יסודות תורת ההסתברות

אוהד נ. פלדהיים
אורי גוראל גורביץ'

מכון איינשטיין למתמטיקה
האוניברסיטה העברית בירושלים

23 בנובמבר 2018

פרק 6

השונות וסטיית התקן

אי-שפיות: לחזור על אותה פעולה שוב ושוב ולצפות לתוצאות שונות.

– פתגם אמריקני

בפרק הקודם התוודענו למושג התוחלת אשר תיאר בעינינו את התנהגותו הממוצעת של משתנה מקרי. ערך זה היווה את "מרכז הכובד" של ההתפלגות, בדומה למרכז הכובד של גוף בעל מסה במרחב. למרות שכבר נוכחנו שמונח זה הנו שימושי כשלעצמו, הרי שמרגע שהוצג ביקשנו לתאר באמצעותו גם את התנהגותו הטיפוסית של המשתנה המקרי.

מובן שלא נוכל לעשות זאת באופן כללי. אם נחשוב למשל על ההשתתפות בהגרלת הלוטו המערבת ניחוש של שישה מספרים שונים מתוך שלושים ושבעה בעבור פרס של שמונה מיליון שקלים. חישוב זריז ילמדנו כי הסתברות לניחוש מדוייק של צירוף המספרים היא כאחד לשני מיליון ותוחלת סכום הזכיה בפרס היא אפוא כארבעה שקלים. המצב הטיפוסי לעומת זאת הוא, בלי צל של ספק, היעדר זכיה. בכדי להבחין בין התנהגות ממוצעת והתנהגות טיפוסית נבקש להגדיר **מדדי פיזור**, שיאמדו את מידת המהימנות של התוחלת כקירוב להתנהגות טיפוסית של המשתנה המקרי. הראשונה והחשובה בין מדדים אלה היא **השונות**.

6.1 שונות

למרות שמושג ה**שונות** בהקשרו הסטטיסטי נטבע על ידי הסטטיסטיקאי רונאלד פישר (Ronald Fisher) ב-1918, הרי שנוסחת השונות שימשה את פאפנוטי צ'בישב ואת תלמידו אנדריי מרקוב עוד ברוסיה הצארית ולפחות החל משנות השמונים של המאה התשע-עשרה. חלק מקסמו של המושג בכך שהוא משתמש בתוחלת עצמה להגדרת אומד פיזור.

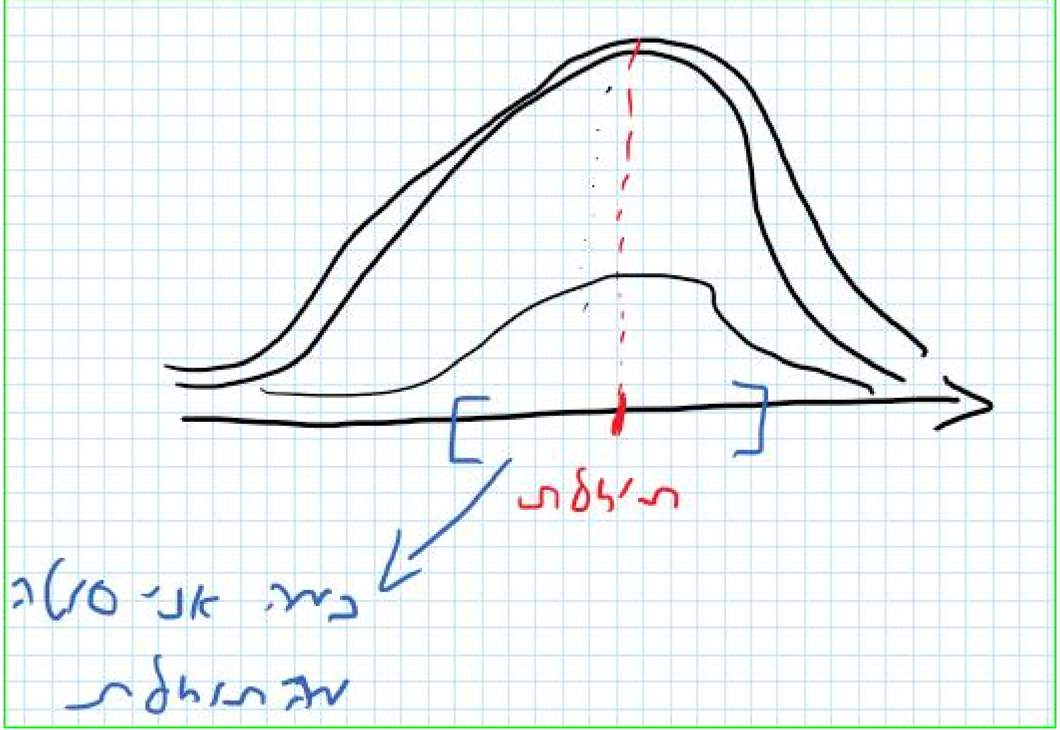
הגדרה 6.1 (שונות וסטיית תקן)**.** יהי X משתנה מקרי, בעל תוחלת סופית. ה**שונות (variance)** של X, מוגדרת כ-

$$\mathrm{Var}(X) := \mathbb{E}\Big(\big(X - \mathbb{E}(X)\big)^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

אם ל-X^2 אומנם יש תוחלת. אחרת נאמר של-X יש שונות אינסופית.

את השורש הריבועי של השונות $\sigma = \sqrt{\mathrm{Var}(X)}$ נכנה בשם **סטיית התקן של** X.

נשים לב שהשוויון בין שתי ההגדרות של השונות נובע מליניאריות התוחלת, טענה 5.5ב, לפי


תוחלת
כמה אני סגורה
שהתוחלת

$$\mathbb{E}\Big(\big(X-\mathbb{E}(X)\big)^2\Big) = \mathbb{E}\big(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2\big)$$
$$= \mathbb{E}(X^2) - 2\mathbb{E}(X)\mathbb{E}(X) + \mathbb{E}(X)^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

ראשית ניווכח כי שונות חיובית מהווה עדות לפיזור כלשהו.

טענה 6.2. יהי X משתנה מקרי בעל תוחלת סופית. אזי X קבוע כמעט תמיד אם ורק אם $\mathrm{Var}(X)=0$.

הוכחה. נסמן $\mu=\mathbb{E}(X)$ ונפתח לפי הגדרה $\mathrm{Var}(X)=\sum_{\omega\in\Omega}(X(\omega)-\mu)^2\mathbb{P}(\omega)$. זהו סכום של מספרים אי-שליליים ולכן שווה לאפס אם ורק אם כל המחוברים שווים ל-0, כלומר, אם ורק אם לכל $\omega\in\Omega$ עבורו $X(\omega)\neq\mu$ מתקיים $\mathbb{P}(\{\omega\})=0$. ■

השונות אינה ליניארית באופן כללי. תחת זאת היא מקיימת את התכונות הבאות.

טענה 6.3 (תכונות השונות). יהי X מ"מ בעל שונות סופית, ויהי $a\in\mathbb{R}$. אזי

(א) $\mathrm{Var}(X+a)=\mathrm{Var}(X)$.

(ב) $\mathrm{Var}(aX)=a^2\,\mathrm{Var}(X)$ (ולכן $\sigma(aX)=|a|\sigma(X)$)

(ג) אם Y משתנה מקרי בעל שונות סופית **בלתי-תלוי** ב-X אז $\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$.

הוכחה. **סעיף א.** לפי ליניאריות התוחלת (טענה 5.5ב) מתקיים $\mathbb{E}(X+a)=\mathbb{E}(X)+a$ ולכן

$$\mathbb{E}\left((X+a)^2\right)=\mathbb{E}\left(X^2+2aX+a^2\right)=\mathbb{E}\left(X^2\right)+2a\mathbb{E}(X)+a^2$$
$$\mathbb{E}(X+a)^2=\mathbb{E}(X)^2+2a\mathbb{E}(X)+a^2$$
$$\mathrm{Var}(X+a)=\mathbb{E}\left((X+a)^2\right)-\mathbb{E}(X+a)^2)=\mathbb{E}\left((X-\mathbb{E}(X))^2\right)=\mathrm{Var}(X).$$

סעיף ב. לפי ליניאריות התוחלת מתקיים

$$\mathrm{Var}(aX)=\mathbb{E}(a^2X^2)-\mathbb{E}(aX)^2=a^2\mathbb{E}(X^2)-(a\mathbb{E}(X))^2=a^2\,\mathrm{Var}(aX).$$

סעיף ג. נשתמש בכפליות התוחלת למשתנים בלתי-תלויים (טענה 5.10) ובליניאריות התוחלת ונחשב

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

כאשר השוויון האחרון מתקבל מהחישוב

$$\mathbb{E}\left((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right)=\mathbb{E}(X-\mathbb{E}(X))\mathbb{E}(Y-\mathbb{E}(Y))=0\cdot 0=0$$

■

דוגמא 6.4 (שונות משתנים מקריים מוכרים).

לאורך הדוגמא נסתמך על חישוב תוחלת מדוגמא 5.3 ועל הגדרת התוחלת.

(א) שונות מ"מ X בעל **התפלגות אחידה** על $[N]$ היא

$$\mathrm{Var}(X)=\sum_{n\in[N]} n^2\mathbb{P}(X=n)-\left(\frac{N+1}{2}\right)^2=\frac{N(N+1)(2N+1)}{6N}-\frac{(N+1)^2}{4}=\mathbf{\frac{N^2-1}{12}}.$$

(ב) שונות משתנה מקרי **ברנולי** $X\sim\mathrm{Ber}(p)$ היא

$$\mathrm{Var}(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2=p-p^2=\mathbf{p(1-p)}.$$

(ג) שונות משתנה מקרי **בינומי** $X\sim\mathrm{Bin}(N,p)$ תחושב בצורה הבאה. ניזכר כי X שווה בהתפלגות לסכום $\{Y_n\}_{n\in\mathbb{N}}\sim\mathrm{Ber}(p)$ כאשר Y_n ב"ת. לפי טענה 6.3 וסעיף ב נקבל

$X\sim\sum_{n\in N} Y_n$
בינומי
ברנולי

$$\mathrm{Var}(X)\overset{\text{טענה 6.3ג}}{=}\sum_{n=1}^{N}\mathrm{Var}(Y_n)\overset{\text{סעיף ב}}{=}\mathbf{Np(1-p)}$$

(ד) שונות משתנה מקרי **פואסוני** $X\sim\mathrm{Po}(\lambda)$ נחשב באמצעות שימוש עקיף בטור טיילור. נרשום

$$e^{\lambda}=\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}=\sum_{n=2}^{\infty}(n)(n-1)\frac{\lambda^{n-2}}{n!}$$

לצרכי ג
את הטור

נכפיל ב-$\frac{\lambda^2}{e^{\lambda}}$ ונקבל

$$\lambda^2=\sum_{n=2}^{\infty}(n)(n-1)\frac{\lambda^n}{n!e^{\lambda}}=\mathbb{E}\Big(X(X-1)\Big)=\mathbb{E}(X^2)-\mathbb{E}(X).$$

מכאן ש $\mathbb{E}(X^2)=\lambda^2+\lambda$ ולכן

$$\mathrm{Var}(X)=\mathbb{E}(X^2)-\mathbb{E}(X)^2=\lambda.$$

(ה) שונות משתנה מקרי **גיאומטרי** $X\sim\mathrm{Geo}(p)$ נחשב באופן הבא. נחזור לנוסחה לסכום של טור הנדסי עבור $|x|<1$:

$$\sum_{n=0}^{\infty}x^n=\frac{1}{1-x}$$

נגזור את שני הצדדים פעמיים ונקבל:

$$\sum_{n=2}^{\infty}n(n-1)x^{n-2}=\frac{2}{(1-x)^3}$$

כאשר הגזירה מוצדקת כיון שטור הנגזרות (הראשונות והשניות) מתכנס במידה שווה בסביבת x. נציב כעת $x=1-p$ ונקבל:

$$\mathbb{E}(X^2)-\mathbb{E}(X)=\mathbb{E}(X(X-1))=\sum_{n=2}^{\infty}n(n-1)p(1-p)^{n-1}=p(1-p)\frac{2}{p^3}=\frac{2(1-p)}{p^2}.$$

לכן,

$$\mathrm{Var}(X)=\mathbb{E}(X(X-1))+\mathbb{E}(X)-\mathbb{E}(X)^2=\frac{2(1-p)}{p^2}+\frac{1}{p}-\frac{1}{p^2}=\mathbf{\frac{1-p}{p^2}}.$$

6.2 שונות ככלי להוכחת ריכוז

לפי הגדרתה השונות היא **תוחלת** של ריבוע הפרש בין מ"מ לתוחלתו. כתוצאה מכך ניתן להפעיל עליה את אי-שוויון מרקוב (משפט 5.15) ולקבל את התוצאה הבאה באופן מיידי.

משפט 6.5 (אי-שוויון צ'בישב). יהי X מ"מ בעל שונות סופית. אזי לכל $a > 0$ מתקיים

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

הוכחה. נגדיר משתנה מקרי חדש $Y = (X - \mathbb{E}(X))^2$. זהו משתנה מקרי אי-שלילי המקיים $\mathbb{E}(Y) = \mathrm{Var}(X)$. על כן, לפי אי-שיוויון מרקוב (משפט 5.15), לכל $b > 0$ מתקיים

$$\mathbb{P}\left(Y \geq b\right) \leq \frac{\mathbb{E}(Y)}{b} = \frac{\mathrm{Var}(X)}{b}.$$

נציב $b = a^2$ ונקבל

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geq a\right) = \mathbb{P}\left((X - \mathbb{E}(X))^2 \geq a^2\right) = \mathbb{P}(Y \geq a^2) \leq \frac{\mathrm{Var}(X)}{a^2}.$$

■

ניסוח שימושי אחר של המשפט מתקבל בהצבת $b = a\sigma_X$,

מסקנה 6.6 (א"ש צ'בישב - ניסוח אחר). יהי X מ"מ אי-שלילי בעל תוחלת סופית וסטיית תקן σ_X. לכל $b > 0$ מתקיים

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq b\sigma_X) \leq \frac{1}{b^2}.$$

אי-שוויון צ'בישב הוא הדרך הבסיסית והנגישה ביותר להראות כי משתנה מקרי מקבל בהסתברות גבוהה ערך קרוב לתוחלתו.

דוגמא 6.7 (ריכוז התפלגות בינומית). מיליון מטבעות הוגנים מוטלים. נחפש חסם מלמטה את ההסתברות שכמות המטבעות שתוצאתם עץ תהייה בין $495,000$ ל-$505,000$.

תשובה: נסמן ב-X את מספר העצים שהתקבלו בניסוי. אזי $X \sim \mathrm{Bin}(10^6, \frac{1}{2})$ ולפי דוגמא 5.3א ודוגמא 6.4ג מתקיים $\mathbb{E}(X) = 500000$ ו-$\mathrm{Var}(X) = 250000$. נפעיל את א"ש צ'בישב, משפט 6.5 ונקבל

$$\mathbb{P}(495000 < X < 505000) = 1 - \mathbb{P}\left(\left|X - \mathbb{E}(X)\right| \geq 5000\right) \geq 1 - \frac{250{,}000}{5000^2} = \frac{99}{100}$$

כלומר תחום הערכים $[495000,505000]$ מתקבל בסיכוי של 99%. בסטטיסטיקה מכונה טווח ערכים כזה רווח בר-סמך של 99%, או טווח מובהקות של 1% (ר' הערה להלן).

הערה: טווח ערכים סימטרי סביב התוחלת אשר בו צפוי משתנה מקרי להימצא בסבירות p מכונה **רווח בר-סֶמֶך confidence interval** והוא מהווה מונח חשוב בסטטיסטיקה, במיוחד בהקשר של סקרים ואומדנים. במדעים שונים נהוג רווח בר-סמך שונה בכדי לקבוע שתגלית סטטיסטית ראויה לפרסום. רווח בר-סמך משמש גם בהפרדת השערות בכדי לשלול השערה מסויימת על סמך האבחנה שמשתנה מקרי מסויים נמצא בפועל מחוץ לרווח בר-סמך לפי המודל המתחייב מאותה השערה.

נכליל תופעה זו ונראה כי באופן כללי ממוצעים של ניסויים חוזרים נוטים להיות מרוכזים סביב הממוצע.

משפט החוק החלש של המספרים הגדולים:

לדוגמה מתוארת כאן מצב בו אם אעשה ניסוי ואמדד את הגובה הממוצע של התלמידים בכיתה ויצא לי

מכן הממוצע הגובה הארצי השונות של הממוצע של ילדי הכיתה ביחס לשונות של ממוצע הגובה הארצי שואפת ל-0 אז

ככל שיהיו לי יותר תלמידים בכיתה כך ביחס לממוצע הארצי השונות תשאף ל-0 במשוואה כאן

$$P_N = P\left(\underbrace{\left|\underbrace{\frac{1}{N}\sum_{n=1}^{N} x_n}_{\text{ממוצע הגובה בכיתה}} - \underbrace{E(x)}_{\text{ממוצע גובה הארצי}}\right| < \varepsilon}_{\text{ההסתברות של השונות קטנה מ[illegible] } \varepsilon}\right)$$

לכן $P_N \xrightarrow{n \to \infty} 1$

משפט 6.8 (החוק החלש של המספרים הגדולים למשתנים בעלי שונות). יהי X משתנה מקרי בעל שונות סופית ויהי $\epsilon > 0$. לכל $N \in \mathbb{N}$, נרשום

$$p_N = \mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \mathbb{E}(X)\right| < \epsilon\right)$$

כאשר $\{X_n\}_{n\in[N]}$ משתנים מקריים בלתי תלויים שווי התפלגות ל-X. אזי

$$\lim_{N\to\infty} p_N = 1.$$

הוכחה. יהי N ונסמן $S_N = \frac{1}{N}\sum_{n=1}^{N} X_n$. לפי ליניאריות התוחלת (טענה 5.5ב), מתקיים

$$\mathbb{E}(S_N) \overset{\text{טענה 5.5ב}}{=} \frac{1}{N}\mathbb{E}\left(\sum_{n=1}^{N} X_n\right) \overset{\text{טענה 5.5ב}}{=} \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}(X_n) \overset{\text{שווי התפלגות}}{=} \mathbb{E}(X).$$

כיוון ש-$\{X_n\}_{n\in[N]}$ ב"ת, לפי טענה 6.3 מתקיים

$$\mathrm{Var}(S_N) \overset{\text{טענה 6.3ג}}{=} \frac{1}{N^2}\,\mathrm{Var}\left(\sum_{n=1}^{N} X_n\right) \overset{\text{טענה 6.3ג}}{=} \frac{1}{N^2}\sum_{n=1}^{N}\mathrm{Var}(X_n) \overset{\text{שווי התפלגות}}{=} \frac{\mathrm{Var}(X)}{N}.$$

נפעיל את אי-שוויון צ'בישב (משפט 5.6) ונקבל

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \mathbb{E}(X)\right| < \epsilon\right) = 1 - \mathbb{P}(|S_N - \mathbb{E}(S_N)| \geq \epsilon) \geq 1 - \frac{\mathrm{Var}(S_N)}{\epsilon^2} = 1 - \frac{\mathrm{Var}(X)}{N\epsilon^2}.$$

ולאחר השאפה של N לאינסוף מתקבל המשפט.

■

כעת נוכל להשלים חוב ישן שמקורו מן ההקדמה לחוברת, ולפתור באופן משביע רצון את בעיית המשחק שנקטע של לוקה פאצ'ולי.

דוגמה 6.9 (המשחק שנקטע).

אביה ובתיה מהמרים על סכום כסף מסויים במשחק מזל. את סכום הכסף מפקידים בקופה ואז מתחילים לשחק מספר סיבובים. בכל סיבוב של המשחק לכל אחד מהשחקנים סיכוי שווה לנצח באופן בלתי תלוי בסיבובים הקודמים. הראשון שזוכה בשישה משחקים מנצח ומקבל את כל הסכום שבקופה. בשעה שבתיה הובילה על אביה עם חמישה נצחונות מול שלושה, ירד הלילה ונוצר הכרח לקטוע את המשחק. כיצד עליהם לחלק את הקופה באופן שיהלום בהסתברות גבוהה את חלוקת הכספים הממוצעת אילו היינו מתחילים מספר רב של משחקים בלתי תלויים ממצב זה (של הובלה חמש-שלוש) וממשיכים עד שאחד הצדדים מנצח.

תשובה: נסמן ב-$\{X_n\}_{n\in[N]}$ אוסף של מ"מ ב"ת שווי-התפלגות שיתארו את החלק מתוך הקופה שמקבלת אביה במשחק ספציפי שהחל ממצב של חמש-שלוש לטובתה. אנו מעוניינים להעריך את $\frac{1}{N}\sum X_n$. לפי החוק החלש של המספרים הגדולים, משפט 6.8, כאשר N ישאף לאינסוף יהיה סכום זה קרוב כרצוננו ל-$\mathbb{E}(X_1)$ בהסתברות קרובה כרצוננו ל-1.

נחשב אפוא את $\mathbb{E}(X_1)$. ניתן דעתנו על כך ש-X_1 הוא משתנה אינדיקטור שכן מדובר במשחק בו לוקח המנצח את כל הקופה. נסמן ב-A_i את המאורע לעובדה שאביה ניצחה בסיבוב ה-i במשחק שמתחיל בתוצאה חמש-שלוש. לפי נוסחת ההסתברות השלמה, טענה 3.7, אם נחלק את המאורע שאביה תנצח, למקרה שבו היא מנצח בסיבוב הראשון, בסיבוב השני לאחר שהובסה בראשון או בסיבוב השלישי לאחר שהובסה בשני קודמיו, נראה כי מתקיים

$$\begin{aligned}\mathbb{E}(X_1) &= \mathbb{P}(\text{אביה תנצח}) = \mathbb{P}(A_1) + \mathbb{P}(A_2 \cap A_1^c) + \mathbb{P}(A_3 \cap A_1^c \cap A_2^c) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \mid A_1^c)\mathbb{P}(A_1^c) + \mathbb{P}(A_3 \mid A_1^c, A_2^c)\mathbb{P}(A_1^c \mid A_2^c)\mathbb{P}(A_1^c) \\ &= \frac{1}{2} + \frac{1}{2\cdot 2} + \frac{1}{2\cdot 2\cdot 2} = \frac{7}{8}\end{aligned}$$

ולכן אילולא המשחק נקטע, ולו היו משחקים שוב ושוב את המשך המשחק הייתה בתיה מקבלת בממוצע $\frac{7}{8}$ מהסכום ואביה $\frac{1}{8}$ ממנו.

דוגמא 6.10 (השונות ובעיית האספן).

נשוב לתיאור בעיית האספן (דוגמא 4.32), לפי דוגמא 5.18. כמקודם אספן רוכש ביצי הפתעה. כל ביצה מכילה אחד מ-n סוגים של בובות אהובות אשר נבחרים באקראי ובאופן ב"ת. ראשית רוכש האספן ביצה אחת. לאחר מכן, הוא רוכש ביצים עד שיקבל בובה השונה מהבובה שהופיעה בביצה הראשונה. בשלב הבא הוא רוכש ביצים עד שיקבל בובה ששונה משני סוגי הבובות שכבר ברשותו וכן הלאה עד שהוא משיג עותק של כל אחת מהבובות. בדוגמא 5.18 ראינו כי תוחלת מספר הרכישות עד להשגת עותק מכל בובה היא $n\sum_{i=1}^n \frac{1}{i}$. השתמש בא"ש צ'בישב (משפט 6.5) על מנת למצא טווח שספר הרכישות הנחוץ להשגת עותק של כל בובה ימצא בתוכו בהסתברות של 75%.

תשובה: כמו בדוגמא 5.18 נתאר את כמות הביצים שנרכשו בשלב ה-i כוקטור של משתנים מקריים גיאומטריים בלתי-תלויים כך ש- $X_i \sim \mathrm{Geo}(\frac{n-i+1}{n})$, ונסמן $X = \sum_{i=1}^n X_i$. נחשב את

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^n X_i\right) \overset{\text{טענה 6.3ג}}{=} \sum_{i=1}^n \mathrm{Var}\,(X_i) \overset{\text{טענה 6.4ה}}{=} \sum_{i=1}^n \frac{\frac{i-1}{n}}{\frac{(n-i+1)^2}{n^2}} = \sum_{i=1}^n \frac{n(i+1)}{(n-i+1)^2}$$

נציב $j = n - i + 1$ ונקבל

$$\mathrm{Var}(X) = \sum_{j=1}^n \frac{n(n-j)}{j^2} = n^2 \sum_{j=1}^n \frac{1}{j^2} - n\sum_{j=1}^n \frac{1}{j} \le n^2 \sum_{j=1}^\infty \frac{1}{j^2} = \frac{\pi n^2}{6}$$

כאשר השוויון האחרון הוא פתרון אויילר לבעיית באסל שהוזכר בדוגמא 2.10.

מכאן ש-$\sigma_X \le n\sqrt{\pi/6}$ נפעיל את ניסוחו השני של משפט צ'בישב (מסקנה 6.6), ונקבל

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \ge 2n\sqrt{\pi/6}\right) \le \frac{1}{4}$$

מספיק הדוק

ניזכר כי

$$n(\log n) < n\sum_{i=1}^n \frac{1}{i} < n(\log n + 1).$$

ונקבל שבהסתברות של 75% לפחות מספר הביצים שנדרש לקנות עד להשגת עותק מכל סוג בובה יהיה בין $n(\log n - 2\sqrt{\pi/6})$ לבין $n(\log n + 2\pi/6 + 1)$. למשל במקרה של 50 סוגי ביצים ידרשו בין 123 ל-317 ביצים.

תכונת הלינ׳ בתוחלת $E(x)$

עזרה לנו לחשב תוחלות מסובכות

גם כאן אנו רוצים שתהיה לנו דרך

נוחה לחשב את השונות (Var)

אז נוכל במקרה גם להתייחס לזוגות

של תוחלות של שמ״מ. שיעזרו לנו לחשב שונות

בדרך קלה יחסית לעשות כך נגדיר

תבנית בילינארית $= cov(x,y) =$ שונות משותפת

6.3 שונות משותפת

כשם שהגדרנו את השונות כמדד לפיזור מ"מ סביב תוחלתו, כן נוכל להגדיר **שונות משותפת** כמדד לפיזור מכפלת משתנים מקריים סביב מכפלת התוחלות.

הגדרה 6.11 (שונות משותפת). יהיו X, Y משתנים מקריים על אותו מרחב הסתברות, בעלי תוחלת סופית. **השונות המשותפת (covariance) של X ו-Y**, מוגדרת על ידי

$$\mathrm{Cov}(X, Y) := \mathbb{E}\Big((X - \mathbb{E}(X)(Y - \mathbb{E}(Y))\Big) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

כאשר תוחלת זו מוגדרת היטב. שני משתנים אשר שונותם המשותפת מתאפסת – נקראים **בלתי-מתואמים (uncorrelated)**.

דרך אחת להבין את השונות המשותפת היא כמדד עד כמה בהינתן $Y = y$ עשויה להשתנות $\mathbb{E}(X \mid Y = y)$. כלומר כמה מידע יש במאורעות על Y לגבי תוחלתו של X בהינתן מאורעות אלה. מנקודת מבט זו נצפה שכאשר ערכו של Y לא נותן לנו כל מידע לגבי התפלגותו של X – תהיה השונות המשותפת 0, כפי שניתן לראות במסקנה הבאה.

מסקנה 6.12 (אי-תלות גוררת חוסר מתואמות). אם X ו-Y בלתי-תלויים אז הם בלתי-מתואמים.

מסקנה זו היא למעשה ניסוח מחודש של טענה 5.10.

נשים לב שהכיוון ההפוך למסקנה 6.12 אינו נכון, כפי שניתן לראות בדוגמא הבאה.

דוגמה 6.13.

מוטלות שלוש קוביות משחק הוגנות. נסמן את תוצאות ההטלה בX, Y ו-Z. נראה כי Z ו-$(X - Y)Z$ תלויים אך בלתי-מתואמים.

תשובה: לפי טענה 4.34 המשתנים $(X - Y)$ ו-Z ב"ת. לכן, לפי טענה 5.10 מתקיים

$$\begin{aligned}\mathrm{Cov}(Z, (X - Y)Z) &= \mathbb{E}((X - Y)Z^2) - \mathbb{E}(X - Y)\mathbb{E}((X - Y)Z)\\ &\overset{\text{טענה 5.10}}{=} \mathbb{E}(X - Y)\,\mathbb{E}(Z^2) - \mathbb{E}(X - Y)\mathbb{E}(Z)^2\\ &= (\mathbb{E}(X) - \mathbb{E}(Y))\mathbb{E}(Z^2) - (\mathbb{E}(X) - \mathbb{E}(Y))\mathbb{E}(Z) = 0\end{aligned}$$

תוך שהשתמשנו בכך ש-X ו-Y שווי התפלגות ולכן $\mathbb{E}(X) - \mathbb{E}(Y) = 0$.

מכאן שהמשתנים Z ו-$(X - Y)Z$ בלתי-מתואמים. בכדי להיווכח בכך שהמשתנים תלויים נשים לב כי המאורעות $\{Z = 2\}$ ו-$\{(X - Y)Z = 1\}$ הם בעלי הסתברות חיובית ואולם חיתוכם בעל הסתברות 0.

מליניאריות התוחלת נוכל להסיק מספר תכונות של השונות המשותפת.

טענה 6.14 (תכונות השונות המשותפת). יהיו X, Y, Z מ"מ בעל שונות סופית, ויהיו $a, b \in \mathbb{R}$. אזי בכל מקרה בו אגף שמאל מוגדר היטב מתקיים.

(א) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(ב) $\operatorname{Cov}(X+Z,Y)=\operatorname{Cov}(X,Y)+\operatorname{Cov}(Z,Y)$.

(ג) $\operatorname{Cov}(aX,bZ)=ab\operatorname{Cov}(X,Z)$.

(ד) $\operatorname{Cov}(X,X)=\operatorname{Var}(X)$.

✎ **בעיה 6.1.** להוכיח את טענה 6.14.

באמצעות השונות המשותפת נוכל לחשב נוסחא לשונות של סכום שני משתנים מקריים כלליים. נוסחא זו היא למעשה הכללה של טענה 6.3ג.

טענה 6.15 (שונות של סכום מ"מ)**.** יהיו X,Y משתנים מקריים. אזי

$$\operatorname{Var}(X+Y)=\operatorname{Var}(X)+\operatorname{Var}(Y)+2\operatorname{Cov}(X,Y)$$

בכל מקרה בו אגף ימין של המשוואה מוגדר היטב.

הוכחה. נשתמש בהגדרת השונות (הגדרה 7.1) ובליניאריות התוחלת (טענה 5.5ב).

$$\begin{aligned}\operatorname{Var}(X+Y)&=\mathbb{E}\left((X+Y)^2\right)-\mathbb{E}(X+Y)^2\\&=\mathbb{E}\left(X^2\right)+2\mathbb{E}(XY)+\mathbb{E}\left(Y^2\right)-\mathbb{E}(X)^2-2\mathbb{E}(X)\mathbb{E}(Y)-\mathbb{E}(Y)^2\\&=\operatorname{Var}(X)+\operatorname{Var}(Y)+2\operatorname{Cov}(X,Y)\end{aligned}$$

■

ניתן להכליל נוסחא זו למספר כלשהו של משתנים מקריים.

טענה 6.16 (נוסחת שונות לסכום)**.** לכל אוסף $(X_n)_{n\in[N]}$ של מ"מ מתקיים

$$\operatorname{Var}\left(\sum_{n=1}^{N}X_n\right)=\sum_{n,k\le N}\operatorname{Cov}(X_n,X_k)=\sum_{n\le N}\operatorname{Var}(X_n)+2\sum_{n<k\le N}\operatorname{Cov}(X_n,X_k).$$

בכל מקרה בו אגף ימין של המשוואה מוגדר היטב.

✎ **בעיה 6.2.** להוכיח את טענה 6.16.

למעשה במסקנה 6.20 להלן נראה שדי בקיום שונות של כל אחד מהמשתנים בכדי להבטיח קיומה של שונות משותפת.

דוגמה 6.17 (מספר הרצפים בסדרת הטלות)**.**

מטבע מוטה בעל סיכוי p לתוצאה של עץ מוטל N פעמים. יהי Y מספר הפעמים שהופיעו שני עצים ברצף (עם חפיפות). חשב את תוחלת ושונות Y.

תשובה: נסמן ב-X_n משתנה אינדיקטור למאורע שהמטבע ה-n יצא עץ, ונשים לב כי $\{X_n\}_{n}\in[N]$ מתפלגים ברנולי עם סיכוי הצלחה p באופן בלתי-תלוי. נסמן $Y_n:=X_nX_{n+1}$ עבור $n\in[N-1]$. נשים לב

כי $Y = \sum_{n\in[N-1]} Y_n$. כעת נחשב לפי ליניאריות התוחלת (טענה 5.5ב):

$$\mathbb{E}(Y) = \sum_{n\in[N-1]} \mathbb{E}(Y_n) = \sum_{n\in[N-1]} \mathbb{E}(X_nX_{n+1}) \overset{\text{טענה } 5.10}{=} \sum_{n\in[N-1]} \mathbb{E}(X_n)\mathbb{E}(X_{n+1}) = (N-1)p^2$$

כמו כן נחשב לפי בעיה 6.2

$$\mathrm{Var}(Y) = \sum_{n,k\le N} \mathrm{Cov}(Y_n, Y_k) = \sum_{n=1}^{N} \mathrm{Var}(Y_n) + 2\sum_{n=1}^{N}\sum_{k=n+1}^{N} \mathrm{Cov}(Y_n, Y_k)$$

המשתנים המקריים Y_n מתפלגים ברנולי עם סיכוי הצלחה p^2 ועל כן

$$\mathrm{Var}(Y_n) = p^2(1-p^2).$$

כאשר $k > n+1$, המשתנים המקריים Y_n ו-Y_k בלתי תלויים לפי טענה 4.34 ולכן $\mathrm{Cov}(Y_n, Y_k) = 0$ במקרה זה. כאשר $k = n+1$, נקבל

$$\mathrm{Cov}(Y_n, Y_{n+1}) = \mathbb{E}(Y_nY_{n+1}) - \mathbb{E}(Y_n)\mathbb{E}(Y_{n+1}) = \mathbb{E}(X_nX_{n+1}X_{n+2}) - p^2\, p^2 = p^3(1-p).$$

סך הכל קיבלנו

$$\mathrm{Var}(Y) = (n-1)p^2(1-p^2) + 2(n-2)p^3(1-p) = p^2(1+2p-3p^2)n - p^2(1+4p-5p^2).$$

נשים לב שלמרות ש-Y_n אינם בלתי תלויים, השונות עדיין גדלה ליניארית ב-n ולכן גם כאן תקפה המסקנה של החוק החלש של המספרים הגדולה, כלומר, לכל $\epsilon > 0$ מתקיים

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\frac{Y}{N-1} - p^2\right| < \epsilon\right) = 1.$$

לסיכום: פירוק השונות השונות המשותפת מאפשרת לנו לחשב שונויות של סכום משתנים מקריים על ידי ניתוח ההתפלגויות המשותפות של זוגות המשתנים בלבד ומבלי להידרש לנתח את ההתפלגות כולה. מבחינה זו, כמו תוחלת סכום משתנים מקריים תלויים, גם שונות של סכום משתנים מקריים תלויים פשוטה יותר לחישוב מאשר התפלגותם המשותפת.

6.4* קשרים לאלגברה ליניארית

התכונות המתוארות בטענות 6.2, 6.3, 6.14 ו- 6.15 מראות על הקבלה מעניינת בין שונות לנורמה במרחב וקטורי ובין שונות משותפת למכפלה סקלרית. השונות היא אי-שלילית למשתנים לא קבועים, ובעלת מבנה של תבנית ריבועית, והשונות המשותפת היא בי-ליניארית. רמז נוסף בהבנת הקשר בין שני האובייקטים ניתן בטענה הבאה

טענה 6.18 (קושי-שוורץ הסתברותי). יהיו X, Y משתנים מקריים בעלי שונויות סופיות. אזי $\mathbb{E}(XY)$ קיימת ומקיימת

$$\mathbb{E}(XY) \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}$$

ושוויון מתקבל רק כאשר, בהסתברות 1, X שווה ל-Ya עבור $a \in \mathbb{R}$ כלשהו.

הוכחה. אם X או Y שווים ל-0 כמעט תמיד הטענה מיידית. אחרת נגדיר

$$\overline{X} = \frac{X}{\sqrt{\mathbb{E}(X^2)}} \qquad \overline{Y} = \frac{Y}{\sqrt{\mathbb{E}(Y^2)}}$$

ונשים לב שמתקיים

$$\mathbb{E}\left(\overline{X}^2\right) = \mathbb{E}\left(\overline{Y}^2\right) = 1.$$

כעת, לכל $x, y \in \mathbb{R}$ מתקיים

$$|xy| \leq \frac{x^2 + y^2}{2}$$

ולכן

$$|\overline{X}\,\overline{Y}| \leq \frac{\overline{X}^2 + \overline{Y}^2}{2}.$$

ממונוטוניות ולינאריות התוחלת נובע כי

$$\mathbb{E}\left(|\overline{X}\,\overline{Y}|\right) \leq \frac{\mathbb{E}\left(\overline{X}^2\right) + \mathbb{E}\left(\overline{Y}^2\right)}{2} = 1.$$

ולכן $\mathbb{E}\left(\overline{X}\,\overline{Y}\right)$ קיימת. נציב את הגדרת $\overline{X}$ ו-$\overline{Y}$ ונקבל

$$\mathbb{E}\left(\overline{X}\,\overline{Y}\right) = \frac{\mathbb{E}(XY)}{\sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}} \leq 1$$

כמבוקש. ■

חישוב זה, שהנו הסתברותי גרידא, הוא למעשה הכללה של אי-שוויון קושי-שוורץ מאלגברה ליניארית. נראה כיצד מסיקים שוויון זה מטענה 6.18.

מסקנה 6.19 (קושי-שוורץ). יהיו v, u וקטורים ב-$\mathbb{R}^n$. אזי

$$\langle u, v\rangle \leq \|v\|\|u\|$$

ושוויון מתקבל רק כאשר $u = av$ עבור $a \in \mathbb{R}$ כלשהו.

הוכחה. נגדיר Z מ"מ אחיד על $[N]$ ויהיו $X = u_Z$ ו- $Y = v_Z$. נחשב

$$\mathbb{E}(X^2) = \sum_{n\in[N]} \frac{(u_n)^2}{N} = \frac{\|u\|^2}{N}, \qquad \mathbb{E}(Y^2) = \sum_{n\in[N]} \frac{(v_n)^2}{N} = \frac{\|v\|^2}{N},$$

$$\mathbb{E}(XY) = \sum_{n\in[N]} \frac{v_n u_n}{N} = \frac{\langle u, v\rangle}{N}$$

וכעת לפי טענה 6.18 מתקיים

$$\frac{\langle u, v\rangle}{N} \leq \frac{||u|| \cdot ||u||}{N}$$

ושוויון מתקיים רק כאשר $u = av$ עבור $a \in \mathbb{R}$ כלשהו. מכאן המסקנה נובעת. ■

נציג מסקנה שימושית נוספת של א"ש קושי-שוורץ.

מסקנה 6.20 (קיום שונות משותפת). אם X ו-Y מ"מ בעלי שונות סופית אז $\mathrm{Cov}(X, Y)$ מוגדר היטב וסופי.

הוכחה. לפי טענה 6.18, $\mathbb{E}(XY)$ קיימת וסופית ולכן כך גם $\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. ■

כמו במקרה של מכפלה פנימית, גם השונות המשותפת מושפעת ליניארית ממתיחה של כל אחד מהמשתנים. במדע הסטטיסטיקה עלה צורך באומדן לעוצמת עוצמת המתאם בין שני מ"מ באופן שינטרל גורם זה. בבעיה 6.4 להלן נמחיש את משמעות טענה 6.18 ביחס למדד זה.

הגדרה 6.21 (מקדם המתאם של פירסון)**. מקדם המתאם (correlation coefficient)** של X ו-Y, משתנים מקריים בעלי שונות סופית. מוגדר כ-

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

תכונות השונות המשותפת משליכות על תכונות מקדם המתאם.

מסקנה 6.22 (תכונות מקדם המתאם). יהיו X, Y, Z מ"מ בעל שונות סופית, ויהיו $a, b \in \mathbb{R}$. אזי

(א) $\mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$.

(ב) $\mathrm{Corr}(aX, bZ) = \mathrm{sgn}(a)\,\mathrm{sgn}(b)\,\mathrm{Corr}(X, Z)$.

(ג) $\mathrm{Corr}(X, X) = 1$.

✎ **בעיה 6.3.** להוכיח את טענה 6.22.

✎ **בעיה 6.4.** להראות כי מקדם המתאם תחום בטווח $[-1, 1]$ וכי הוא מקבל את ערכי הקצה רק כאשר $X = aY$ בהסתברות 1. מהו מקדם המתאם של משתנים מקריים בלתי תלויים?

מקדם המתאם משמש אפוא מדד תחום לעוצמת הקשר הליניארי בין משתנים מקריים ויש דמיון רב בינו לבין קוסינוס הזווית בין ווקטורים כאומדן לתלות הליניארית ביניהם.

לאור הסתכלות גיאומטרית זו נוכל לפתוח את השונות המשותפת לפעולת הטלה באופן הבא,

טענה 6.23 (רגרסיה ליניארית). יהיו X, Y משתנים מקריים בעלי שונויות סופיות. אזי המשתנה המקרי $Z = X - \frac{\mathrm{Cov}(X,Y)Y}{\mathrm{Var}(Y)}$ בלתי מתואם עם Y. כמו כן, לכל $a \in \mathbb{R}$ מתקיים $\mathrm{Var}(X - aY) \geq \mathrm{Var}(Z)$.

הוכחה. החלק הראשון נובע מבי-ליניאריות השונות המשותפת (טענה 6.14):

$$\mathrm{Cov}(Z, Y) = \mathrm{Cov}\left(X - \frac{\mathrm{Cov}(X, Y)Y}{\mathrm{Var}(Y)}, Y\right) = \mathrm{Cov}(X, Y) - \frac{\mathrm{Cov}(X, Y)\,\mathrm{Cov}(Y, Y)}{\mathrm{Var}(Y)} = 0$$

עבור החלק השני נציב $b = a + \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}$ ונקבל

$$\begin{aligned}\mathrm{Var}(X + aY) &= \mathrm{Var}\left(X - \frac{\mathrm{Cov}(X,Y)Y}{\mathrm{Var}(Y)} + bY\right)\\ &= \mathrm{Var}(Z) + 2\,\mathrm{Cov}(Z, bY) + b^2\,\mathrm{Var}(Y) = \mathrm{Var}(Z) + b^2\,\mathrm{Var}(Y)\end{aligned}$$

ולכן ביטוי זה מינימלי עבור $b = 0$, כנדרש. ■

הערה: מנקודת מבט סטטיסטית הישר $\mathbb{E}(X) + \frac{\mathrm{Cov}(X,Y)Y}{\mathrm{Var}(Y)}$ הוא הקירוב הליניארי במונחי Y של X שממזער את סכום ריבועי הסטיות. באופן דומה ניתן לבטא את הקירוב הליניארי הטוב ביותר (במונחי סטייה ריבועית) של X על ידי Y ו-Z באמצעות

$$\mathbb{E}(X) + \frac{\mathrm{Cov}(X,Y)Y}{\mathrm{Var}(Y)} + \frac{\mathrm{Cov}(X,Z)Z}{\mathrm{Var}(Z)} - \frac{\mathrm{Cov}(Y,Z)Z}{\mathrm{Var}(Y)\,\mathrm{Var}(Z)}$$

אם נציב $Z = Y^2$, נקבל את הקירוב הפולינומי של X שימזער את הסטיה הריבועית. על ידי הכללה של רעיון זה לממדים גבוהים יכולים סטטיסטיקאים להפחית את השפעתו של משתנה אחד על התפלגותו של משתנה אחר ולזקק את השפעתם של הגורמים שאינם תלויים בו. ביטוי אחר לרעיון זה מתגלם ברעיון של הפרדת שונויות המוצג בתת-פרק הבא.

לסיום חלק זה נותיר לקורא את הצעד האחרון בקשירת השונות המשותפת למרחב מכפלה פנימית.

✎ **בעיה 6.5.** הראה כי אוסף המשתנים המקריים הבדידים בעלי תוחלת 0 ושונות סופית הוא מרחב ליניארי, וכי השונות המשותפת של שני משתנים במרחב זה היא פעולה חילופית, בי-ליניארית, חיובית לחלוטין. הסק, כי השונות המשותפת היא מכפלה פנימית במרחב זה. (תזכורת: פעולה f היא חיובית לחלוטין אם $f(x,x) \geq 0$ ושוויון מתקבל רק עבור $x = 0$)

6.5* שונות מותנית

ראשית נשים לב שהתוחלת המותנית בלתי-מתואמת עם השארית. תיאור זה מתאים לרעיון שהיא מהווה את הקירוב הטוב ביותר של X בהינתן Y (אילולא הייתה בלתי מתואמת יכולנו לשפר אותה על ידי רגרסיה ליניארית).

טענה 6.24. יהי X משתנה מקרי בעל תוחלת על $(\Omega, \mathcal{F}, \mathbb{P})$, ויהי Y משתנה מקרי אותו מרחב הסתברות. אז

$$\mathrm{Cov}(\mathbb{E}[X \mid Y], X - \mathbb{E}[X \mid Y]) = 0$$

זה בדיוק המשלים החלק שאינו תלוי

הוכחה. נשים לב כי $\mathbb{E}(X - \mathbb{E}[X \mid Y]) = \mathbb{E}(X) - \mathbb{E}(\mathbb{E}[X \mid Y]) = 0$ ולכן

$$\begin{aligned}
\mathrm{Cov}(\mathbb{E}[X \mid Y], X - \mathbb{E}[X \mid Y]) &= \mathbb{E}(\mathbb{E}[X \mid Y](X - \mathbb{E}[X \mid Y])) \\
&\overset{\text{תכונה ההרכבה}}{=} \mathbb{E}(\mathbb{E}[X \mid Y](X - \mathbb{E}[X \mid Y]) \mid Y]) \\
&\overset{\text{הוצאת החלק הידוע}}{=} \mathbb{E}(\mathbb{E}[X \mid Y]\mathbb{E}[X - \mathbb{E}[X \mid Y]) \mid Y]) \\
&\overset{\text{ליניאריות}}{=} \mathbb{E}(\mathbb{E}[X \mid Y](\mathbb{E}[X \mid Y] - \mathbb{E}[X \mid Y])) = \mathbb{E}(0) = 0
\end{aligned}$$

■

לאור תכונה זו ניתן לתאר את התוחלת המותנית במונחים של מזעור סטיה ריבועית ממוצעת.

✎ **בעיה 6.6.** להוכיח כי לכל $f : \mathbb{R} \to \mathbb{R}$ מתקיים $\mathbb{E}\left((X - f(Y))^2\right) \le \mathbb{E}\left(X - \mathbb{E}(X \mid Y))^2\right)$.

כשם שהגדרנו בפרק 5.4 את התוחלת המותנית $E[Y \mid X]$ כפונקציה על המשתנה המקרי X המקבלת לכל $x \in \mathbb{R}$ את $\mathbb{E}(Y \mid X = x)$ - כן נוכל לתאר את **השונות המותנית** כמשתנה מקרי המקבל לכל $x \in \mathbb{R}$ את שונותו של Y בהינתן $X = x$.

הגדרה 6.25 (תוחלת מותנית)**.** יהיו X, Y משתנים מקריים המוגדרים על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$ כך ש-X בעל שונות סופית. נרשום

$$f(a) := \begin{cases} \mathrm{Var}(Y \mid X = a) & \mathbb{P}(X = a) > 0 \\ 0 & \mathbb{P}(X = a) = 0 \end{cases}.$$

השונות המותנית (conditional variance) של Y בהינתן X, שנסמנה $\mathbb{E}(Y \mid X)$ היא המשתנה המקרי Z המוגדר על ידי

$$Z = \mathrm{Var}[Y \mid X] := f(X),$$

בכל מקרה בו $f(a)$ מוגדרת לכל a בתומך של X.

נשים לב כי השונות המותנית מקיימת נוסחא דומה לנוסחת השונות.

אבחנה 6.26 (שונות מותנית)**.** אם למשתנה מקרי X קיימת שונות מותנית, אז מתקיים

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

טענה 6.27. יהי X משתנה מקרי בעל שונות סופית על $(\Omega, \mathcal{F}, \mathbb{P})$ ויהיו Y משתנה מקרי על אותו מרחב הסתברות. אזי

$$\mathbb{E}(\mathrm{Var}[X \mid Y]) = \mathrm{Var}(X - \mathbb{E}[X \mid Y]).$$

הוכחה. ראשית ניווכח כי

$$\mathbb{E}(X\mathbb{E}[X\,|Y]) \overset{\text{תכונה ההרכבה}}{=} \mathbb{E}(\mathbb{E}[X\mathbb{E}[X\,|Y]\,|\,Y]) \overset{\text{הוצאת החלק הידוע}}{=} \mathbb{E}(\mathbb{E}[X\,|\,Y]^2)$$

ולכן נוכל לוודא כי

$$\begin{aligned}
\mathrm{Var}(X-\mathbb{E}[X\,|Y]) &= \mathbb{E}\left((\mathbb{E}[X\,|Y]-X)^2\right)\\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}[X\,|Y]) + \mathbb{E}(\mathbb{E}[X\,|Y]^2)\\
&= \mathbb{E}(X^2) - \mathbb{E}(\mathbb{E}[X\,|\,Y]^2)\\
&\overset{\text{תכונה ההרכבה}}{=} \mathbb{E}(\mathbb{E}[X^2\,|\,Y]) - \mathbb{E}(\mathbb{E}[X\,|\,Y]^2)\\
&= \mathbb{E}(\mathrm{Var}[X\,|Y]),
\end{aligned}$$

זה בדיוק השונות החלק שאינו תלוי

כאשר במעבר האחרון השתמשנו בכך ש-$\mathbb{E}(X-E[X\,|Y]) = \mathbb{E}(X) - \mathbb{E}(E[X\,|Y]) = 0$. ■

כוחה של השונות המותנית בכך שהיא מאפשר לנו לפרק את השונות של X לתרומה לשונות של Y ולתרומה לשונות של X בהינתן Y.

טענה 6.28 (ניתוח שונויות). יהי X מ"מ בעל שונות סופית על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ ויהי Y משתנה מקרי על אותו מרחב הסתברות. אז $\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}[X\,|\,Y]) + \mathrm{Var}(\mathbb{E}[X\,|\,Y])$.

$\overset{\shortparallel}{\mathrm{Var}(X-E[X|Y])}$

הוכחה. נחשב:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{Var}(X-\mathbb{E}[X\,|\,Y] + \mathbb{E}[X\,|\,Y]) =\\
&= \mathrm{Var}(X-\mathbb{E}[X\,|\,Y]) + \mathrm{Var}(\mathbb{E}[X\,|\,Y]) + 2\,\mathrm{Cov}(X-\mathbb{E}[X\,|\,Y]), \mathbb{E}[X\,|\,Y])\\
&\overset{\text{טענה 6.24}}{=} \mathrm{Var}(X-\mathbb{E}[X\,|\,Y]) + \mathrm{Var}(\mathbb{E}[X\,|\,Y]) \overset{\text{טענה 6.27}}{=} \mathbb{E}(\mathrm{Var}[X\,|\,\mathcal{G}]) + \mathrm{Var}(\mathbb{E}[X\,|\,Y]).
\end{aligned}$$

■

הערה: לטענה 6.28 חשיבות מכרעת בסטטיסטיקה, שכן היא מספרת לנו שאת הסטיה הריבועית של נעלם כלשהו נוכל לפרק לסטיה הנובעת מההבדל בערכים של משתנה תלוי Y, ומההבדל הממוצע בין ערכי X לכל ערך Y. אם $\mathbb{E}(\mathrm{Var}[X\,|\,Y])$ גדול יותר - נאמר כי באופן טיפוסי Y מסביר את מרבית הפיזור של X, ואילו אם $\mathrm{Var}(\mathbb{E}[X\,|\,Y])$ גדול יותר - נאמר שבאופן טיפוסי Y אינו מסביר את הפיזור של X. ניתוח כזה, מכונה ניתוח שונויות (ANOVA בלע"ז) והוא משמש, למשל, בכדי לקבוע האם כדאי לפלח את מרחב המדגם לפי ערכי Y כאשר מבצעים ניתוח סטטיסטי.

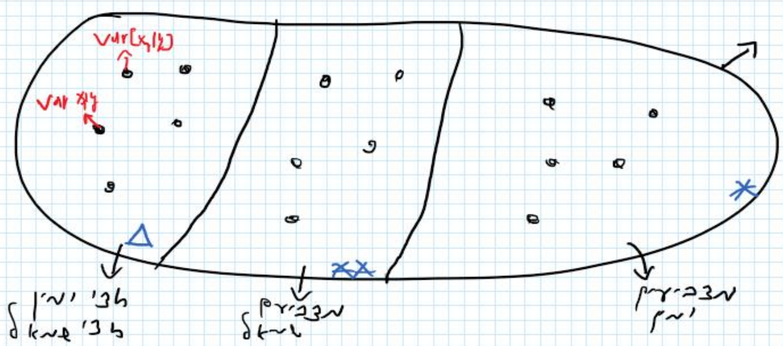
6.6* החוק החלש של המספרים הגדולים

לאמיתו של דבר הדרישה לקיומה של שונות במשפט 6.8 לא הייתה נחוצה כלל ועיקר. נציג את ההכללה הבאה.

אינטואיציה למשוואה $Var(x)=E(Var[X|Y])+Var(E[X|Y])$

נקח את X המשתנה המקרי שלי ונראה איך מתנהג X|Y

משתנה מקרי X



• משתנים מקריים X|Y
נשים לב כי ל X יש תוחלת
$E[X]$ ובנוסף עבור כל X|Y
מה הסטיה שלי מהתוחלת
$var(X|Y)$

דוגמא: כנסת ויש בחירות במחוזים שונים

ברור שבכל אזור אם התוחלת להצביע היא חצי חצי הסטייה שלי מהתוחלת היא רצינית ב(*)
וב(**) אבל ב(Δ) אין סטייה כלומר אם אני רוצה למצוא את הסטייה של X מהתוחלת אז עלי
לכאורה לבדוק מה הסטייה בכל אזור וזהו $\boxed{\text{אבל}}$ אם אני מחלק לגמרי $Var(X)=E[Var(X|Y)]+(\ )$

[(*) + (**) + Δ]

חסר לי כאן משהו

מה בעצם חסר לי בבדיקת את הסטייה של התוחלת $E[X|Y]$ מהתוחלת המקורית $E[X]$ ולכן
מתקיים בעצם שוויון של ממש

משפט 6.29 (החוק החלש של המספרים הגדולים). יהי X משתנה מקרי בעל תוחלת סופית ויהי $\epsilon > 0$. לכל $N \in \mathbb{N}$, נרשום

$$p_n = \mathbb{P}\left(\left|\frac{1}{N}\sum_{n=1}^{N} X_n - \mathbb{E}(X)\right| < \epsilon\right)$$

כאשר $\{X_n\}_{n\in[N]}$ משתנים מקריים בלתי תלויים שווי התפלגות ל-X. אזי

$$\lim_{N\to\infty} p_n = 1.$$

לצורך הוכחת המשפט נעזר בטענה הבאה.

טענה 6.30. אם למשתנה מקרי בדיד X תוחלת סופית, אזי $\lim_{K\to\infty} \mathbb{E}\left(|X|\mathbb{1}_{|X|\geq K}\right) = 0$.

הוכחה. לפי הגדרת התוחלת, הטור $\sum_{s\in\mathbb{R}} s\,\mathbb{P}(X = s)$ מתכנס בהחלט, כלומר הטור $\sum_{s\in\mathbb{R}} |s|\,\mathbb{P}(X = s)$ מתכנס לערך סופי שנסמנו a. לכן, לכל $\epsilon > 0$ קיימת קבוצה סופית $S \subset \mathbb{R}$ כך שמתקיים

$$\sum_{s\in S} |s|\,\mathbb{P}(X = s) > a - \epsilon.$$

לכל $K > max\{|s| \ : \ s \in S\}$ נקבל

$$\mathbb{E}\left(|X|\mathbb{1}_{|X|\geq K}\right) = \mathbb{E}(|X|) - \mathbb{E}\left(|X|\mathbb{1}_{|X|<K}\right) = a - \sum_{|s|<K} |s|\,\mathbb{P}(X = s) \leq a - \sum_{s\in S} |s|\,\mathbb{P}(X = s) \leq \epsilon$$

כנדרש. ■

הוכחת משפט 6.29. עבור K נגדיר $a_K = \mathbb{E}\left(|X|\mathbb{1}_{|X|\geq K}\right)$. לפי טענה 6.30, עבור K גדול מספיק מתקיים $a_K \leq \frac{\epsilon}{3}$. כעת, נגדיר $Y_n = X_n\mathbb{1}_{|X_n|<K}$ ו-$Z_n = X_n\mathbb{1}_{|X_n|\geq K}$, כך שמתקיים $X_n = Y_n + Z_n$. נגדיר

$$R_N = \frac{1}{N}\sum_{n=1}^{N} X_n \qquad S_N = \frac{1}{N}\sum_{n=1}^{N} Y_n \qquad T_N = \frac{1}{N}\sum_{n=1}^{N} Z_n$$

כך שמתקיים $R_N = S_N + T_N$. את נכונותו של החוק החלש של המספרים הגדולים עבור S_N לכל K מסויים, הראינו כבר במשפט 6.8, נבקש אפוא לחסום את הסטיה שעלולה להיגרם כתוצאה מ-T_N. נחשב

$$|\mathbb{E}(Z_1)| \leq \mathbb{E}(|Z_1|) = a_K \leq \frac{\epsilon}{3}.$$

כמו כן

$$\mathbb{E}(|T_N|) = \frac{1}{N}\mathbb{E}\left(\left|\sum_{n=1}^{N} Z_n\right|\right) \leq \frac{1}{N}\mathbb{E}\left(\sum_{n=1}^{N} |Z_n|\right) = \mathbb{E}(|Z_1|) = a_K$$

לבסוף נוכל לחסום את ההסתברות המבוקשת באופן הבא

$$\begin{aligned}\mathbb{P}(|R_N - \mathbb{E}(X)| \geq \epsilon) &= \mathbb{P}(|S_N - \mathbb{E}(Y_1) + T_N - \mathbb{E}(Z_1)| \geq \epsilon) \\ &\leq \mathbb{P}(|S_N - \mathbb{E}(Y_1)| + |T_N| + |\mathbb{E}(Z_1)| \geq \epsilon) \\ &\leq \mathbb{P}(|S_N - \mathbb{E}(Y_1)| + |T_N| \geq 2\epsilon/3) \\ &\leq \mathbb{P}(|S_N - \mathbb{E}(Y_1)| \geq \epsilon/3) + \mathbb{P}(|T_N| \geq \epsilon/3) \\ &\overset{\text{מרקוב}}{\leq} \mathbb{P}(|S_N - \mathbb{E}(Y_1)| \geq \epsilon/3) + \frac{a_K}{\epsilon/3}.\end{aligned}$$

לפי משפט 6.8 המחובר הראשון שואף לאפס כאשר N שואף לאינסוף ולכן

$$\lim_{N\to\infty} \mathbb{P}(|R_N - \mathbb{E}(X)| \geq \epsilon) \leq \frac{a_K}{\epsilon/3}.$$

נשאיף את K לאינסוף ולפי טענה 6.30 נקבל כי

$$\lim_{N\to\infty} \mathbb{P}(|R_N - \mathbb{E}(X)| \geq \epsilon) = 0,$$

כנדרש.

■

בעיות הרחבה והעשרה

✎ **בעיה 6.7** (התפלגות בינומית שלילית). מטילים מטבע הוגן עד אשר שמתקבלת תוצאה של עץ k פעמים. יש למצא את תוחלת ושונות מספר ההטלות, ומצא טווח שבו מספר ההטלות יימצא בסבירות של 90%.

✎ **בעיה 6.8.** בגן טרום-חובה n ילדים. כל ילד חבר של כל ילד אחר בהסתברות p_n באופן בלתי תלוי. בהסתמך על אי-שוויון צ'בישב יש להראות שאם $\lim_{n\to\infty} \frac{p_n}{n} = \infty$ אז לכל $\epsilon > 0$ קיים m כך שההסתברות שיש בגן שלושה ילדים שכל שניים מביניהם חברים גדולה מ-$1 - \epsilon$.

✎ **בעיה 6.9** (התפלגות היפר-גיאומטרית). נתון שבהטלת n מטבעות התקבלו m תוצאות של עץ. יש לחשב את שונות מספר העצים בk ההטלות הראשונות.

✎ **בעיה 6.10.** חצי מחיות המחמד בתל-אביב הן כלבים, שליש הן חתולים ושישית הן חמוסים. נאספו n חיות מקריות. מה תוחלת ושונות ההפרש בין מספר החתולים למספר החמוסים?

✎ **בעיה 6.11.** יש להראות כי אוסף המשתנים המקריים הבדידים בעלי שונות סופית הוא מרחב ליניארי, וכי הפעולה $g(x,y) = \mathbb{E}(XY)$ במרחב זה היא פעולה חילופית, בי-ליניארית, חיובית לחלוטין. הסק, כי פעולה זו מגדרה מכפלה פנימית במרחב זה. (מרחב זה שימושי הרבה פחות מאשר המרחב שהוצג בבעיה 6.5)

✎ **בעיה 6.12** (פרדוקס ברטרנד). ויינוגרד ובווין הם שני חובבי יין, אלא שוינוגרד שותה רק יין לבן ובווין שותה רק יין אדום. כיום בקבוק יין עולה 100 שקלים בלי קשר לצבעו, אך ידוע שבעוד עשר שנים אחד מסוגי היין יאבד מחצית מערכו והשני יכפיל את ערכו. כל אחד מחובבי היין החליט להקצות 1000 שקלים לרכישת בקבוקים ובסוף התקופה למכור אותם במחיר השוק ולרכוש את הבקבוקים האהובים עליו (אותם ישתה להנאתו). כיצד עליהם להשתמש בכספם אם ברצונם שתוחלת כמות הבקבוקים שיוכלו לרכוש תהיה מירבית?

✎ **בעיה 6.13** (*). לתוך שתי מעטפות אטומות מוכנסים סכומי כסף X ו-$10X$ עבור $X \sim \mathrm{Geo}(1/2)$. אהוד מקבל מעטפה מקרית, מציץ בתוכה ומחליט האם לקחת אותה או להחליפה במעטפה השניה.

(א) יש להראות כי בלי תלות בסכום, בהינתן הסכום שבמעטפה, תוחלת הכסף שיקבל אהוד תגדל אם יחליט להחליף בין המעטפות.

(ב) מסעיף (א') ניתן להסיק כי עוד לפני שנפתחה המעטפה כדאי לאהוד להחליף בין המעטפות, אך מטעמי סימטריה הדבר כמובן לא יתכן. כיצד נפתור את הפרדוקס?