(X_i; \theta) = \prod_{i=1}^n \theta^{X_i} (1 - \theta)^{1-X_i}$$

- כעת כדי למצוא את המקסימום נוציא לוג, נגזור, נשווה ל-0, ונחליץ את θ :

$$\operatorname{argmax}_{\theta} \prod_{i=1}^n \theta^{X_i} (1-\theta)^{1-X_i} = \operatorname{argmax}_{\theta} \sum_{i=1}^n (X_i \log(\theta) + (1-X_i) \log(1-\theta))$$

$$\frac{d}{d\theta} \left\{ \sum_{i=1}^n (X_i \log(\theta) + (1-X_i) \log(1-\theta)) \right\} = \sum_{i=1}^n \left(\frac{X_i}{\theta} - \frac{1-X_i}{1-\theta} \right) = 0 \Leftrightarrow \sum_{i=1}^n \frac{X_i}{\theta} = \sum_{i=1}^n \frac{1-X_i}{1-\theta} \Leftrightarrow \theta = \frac{1}{n} \sum_{i=1}^n X_i$$

- וזו התוצאה שגם הצענו אינטואיטיבית כדי לאמוד הטלות מטבע.

- יהיו $X_i \sim U(0, \theta^*)$ ונמצא אומד ניראות מירבית ל- θ^* :

$$\hat{\theta}(X) = \operatorname{argmax}_{\theta} \log(P(X_1, \dots, X_n; \theta)) = \operatorname{argmax}_{\theta} \sum_{i=1}^n \log(P(X_i; \theta)) = \operatorname{argmax}_{\theta} \sum_{i=1}^n \log\left(\frac{1}{\theta}\right)$$

- ברור שהביטוי הזה מקבל מקסימום כאשר θ קטן ככל האפשר. עם זאת, אנו מוגבלים ל- $\theta \geq \max\{X_i\}$, אחרת הוא המדגם לא אפשרי עבורו. אם כן, ה- θ הקטן ביותר שעבורו הביטוי הזה הוא לא 0 הוא $\max\{X_i\}$, וזה צריך להיות האומד שלנו:

$$\hat{\theta}(X) = \max\{X_i\}$$

- נשווה אומדים ל- $X_i \sim U(0, \theta^*)$: $\hat{\theta}_1(X) = \max\{X_i\}$, $\hat{\theta}_2(X) = 2\bar{X} = \frac{2}{n} \sum_{i=1}^n X_i$, $\hat{\theta}_3(X) = \frac{n}{n+1} \hat{\theta}_1(X)$, $\hat{\theta}_4(X) = \frac{n+1}{n+2} \hat{\theta}_1(X)$

- עבור $\hat{\theta}_2$, מתקיים: $\theta^* = \frac{2}{n} n \frac{\theta^*}{2} = \theta^*$ $E[\hat{\theta}_2(X)] = E[2\bar{X}] = \frac{2}{n} \sum_{i=1}^n E[X_i]$ ולכן הוא חסר הטייה.

- עבור $\hat{\theta}_1$ (שהוא אומד הניראות המירבית שקיבלנו קודם), מתקיים:

$$\begin{aligned} F_{\hat{\theta}_1(X)}(u) &= P(\hat{\theta}_1(X) \leq u) = P(\max\{X_i\} \leq u) = P(X_1 \leq u, \dots, X_n \leq u) = P(X_1 \leq u) \dots P(X_n \leq u) = \\ &= P(X_1 \leq u)^n = \begin{cases} 0 & u \leq 0 \\ \left(\frac{u}{\theta^*}\right)^n & 0 \leq u \leq \theta^* \\ 1 & u \geq \theta^* \end{cases} \Rightarrow f_{\hat{\theta}_1(X)}(u) = \frac{d}{du} F_{\hat{\theta}_1(X)}(u) = \frac{nu^{n-1}}{(\theta^*)^n} \end{aligned}$$

- באמצעות מה שקיבלנו אפשר גם לחשב תוחלת:

$$E[\hat{\theta}_1(X)] = \int_0^{\theta^*} \frac{nu^{n-1}}{(\theta^*)^n} du = \frac{n}{(\theta^*)^n} \int_0^{\theta^*} u^{n-1} du = \frac{n}{(\theta^*)^n} \left[\frac{u^n}{n} \right]_0^{\theta^*} = \frac{n}{(\theta^*)^n} \frac{(\theta^*)^n}{n+1} = \frac{n}{n+1} \theta^*$$

- כלומר, זהו אומד בעל הטייה (למרות שכאשר מספר הדגימות גדל, התוחלת מתקרבת לפרמטר האמיתי).

- עבור $\hat{\theta}_3$, מתקיים: $\theta^* = \frac{n+1}{n} \left(\frac{n}{n+1} \theta^* \right) = \theta^*$ $E[\hat{\theta}_3(X)] = E\left[\frac{n+1}{n} \hat{\theta}_1(X)\right] = \frac{n+1}{n} \left(\frac{n}{n+1} \theta^* \right) = \theta^*$

להפוך את $\hat{\theta}_1$ לחסר הטייה.

- לגבי שונות, עבור $\hat{\theta}_1$: $\hat{\theta}_1(X) = \max\{X_i\}$ $Var(\hat{\theta}_1(X)) = \int_0^{\theta^*} \frac{nu^{n-1}}{(\theta^*)^n} u^2 du - \left(\frac{n}{n+1} \theta^* \right)^2 = \frac{n}{(n+1)^2 (n+2)} (\theta^*)^2$

- עבור $\hat{\theta}_2$, השונות: $\hat{\theta}_2(X) = \frac{2}{n} \sum_{i=1}^n X_i$ $Var(\hat{\theta}_2(X)) = Var\left(\frac{2}{n} \sum_{i=1}^n X_i\right) = \frac{4}{n^2} \sum_{i=1}^n Var(X_i) = \frac{4}{n^2} n Var(X_i) = \frac{4}{n} Var(X_i) = \frac{4}{n} \frac{(\theta^*)^2}{12} = \frac{(\theta^*)^2}{3n}$

- עבור $\hat{\theta}_3$, השונות: $\hat{\theta}_3(X) = \frac{n+1}{n} \hat{\theta}_1(X)$ $Var(\hat{\theta}_3(X)) = \left(\frac{n+1}{n}\right)^2 Var(\hat{\theta}_1(X)) = \frac{1}{n(n+2)} (\theta^*)^2$

- מכל אלה אפשר להשוות את ה-MSE. מקבלים: $MSE(\hat{\theta}_2(X)) = \frac{(\theta^*)^2}{3n}$, $MSE(\hat{\theta}_3(X)) = \frac{(\theta^*)^2}{n(n+2)}$

הראשון, $MSE(\hat{\theta}_1(X)) = \frac{(\theta^*)^2}{(n+1)^2} \left(\frac{n}{n+2} + 1 \right)$. המסקנה הסופית היא שה-MSE של האומד השלישי הוא הטוב ביותר מהשלושה.

- יהיו $X_i \sim f_X(x)$ ורוצים אומד לתוחלת $\theta^* = E[X_i]$. נגביל את עצמנו לסטטיסטיים מהצורה $T(X) = \sum_{i=1}^n c_i X_i$ כאשר $\{c_i\}$ קבועים ממשיים.

- נראה ש- T הוא לא מוטה אם $\sum_{i=1}^n c_i = 1$. ובכן:

$$Bias(T) = E[T(X) - \theta^*] = E\left[\sum_{i=1}^n c_i X_i - \theta^*\right] = \sum_{i=1}^n c_i E[X_i] - \theta^* = \sum_{i=1}^n c_i \theta^* - \theta^* = 0 \Leftrightarrow \sum_{i=1}^n c_i = 1$$

- נראה ש- $T(X) = \sum_{i=1}^n \frac{1}{n} X_i$ (ובכך גם את ה-MLE):

$$Var(T(X)) = Var\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 Var(X_i) = \sigma^2 \sum_{i=1}^n c_i^2$$

- כלומר, אנחנו צריכים להביא למינימום את $\sum_{i=1}^n c_i^2$ תחת האילוץ שסכומם הוא 1. מא"ש קושי-שורץ או פתרון ישיר ע"י

$$c_1 = \dots = c_n = \frac{1}{n}$$

- יהיו $X_i \sim \exp(\lambda)$ ורוצים אומד ניראות מירבית ל- λ .

- נגדיר את פונקציית הניראות: $L(\lambda; X_1, \dots, X_n) = \prod_{i=1}^n f_{X_i}(X_i; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda X_i}$

- נוציא לוג לנוחות: $\log L(\lambda; X_1, \dots, X_n) = n \log \lambda - \lambda \sum_{i=1}^n X_i$

נגזור ונשווה ל-0 כדי למצוא את המקסימום:

$$\frac{d}{d\lambda} \log L(\lambda; X_1, \dots, X_n) = \frac{n}{\lambda} - \sum_{i=1}^n \frac{1}{X_i} = 0 \Leftrightarrow \lambda = \frac{n}{\sum_{i=1}^n \frac{1}{X_i}} = \frac{1}{\bar{X}}$$

כלומר, אומד הניראות המירבית כאן הוא פשוט $\frac{1}{\bar{X}}$.

• יהיו $X_i \sim U(\theta, \theta + 1)$ ורוצים אומד ניראות מירבית ל- θ .

נגדיר את פונקציית הניראות: $L(\theta; X_1, \dots, X_n) = \prod_{i=1}^n f_{X_i}(X_i; \theta) = 1$ (פונקציית הצפיפות כאן היא פשוט 1 על הקטע

$[\theta, \theta + 1]$ וכמובן כל ה- X_i ים לקוחים מתוך הקטע הזה).

נשים לב ש- $\theta \leq \min\{X_i\}$, $\max\{X_i\} \leq \theta + 1$ וכיוון שפונקציית הניראות קבועה בקטע $[\max\{X_i\} - 1, \min\{X_i\}]$, בעצם כל ערך בקטע זה מספק את המקסימום.

למשל, $\hat{\theta}(X) = \min\{X_i\}$ הוא בעצמו בתוך הקטע ולכן מהווה אומד ניראות מירבית, אבל הוא לא חסר הטיה (לפחות, הוא קונסיסטנטי): $E[\min\{X_i\}] = \theta + \frac{1}{n+1}$.

אפשר גם למצוא אומד חסר הטיה: $\hat{\theta}(X) = \frac{\min\{X_i\} + \max\{X_i\} - 1}{2}$ שהוא גם כן אומד ניראות מירבית.

• יהיו $X_i \sim N(\mu, 1)$ ורוצים לאמוד את $\theta = e^\mu$. מציעים אומד $\hat{\theta}(X) = e^{\bar{X}}$, שנשמע הגיוני כיוון ש- $\bar{X}$ הוא אומד ניראות מירבית לתוחלת μ . נחשב את ה-MSE שלו:

תחילה נחשב את התוחלת:

$$E[\hat{\theta}(X)] = E[e^{\bar{X}}] = M_{\bar{X}}(1) = e^{\mu + \frac{1}{2n}} = e^\mu e^{\frac{1}{2n}}$$

כאשר השתמשנו בכך ש- $\bar{X} \sim N(\mu, \frac{1}{n})$ ולכן אנו יודעים את הפונקציה יוצרת מומנטים שלו. קיבלנו שהאומד אינו חסר הטיה,

אבל הגורם $e^{\frac{1}{2n}}$ דועך אקספוננציאלית (אפשר גם להציע אומד חסר הטיה ע"י נירמול בהופכי של גורם זה).

כעת נחשב את השונות:

$$Var(\hat{\theta}(X)) = Var(e^{\bar{X}}) = E[(e^{\bar{X}})^2] - E[e^{\bar{X}}]^2 = M_{\bar{X}}(2) - (M_{\bar{X}}(1))^2 = e^{2\mu + \frac{1}{n}} - e^{2\mu} e^{\frac{1}{n}}$$

וקיבלנו $MSE(\hat{\theta}) = e^{2\mu} (e^{\frac{1}{n}} - 1)$. גם ערך זה שואף ל-0 כאשר n גדל, שזה "טוב".

• יהיו $X_i \sim N(\mu, \sigma^2)$ ורוצים אומדי ניראות מירבית (הוכחות בפתרון תרגיל 11):

כאשר σ^2 ידוע, $\hat{\mu}_{MLE}(X) = \bar{X}$

כאשר μ ידוע, $\hat{\sigma}^2_{MLE}(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$

כאשר שני הפרמטרים לא ידועים, האומד לתוחלת הוא כמו קודם והאומד לשונות הוא $\hat{\sigma}^2(X) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$.

משפט: (ללא נוכחה) אומד ניראות מירבית הוא קונסיסטנטי.

הגדרה: אומד $\hat{\theta}(X)$ נקרא אסימפטוטי נורמלי אם $\sqrt{n}(\hat{\theta}(X) - \theta^*) \xrightarrow{D} N(0, \sigma_{\hat{\theta}}^2)$

הכוונה כאן ב- $\sigma_{\hat{\theta}}^2$ לשונות של ה"טבע", כלומר של ה- X_i של המדגם. זה דבר טוב כיוון שהתפלגות נורמלית היא בעלת דעיכה אקספוננציאלית משני הצדדים של התוחלת, כלומר ההסתברות להיות רחוקים מהממוצע הולכת וקטנה. למשל, ההבדל בין $n = 100$ ל- $n = 200$ יהיה מאוד גדול, וזה קצב "טוב".

דוגמאות:

• יהי $\hat{\theta}(X) = \bar{X}$, ונניח $E[X_i] = \theta^*$ ו- $Var(X_i) = \sigma_{\theta^*}^2$. כלומר אומדים את התוחלת של ה"טבע". כעת לפי משפט הגבול המרכזי, $\frac{\sqrt{n}(\bar{X} - \theta^*)}{\sigma_{\theta^*}} \xrightarrow{D} N(0, 1) \Rightarrow \sqrt{n}(\hat{\theta}(X) - \theta^*) \xrightarrow{D} N(0, \sigma_{\theta^*}^2)$ (לתוחלת).

משפט: (ללא הוכחה) אומד ניראות מירבית הוא אסימפטוטי נורמלי.

רווח סמך

הגדרה: יהי X מדגם מקרי $X_1, \dots, X_n$ הנדגמים באופן ב"ת ש"ה מתוך הטבע, $f_X(x; \theta)$. רווח סמך (Confidence Interval) הוא זוג פונקציות U, V המספקות א"ש $U(X) \leq \theta \leq V(X)$.

כלומר, האומדן שלנו במקרה זה הוא לא ערך ספציפי של θ , אלא טווח ערכים שבו θ נמצאת. כמובן, אי-השוויון יהיה נכון רק בהסתברות מסוימת, ונרצה שהיא תהיה קטנה, כלומר $1 - \alpha = P(U(X) \leq \theta \leq V(X))$.
(הטעות) הוא קטן.

דוגמאות:

- קוקה-קולה בונה מכונה למילוי פחיות, היינו רוצים $\mu = 330$ מ"ל קולה בכל פחית. נניח שניתן לכייל את σ^2 והכמות הנשפכת לכל פחית היא $N(\mu, \sigma^2)$. מטרתנו: ב-95% מהפחיות להשיג בין 328 ל-332 מ"ל קולה.
- יהיו $X_i \sim N(\mu, 1)$ ונגדיר רווח סמך $[\bar{X} - 1, \bar{X} + 1]$. כמה הדבר הזה טוב? נחשב את ההסתברות שאכן התוחלת נופלת בתוך רווח הסמך, כאשר נזכור ש- $\bar{X} \sim N(\mu, \frac{1}{n})$:

$$\begin{aligned} P(\bar{X} - 1 \leq \mu \leq \bar{X} + 1) &= P(\mu - 1 \leq \bar{X} \leq \mu + 1) = P\left(\frac{(\mu - 1) - \mu}{\sqrt{\frac{1}{n}}} \leq \frac{\bar{X} - \mu}{\sqrt{\frac{1}{n}}} \leq \frac{(\mu + 1) - \mu}{\sqrt{\frac{1}{n}}}\right) = \\ &= P(-\sqrt{n} \leq (\bar{X} - \mu)\sqrt{n} \leq \sqrt{n}) = \Phi(\sqrt{n}) - \Phi(-\sqrt{n}) = 1 - 2\Phi(-\sqrt{n}) \end{aligned}$$

○ כאן המעבר האחרון הוא מסימטריה של ההתפלגות הנורמלית, והמעבר ל- Φ הוא כיוון ש- $(\bar{X} - \mu)\sqrt{n} \sim N(0, 1)$.

○ למשל, עבור $n = 4$ מקבלים שההסתברות להיות בטווח הנתון היא 0.954.

כיצד נחשב רווח סמך? פועלים לפי שלושה שלבים:

- (1) מוצאים פונקציה $g(X; \theta)$ כך שההתפלגות של g אינה תלויה ב- θ ; פונקציה זו נקראת *pivotal quality* או *pivot function*
- (2) נחפש a, b כך ש- $P(a \leq g(X_i; \theta) \leq b) = 1 - \alpha$ – זה אמור להיות קל כי g לא תלויה ב- θ ובמיוחד קל אם היא באה מהתפלגות מוכרת
- (3) מחלצים רווח סמך על θ

דוגמאות:

- עבור $X_i \sim N(\mu, \sigma^2)$ – כאשר אומדים את μ (σ^2 ידוע):
 - נגדיר $g(X; \theta) = \frac{\bar{X} - \mu}{\sigma} \sqrt{n}$ שהוא מ"מ נורמלי סטנדרטי ואינו תלוי בפרמטר.
 - נסתכל על $1 - \alpha = P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\sigma} \sqrt{n} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$ – ההסתברות שווה ל- $1 - \alpha$ כיוון ש- $1 - 2P\left(Z \geq z_{1-\frac{\alpha}{2}}\right) = 1 - 2\left(1 - P\left(Z \leq z_{1-\frac{\alpha}{2}}\right)\right) = 1 - 2 + 2\left(1 - \frac{\alpha}{2}\right) = 1 - \alpha$
 - כשמחלצים את μ , מתקבל רווח הסמך $\mu \in \left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ (הגיוני שהוא סימטרי כי מדובר בהתפלגות נורמלית, שהיא סימטרית).
- יהיו $X_i \sim U(0, \theta)$ ורוצים לאמוד את θ .
 - אפשר לקחת $g(X_i; \theta) = \frac{X_i}{\theta}$ ואז $g(X_i; \theta) = \frac{X_i}{\theta} = u$ ו- $F_{g(X_i; \theta)}(u) = P\left(\frac{X_i}{\theta} \leq u\right) = P(X_i \leq \theta u) = F_{X_i}(\theta u) = \frac{\theta u}{\theta} = u$ – תלויה ב- θ .
 - אפשר גם להשתמש במה שקיבלנו עבור אומד ניראות מירבית ($\max\{X_i\}$) ולקחת $g(X; \theta) = \frac{\max\{X_i\}}{\theta}$. גם ההתפלגות שלה אינה תלויה ב- θ , נראה זאת:

$$F_{\frac{\max\{X_i\}}{\theta}}(u) = P\left(\frac{\max\{X_i\}}{\theta} \leq u\right) = P(\max\{X_i\} \leq \theta u) = P(X_i \leq \theta u)^n = F_{X_i}(\theta u)^n = \left(\frac{\theta u}{\theta}\right)^n = u^n$$
 - כעת ננסה למצוא חסמים עבור $1 - \alpha = P\left(k \leq \frac{\max\{X_i\}}{\theta} \leq l\right)$. כיוון ש- $g(X; \theta)$ מהר מאוד מתקרבת ל-1, סביר לבחור $l = 1$ ו"לשחק" רק עם ערכו של k :

$$P\left(k \leq \frac{\max\{X_i\}}{\theta} \leq 1\right) = F_{\frac{\max\{X_i\}}{\theta}}(1) - F_{\frac{\max\{X_i\}}{\theta}}(k) = 1^n - k^n = 1 - k^n$$
 - אם אנו רוצים $1 - \alpha = 1 - k^n$, מכאן שאנחנו צריכים לבחור $k = \sqrt[n]{1 - \alpha}$.

- כדי לקבל רווח סמך על θ אנחנו צריכים לבודד אותו מהא"ש:

$$P\left(\sqrt[n]{\alpha} \leq \frac{\max\{X_i\}}{\theta} \leq 1\right) = P\left(\max\{X_i\} \leq \theta \leq \frac{\max\{X_i\}}{\sqrt[n]{\alpha}}\right)$$
- וקיבלנו רווח סמך מהצורה $\theta \in \left[\max\{X_i\}, \frac{\max\{X_i\}}{\sqrt[n]{\alpha}}\right]$ המספק רמת שגיאה α (הגבול השמאלי של הרווח ברור מאילו, כי במילא לא ייתכן $\max\{X_i\} < \theta$ כי הוא לא יכול לייצר דגימות כאלה).
- נניח שרוצים רמת ביטחון של 95%-0.9 k . מהו ה- n הדרוש? פותרים ומקבלים $n > 28 \Rightarrow 0.05^{\frac{1}{n}} = 0.9$.
- יהיו $X_i \sim \text{Ber}(p)$ ורוצים רווח סמך ל- p .
 ○ לפי משפט הגבול המרכזי, עבור n גדול, $\bar{X} \sim N\left(p, \frac{p(1-p)}{n}\right)$ (בקירוב). לכן, לפי מה שראינו קודם, רווח הסמך (לנורמלי) הוא

$$p \in \left[X \pm \frac{z_{1-\frac{\alpha}{2}} \sqrt{p(1-p)}}{\sqrt{n}}\right]$$
- הבעיה כאן היא שהרווח סמך תלוי ב- p . פתרון אפשרי אחד הוא לחסום: $\frac{1}{4} \geq p(1-p)$ ולכן אפשר להחליף את הגורם הזה ב- $\frac{1}{4}$ בתוך רווח הסמך.
- פתרון אפשרי אחר הוא להעריך את p באמצעות $\bar{X}$, כלומר להסתכל על $\bar{X}(1-\bar{X})$ במקום על $p(1-p)$.

בדיקת השערות

במקום לספק רווח סמך או לאמוד ממש את הפרמטר, כעת אנו רק מתעניינים בשתי השערות על מנת להכריע איזו מהן נכונה לגבי הטבע על סמך המדגם שקיבלנו.

הגדרה: בבדיקת השערות עלינו לבנות מבחן מהצורה $T(X) \geq c$ כאשר X מדגם מקרי $(X_1, \dots, X_n)$ ב"ת ש"ה) עבור שתי השערות - H_0 השערת האפס, H_1 השערת האלטרנטיבה, כאשר המבחן דוחה את השערת האפס.

הגדרה: נסמן ב- R את קבוצת המדגמים שעבורם המבחן דוחה את H_0 . אז רמת המובהקות (significance) תסומן $\alpha := P_{H_0}((X_1, \dots, X_n) \in R)$ ועוצמת המבחן (power) תסומן $\beta := P_{H_1}((X_1, \dots, X_n) \in R)$ והיא $1 - \beta$.

כאשר כותבים P_{H_0} , הכוונה להסתברות של המאורע כאשר ההשערה H_0 היא הנכונה. כלומר, רמת המובהקות מגדירה את ההסתברות לדחות את H_0 כאשר היא נכונה (על סמך מדגם שעבר את הסף שלנו), ועוצמת המבחן מגדירה את ההסתברות לקבל את H_0 כאשר היא אינה נכונה.

בהינתן מבחן מסוים ו- n (גודל המדגם), כאשר קובעים את α , זה קובע גם את ערכו של β . אם רוצים להקטין את שני סוגי הטעות, אין ברירה אלא להגדיל את n .

דוגמאות:

- יהיו $X_i \sim \text{Ber}(\theta)$ ואנו רוצים לבדוק את ההשערות H_0 - המטבע לא מוטא, $\theta = \frac{1}{2}$, וההשערה המשלימה H_1 - המטבע מוטא, $\theta \neq \frac{1}{2}$. נניח שמטילים את המטבע 1000 פעמים, ונסמן $S = \sum_{i=1}^{1000} X_i$ סכום ההטלות. כאשר המטבע לא מוטא, התוחלת של הסכום היא 500, ולכן נגדיר מבחן מהצורה: $|S - 500| \geq \delta$, כאשר כאן δ הוא סף המבחן, ורוצים שכאשר המטבע לא מוטא, ההסתברות $P_{H_0}(|S - 500| \geq \delta) = \alpha$ (כלומר הטעות) תהיה קטנה.
- יהיו $X_i \sim N(\mu, 1)$ והשערות $H_0: \mu = 3$, $H_1: \mu = 5$. נשתמש במבחן הבא: אם $\bar{X} \geq 3.5$, דוחים את H_0 .
 ○ נבדוק מהם α, β כאן (כאשר $\bar{X} \sim N\left(\mu, \frac{1}{n}\right)$ כמו תמיד):

$$\alpha = P_{H_0}(\bar{X} \geq 3.5) = P\left(\frac{\bar{X} - 3}{\frac{1}{\sqrt{n}}} \geq \frac{3.5 - 3}{\frac{1}{\sqrt{n}}}\right) = P\left(Z \geq \frac{\sqrt{n}}{2}\right) = 1 - \Phi\left(\frac{\sqrt{n}}{2}\right)$$

$$\beta = P_{H_1}(\bar{X} \leq 3.5) = P\left((\bar{X} - 5)\sqrt{n} \leq (3.5 - 5)\sqrt{n}\right) = P\left(Z \leq -\frac{3}{2}\sqrt{n}\right) = \Phi\left(-\frac{3}{2}\sqrt{n}\right)$$

- נניח כעת שרוצים $\alpha = 0.05$, ונחשב את c (עבור המבחן $\bar{X} \geq c$) שיספק אותו:
 $0.05 = \alpha = P_{H_0}(\bar{X} \geq c) = P\left((\bar{X} - 3)\sqrt{n} \geq (c - 3)\sqrt{n}\right) = 1 - P(Z \leq (c - 3)\sqrt{n}) \Rightarrow (c - 3)\sqrt{n} = z_{0.95}$

○ עבור אותם α, c שמצאנו, מה קורה עם β ?

$$\beta = P_{H_1}((\bar{X} - 3)\sqrt{n} \leq z_{0.95}) = P((\bar{X} - 5)\sqrt{n} \leq z_{0.95} - 2\sqrt{n}) = \Phi(z_{0.95} - 2\sqrt{n})$$

• עבור $X_i \sim N(\mu, \sigma^2)$ וההשערות $H_0: \mu = \mu_0, H_1: \mu = \mu_1$ (כאשר $\mu_0 < \mu_1$) נתסכל על מבחן מהצורה $\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \geq c(n)$ (בצד שמאל יש

נורמלי סטנדרטי, שזה נוח) ונרצה למצוא כמה גדול צריך להיות המדגם על מנת לקיים תנאים גם על α וגם על β .

○ ראשית נדרוש $c(n) \geq z_{1-\alpha}$ כדי שהטעות מהסוג הראשון לא תעלה על α .

○ שנית נדרוש $c(n) \leq z_\beta - \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}}$, שהרי $\beta = P_{H_1}\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} + \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}} \leq c(n) + \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}}\right)$

○ משני התנאים יחד אנו מקבלים: $z_\beta \leq c(n) + \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\alpha} + \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}}$, ואם מתעלמים מהחלק האמצעי של הא"ש,

מקבלים $z_{1-\alpha} + \frac{(\mu_0 - \mu_1)}{\frac{\sigma}{\sqrt{n}}} \leq z_\beta$ ומכאן אפשר להפיק תנאי על n (שהוא המשתנה האמיתי כאן). ובכן:

$$\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} \leq z_\beta - z_{1-\alpha} \Leftrightarrow \sqrt{n} \geq \frac{\sigma(z_\beta - z_{1-\alpha})}{(\mu_0 - \mu_1)}$$

○ ובכל מקרה כדי שאפשר יהיה לקיים את התנאי הזה צריך $z_\beta \leq z_{1-\alpha}$.

• יהי X מ"מ רציף עם הצפיפות $f_X(x) = \frac{x+\theta}{\frac{1}{2}+\theta}$ בקטע $[0,1]$ ו-0 בשאר הישר. ההשערות: $H_0: \theta = 0, H_1: \theta = 2$. נגדיר מבחן המתבסס

על תצפית אחת, X , והדוחה את H_0 אם $\frac{1}{2} > X$. מהי רמת המובהקות והעוצמה?

$$\alpha = P_{H_0}\left(X \leq \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \frac{x}{\frac{1}{2}} dx = [x^2]_0^{\frac{1}{2}} = \frac{1}{4}$$

$$1 - \beta = 1 - P_{H_1}\left(X > \frac{1}{2}\right) = P_{H_1}\left(X \leq \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \frac{\frac{1}{2}x + 2}{\frac{1}{2} + 2} dx = 0.45$$

עד כה בבדיקת השערות ידענו רק להפריד בין שתי השערות, אך לא ידענו לומר "עד כמה" אנחנו מאמינים שהשערה מסוימת היא נכונה. ובכן, נגדיר מדד כזה:

הגדרה: יהי $x_1, \dots, x_n$ מדגם ספציפי ונסמן ב- $T(x_1, \dots, x_n)$ את ערך הסטטיסטי עבור המדגם (המבחן), בבדיקת השערות). p-value של המדגם הוא ההסתברות $P_{H_0}(T(X) \geq T(x_1, \dots, x_n))$ (כאן X הוא מ"מ של כל המדגם, כלומר $(X_1, \dots, X_n)$).

ככל שה-p-value גדול יותר, כך המדגם יותר סביר בהינתן H_0 . כלומר, אם רוצים לאשש את השערת האפס, רוצים p-value גבוה ככל האפשר. בעצם, ה-p-value של המדגם $x_1, \dots, x_n$ הוא הערך של α אילו היינו בוחרים לקבוע את הסף של המבחן ב- $c = T(x_1, \dots, x_n)$.

טענה: ה-p-value מתפלג אחיד בין 0 ל-1.

הוכחה: למעשה, הטענה נכונה לכל CDF שהיא מונוטונית עולה ממש. תהי $F_X(x)$ CDF מונוטונית עולה ממש, ונגדיר $Y = F_X(X)$ מ"מ חדש. כעת מתקיים:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(F_X(X) \leq y) = P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) = \\ &= P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y \end{aligned}$$

וקיבלנו שפונקציית ההתפלגות המצטברת היא כמו של $U(0,1)$, כאשר בדרך השתמשנו במונוטוניות ממש כדי שתהיה לרשותנו הפונקציה ההופכית. במקרה של p-value, מדובר ב-CDF מונוטונית ממש (למעט מקרים של מבחנים מנוונים ולא מעניינים) ולכן הטענה הנ"ל נכונה עבורה כמקרה פרטי. ■

דוגמאות:

- יהי $X \sim N(\mu, 1)$ וההשערות $H_0: \mu = 3, H_1: \mu = 5$, וכלל הדחיה הבא: $T(X) = \frac{\bar{X}-3}{\frac{1}{\sqrt{n}}} \geq c$ בחרנו מבחן זה כיוון שכעת צד שמאל מתפלג נורמלי סטנדרטי.

- נניח שקיבלנו $T(x_1, \dots, x_n) = 0.7$. מה ה-p-value?
 $P_{H_0}(T(X) \geq 0.7) = 1 - P_{H_0}(T(X) \leq 0.7) = 1 - \Phi(0.7) \approx 0.242$
- כלומר, תחת ההנחה $\mu = 3$, 75% מהמדגמים יקבלו ציון גבוה מ-0.7 מהמבחן שלנו.

מבחן יחס הנראות

הגדרה: מבחן יחס הנראות לבדיקת השערות הוא מבחן מהצורה $L(X) = \frac{P_{H_1}(X_1, \dots, X_n)}{P_{H_0}(X_1, \dots, X_n)} \geq c$

דוגמאות:

- מטילים קוביה, וההשערות: $H_0: P_X(x) = \frac{1}{6}$ ו- $H_1: P_X(x) = \begin{cases} \frac{1}{4} & x = 1, 2 \\ \frac{1}{8} & x = 3, 4, 5, 6 \end{cases}$
 - נפעיל את מבחן יחס הנראות על הטלה אחת:
- $$L(X) = \begin{cases} \frac{\frac{1}{4}}{\frac{1}{6}} = \frac{3}{2} & X = 1, 2 \\ \frac{\frac{1}{8}}{\frac{1}{6}} = \frac{3}{4} & X = 3, 4, 5, 6 \end{cases}$$
- יש לנו כעת כמה אפשרויות לקבוע סף: אם נבחר מבחן $L(X) \geq \frac{3}{2} + \varepsilon$, אז אנו תמיד מקבלים את השערת האפס. אם נבחר מבחן $L(X) \geq \frac{3}{4} + \varepsilon$, אז אנו דוחים עם טעות $\alpha = \frac{1}{3}$. לבסוף, אם נבחר מבחן $L(X) \geq \frac{3}{4}$ אז אנו תמיד דוחים את השערת האפס.
 - מצלמת מעקב מחזירה סיגנל $X = \omega$ אם אין פורץ, או $X = \omega + 1$ אם יש פורץ. נסמן את ההשערה שאין פורץ ב- H_0 וההשערה שיש פורץ ב- H_1 , ונתון $\omega \sim N(0, \sigma^2)$
 - נשים לב שתחת ההנחה שיש פורץ, $X \sim N(1, \sigma^2)$ וותחת ההנחה שאין פורץ $X \sim N(0, \sigma^2)$.
 - נבנה מבחן יחס נראות:

$$\frac{f_{H_1}(X)}{f_{H_0}(X)} = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-1)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{X^2}{2\sigma^2}}} = e^{\frac{2X-1}{2\sigma^2}} \geq c$$

- אפשר לפרק את זה ולקבל $X \geq \sigma^2 \log c + \frac{1}{2}$, וכיוון שצד ימין הוא קבוע, בעצם המבחן מצטמצם לצורה מאוד פשוטה: האם $X \geq c'$

- כעת, נניח שרוצים $\alpha = 0.025$. מהו ה- c' הדרוש לשם כך?
 $P_{H_0}(X \geq c') = P_{H_0}\left(\frac{X}{\sigma} \geq \frac{c'}{\sigma}\right) = 1 - \Phi\left(\frac{c'}{\sigma}\right) = 0.025 \Rightarrow c' \approx 1.96\sigma$

- יהיו $X_i \sim \text{Pois}(\theta)$ וההשערות: $H_0: \theta = \theta_0, H_1: \theta = \theta_1$ עבור $\theta_0 < \theta_1$
 - נבנה כלל דחיה על פי מבחן יחס הנראות:

$$L(X_1, \dots, X_n) = \frac{P_{H_1}(X_1, \dots, X_n)}{P_{H_0}(X_1, \dots, X_n)} = \frac{\prod_{i=1}^n e^{-\theta_1} \frac{\theta_1^{X_i}}{X_i!}}{\prod_{i=1}^n e^{-\theta_0} \frac{\theta_0^{X_i}}{X_i!}} = \prod_{i=1}^n e^{-(\theta_1 - \theta_0)} \left(\frac{\theta_1}{\theta_0}\right)^{X_i} = e^{(-\theta_1 - \theta_0)n} \left(\frac{\theta_1}{\theta_0}\right)^{\sum_{i=1}^n X_i} \geq c$$

- כל מה שקיבלנו זו פונקציה מונוטונית עולה של $\sum_{i=1}^n X_i$ (כי $\theta_0 < \theta_1$) ולכן הבסיס של החזקה גדול מ-1), וכך אפשר לומר שהמבחן הזה שקול למבחן $\sum_{i=1}^n X_i \geq c'$ שהוא קל יותר לחישוב.
- מתרגיל שהיה קודם, אנו יודעים ש- $\sum_{i=1}^n X_i \sim \text{Pois}(n\theta)$, מה שאומר שגם קל לנו לחשב הסתברויות על תנאי הדחייה, ולחשב p-value.

- יהיו $X_i \sim \text{Ber}(p)$ עם ההשערות $H_0: p = p_0, H_1: p = p_1 > p_0$
 - נבנה את מבחן יחס הנראות:

$$L(X_1, \dots, X_n) = \frac{P_{H_1}(X_1, \dots, X_n)}{P_{H_0}(X_1, \dots, X_n)} = \frac{\prod_{i=1}^n p_1^{x_i} (1-p_1)^{1-x_i}}{\prod_{i=1}^n p_0^{x_i} (1-p_0)^{1-x_i}} = \left(\frac{p_1}{p_0}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-p_1}{1-p_0}\right)^{n-\sum_{i=1}^n x_i} = \left(\frac{p_1(1-p_1)}{p_0(1-p_0)}\right)^{\sum_{i=1}^n x_i} \left(\frac{1-p_1}{1-p_0}\right)^n \geq c$$

- הביטוי שקיבלנו הוא שוב פונקציה מונוטונית עולה של $\sum_{i=1}^n X_i$, ולכן המבחן שוב שקול ל- $\sum_{i=1}^n X_i \geq c'$.
- נשים לב שכיוון שהמ"מ כאן בדיד, יש טעם לדבר רק על $\{0, 1, \dots, n+1\}$. אם רוצים α מסוים ספציפי מאוד, יכול להיות שלא ניתן להשיג אותו באמצעות c' כלשהו. במקרה כזה, נצטרך גרסה מתוחכמת של המבחן: יהיה לנו סף דחייה וודאית וסף קבלה וודאית, ובאיזור החיץ, המבחן ידחה את H_0 בהסתברות מסוימת (לא 0 ולא 1).
- יהיו $X_i \sim \exp(\theta)$ והשערות $H_0: \theta = \theta_0, H_1: \theta = \theta_1 > \theta_0$.
- מבחן יחס הנראות המתקבל הוא מהצורה $\sum_{i=1}^n X_i \leq c$ (הוכחה בפתרון תרגיל 13).

משפט (ניימן-פירסון): יהי $L(X) \geq c$ מבחן יחס הנראות עבור שתי השערות, R – תחום דחיית H_0 , α, β כרגיל. יהי $X \in S$ מבחן כלשהו, S – תחום דחיית H_0 , α', β' כרגיל. אזי $\alpha' \leq \alpha \Rightarrow \beta' \leq \beta$.

הוכחה: מההגדרות,

$$\begin{aligned}\alpha &= P_{H_0}(X \in R) = P_{H_0}(X \in R \cap S) + P_{H_0}(X \in R \cap \bar{S}) \\ \alpha' &= P_{H_0}(X \in S) = P_{H_0}(X \in R \cap S) + P_{H_0}(X \in \bar{R} \cap S) \\ 0 \leq \alpha - \alpha' &= P_{H_0}(X \in R \cap \bar{S}) - P_{H_0}(X \in \bar{R} \cap S) \Rightarrow P_{H_0}(X \in \bar{R} \cap S) \leq P_{H_0}(X \in R \cap \bar{S})\end{aligned}$$

כעת נסתכל על H_1 : ממבחן יחס הנראות, $\frac{P_{H_1}(X \in R \cap \bar{S})}{P_{H_0}(X \in R \cap \bar{S})} \geq c$, שכן כל המדגמים הם ב- R ולכל אחד המנה $\leq c$. לכן מקבלים:

$$\frac{P_{H_1}(X \in R \cap \bar{S})}{c} \geq P_{H_0}(X \in R \cap \bar{S}) \geq P_{H_0}(X \in \bar{R} \cap S)$$

ובאותו אופן, $\frac{P_{H_1}(X \in \bar{R} \cap S)}{P_{H_0}(X \in \bar{R} \cap S)} \leq c$, כי כל המדגמים אינם ב- R ולכן עבורם המבחן החזיר $c \geq$. מכאן שוב מקבלים:

$$\frac{P_{H_1}(X \in \bar{R} \cap S)}{c} \leq P_{H_0}(X \in \bar{R} \cap S) \leq P_{H_0}(X \in R \cap \bar{S})$$

כעת נסתכל על ה- β ו- β' :

$$\begin{aligned}\beta &= P_{H_1}(X \in \bar{R}) = P_{H_1}(X \in \bar{R} \cap S) + P_{H_1}(X \in \bar{R} \cap \bar{S}) \\ \beta' &= P_{H_1}(X \in \bar{S}) = P_{H_1}(X \in R \cap \bar{S}) + P_{H_1}(X \in \bar{R} \cap \bar{S})\end{aligned}$$

כמו שראינו קודם,

$$\frac{P_{H_1}(X \in \bar{R} \cap S)}{c} \leq P_{H_0}(X \in \bar{R} \cap S) \leq P_{H_0}(X \in R \cap \bar{S}) \leq \frac{P_{H_1}(X \in R \cap \bar{S})}{c}$$

ומכאן המסקנה: $P_{H_1}(X \in \bar{R} \cap S) \leq P(X \in R \cap \bar{S})$, ולכן $\beta \leq \beta'$ כנדרש. ■

במובן מסוים המשפט אומר שמבחן יחס הנראות הוא הכי טוב שאפשר להשיג – אי אפשר לשפר את המובהקות (α) מבלי לפגוע בעוצמת המבחן ($1 - \beta$).

מבחן יחס הנראות המוכלל

כאן נרצה לעסוק במקרה שההשערות אינן שתי נקודות (למשל, שני ניחושים של הפרמטר) אלא משהו מורכב יותר. נתחיל בדוגמא ואז נבנה מבחן שפועל על כל שתי השערות.

דוגמאות:

- יהיו $X \sim N(\mu, \sigma^2)$, לרשותנו מדגם של דגימה אחת, וההשערות: $H_0: \mu = \mu_0, H_1: \mu > \mu_0$. נתחיל בניחוש: כיוון שאנו רוצים לבדוק השערה לגבי התוחלת, יהיה הגיוני להסתכל על המדגם עם נרמול, כלומר על מבחן מהצורה $\frac{X - \mu_0}{\sigma} \geq c$.
- נתבונן בטעויות: כיוון ש- α תלוי ב- H_0 , זה חישוב שכבר עשינו. לעומת זאת, β כאן היא כבר פונקציה של μ :

$$\beta(\mu) = P_{H_1}\left(\frac{X - \mu_0}{\sigma} \leq c\right) = P_{H_1}\left(\frac{X - \mu}{\sigma} \leq c + \frac{\mu_0 - \mu}{\sigma}\right) = \Phi\left(c + \frac{\mu_0 - \mu}{\sigma}\right)$$
- ככל ש- μ גדל, $(1 - \beta(\mu))$ מתקרב ל-0 בהדרגה. בנוסף, כאשר $\mu = \mu_0$, מקבלים $\beta(\mu) = \alpha$, כי אז:

$$P_{H_1}\left(\frac{X - \mu}{\sigma} \leq c + \frac{\mu_0 - \mu}{\sigma}\right) = P_{H_1}\left(\frac{X - \mu_0}{\sigma} \leq c\right) = \alpha$$

הגדרה: מבחן יחס ניראות מוכלל (Generalized Likelihood Ratio Test) הוא מבחן בדיקת השערות כאשר

$$L(X) = \frac{\max_{\theta \in \Theta_1} P(X_1, \dots, X_n; \theta)}{\max_{\theta \in \Theta_0} P(X_1, \dots, X_n; \theta)} \geq c$$

ההשערות מהצורה $H_0: \theta \in \Theta_0, H_1: \theta \in \Theta_1$ שסוף הדחייה שלו הוא c

דוגמאות:

- מטילים n הטלות קוביה בעלת הסתברויות $\theta = (\theta_1, \theta_2, \dots, \theta_6)$, וההשערות H_0 – הקוביה הוגנת, H_1 – הקוביה לא הוגנת. נבנה מבחן יחס ניראות מוכלל לבדיקת השערות אלה:
 - ברור מה צריך להיות במכנה - $P_{H_0}(X) = \left(\frac{1}{6}\right)^n$, זה נכון לכל מדגם. לעומת זאת, במונה אנו רוצים "תחרות" – המודל הכי סביר תחת H_1 מכל האפשרויות על כל הפרמטרים של הבעיה. נסמן n_i מספר הפעמים ש- i יצא במדגם, ונקבל:

$$P(X_1, \dots, X_n; \theta) = \prod_{i=1}^n P(X_i; \theta) = \theta_1^{n_1} \dots \theta_6^{n_6} = \theta_1^{n_1} \dots \theta_5^{n_5} \left(1 - \sum_{i=1}^5 \theta_i\right)^{n - \sum_{i=1}^5 n_i}$$
 - זה ביטוי שאפשר להוציא מתוכו לוג ואז לגזור לפי 5 הפרמטרים כדי למצוא את המקסימום:

$$\log P(X_1, \dots, X_n; \theta) = \sum_{i=1}^5 n_i \log \theta_i + \left(n - \sum_{i=1}^5 n_i\right) \log \left(1 - \sum_{i=1}^5 \theta_i\right)$$
 - נחשב את הנגזרת:

$$\frac{d}{d\theta_i} \log P(X_1, \dots, X_n; \theta) = \frac{n_i}{\theta_i} + \left(n - \sum_{j=1}^5 n_j\right) \left(-\frac{1}{1 - \sum_{j=1}^5 \theta_j}\right) = 0 \Rightarrow \frac{n_i}{\theta_i} = \frac{1 - \sum_{j=1}^5 n_j}{1 - \sum_{j=1}^5 \theta_j}$$
 - נשים לב גם שזה המקסימום ולא המינימום (כי המינימום – שהוא 0 – מושג כאשר איזשהו $\theta_i = 0$ וה- n_i המתאים לו אינו 0).
 - מכאן מבחן יחס הניראות המוכלל הוא (נשים לב שכאן המקריות היא כבר ב- n_i – אין X_i בביטוי):

$$L(X) = \prod_{i=1}^5 \left(\frac{n_i}{n}\right)^{n_i} \left(\frac{1}{1 - \sum_{j=1}^5 \frac{n_j}{n}}\right)^{n - \sum_{j=1}^5 n_j} \geq c$$

- יהיו $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$. מקבלים מדגמים מצומדים $(X_1, \dots, X_n), (Y_1, \dots, Y_n)$. נגדיר $Z = X - Y$ ושתי השערות $H_0: \mu_1 = \mu_2, H_1: \mu_1 \neq \mu_2$.

נשים לב לכך ש- $Z \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. נכתוב את מבחן יחס הניראות, כאשר $Z_i = X_i - Y_i$

$$L(X) = \frac{\max_{\mu \in \Theta_1} \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{(Z_i - \mu)^2}{2\sigma_z^2}}}{\prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_z^2}} e^{-\frac{Z_i^2}{2\sigma_z^2}}} \geq c$$

- כיוון ש- $\bar{X}$ הוא אומד הניראות המירבית לתוחלת של "מ"מ נורמלי", כל הביטוי ניתן לצמצום – הגורם המונה הוא זה שבו $\mu = \bar{X}$.

