יסודות תורת ההסתברות

אוהד נ. פלדהיים
אורי גוראל גורביץ'

מכון איינשטיין למתמטיקה

האוניברסיטה העברית בירושלים

27 בינואר 2019

מבוא

"הסתברות היא מורת הדרך של החיים עצמם."
– מרכוס טוליוס קיקרו, על הטבע, 5, 12.

ספר זה מגיש לקורא העברי מבוא מקיף לתורת ההסתברות ברמה העיונית הגבוהה הנחוצה לתלמידי המדעים הטהורים (מתמטיקה, מדעי-המחשב ופיסיקה עיונית), המבקשים להגיע למומחיות הנחוצה לצורך עריכת מחקר עיוני. הוא מיועד לתלמיד אשר התוודע ליסודות החשבון האינפיניטיסימלי והקומבינטוריקה אך טרם למד את תורת המידה. מטרתו להעניק לתלמיד כזה היכרות עם המושגים, התוצאות ודרך החשיבה המאפיינים את תורת ההסתברות – כל זאת ברמת הדיוק הגבוהה ביותר.

בראש ובראשונה נועד הספר ללוות קורס מבוא להסתברות לתלמידי תואר בוגר במסגרת לימודים אקדמיים. עם זאת, בתכנון הספר הובאה בחשבון גם למידה עצמאית בלי כל הנחיה חיצונית. לצורך כך צרפנו פרק מקדים הסוקר חומרי רקע ותוצאות מתחומים אחרים כמו החשבון האינפיניטיסימלי והקומבינטוריקה אשר תהיינה נחוצות להבנת החומר. יתר הספר מחולק לשלושה חלקים. בחלק הראשון נלמדת תורת ההסתברות הבדידה בצורה מלאה ומדוייקת, בחלק השני מוכללות התוצאות למשתנים מקריים רציפים בהחלט, תוך הסתמכות על מספר מצומצם של טענות מתקדמות שניתנות בלא הוכחה. בפרק השלישי מוצגים נושאים מתקדמים נבחרים ויישומים מורכבים של התיאוריה אשר יכולים לשמש כפרק בחירה או כדוגמא מסכמת. כל אחד מהפרקים סוקר נושא מרכזי בתורת ההסתברות, מציג את המושגים והמשפטים החשובים ביותר ומלווה אותם במספר דוגמאות. לאורך הפרק מוצגים תרגילים בסיסיים שמטרתם לוודא הבנה של החומר הנלמד. נושאים מתקדמים יותר, אשר אינם בהכרח חלק מקורס יסודי בתורת ההסתברות מופיעים בחלקים האחרונים של כל אחד מהפרקים והם מסומנים ב-$(*)$, בקורס ברמה נמוכה יותר ניתן לדלג על פרקים אלו. כל פרק מסתיים באוסף של תרגילים מתוחכמים או עמוקים יותר, אשר נועדו לפתח את המיומנות המתמטית, לקשור את הנושא לנושאים אחרים במתמטיקה או במדעים, או לרמז על היבטים מתקדמים יותר של תורת ההסתברות שאינם מכוסים על ידי ספר זה. תרגילים קשים במיוחד סומנו ב-$(*)$ ובמקרים חריגים ב-$(**)$.

היה זה הרצון לשמור על דיוק שהביאנו להימנע מעיסוק ישיר במרחבי הסתברות רציפים וכאלה שאינם בני-מניה. זאת משום שהרחבת תורת ההסתברות למרחבים אלה באופן מדויק נדרשת לתוצאות מתורת המידה אשר אינן נכללות בחומר הרקע שעליו רצינו להתבסס. תחת זאת, ניסינו להמחיש את מרבית התופעות והכלים ההסתברותיים במרחב הבדיד באופן שיקל את הרחבתם למרחבים רציפים.

כדי לשפר את הבנתו של תחום הידע ההסתברותי ושל צורת החשיבה המאפיינת אותו ולעניין את הקורא בעיסוק ההסתברותי ובתולדותיו מלווה הספר בהערות היסטוריות ובאנקדוטות. בנוסף השתדלנו לצרף לחלקים

הקשים יותר הסברים מילוליים מפורטים שיקלו על הקורא להסתגל לצורת החשיבה.

בכתיבת הספר נסמכנו על חומר הקורס תורת ההסתברות, כפי שנלמד באוניברסיטה העברית בשנים 2015-2018, וברצוננו להודות לפרופסור רז קופרמן שערך ועיבד חומר זה. כמו כן סייעו בידנו ..., בקוראם את הספר ובהערותיהם ואנו אסירי תודה להם.

תוכן העניינים

מושג ההסתברות

"מופלא הדבר, שמדע שהחל בעיון במשחקי מזל הפך למושא החשוב ביותר של הידע האנושי."
– פְּיֵיר סִימוֹן לָפְּלַס, תיאוריה אנליטית של ההסתברות, 1812.

תורת ההסתברות עוסקת בחקר תחזיות והערכות הנוגעות ל**בלתי-נודע**. נוכחותו הרבה של הבלתי-נודע בחיינו הופכת את התורה ליישומית בהקשרים רבים. היא משמשת אותנו לניבוי התרחשויות עתידיות באומדנו את הסתברות הצלחתו של ניתוח, את ההסתברות שירד גשם ביום המחרת ואת ההסתברות שמועמד זה או אחר יזכה בבחירות הכלליות. היא משמשת אותנו בנתחנו את העבר שעה שאנו מעריכים את סבירותה של השערה בנוגע לשכיחות כדי חרס ממין מסויים ברחבי האמפריה הרומיות, או בנוגע לדרך שבה התפזרו מחצבים על כוכב הלכת שלנו בתהליך היווצרותו. והיא אף משמשת אותנו לגבש הערכות בנוגע להווה, כאשר אנו אומדים את הסבירות שאחד ממכרינו נמצא בביתו כעת, או את הסבירות שארגז שנתקלנו בו ברחוב הונח שם בכוונה. נוסף על כך, אנו משתמשים בהסתברות בתכנון מכשירים ותהליכים - ובכלל זה פרוטוקולים לבדיקות רפואיות, כורים גרעיניים ומשחקי מזל.

אוצר המונחים של עולם ההסתברות, מורכב ממילים כמו "סבירות", "מקריות" (ואחותה הארמית "אקראיות"). מקורן של מילים אלו בשפות העולם השונות משורשים המבטאים ניחוש, פליאה, נדירות, יכולת לשאת חן, או נכונות להאמין. בכך מתבטא הלך רוחנו כאשר אנו נחשפים למידע חדש (ואומנם תורת ההסתברות קשורה קשר הדוק לתורת המידע). כאשר המידע החדש בו אנו נתקלים שונה ממרבית חוויותינו עד כה - אנו מגיבים בהפתעה ובפליאה ואילו כאשר הוא תואם אותן - אנו חווים שביעות רצון מכך שהמציאות מגשימה את ציפיותינו. ניכר כאן גם קשר בין שכיחותה של תופעה לבין ההסתברות שאנו מייחסים לה, זאת מתוך ציפיה שהעתיד יהיה דומה במידה רבה לעבר. תכונה זו, המכונה בעולם המדע "הומוגניות", הנה חיונית לצורך יישום תורת ההסתברות. לא ניתן ליחס הסתברות למאורע יחידני שאין דומה לו.

ביסוד תורת ההסתברות עצמה עומדת מטאפורה המשווה את תוצאתה של כל מדידה שתוצאתה אינה ידועה – לשליפת פתק מתוך שק. השק מלא במספר רב של פתקים זהים בצורתם ועל כל אחד מהם כתובה תוצאה אפשרית של הבדיקה. יתכן כי על פתקים שונים כתובה אותה התוצאה וכי תוצאות שונות נבדלות אלו מאלו בשכיחותן. הסתברותה של תוצאה מסויימת מייצגת את אחוז הפתקים שעליהם היא כתובה. על כן הסתברותה של תוצאה עומדת ביחס ישר לאיכות הניבוי (או הניחוש) שזו התוצאה שתתקבל וביחס הפוך לפליאה שנחוש שעה שנחזה בתופעה אם תתממש. תוצאות בלתי-סבירות (כלומר בעלות הסתברות נמוכה) תחשבנה בעינינו לפלא בלתי-מצוי ואילו תוצאות סבירות (כלומר בעלות הסתברות גבוהה) - לעניין שבשגרה. על תורת ההסתברות מוטלת המשימה לתאר את הרכב השק המתאים לתיאור תוצאותו המקרית של ניסוי אידיאלי.

תורת הסטטיסטיקה, אחותה, מסייעת בבחירת הניסוי האידיאלי המתאים להתרחשות מסויימת במציאות.

הכלתה של השקפה הסתברותית זו כמודל אוניברסלי לתיאור הבלתי-נודע והשתתתה על עקרונות מתמטיים הייתה תהליך איטי והדרגתי. ראשיתה של ההסתברות, ככל הנראה, בפיתוח אסטרטגיות לניצחון במשחקי מזל. הללו, עתיקים כמו ההסטוריה עצמה, היו רווחים לפחות למן המאה השלושים וחמש לפני-הספירה במצרים הפרעונית של השושלת הראשונה[1]. ניתוחים אינטואיטיביים והמלצות לאסטרטגיה מנצחת נמצאו על פפירוסים המתארים את משחק המזל המצרי סנת, אך בתוך אלו שזורה ספקנות לגבי שליטת המזל בתוצאת המשחק ואמונה שידם של כוחות על-טבעיים מכתיבה בסופו של דבר את זהות המנצח. ספקנות זו בנוגע לתוקפן של טענות סטטיסטיות והסתברויות הייתה רווחת גם בקרב המלומדים עד לימי-הביניים המאוחרים. אותה ספקנות הייתה שרירה פחות כאשר ניתחו הקדמונים סברות הנוגעות להווה ולעבר, ואכן מושגים כמו סבירותן של טענות בבית דין והאפשרות שנפלה בהן טעות מופיעה כבר בכתבים בבליים מהמאה השמינית לפני הספירה ואף הללו עדיין לא היו מושתתים על היקש פורמלי ועל חישובים מתמטיים, אלא על השכל הישר. טענות הסתברותיות עורבו במידת מה גם בדיונים דתיים כאשר השוו הקדמונים את עוצמתם של ניסים שונים על פי שכיחותם.

עם התרופפות אחיזתה של הכנסייה בחיים האינטלקטואליים באירופה של ראשית הרנסנס, פנו לראשונה מלומדים לבחון את סוגיית המשחקי המזל מנקודת מבט פורמלית. בשנת 1494 הנזיר לוּקָה פָּאצִ'וֹלִי הציג את הבעיה הבאה שנודעה בשם בעיית המשחק שנקטע.

סוגיה 1. [המשחק שנקטע] נשוה בנפשינו משחק בין שני שחקנים, אביה ובתיה, המהמרים על סכום כסף שווה. שניהם מטילים שוב ושוב מטבע הוגן; בכל פעם שיוצאת בהטלה תוצאה של **עץ**, מקבל אביה נקודה, ובכל תוצאה של **פלי** מקבלת בתיה נקודה. הזוכה הוא הראשון שיזכה בשש הטלות והוא יקבל את סך ההימור של שניהם. כעת, נניח שאביה מוביל חמש כנגד שלוש כאשר היום מחשיך והמשחק נקטע באיבו. כיצד עליהם לחלק את הקופה?

עצם הצגת הבעיה, מעידה על התקדמות כבירה. ההכרה (המובלעת באותה העת) באידיאל המופשט של מטבע הוגן והעיון במצב שבו חלק מן המידע בנוגע לתוצאת המשחק נחשף לא הוצגו בבהירות כזו קודם לכן.

מספר מתמטיקאים רבי-חשיבות (כדוגמת גְ'רוֹנִימוֹ קַרְדָאנוֹ וְנִיקוֹלוֹ טַרְטָלְיָה) ניסו לפתור את הבעיה במשך כמאתיים שנה, וכולם הציגו פתרונות שגויים. הטעות שחזרה על עצמה הייתה עיון מדוקדק בהשתלשלות מאורעות העבר שבאה במקום התמקדות בעתידים האפשריים (אשר נדמו בעיניהם כבלתי ניתנים לחיזוי). עוד מעניין שאיש מבין כל אלא לא טרח לבצע את הניסוי המתבקש ולאמוד את סיכויי הניצחון של שני המתמודדים. זאת ללמדנו כי תורת הסטטיסטיקה אחרה להגיח לאוויר העולם אפילו יותר מאחותה, תורת ההסתברות.

נקודת המפנה בהתמודדות עם בעיית המשחק שנקטע נולדה במהלך חליפת מכתבים ממושכת בשנת 1654 בין בְּלֶז פַּסְקָל (1623-1662) לפְּיֶיר דֶה-פֶּרְמָה (1601-1665). במסגרתה ניתחו השניים את המשחק ומצאו לבסוף את הפתרון הנכון - עליהם לחלק את הקופה ביחס של שבע לאחד (ראה דוגמא 5.9 להלן). חשוב מכך - אותה חליפת מכתבים הניחה את יסודותיה של תורת ההסתברות. קשה להפריז בחשיבות התפקיד שהייתה עתידה תורה זו לגלם בחיינו – כיום היא עומדת ביסוד כל ניהול סיכונים, ובפרט מהווה נדבך יסודי בבנקאות, בהנדסה ובמדעי החברה. עבודותיהם של פסקל ופרמה נאלצו להמתין כמחצית-המאה, בטרם יַעֲקֹב בֶּרְנוּלִי, טבע ב-1713

[1] Sebbane, Michael. ”Board games from Canaan in the Early and Intermediate Bronze Ages and the origin of the Egyptian Senet game.” Journal of the Institute of Archaeology of Tel Aviv University 28.2 ,(2001) .213-230

את המונח **הסתברות**. הוא גם טבע את המונחים העיקריים של תורת ההסתברות ובעקבותיו הרחיבו את התורה מתמטיקאים כמו לְאוֹנַרְד אוֹילֶר, קָרְל פְרִדְרִיךְ גָאוּס ופְּיֵיר סִימוֹן לַפְּלַס. הם גם היו הראשונים להבחין ברבים מישומייה האפשריים של התורה אשר התאגדו במדע שלימים נודע בשם "אקטואריה".

הקפיצה הבאה בתורת ההסתברות מיוחסת לאַנְדְרֵיי נִיקוֹלָאבִיץ' קוֹלְמוֹגוֹרוֹב אשר ייסד את **תורת ההסתברות המודרנית** על כתפיה של תורת המידה. ביחד עם אַנְדְרֵיי מַרְקוֹב, פָּאפְנוּטִי צֶ'בִישֶׁב, אֶמִיל בּוֹרֶל וְוִילְיָאם פֶּלֶר, הורחבה התורה למרחבים חדשים ואפשרה התמודדות עם שאלות שעד אותה העת לא ניתן היה אפילו לנסחן. כרך זה יסקור בעיקר את תורת ההסתברות של המאה התשע עשרה ורק לקראת סופו נחשף למקצת הבעיות והפתרונות שהציעה המאה העשרים.

"גם המזל, אשר נדמה כמסתער רפוי מושכות, נשלט ומרוסן בידי חוקים."

– אָנְקִיוּס מָאנְלִיוּס סֶוֶרִינוּס בּוֹאֶתְיוּס, נחמת הפילוסופיה, 523 לסה"נ.

הסתברות כתורה מתמטית. כיתר התורות המתמטיות – מהווה תורת ההסתברות מערכת של יחסים בין אובייקטים מופשטים המצייתת לכללי ההגיון הפורמלי. אלא שתורת הסתברות, בדומה לגיאומטריה הקלאסית, משמרת קרבה רבה בין השפה המשמשת את יישומיה לתיאור וניתוח מודלים מוחשיים ובין השפה הפורמלית המשמשת להגדרתה. קרבה זו אינה רק שריד הסטורי שנשתרש בשפה. גם כיום שימושי ההסתברות משמשים כמצפן המכוון את פיתוח התורה. המה מנחים אותנו בבחירת הכלים שנבחר לפתח והתופעות אחריהן נבקש להתחקות. ביטוי עז לקרבה זו בין הסתברות לשימושיה תואר בשנת 1900 על ידי המתמטיקאי הנודע דֵיוִיד הִילְבֶּרְט כאשר תאר בקונגרס המתמטי הבין-לאומי בסורבון את השישית מבין "עשרים ושלוש הבעיות של הילברט", אשר התוו את דרכה של המתמטיקה של המאה העשרים. זאת עוד לפני שתורת ההסתברות זכתה לביטוי פורמלי בהיר ומדויק. בעיה זו נוסחה בזו הלשון:

> **הבעיה השישית של הילברט:** חקר יסודות הגיאומטריה מעלה את הבעיה הבאה: לטפל בצורה דומה, באמצעות אקסיומות, במדעי הטבע אשר בהם כבר היום מתמטיקה משחקת תפקיד חשוב; ובראש ובראשונה - תורת ההסתברות והמכאניקה.

התקדמות אדירה לקראת מטרה זו של התחקות אחר האקסיומות העומדות ביסודה של תורת ההסתברות בהקשרה של המכאניקה הסטטיסטית, אותו מדע המתאר את הקשר בין חוקי הטבע היסודיים לבין ביטוים בהתנהגות צבירי המולקולות של חומר מעובה, התחוללה לאורך המאה העשרים, והקורא אשר יוסף ללמוד את תורת ההסתברות לאחר שיקנה שליטה בחומר המופיע בכרך זה, יווכח עד כמה התקדמות זו עיצבה את תורת ההסתברות המודרנית.

השקפה פילוסופית על הסתברות. ישנן לפחות שלוש צורות שונות לפרש את תורת ההסתברות מנקודת מבט פילוסופית. מכל אחת מצורות אלו נגזרים מושגים ושימושים שונים בהם ניתקל לאורך הספר. ההשקפה הראשונה מכונה הסתברות **הכרתית**, כזו העוסקת בהסתברות ככלי למדידת ודאות של אמונות. בהשקפה זו אנו נוקטים כאשר אנו מדברים על ה**סבירות** שסברה היא נכונה. השניה היא הסתברות **שכיחותנית**, כזו הרואה הסתברות כמבטאת **שכיחות** מוחשית של התרחשויות, ביטוי של ההומוגניות של היקום בזמן ובמרחב. תפיסה זו נשענת על האפשרות לחזור בקירוב ניכר על אותו ניסוי מספר רב של פעמים ועל הנחה של סימטריה בין חזרות

שונות. כך למשל, נוכל לדבר על ההסתברות שתינוק בלידתו ישקול למעלה משלושה קילוגרמים, שכן נתקלנו בלידות רבות ואנו סוברים שאין הבדל מהותי בין התפלגות משקלו ילוד אחד למשנהו. שלישית היא הסתברות **מהותית**, כזו שרואה בחוסר הוודאות תכונה יסודית של המציאות והיא רווחת בפרשנויות שונות של מכאניקת הקוואנטים. תפיסה כזו רואה באי-וודאות תכונה מובנה של המציאות, ובמידע - כמות פיסיקלית אובייקטיבית המצייתת לכללים מתמטיים.

מטבע הדברים אנו נתמקד בהיבטיה המתמטיים של תורת ההסתברות. לצורך כך נניח אקסיומות ונסיק מסקנות אשר תהיינה בעלות ערך לכל נקודות המבט הפילוסופיות על הסתברות באשר הן. עם זאת קרבה יתירה זו בין היישום והתיאוריה תגלם עצמה במושגים גשמיים שיופיעו לאורך הספר כמו **הטלת מטבע הוגן** ו**הטלת כדורים לתאים** באופן מקרי, כמו גם בשאלות שתתוארנה בשפת המודל ולא תמיד בשפה פורמלית. במידה רבה תדמנה השאלות בהן ניתקל לאותה שאלה על המשחק שנקטע ששאל הנזיר לוּקָה פָּאצ׳וֹלִי לפני למעלה מחמש מאות שנה. גם בשעה שיתקל הקורא בדוגמאות מוחשיות כאלו, עליו תמיד לתת דעתו שמאחורי כל דוגמא כזו נמצאות הנחות יסוד מתמטיות ושאלה פורמלית הנוגעת לגזירת מסקנות מהנחות יסוד באמצעות כללי הסק. מתוך הרצון לבסס תורה פורמלית כזו נניח תמיד כי הסיטואציה ההסתברותית מתוארת באופן מדוייק וחד-משמעי, כלומר, כי אנו חוקרים תוצאה של **ניסוי**, שאת הסתברותה של כל אחת מתוצאותיו אנו יודעים מראש. כדי להביא את המציאות למצב שניתן לנתחה בכלים נוקשים כאלו, דרושה יכולת להמיר מצבי אי-ודאות שבהם אפילו אופיו של הניסוי אינו מוגדר היטב לתצורה הסתברותית. מלאכה זו מוטלת על כתפי הסטטיסטיקה כפי שמוסבר להלן.

הסתברות וניבוי סטטיסטי. תורת ההסתברות הנלמדת בקורס זה הנה תורה מתמטית טהורה, בדיוק כמו הגאומטריה, החשבון האינפיניטיסימלי והאלגברה הלינארית. כמו במקרה של תורות אלה, כך גם בעזרת תורת ההסתברות ניתן לבנות מודלים לתיאר המציאות ולנבא ניבויים הנוגעים אליה. התורה העוסקת באומדן המתאם בין מודלים הסתברותיים למציאות מכונה **סטטיסטיקה היסקית (Statistical Inference)**. במקרה של החשבון האינפיניטיסימלי, אנו יכולים לחזות, בעזרת מודל מתאים, את מסלולו של כדור תותח. כל יריה אשר החיזוי מוצלח לגביה – מאששת את המודל ובעקיפין את ביטחוננו שהתורה המתמטית העומדת מאחורי המודל היא כלי הולם לבניית מודלים המתארים נכוחה את העולם.

מהן התחזיות המתקבלות ממודל הסתברותי? עצם השאלה מסגירה עד כמה המדעים הגבוהים השליטו בתקופתנו אמות מידה קרטזיות של חיפוש אחר אמת מוחלטות. מנקודת מבט זו הטענה ״המטבע יפול על עץ בהסתברות חצי״ אינה יכולה להחשב תחזית - שכן אין ניסוי היכול להפריכה - אם נטיל מטבע נקבל או עץ או פלי ושתי התוצאות אינן סותרות את התחזית. לעומת זאת, ״אם נטיל מטבע אלף פעמים יצאו בין 400 ל-600 עצים״ זו תחזית שניתן לאשש בניסוי. נשים לב כי תחזית זו אינה כוללת בתוכה את המילה ״הסתברות״. אמנם, התורה המתמטית שלנו אינה גורסת שתוצאה של למעלה מ-600 עצים הנה בלתי-אפשרית, אלא רק שהסתברותה קטנה עד-מאוד (פחות מאחד למיליארד!). בפירוש התורה לצורכי ניבוי קיבלנו החלטה מושכלת להזניח הסתברות הזו בכדי להתאים עצמנו לאמות מידה של תחזית מוחלטת. ההחלטה מתי הסתברויות מסוימות הן זניחות לצורך מסוים אינה החלטה מתמטית טהורה והיא מושפעת מגורמים חיצוניים רבים. עצם ההחלטה להחזיק באמונות ״מוחלטות״ היא החלטה פילוסופית אשר ניתן לערער עליה. למעשה גם תחזיות לגבי מעופו של כדור התותח אינן מתמטיות-טהורות ומצריכות שלב של התאמה בין המודל למציאות, למשל, החלטה מתי טעות מסוימת

בחיזוי מסלול הכדור היא סבירה ומתי היא מעידה שהמודל בו אנו משתמשים שגוי. לימוד הסטטיסטיקה מועיל ללימוד ההסתברות, מרחיב אותו ומאיר אותו מזויות חדשות. מסיבה זו החלטנו לכלול בספר פרקים העוסקים בסטטיסטיקה היסקית אשר יעניקו לנו מושג מסויים על השיקולים המתמטיים בהכרעה באשר לנכונותו של מודל נתון. עיון נרחב יותר בסוגיה זו יתבע מהקורא עיון בספרים שעניינם המרכזי הוא התורה הסטטיסטית. תורה זו מרחיבה את עיסוקיה גם למצבים שבהם לא נערך ניסוי כלל, אלא המידע נופל לרשותו של החוקר מתוך צפייתו בטבע. כאשר ברצוננו להתחקות אחר ההצדקה לכך שהעיון בטבע יביא אותנו לנסח השערות מסויימות, חורג העיון בבעיית אי-הוודאות גם מתחומה של הסטטיסטיקה והוא גולש לתחום הפילוסופיה ההכרתית המבקשת לבאר "מה היא ידיעה".

סימונים

בפרק קצר זה נסקור ונבאר את הסימנים הנהוגים בספר. ככלל ננהיג את הסימנים הבאים.

קבועים טבעיים	a, b, d, n, N
שברים	γ, θ
קבועים ממשיים	s, t, r
קבועים שערכם אינו משמעותי	C, c
קבוצות	A, B, D, E
אוספי קבוצות	$\mathcal{A}, \mathcal{B}, \mathcal{S}$
משתנים	x, y, z, w
אינדקסים	i, j, k, n
משתנים מקריים	X, Y, Z, W
התפלגויות	$\mathcal{D}$
קבוצות של משתנים מקריים ושל התפלגויות	$\mathcal{X}, \mathcal{Y}, \mathcal{F}, \mathcal{H}$
פונקציות	f, g, h, ϕ, ψ
ערכים קטנים השואפים לאפס	ϵ, η
הפרשים או ערכים קטנים כרצוננו	Δ, δ
תוחלת	μ
סטיית-תקן	σ
מבחן סטטיסטי, או קבוצת אינדקסים	Λ
שגיאות סטיסטיות	α, β

פעולות מתמטיות:

הגדרה של מושג	הגדרה $:=$ מושג
חיבור של סדרה של מספרים $\{a_i\}_{i\in[n]}$	$\sum_{i=1}^n a_i$
מכפלה של סדרה של מספרים $\{a_i\}_{i\in[n]}$	$\prod_{i=1}^n a_i$
איחוד של אוסף קבוצות $\{A_i\}_{i\in[n]}$	$\bigcup_{i=1}^n A_i$
איחוד של אוסף קבוצות זרות $\{A_i\}_{i\in[n]}$	$\dot{\bigcup}_{i=1}^n A_i$
חיתוך של אוסף קבוצות $\{A_i\}_{i\in[n]}$	$\bigcap_{i=1}^n A_i$
קבוצה A מכילה את B	$B \subset A$
קבוצה A מכילה ממש את B	$B \subsetneq A$
ההפרש בין קבוצה A לקבוצה B	$A \setminus B$
קבוצה משלימה ל-A (ביחס לקבוצה אוניברסלית Ω)	A^c
N עצרת	$N! := \prod_{n=1}^N n$
n מעל k	$\binom{n}{k} := \frac{n!}{k!(n-k)!}$
עוצמה (גודל) של קבוצה A	$\|A\|$
קבוצת חזקה של A	$2^A := \{B \ : \ B \subseteq A\}$

מושגים מלוגיקה:

שלילה של טענה ψ	$\neg\psi$
ψ וגם ϕ	$\psi \wedge \phi$
ψ או ϕ	$\psi \vee \phi$
ψ גורר ϕ	$\psi \Rightarrow \phi$
ψ שקול ל-ϕ	$\psi \Leftrightarrow \phi$
לכל x מתקיים $\psi(x)$	$\forall x\, \psi(x)$
קיים x עבורו מתקיים $\psi(x)$	$\exists x\, \psi(x)$

קבוצות מיוחדות:

קבוצת הטבעיים	$\mathbb{N} := \{1, 2, \ldots, \}$
קבוצת השלמים	$\mathbb{Z} := \{\ldots, -1, 0, 1, \ldots\}$
קבוצת המספרים הממשיים	$\mathbb{R} := (-\infty, \infty)$
קבוצת המספרים האי-שליליים	$\mathbb{R}_+ := [0, \infty)$
קבוצת התמורות על המספרים עד N	$S_N := \{a_1, \ldots, a_N \ : \ a_i \in [N], a_i \neq a_j\}$
קבוצת התמורות על המספרים הטבעיים	$S_{\mathbb{N}} := \{a_1, a_2, \ldots \ : \ a_i \in \mathbb{N}, a_i \neq a_j\}$
מספרים טבעיים קטנים מ-$k \in \mathbb{N}$	$[k] := \{1, 2, \ldots, k\}$
קטעים פתוחים	$(a, b) := \{x \ : \ a < x < b\}$
קטעים חצי-פתוחים חצי-סגורים	$[a, b) := \{x \ : \ a \leq x < b\}, \quad (a, b] = \{x \ : \ a < x \leq b\}$
קטעים סגורים	$[a, b] := \{x \ : \ a \leq x \leq b\}$
קטעים המכילים את אינסוף	$(a, \infty] := (a, \infty) \cup \{\infty\}$

כללי-כתיבה:

נהיג את הכלל לפיו טור מתכנס בהחלט שאבריו הם $\{a_i\}_{i\in\mathbb{N}}$ יסומן באמצעות $\sum_{i\in\mathbb{N}} a_i$ ואילו טור שאינו

בהכרח מתכנס בהחלט - יסומן $\sum_{i=1}^{\infty} a_i$. זאת כדי להזכיר לקורא שאין חשיבות לסדר הסכימה בטור מתכנס בהחלט (כמפורט בפרק א.4.1 להלן).

פרק א

רקע מתמטי

ספר זה נועד לקוראים הבקיאים ברזי החשבון האינפיניטיסימלי, אשר רכשו ניסיון בסיסי בשימוש בלוגיקה, בקומבינטוריקה ובתורת הקבוצות. בפרק זה מצויות כל התוצאות מתחומים אלו שנעשה בהן שימוש והן משמשות כמובאות להוכחות ביתר הפרקים. בניגוד ליתר פרקי הספר, נושאי פרק זה יוצגו בקיצור וללא הוכחות ולקורא שטרם נתקל בהם המבקש לקנות עליהם שליטה מומלץ לפנות לספרים יעודיים בנושאים אלה.

א.1 לוגיקה מתמטית

מאז ימי קדם נשענת המתמטיקה על הלוגיקה – **תורת ההיסק**. תורה זו עוסקת בטענות, בפעולות לוגיות ובכללי היסק. **תחשיב הפסוקים** הלוגי עוסק ביחסים בין טענות מוחלטות המערבות פעולות כמו וגם ($\wedge$), או ($\vee$) ושלילה ($\neg$) ואילו **תחשיב הפרדיקטים** מרחיב עיסוק זה לטענות המכילות את הכמתים **לכל** ($\forall$) **וקיים** ($\exists$). טענות לוגיות יכולות להכיל **משתנים** (למשל X) המקבלים ערכים בקבוצה מסויימת. לכל השמה של ערכים במשתנים הטענה מקבלת את הערכים **אמת** (T) **ושקר** (F). נאמר כי שתי טענות הן **שקולות** ($\Leftrightarrow$) אם הן מקבלות את אותם ערכי אמת לכל הצבה של ערכים במשתנים. נאמר כי טענה ϕ **גוררת** ($\Leftarrow$) את טענה ψ, אם בכל הצבה שבה ϕ מקבלת ערך אמת, גם ψ מקבלת ערך אמת. טענה שהנה אמיתית לכל הצבה של ערכים במשתנים נקראת טָאוטולוֹגְיָה או טענת אמת. מטרתנו בעיסוק המתמטי היא להוכיח באמצעות שרשרת של שקילויות לוגיות כי טענות בעלות עניין עבורנו הנן טענות אמת. זאת נעשה על ידי כך שנראה באמצעות שרשרת של מעברים לוגיים כי **הנחות יסוד** וטענות שכבר השתכנענו באמיתותן גוררות את הטענה שברצוננו להוכיח. שרשרת מעברים זו מכונה **הוכחה**.

ניתן לנסח טענות לוגיות הן בכתיב פורמלי משמאל לימין והן בניסוח מילולי בשפה פורמלית. ברוב המקרים נשתמש בהוכחותינו בשפה פורמלית בכדי להקל על קריאתן, כאשר סימנים לוגיים ישמשו רק במשוואות, לשם קיצור. הוכחותינו עצמן נשענות על מספר טענות לוגיות מוכרות אשר החשובות שבהן מובאות בפרק זה.

כללי דה-מורגן (De Morgan's laws). כללים אלו קובעים את היחסים בין שלילה לבין הפעולות "וגם" ו-"או" באופן הבא.

טענה א.1. תהיינה ϕ ו-ψ טענות. אזי
$$\neg(\psi \vee \phi) = (\neg\psi) \wedge (\neg\phi)$$
$$\neg(\psi \wedge \phi) = (\neg\psi) \vee (\neg\phi)$$

תיאור מילולי של הכלל השני למשל, הוא "אי קיום הטענה (ψ וגם ϕ) שקול לקיום הטענה (לא התקיימה ψ או לא התקיימה ϕ)".

שיטות הוכחה. מלבד הוכחה טיפוסית של טענות גרירה לוגית על ידי גזירת המסקנה מההנחות, נשתמש בשלושה כללי גזירה נוספים. **הוכחה בדרך השלילה**: כלל זה קובע כי הטענה (טענה ψ גוררת את טענה ϕ) שקולה לטענה (שלילת טענה ϕ גוררת את שלילת טענה ψ). באמצעותו נוכל כאשר נבקש להוכיח כי "טענה ψ גוררת את טענה ϕ", להניח דווקא את שלילת טענה ϕ ולהוכיח על סמך הנחה זו את שלילת טענה ψ.
שקילות היא שרשרת מעגלית של גרירות: כלל זה יאפשר לנו להוכיח שקילות בין טענות במונחים גרירות.

טענה א.2 (שקילות וגרירה מעגלית)**.** תהיינה $\psi_1, \ldots, \psi_k$ טענות לוגיות. השקילות
$$\psi_1 \Leftrightarrow \psi_2 \Leftrightarrow \cdots \Leftrightarrow \psi_k$$

שקולה לטענה
$$(\psi_1 \Rightarrow \psi_2) \wedge (\psi_2 \Rightarrow \psi_3) \wedge \cdots \wedge (\psi_{n-1} \Rightarrow \psi_n) \wedge (\psi_n \Rightarrow \psi_1)$$

אינדוקציה מתמטית: כלל זה מאפשר להוכיח את נכונותה של טענה $\psi(n)$ עבור כל המספרים הטבעיים הגדולים או שווים ל-n_0. ראשית נוכיח את נכונות הטענה עבור n_0 (בסיס האינדוקציה), ולאחר מכן נוכיח כי $\psi(n)$ גוררת את נכונות $\psi(n+1)$. לעיתים נרצה בהוכחת $\psi(n+1)$ להסתמך על נכונות כל הטענות $\psi(m)$ עבור $n_0 \le m \le n$. הוכחה כזו נקראת **אינדוקציה שלמה**.

שימוש בכמתים. נדגים שימוש בפסוקים לוגיים המערבים **כמתים**, כלומר את הסימנים $\forall$ (לכל) ו-$\exists$ (קיים). כרגיל, יש לקרא טענות בכתיב פורמלי משמאל לימין. כך למשל נכתוב באופן פורמלי את הגדרת התכנסותה של סדרה $\{a_n\}$ לאפס באופן הבא:

"לכל $\epsilon > 0$ קיים $n \in \mathbb{N}$ כך שלכל $m > n$ מתקיים $|a_m| < \epsilon$" $\qquad$ $\forall\epsilon > 0\ \exists n \in \mathbb{N}\ \forall m > n\ :\ |a_m| < \epsilon$

שלילת כמתים. כאשר שוללים טענה ניתן להחליף את הסדר בין שלילה לבין כמת על ידי החלפת הכמת "לכל" בכמת "קיים" ולהפך. כך מתקיים עבור סדרת מספרים $\{a_n\}_{n\in\mathbb{N}}$:

לא כל מספר a_n הינו שלם. $\Longleftrightarrow$ קיים מספר a_n שאינו שלם

$$\neg\forall\epsilon > 0\ \exists n \in \mathbb{N}\ \forall m > n\ :\ |a_m| < \epsilon \Longleftrightarrow \exists\epsilon > 0\ \forall n \in \mathbb{N}\ \exists m > n\ :\ \neg(|a_m| < \epsilon).$$

יחסים. יחס הנו ביטוי ψ המקבל שני איברים מקבוצה מסויימת Ω, כלומר מבנה המקבל שני אובייקטים ומחזיר ערך "אמת" או "שקר" המייצגים את התשובה לשאלה "האם שני האובייקטים מקיימים את היחס". ליחסים יכולות להיות תכונות שונות. הבה נמנה כמה מן החשובות שבהן.

בהינתן יחס $\psi(x,y)$ נאמר כי הוא **רפלקסיבי** אם כל איבר מקיים אותו ביחס לעצמו כלומר $\forall x \in \Omega\ :\ \psi(x,x) = T$. כך למשל היחס "$x$ מתחלק ב-y" הוא רפלקסיבי. $\psi(x,y)$ הוא **אנטי-רפלקסיבי** אם אף איבר בקבוצה אינו מקיים אותו על עצמו, כלומר $\forall x \in \Omega\ :\ \psi(x,x) = F$. כך למשל היחס "$x$ גדול ממש מ-y"

הוא אנטי-רפלקסיבי. $\psi(x,y)$ הוא **סימטרי** אם התקיימותו אינה תלויה בסדר האיברים, כלומר $\forall x,y\in\Omega:$ $\psi(x,y)\Leftrightarrow\psi(y,x)$. כך למשל היחס "$x$ ו-y בעלי אותה שארית חלוקה ב-3" הוא סימטרי. $\psi(x,y)$ הוא **אנטי-סימטרי** אם לא יתכן שהוא יתקיים באופן סימטרי אפילו על זוג איברים יחיד שאינם זהים. כך למשל "x גדול או שווה ל-y" הוא אנטי-סימטרי. $\psi(x,y)$ הוא **עובר** אם התקיימותו על הזוג x ו-y ועל הזוג y ו-z מבטיחה את התקיימותו על x ו-z. כלומר $\forall x,y,z\in\Omega:\psi(x,y)\wedge\psi(y,z)\Rightarrow\psi(x,z)$. כך למשל היחס "$x$ גדול או שווה ל-y" הוא יחס עובר.

יחס שקילות. יחס נקרא יחס שקילות אם הוא רפלקסיבי, סימטרי ועובר. יחס שקילות מחלק את אברי הקבוצה עליה הוא פועל למחלקות. כל שני איברים באותה מחלקה מקיימים את היחס, ואילו איברים ממחלקות שונות אינם מקיימים אותו.

א.2 תורת הקבוצות

תורת הקבוצות עוסקת ביחסים בין קבוצות של איברים. מושגי היסוד של התורה הם **איבר**, **קבוצה** ויחס ה-**שייכות**. שייכותו של איבר a לקבוצה A מסומנת על ידי $a\in A$. נשים לב כי כל איבר יכול להימצא בקבוצה פעם אחת בלבד. אם קבוצה A מכילה את כל האיברים של קבוצה B אזי נאמר כי A **חלקית** ל-B (או **מוכלת** ב-B) ונסמן $A\subset B$. **חיתוך** של שתי קבוצות A ו-B, המסומן $A\cap B$, הוא אוסף כל האיברים המצויים בשתיהן ואילו **איחוד** של שתי קבוצות, המסומן $A\cup B$, הוא אוסף כל האיברים המצויים לפחות באחת מהן. איחוד וחיתוך מקיימים את כלל הפילוג כמו חיבור וכפל בהתאמה. כך למשל $(A\cup B)\cap C=(A\cup C)\cap(B\cup C)$. **ההפרש** בין קבוצה A לקבוצה B המסומן $A\setminus B$ הוא אוסף כל האיברים המצויים ב-A ואינם מצויים ב-B. לעיתים קרובות אנו מגבילים את עצמנו לדיון בתת-קבוצות של קבוצה אוניברסלית Ω. במקרה כזה נסמן את ההפרש בין Ω ל-A ב-A^c. באמצעות כללי דה-מורגן, נוכל להראות את הקשרים הבאים בין פעולות איחוד, חיתוך ומשלים.

טענה א.3 (כללי דה-מורגן לקבוצות)**.** תהיינה A ו-B קבוצות חלקיות לקבוצה אוניברסלית Ω. אזי

$$(A\cup B)^c=(A^c\cap B^c)$$
$$(A\cap B)^c=(A^c\cup B^c)$$

ה**עוצמה** של קבוצה מייצגת את גודלה. עוצמתה של קבוצה סופית היא מספר האיברים בה. עוצמתה של קבוצה A גדולה או שווה לעוצמת הקבוצה B אם קיימת התאמה **חד-חד-ערכית** מאברי A לאברי B, כלומר לכל איבר ב-B מותאם לכל היותר איבר יחיד ב-A. אם עוצמה של קבוצה A גדולה או שווה לעוצמת הקבוצה B ובנוסף עוצמת B גדולה או שווה לעוצמת A – אז נאמר כי שתי הקבוצות שוות עוצמה.

קבוצה שוות עוצמה ל-$\mathbb{N}$, קבוצת המספרים הטבעיים, נקראת **בת-מניה**. קבוצה שהנה בת-מניה או סופית נקראת **בדידה**. הקטע $[0,1]$, קבוצת כל המספרים הממשיים $\mathbb{R}$, וכן אוסף כל הסדרות המורכבות מהספרות 0 ו-1 מהוים דוגמאות לקבוצות שאינן בנות מניה, כלומר לא קיימת התאמה חד-חד-ערכית מקבוצות אלו ל-$\mathbb{N}$.

קבוצת חזקה של קבוצה נתונה A היא קבוצת כל תת-הקבוצות שלה שתסומן בסימון $2^A=\{B\ :\ B\subset A\}$ וזאת כדי לציין שניתן לראות אותה כקבוצה של כל ההתאמות של אברי A לקבוצה בת שני האיברים הנותנה על ידי {נכלל, לא-נכלל} כאשר הקבוצה המתאימה להתאמה היא קבוצת כל אברי A שהותאמו לאיבר נכלל. נשים

לב שכאשר A סופית ישנן בדיוק שתי תמונות אפשרויות לכל איבר ב-A ולכן אומנם מתקיים $|2^A| = 2^{|A|}$. **משפט האלכסון של קנטור** מבטיח לנו שעוצמתה של קבוצת החזקה תמיד גדולה ממש מעוצמת הקבוצה המקורית.

פעולת חזקה בין שתי קבוצות עבור קבוצות A ו-B, הקבוצה A^B מסמלת את כל ההתאמות מהקבוצה B לקבוצה A. כך למשל הקבוצה $\{H, T\}^{\mathbb{N}}$ מיצגת את אוסף ההתאמות בין כל מספר טבעי n לערך (עץ או פלי - ובלעז Head או Tails). קבוצה זו מתאימה למשל לתאר תוצאה של סדרה אינסופית של הטלות מטבע, כאשר תוצאת ההטלה ה-n-ית מיוצגת על ידי האיבר אליו מותאם הערך n.

3.א קומבינטוריקה

עוד בטרם נתוודענו להגדרות הפורמליות של תורת ההסתברות, יש בכוח מחשבתנו להכריע מטעמי סימטריה מספר שאלות הסתברותיות. אם נסתכל, למשל, על הטלת קובייה סימטרית בעלת N פאות, בחירה של כדור אחד מתוך N כדורים המצויים בכד אטום, או שליפה באקראי של קלף מתוך חפיסת קלפים מעורבבת בת N קלפים. בכל אחד ממצבים אלו, מטעמי סימטריה הסתברותה של כל תוצאה חייבת להיות שווה, והיות שסך ההסתברות של כל התוצאות צריך להיות 1 - הרי שהסתברותה של כל תוצאה היא $1/N$ והסתברותה של קבוצת תוצאות בת k איברים היא k/N. במצבים שכאלה מפושטת הבעיה ההסתברותית לחישוב ערכי N ו-k. מציאת ערכים אלו היא **בעיית ספירה** ועל כן פתרונה נמצא לעיתים קרובות באמצעות תורת הספירה - ה**קומבינטוריקה**. הקומבינטוריקה תשמש אותנו הן לבעיות **מניה**, כאשר נבקש לחשב במדוייק את ערכי n ו-k, והן לבעיות **אומדן** – כאשר נבקש להעריך את n ו-k באופן מקורב. סימונים ותוצאות הנוגעות לאומדנים מובאות להלן לאחר העיסוק בפונקציות.

3.1.א בעיות מניה

הבעיה הבסיסית של הקומבינטוריקה היא "בהינתן קבוצה A לחשב או לאמוד את גודל הקבוצה $|A|$". בנסיבות הקומבינטוריות הפשוטות ביותר, גודל זה מחושב תוך שימוש בתכונות סימטריה של הקבוצה. לאורך הספר נרבה להשתמש בסימון הקומבינטורי $[n] := \{1, \ldots, n\}$.

מכפלות קרטזיות. במקרים רבים תיאורה של הקבוצה A יתבצע באמצעות תיאור של אוסף של בחירות. כך למשל קבוצת האפשרויות לחלק כובעים שחורים או לבנים לחמישה אנשים ניתנת לתיאור כחמש בחירות שבכל אחת מהן עלינו לקבוע האם איש מסויים יקבל כובע שחור או לבן. את קבוצת האפשרויות לחלק את המספרים מאחד עד ארבע לארבעה סוסי-מרוץ אפשר לתאר כארבע בחירות. לסוס הראשון נוכל לתת אחד מארבעה מספרים, לסוס השני אחד משלושה מספרים, לשלישי אחד משני מספרים ולאחרון נותרת רק אפשרות יחידה. מרחבים מסוג זה ניתן להתאים לאוסף כל ה-k-יות $(a_1, a_2, \ldots, a_k)$ כאשר $a_i \in A_i$. כך למשל בדוגמת הכובעים $k = 5$ ו- $A_i = \{\text{לבן}, \text{שחור}\}$, ובדוגמת המספרים $k = 4$ ו- $A_i = [4 - i]$ כאשר a_i מייצג את מיקומו הסידורי של המספר שנבחר לסוס ה-i מבין המספרים שעדיין לא חולקו לסוסים הקודמים. מרחב כזה מכונה **מכפלה קרטזית** של הקבוצות A_i והוא מסומן $A = \times_{i=1}^{k} A_i := \{(a_1, \ldots, a_k) \ : \ a_i \in A_i\}$.

גודלה של מכפלה קרטזית הוא מכפלת גדלי הקבוצות המרכיבות אותה, כלומר $\left|\times_{i=1}^{k} A_i\right| = \prod_{i=1}^{k} |A_i|$. כאשר כל הקבוצות זהות נשתמש בכתיב חזקה $A^k := \left|\times_{i=1}^{k} A\right|$. מתוך נוסחא זו נגזרות מספר נוסחאות קומבינטוריות הנוגעות למדגמים סדורים.

מדגם סדור עם החזרה. מדגם סדור עם החזרה של k איברים מתוך קבוצה A בגודל $|A| = n$ הוא שם אחר לקבוצת החזקה A^k. מקור המונח במצב שבו אנו בוחרים איזה איבר ב-A להתאים לכל אחד מ-k גורמים. המילים "עם החזרה" מתייחסות לכך שלאחר שהותאם לגורם כלשהו איבר מסויים ב-A הוא "מוחזר לקבוצה A" וניתן לשוב ולהתאים אליו גורמים נוספים. נוכל לספור את האפשרויות למדגם סדור באמצעות הנוסחה:
גודל קבוצת המדגמים הסדורים בגודל k עם החזרה מתוך n איברים שונים הוא n^k.

דוגמאות למדגם סדור עם החזרה הן מספר האפשרויות לחלק k כדורים ל-n תאים ומספר הדרכים לצבוע את המספרים $1, \ldots, n$ ב-k צבעים.

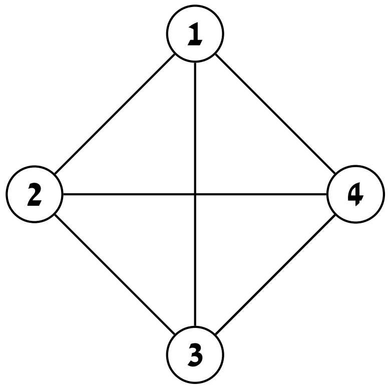
מדגם סדור ללא החזרה. מדגם סדור ללא החזרה של k איברים מתוך קבוצה A בגודל $|A| = n$ (עבור $n \geq k$). מתאים לאוסף ה-k-יות של איברים מתוך A בהן לא מופיע אותו איבר פעמיים. המילים "ללא החזרה" מתייחסות לכך שלאחר שהותאם לגורם כלשהו איבר מסויים ב-A הוא "מוצא מן הקבוצה A" ולא ניתן לשוב ולהתאים אליו גורמים נוספים. קבוצה זו ניתן להתאים לסדרה של בחירות בבין כמות חלופות הולכת ופוחתחת, בתחילה עומדים לרשותנו n חלופות התאמה לאיבר הראשון, $n-1$ לאיבר השני וכן הלאה עד שלאיבר ה-k נותרו רק $n-k+1$ חלופות. גודל קבוצת המדגמים הסדורים בגודל k ללא החזרה מתוך n איברים הוא $n!/(n-k)!$.

תמורות. המקרה הקיצוני ביותר של מדגם סדור ללא החזרה, הוא המקרה $k = n$ המתאים לאוסף הדרכים לסדר n איברים בשורה. סידור כזה מכונה **תמורה על n איברים**. אוסף כל התמורות על n איברים מסומן ב-S_n וכל תמורה בו מסומנת על פי רוב ב-$\sigma \in S_n$. את התמורה אנו רואים כפונקציה המתאימה לכל איבר $a \in [n]$ את מיקומו $\sigma(a)$. לתיאור התמורה כולה נרשום סדרת ערכים שיתאימו למיקומי האיברים לפי סדרם. כך למשל $\sigma = (1\ 3\ 4\ 2)$ מקיימת $\sigma(1) = 1$, $\sigma(2) = 3$, $\sigma(3) = 4$ ו-$\sigma(4) = 2$. **נקודת שֶׁבֶת** של תמורה היא ערך $a \in [n]$ עבורו מתקיים $\sigma(a) = a$. על כן בתמורה שהצגנו 1 היא נקודת השבת היחידה. בייצוג של תמורה כללית נפריד בין ערכי התמורה באמצעות פסיקים, כך נרשום $S_n = \{(x_1, \ldots, x_N) \ : \ x_i \neq x_j, \forall i < j\}$ כאשר $\sigma = (x_1, \ldots, x_N)$ מקיימת $\sigma(1) = x_1$.

חלוקה למחלקות. לעיתים איננו מבחינים בין תצריפים מסויימים במדגם שלנו. כך למשל נוכל לשאול בכמה תצורות נוכל לסדר n ילדים סביב שולחן עגול כאשר רק המיקום היחסי בין הילדים מייצג תצריף. כבר נוכחנו שאילו רצינו לסדר את הילדים בשורה הרי שהיו $n!$ דרכים לעשות כן, אלא שכעת איננו מבחינים בין כל אחד מ-n סידורים שבהם סדר הילדים זהה עד כדי סיבוב סביב השולחן. נוכל אפוא לחשוב על השאלה כשאלה של ספירת מחלקות של סידור ילדים בשורה כאשר בכל מחלקה מצויים כל n הסידורים המזדהים עד כדי סיבוב סביב שולחן. מכיוון שהמחלקות זרות ושוות גודל, מתקיים העקרון הבא:
אם מחלקים מרחב בגודל n למחלקות זרות אשר כולן בגודל k אז מספר המחלקות הוא n/k. נסיק כי מספר הדרכים לסדר את הילדים במעגל היא $n!/n = (n-1)!$.

בחירת k מתוך n איברים. בחירת k איברים מתוך קבוצה A בגודל $|A| = n$ הוא כינוי לאוסף כל תת הקבוצות בגודל k של A. באמצעות בחינת מדגמים סדורים ללא החזרה ושימוש בעקרון החלוקה למחלקות ניתן להראות כי: מספר האפשרויות לבחירת k איברים מתוך n הוא $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ (קרי n מעל k).

זהויות קומבינטוריות. לעיתים נוכל לקבל זהות בין שתי נוסחאות על ידי כך שנראה כי הן מתארות שתי



תרשים 1: הגרף השלם בן 4 קודקודים הנתון על ידי $G = ([4], [4] \times [4])$.

ספירות שונות של אותה כמות גורמים. כך למשל

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

זאת מפני שאת כמות הדרכים לבחור k גורמים מתוך המספרים ב-$[n]$ הנתונה באגף שמאל, ניתן לבטא כסך האפשרויות לבחור k גורמים מתוך $[n-1]$ ועוד סך האפשרויות לבחור את האיבר האחרון ולהוסיף לו $k-1$ גורמים מתוך $[n-1]$.

נוסחת הבינום. משילוב הכלים שהצגנו נסיק את הנוסחא הבאה: $(b + aX)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} X^k$. נוסחא זו מכונה נוסחאת הבינום, שכן היא מאפשרת חישוב של חזקות פולינום בן שני איברים (בלטינית בִּיס - שתיים, נוֹמֶן - שם, תואר). בשל נוסחא זו מכונה המקדם $\binom{n}{k}$ בשם מקדם בינומי.

3.2.א תורת הגרפים

גרף $G = (V, E)$ הוא זוג המורכב מקבוצה בדידה של קודקודים V (vertices) ואוסף $E \subset V \times V$ של קשתות (edges) המחברות ביניהם. לעיתים מסמנים את קודקודיו וקשתותיו של גרף G ב-$V(G)$ וב-$E(G)$ בהתאמה. גרפים יכולים לשמש לתיאור יחסים סימטריים בין אובייקטים, כאשר בין שני קדקודים יש קשת אם ורק אם מתקיים ביניהם היחס. מבנה זה שימושי בתחומים רבים ומגוונים ובמיוחד במדעי המחשב. הוא יכול לתאר קשרים חברתיים בין אנשים, יחסים בין מספרים או חיבורים ברשת מחשבים, ברשת חשמלית או ברשת כבישים. מסיבה זו התפתחה תורת הגרפים כענף שלם של הקומבינטוריקה המבקש להבין את תכונותיהם. כחלק מעיסוק זה נחקרים גם גרפים מקריים והללו יהוו עבורנו מקור לדוגמאות מעניינות.

גרף שלם. גרף שלם בן n קודקודים הוא הגרף $K_n = \{[n], [n] \times [n]\}$ שבו כל זוג קודקודים מחוברים זה לזה. **מסלול.** מסלול בגרף G הוא סדרה של קודקודים $\{a_i\}_{i \in \{0,\ldots,N\}}$ כך שמתקיים $(a_{i-1}, a_i) \in E(G))$ לכל $i \in [N]$. נאמר כי המסלול מחבר את v_0 ל-v_N. גרף שכל שני קודקודים בו מחוברים זה לזה נקרא גרף **קשיר**. נאמר כי גרף G' הוא **תת-גרף** של G אם $V(G') \subset V(G)$ ו-$V(E') \subset V(E)$.

4.א פונקציות ממשיות וחשבון אינפיניטיסימלי

לעיתים קרובות אנו מתעניינים במודלים הסתברותיים לתיאור תוצאות של מדידות. לצורך ניתוח מודלים אלו נידרש לעבודה עם פונקציות, עם סדרות ועם טורים של מספרים ממשיים. כמו כן, בבואנו לחסום את ערכם של טורים, נשתמש לעיתים בתכונות של פונקציות רציפות ובנגזרות. לצורך כך תשמשנה אותנו מגוון תוצאות מתחום הפונקציות הממשיות והחשבון הדיפרנציאלי והאינטגרלי.

4.1.א סדרות

סדרה. סדרה של מספרים ממשיים $\{a_n\}_{n\in\mathbb{N}}$ הנה התאמה בין המספרים הטבעים לבין איברים ב-R_n. האיבר a_n נקרא האיבר ה-n בסדרה. באופן דומה ניתן להגדיר סדרות של אובייקטים מתמטיים אחרים כמו מספרים מרוכבים או מטריצות או להתעניין בסדרות המתחילות באינדקס שאינו 0 או בסדרות דו-אינסופיות אשר האינדקסים שלהן רצים על אברי $\mathbb{Z}$. כאשר קבוצת האינדקסים מובנת מן ההקשר, משמיטים אותה ומסמנים את הסדרה על ידי $\{a_n\}$. נאמר ש**סדרה** $\{a_n\}$ **מתכנסת** ונסמן $\lim_{n\to\infty} a_n = L$ אם לכל $\epsilon > 0$ קיים n_0 כך שלכל $n > n_0$ מתקיים $|a_n - L| < \epsilon$. נאמר כי $\{a_n\}$ **מתכנסת בהחלט** אם $\{|a_n|\}$ מתכנסת. לעיתים נאפשר לסדרה להתכנס גם לאינסוף. נסמן זאת $\lim_{n\to\infty} a_n = \infty$. טענה זו תתקיים אם לכל $M > 0$ קיים n_0 כך שלכל $n > n_0$ מתקיים $a_n > m$. בדומה, נאמר כי $\{a_n\}$ מתכנסת למינוס אינסוף, אם $\{-a_n\}$ מתכנסת לאינסוף. סדרה המתכנסת לאינסוף, למינוס אינסוף או למספר תיקרא **מתכנסת במובן הרחב**.

לעיתים נרצה להראות כי סדרה מתכנסת מבלי לחשב את גבולה. לשם כך הוכיח קושי את הקריטריון הבא:

טענה א.4 (קריטריון קושי)**.** סדרה $\{a_n\}$ מתכנסת אם ורק אם לכל $\epsilon > 0$ קיים n_0 כך שלכל $m, n > n_0$ מתקיים $|a_n - a_m| < \epsilon$

4.2.א פונקציות

פונקציות. פונקציה $f : D \to R$ היא העתקה מקבוצה D (**התחום** של f) לקבוצה R (**הטווח** של f). לכל איבר $x \in D$ מתאימה הפונקציה איבר יחיד $f(x)$ המכונה ה**תמונה** של x. קבוצת **המקורות** של $x \in R$ דרך f המכונה גם **התמונה ההפוכה של** x מוגדרת באמצעות $f^{-1}(x) = \{d \in d : f(d) = x\}$. אם לכל $x \in R$ קבוצת מקורות בת לכל היותר איבר יחיד – נאמר כי הפונקציה **חד-חד-ערכית (חח"ע)**. אם לכל $x \in R$ קבוצת מקורות בת לכל הפחות איבר יחיד – נאמר כי הפונקציה היא **על**. פונקציה שהנה חח"ע ועל היא פונקציה **הפיכה**, ונגדיר את $f^{-1} : D \to R$ בתור הפונקציה שמתאימה לכל איבר ב-D את האיבר היחיד באוסף המקורות שלו, איבר זה יקרא **המקור של** x. כשנרצה לתאר שוויון בין שתי פונקציות $f, g : R \to D$ המתקיים לכל $x \in R$, נשתמש בסימון $f \equiv g$. זאת בכדי להבדיל אותו משוויון $f(x) = g(x)$ המתקיים לערך x מסויים.

פונקציות ממשיות. פונקציה ממשית היא פונקציה מ-$D \subset \mathbb{R}$ ל-$\mathbb{R}$. פונקציה ממשית נקראת **קמורה** אם D

הנה קטע במובן הרחב ולכל $\alpha \in [0,1]$ ולכל $x, y \in D$ מתקיים

$$f(\alpha x + (1-\alpha)y) \le f(\alpha x) + f((1-\alpha)y).$$

בפרט כל משיק לפונקציה קמורה עובר מתחת לגרף הפונקציה כלומר – לכל $x_0 \in D$ קיים $\beta \in \mathbb{R}$ כך שמתקיים $f(x) \ge f(x_0) + \beta(X - X_0)$ לכל $x \in D$. דוגמאות מרכזיות לפונקציות קמורות על $\mathbb{R}$ הן הפונקציות X^{1+a} עבור $a \ge 0$ ו-$\exp(X)$, וכן X^{-1-a} בתחום $(0,\infty)$. פונקציה ממשית f נקראת **מונוטונית עולה** אם לכל $x \le y$ מתקיים $f(x) \le f(y)$. אם בנוסף לכל $x < y$ מתקיים $f(x) < f(y)$ אז הפונקציה נקראת **מונוטונית עולה ממש**. בדומה , פונקציה ממשית f נקראת **מונוטונית יורדת** אם לכל $x \le y$ מתקיים $f(x) \ge f(y)$ **ומונוטונית יורדת ממש** אם לכל $x < y$ מתקיים $f(x) > f(y)$. אם f מונוטונית עולה או מונוטונית יורדת נאמר כי היא **מונוטונית**.

4.3.א גבולות

גבול של סדרה. עבור קבוצה איברים $A \subset \mathbb{R}$ נגדיר את **החסם העליון** של A, שיסומן על ידי $\sup(A)$, באמצעות ה-a הקטן ביותר כך שכל אברי A קטנים או שווים לו. בדומה נגדיר את **החסם התחתון** של A, שיסומן על ידי $\inf(A)$, באמצעות ה-a המירבי כך שכל אברי A גדולים או שווים לו. אם A אינה חסומה מלעיל, נאמר כי $\sup(A) = \infty$ ובדומה - אם אינה חסומה מלרע - נאמר כי $\inf(A) = -\infty$. לעיתים נסמן:

$$\inf_{x\in A} b_x = \inf\{b_x \ : \ x \in A\}, \qquad \sup_{x\in A} b_x = \sup\{b_x \ : \ x \in A\}$$

בהינתן סדרה $\{a_n\}_{n\in\mathbb{N}}$ מוגדרים ה**גבול העליון** וה**גבול התחתון** של הסדרה על ידי

$$\liminf_{n\to\infty} a_n = \sup_{n\in\mathbb{N}} \inf\{a_m \ : \ m > n\} \qquad \limsup_{n\to\infty} a_n = \inf_{n\in\mathbb{N}} \sup\{a_m \ : \ m > n\}$$

אם $\liminf_{n\to\infty} a_n = \limsup_{n\to\infty} a_n$, נגדיר ערך להיות **הגבול של** a_n ונסמנו על ידי $\lim_{n\to\infty} a_n$.

באופן דומה נוכל לדבר על גבול של סדרת וקטורים, למשל $(a_n, b_n)_{n\in\mathbb{N}}$ אם כל אחת מהקואורדינטות משל וקטורי הסדרה מתכנסת.

גבול של פונקציה. לכל איבר $x \in \mathbb{R}$ ולכל $\epsilon > 0$ נגדיר את **סביבת** ϵ **של** x להיות הקטע $(x-\epsilon, x+\epsilon)$ ונסמן $B_\epsilon(X)$. בהינתן פונקציה $f : \mathbb{R}^d \to \mathbb{R}$ נגדיר

$$\liminf_{x\to x_0} f(x) = \sup_{\epsilon>0} \inf\{f(x) \ : \ x \in B_\epsilon(x_0)\} \qquad \limsup_{x\to x_0} f(x) = \inf_{\epsilon>0} \sup\{f(x) \ : \ x \in B_\epsilon(x_0)\}.$$

ה-$\limsup$ הוא הערך הגדול ביותר שהנו קטן יותר מכל חסם מלעיל של הפונקציה בסביבת x_0.

אם $\liminf_{x\to x_0} f(x) = \limsup_{x\to x_0} f(x)$, נכנה ערך זה **הגבול של** f **ב-**x_0 ונסמנו $\lim_{x\to x_0} f(x)$. כך גם נגדיר גבולות עליונים ותחתונים באינסוף ובמינוס אינסוף על ידי

$$\liminf_{x\to\infty} f(x) = \sup_{m>0}\{\inf\{f(x) \ : \ x > m\}\} \qquad \limsup_{x\to\infty} f(x) = \inf_{m>0}\{\sup\{f(x) \ : \ x > m\}\}$$

$$\liminf_{x\to-\infty} f(x) = \sup_{m<0}\{\inf\{f(x) \ : \ x < m\}\} \qquad \limsup_{x\to-\infty} f(x) = \inf_{m<0}\{\sup\{f(x) \ : \ x < m\}\}.$$

רציפות. פונקציה ממשית $f(x)$ נקראת **רציפה** בנקודה x_0 אם מתקיים $\lim_{x\to x_0} f(x) = f(x_0)$. את מושג הרציפות ניתן להכליל באופן טבעי לפונקציה ממשית מ-$\mathbb{R}^d$ ל-$\mathbb{R}$. הפונקציה נקראת **רציפה למקוטעין** אם היא רציפה בכל נקודה מלבד לכל היותר אוסף בן-מניה של נקודות.

טענה א.5 (פונקציה מונוטונית היא רציפה למקוטעין)**.** כל פונקציה מונוטונית הנה רציפה למקוטעין.

נגזרות. פונקציה ממשית $f(x)$ נקראת **גזירה** בנקודה x אם היא רציפה בה וקיים הגבול

$$f'(x) := \lim_{x\to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

גבול זה מכונה ה**נגזרת** של f בנקודה x. כאשר f היא פונקציה של מספר משתנים, נוכל להגדיר נגזרת לפי כל אחד מהם בנפרד על ידי

$$\frac{\partial f(x_1,\ldots,x_n)(a_1,\ldots,a_n)}{\partial x_1} := \lim_{x_1\to a_1} \frac{f(x_1,a_2,\ldots,a_n) - f(a)}{x_1 - a_1}.$$

ניתן לחשב בקלות נגזרת של פונקציה הופכית באמצעות נגזרת של הפונקציה המקורית.

טענה א.6 (נגזרת של פונקציה הופכית)**.** תהי $g(x)$ פונקציה כלשהי, גזירה ב-x_0. נניח כי $y_0 = g(x_0)$ ונסמן $f(x) = g^{-1}(x)$. אזי

$$f'(x_0) = \frac{1}{g'(y_0)}.$$

א.4.4 אינטגרלים

אינטגרל רימן. ה**פונקציה הקדומה** של פונקציה ממשית $f(x)$ רציפה למקוטעין היא פונקציה $F(x)$ המקיימת $F'(x) = f(x)$ בכל נקודות הגזירות שלה. כל הקדומות של פונקציה מסויימת נבדלות זו מזו בקבוע. נגדיר סדרת ϵ בתחום $[a,b]$ להיות סדרת נקודות $a = x_0 \le x_1 \le \cdots \le x_n = b$ המקיימת $x_{i+1} - x_i < \epsilon$ לכל $i \le n$. נסמן אוסף סדרות אלו ב-$\mathcal{A}_\epsilon$. **אינטגרל רימן** של $f(x)$ בתחום $[a,b]$ הוא הגבול

$$\int_a^b f(X)dx := \lim_{\epsilon\to 0} \sum_{i\in[n]} f(x_{i,\epsilon})(x_{i,\epsilon} - x_{i-1,\epsilon})\}$$

כאשר $(x_{0,\epsilon},\ldots x_{n_\epsilon,\epsilon}) \in \mathcal{A}_\epsilon$. בתנאי שגבול זה אינו תלוי בבחירת הסדרות $(x_{0,\epsilon},\ldots x_{n_\epsilon,\epsilon})$. פונקציה שעבורה לכל $a < b$ האינטגרל $\int_a^b f(X)dx$ מוגדר היטב נקראת **אינטגרבילית**. כל פונקציה $F(y) = \int_a^y f(x)dx$ עבור $a \in R$ מכונה **אינטגרל בלתי-מסויים** של $f(x)$.

משפט א.7 (ניוטון-לייבניץ)**.** אם $f(x)$ רציפה למקוטעין בקטע $[a,b]$ אז

$$\int_a^b f(x)dx = F(b) - F(a).$$

לכל $F(x)$ - פונקציה קדומה של $f(x)$.

בדומה אם $F(b)$ גזירה חוץ מאשר בקבוצה בת-מניה של נקודות, ומתקיים $f(x) = F'(x)$ בכל נקודת

גזירות, אז מתקיים

$$\int_a^y f(x)dx = F(y) - F(a).$$

האינטגרל הבלתי-מסויים איננו בהכרח גזיר בכל מקום, אבל הוא רציף בכל מקום.

טענה א.8 (רציפות פונקציית האינטגרל)**.** תהי $f(x)$ פונקציה אינטגרבילית בקטע $[a,b]$ אזי מתקיים כי $F(y) = \int_a^y f(x)dx$ היא פונקציה רציפה ב-$[a,b]$.

חישובי אינטגרלים. היכרות עם פונקציות קדומות מאפשרת לנו, באמצעות משפט א.7 לחשב אינטגרלים של פונקציות רבות על ידי היכרות עם הקדומה שלהן.

באמצעות פונקציות אלו נוכל לחשב אינטגרלים של פונקציות רבות באמצעות שימוש ב**ליניאריות**, **הצבה** ו-**אינטגרציה בחלקים**.

משפט א.9. לכל פונקציה חיובית $f(x,y)$

$$\int_{a_0}^{a_1}\int_{b_0}^{b_1} f(x,y)dydx = \int_{b_0}^{b_1}\int_{a_0}^{a_1} f(x,y)dxdy$$

בכל מקרה בו לפחות אחד האגפים מוגדר היטב.

אינטגרלים חוזרים. לעיתים, במיוחד כאשר נבקש לחשב שטחים של תחומים דו-ממדיים, נתעניין במצב שבו הפונקציה שעליה אנו מבצעים את פעולת האינטגרציה היא בעצמה אינטגרל בלתי-מסויים. מצב זה מכונה **אינטגרל חוזר**.

משפט א.10 (משפט פוביני לאינטגרל חוזר)**.** לכל פונקציה חיובית $f(x,y)$

$$\int_{a_0}^{a_1}\int_{b_0}^{b_1} f(x,y)dydx = \int_{b_0}^{b_1}\int_{a_0}^{a_1} f(x,y)dxdy$$

בכל מקרה בו לפחות אחד האגפים מוגדר היטב.

בפרקים המתקדמים של הספר נשתמש גם בהכללה של משפט זה לתחומים מורכבים יותר.

טענה א.11 (משפט פוביני)**.** תהי $f:\mathbb{R}^2 \to \mathbb{R}$ פונקציה אינטגרבילית ותהי $A \subset \mathbb{R}^2$ קבוצה מהצורה

$$A = \{(x,y) \mid a \le x \le b, g(x) \le y \le h(x)\},$$

עבור a ו-b כלשהם ו-g ו-h פונקציות אינטגרביליות המקיימות $g(x) \le h(x)$ לכל $a \le x \le b$. אזי

$$\iint_A f(x,y)d(x,y) = \int_a^b \left(\int_{g(x)}^{h(x)} f(x,y)dy\right)dx.$$

יתר על כן, a, b, g ו-h יכולים לקבל גם את הערכים ∞ או $-\infty$.

4.5.א טורים

טור. הטור $\sum_{i=1}^{\infty} a_i$ מייצג את הסדרה $\{b_n\}$ עבור $b_i = \sum_{i=1}^{n} a_i$. אם הסדרה מתכנסת במובן הרחב לערך L נרשום $\sum_{i=1}^{\infty} a_i = L$ ונאמר כי זהו ערכו של הטור. טור $\sum_{i=1}^{\infty} a_i$ נקרא **מתכנס בהחלט** אם הטור $\sum_{i=1}^{\infty} |a_i|$ מתכנס.

טורים סופיים חשובים. נציג מספר טורים סופיים אשר ניתן לחשב את ערכם המדוייק בכלים אלגבריים.
סכום סדרה חשבונית. יהיו $a, b \in \mathbb{R}, n \in \mathbb{N}$. אזי מתקיים

$$\sum_{k=1}^{n}(a+kb) = an + b\binom{n+1}{2}$$

ובפרט

$$\sum_{k=m}^{n} k = \frac{(m+n)(n-m+1)}{2}.$$

סכום סדרת ריבועים. לפי נוסחת פלבר (Faulhaber) מתקיים:

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

סכום סדרה הנדסים. יהיו $a, r \in \mathbb{R}, n \in \mathbb{N}$. אזי מתקיים:

$$\sum_{k=1}^{n} ar^k = \frac{a(1-\alpha^{k+1})}{1-r},$$

בפרט עבור $r < 1$ מתקיים

$$\sum_{k=1}^{\infty} ar^k = \frac{a}{1-r}.$$

טורים אינסופיים. טור מתכנס שאינו מתכנס בהחלט נקרא **מתכנס בתנאי**. שינוי סדר סכימה של טור, היא העתקה המקבלת תמורה על המספרים הטבעיים $\sigma \in S_{\mathbb{N}}$ וטור $\sum_{i=1}^{\infty} a_i$ ומחזירה את הטור $\sum_{i=1}^{\infty} a_{\sigma_i}$. משפט רימן על שינוי סדר סכימה קושר בין אדישות לשינוי סדר סכימה והתכנסות בהחלט.

טענה א.12 (משפט רימן על שינוי סדר סכימה)**.** יהי $\sum_{i=1}^{\infty} a_i$ טור מתכנס.

(א) אם הטור מתכנס בהחלט במובן הרחב אז כל שינוי סדר סכימה שלו מתכנס לאותו ערך.

(ב) אם הטור מתכנס בתנאי במובן הרחב אז לכל $L \in [-\infty, \infty]$ קיים שינוי סדר סכימה $\sigma \in S_{\mathbb{N}}$ כך ש-$\sum_{i=1}^{\infty} a_{\sigma_i} = L$, ובנוסף קיים $\sigma' \in S_{\mathbb{N}}$ כך ש-$\sum_{i=1}^{\infty} a_{\sigma_i}$ אינו מתכנס כלל.

מכאן נובע כי, מחד גיסא, אם טור מתכנס בהחלט במובן הרחב - אז כל שינוי סדר סכימה שלו מתכנס לאותו ערך, ומאידך גיסא - אם טור מתכנס בתנאי אז לכל $L \in [-\infty, \infty]$ קיים שינוי סדר סכימה של אברי הטור שמתכנס לL (וכן קיים לטור כזה שינוי סדר סכימה שאינו מתכנס כלל). נשים לב שהתכנסותם של טורים חיוביים היא, מן הסתם התכנסות בהחלט. לפיכך נוכל לסכום טורים כאלו גם על קבוצות אינדקסים שאינן תת-קבוצות של $\mathbb{N}$. ננהיג אפוא את ההגדרה הבאה:

הגדרה א.13 (טורים חיוביים מוכללים)**.** תהי Ω קבוצה כלשהי ותהי $f : \Omega \to \mathbb{R}_+$, נכנה את הקבוצה $A = \{\omega : f(\omega) > 0\}$ בשם התומך של f. אם A לכל היותר בת-מניה – נגדיר

$$\sum_{\omega\in\Omega} f(\omega) := \sum_{\omega\in A} f(\omega)$$

כאשר, לפי טענה א.12, אגף ימין הוא סכום האיברים ללא קשר לסדר בו נסכום אותם. אם A אינה בת-מניה, נגדיר $\sum_{\omega\in\Omega} f(\omega) = \infty$.

בכל מקרה בו נסכום טור מתכנס בהחלט, נעדיף את הכיתוב $\sum_{i\in\mathbb{N}} a_i$ על פני הכיתוב $\sum_{i=1}^{\infty} a_i$ בכדי להדגיש את התכנסותו בהחלט של הטור. באמצעות הגדרה א.13 נוכל להגדיר טורים מוכללים שאינם בהכרח חיוביים:

הגדרה א.14 (טורים מוכללים)**.** תהי Ω קבוצה כלשהי ותהי $f : \Omega \to \mathbb{R}$, נגדיר

$$\sum_{\omega\in\Omega} f(\omega) := \sum_{\omega\in\Omega} \min(f(\omega), 0) - \sum_{\omega\in\Omega} (-\max(f(\omega), 0)),$$

ונאמר כי זהו הערך אליו הטור **מתכנס בהחלט** בכל מקרה שבו לפחות שני הטורים באגף ימין מתכנסים. אחרת נאמר שהטור $\sum_{\omega\in\Omega} f(\omega)$ אינו **מתכנס בהחלט**.

לצורך חישוב הסתברויות נשתמש גם בגרסא החלשה הבאה של משפט פוביני.

טענה א.15 (משפט פוביני לטורים חיוביים)**.** לכל אוסף מספרים חיוביים $\{a_{ij}\}_{i,j\in\mathbb{N}}$ מתקיים

$$\sum_{i\in\mathbb{N}} \sum_{j\in\mathbb{N}} a_{ij} = \sum_{j\in\mathbb{N}} \sum_{i\in\mathbb{N}} a_{ij}.$$

לעיתים נבקש להוכיח התכנסות של טור מספרים חיוביים באמצעות מבחני התכנסות. כך למשל

טענה א.16 (מבחן האינטגרל)**.** תהי סדרה חיובית $\{a_n\}$ המקיימת $a_n < f(n)$ עבור פונקציה מונוטונית יורדת על $\mathbb{R}$. אז אם $\int_0^\infty f(x)dx$ הרי שגם הטור $\sum_{n\in\mathbb{N}} a_n$ מתכנס.

פולינום טיילור. כיוון שעבודה עם פולינומים היא לעיתים קרובות נוחה יותר מאשר עבודה עם פונקציות כלליות, נרצה לעיתים לקרב פונקציה באמצעות פולינומים. עבור פונקציה f גזירה k פעמים בנקודה x_0 נגדיר את **פולינום טיילור מסדר** k להיות

$$P^k_{f,x_0}(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)x^k}{k!},$$

כאשר $f^{(k)}$ היא הנגזרת מסדר-k של f. פולינומים אלו חוסמים את ערכי הפונקציה f באופן הבא:

טענה א.17 (שארית לגרנג' לפולינום טיילור)**.** עבור פונקציה f גזירה $k+1$ פעמים בנקודה x_0 אשר פולינום טיילור שלה מסדר k יסומן $P^k_{f,x_0}(x)$, ונקודה x, קיימת נקודה c בין x_0 ל-x המקיימת

$$f(x) = P^k_{f,x_0}(x) + \frac{f^{(k+1)}(c)c^{k+1}}{(k+1)!}$$

כך למשל נוכל לקבל מטור טיילור סביב 0 של $\log(X+1)$ כי $\log(x+1) = x^{-1} - \frac{c^{-2}}{2}$ לכל $x > -1$ ולכן $\log(x+1) < x$.

התכנסות במידה שווה של סדרות. נאמר כי סדרת פונקציות $\{f_n(x)\}_{n\in\mathbb{N}}$ **מתכנסת במידה שווה** לפונקציה-f בקבוצה A אם לכל $\epsilon > 0$ קיים $n_0 \in \mathbb{N}$ כך שלכל $n > n_0$ ולכל $x \in A$ מתקיים $|f_n(x) - f(x)| \le \epsilon$. כלומר סדרת פונקציות מתכנסת במידה שווה אם קצב ההתכנסות שלה חסום מלמטה עבור האיברים ב-A. טור פונקציות $\{f_n(x)\}_{n\in\mathbb{N}}$ מתכנס במידה שווה אם הסדרה $\sum_{k=1}^n f_k(x)$ מתכנסת במידה שווה כאשר n שואף לאינסוף. התכנסות במידה שווה מאפשרת לנו לבצע פעולות של גזירה ואינטגרציה על טורים.

טענה א.18 (גזירה איבר-איבר)**.** יהי $S(x) = \sum_{i\in\mathbb{N}} f_i(x)$ טור פונקציות המתכנס במידה שווה בסביבת x_0. אז מתקיים

$$\frac{dS(x)}{dx} = \sum_{i\in\mathbb{N}} \frac{df_i(x)}{dx}.$$

כלומר ניתן לגזור את אברי הטור, לסכם אותם ולקבל את נגזרת נוסחת הטור כולו.

טענה א.19 (אינטגרציה איבר-איבר)**.** יהי $S(x) = \sum_{k=1}^\infty f_k(x)$ טור פונקציות ותהי $F_k(x)$ פונקציה קדומה של $f_k(x)$. אם הטור $\sum_{k=1}^\infty F_k(x)$ מתכנס במידה שווה בסביבת x_0 אז טור זה הוא פונקציה קדומה של $S(x)$. כלומר ניתן לבצע אינטגרציה לאברי הטור, לסכם את האינטגרלים ולקבל את אינטגרל נוסחת הטור כולו.

טור חזקות. טור פונקציות מהצורה $S(x) = \sum_{k=1}^\infty a_k x^k$ נקרא **טור חזקות**. לטור חזקות תמיד קיים **רדיוס התכנסות** r כך ש-$S(x)$ מתכנס עבור $|x| < r$ ואינו מתכנס עבור $|x| > r$. טורים אלו מתכנסים במידה שווה בפנים של רדיוס התכנסותם (כלומר עבור $|x| < r$). קיומה של התכנסות בשולי רדיוס ההתכנסות משתנה מטור לטור ואינה מובטחת.

טור טיילור. עבור פונקציות מסויימות המכונות **אנליטיות**, ניתן להרחיב את פולינומי טיילור לטור טילור. טור חזקות המתכנס לערכי הפונקציה עצמה, לפחות בטווח מסויים. טור זה מוגדר בנקודה בה הפונקציה גזירה אינסוף פעמים על ידי:

$$P_{f,x_0}(x) = \sum_{n=0}^\infty \frac{f^{(n)}(x_0)x^n}{n!}.$$

אנליטיות בקטע I סגור מסויים המכיל את x_0 שקולה לקיום קבוע אוניברסלי C , כך שלכל $x \in I$ הפונקציה f גזירה אינסוף פעמים ב-x ומקיימת

$$f^{(k)}(x) \le C^{k+1}k!.$$

חשוב במיוחד עבורנו יהיה טור טיילור המתכנס של הפונקציה המעריכית $e^x = \sum_{n=1}^\infty \frac{x^n}{n!}$.

א.5 אומדנים והערכות

לעיתים קרובות נבקש לאמוד בקירוב את התנהגותה של פונקציה עבור ערכי פרמטר גדולים או שואפים לאינסוף. לצורך כך נציג מספר סימונים מיוחד.

א.5.1 סימוני בוכמן-לנדאו והכללותיהם

המתמטיקאי הגרמני פאול בוכמן הציג מערכת סימנים להשוואת סדרי גודל בין פונקציות. מערכת זו שפותחה על ידי אדמונד לנדאו, דונלד קנות׳ ואחרים וזכתה לשימוש נרחב במדעי המחשב מקלה עלינו להשוות בין סדרי הגודל של פונקציות שונות. להלן הסימונים והגדרותיהם.

הגדרה	סימון
$\limsup_{n\to\infty} \frac{\lvert f(n)\rvert}{g(n)} < \infty$	$f(n) = O(g(n))$
$\lim_{n\to\infty} \frac{\lvert f(n)\rvert}{g(n)} = 0$	$f(n) = o(g(n))$
$\liminf_{n\to\infty} \frac{\lvert f(n)\rvert}{g(n)} > 0$	$f(n) = \Omega(g(n))$
$\lim_{n\to\infty} \frac{\lvert f(n)\rvert}{g(n)} = \infty$	$f(n) = \omega(g(n))$
$f = O(g)$ וגם $f = \Omega(g)$	$f(n) = \Theta(g(n))$

א.5.2 אומדן מקדמים בינומים.

נוסחת סטרלינג. המתמטיקאי הצרפתי אברהם דה-מואבר הוכיח לראשונה את האומדן מדוייק הבא לקצב הגידול של פונקציית העצרת, אשר זכתה, שלא בצדק, להיקרא על שם המתמטיקאי הסקוטי ג׳יימס סטרלינג.

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \Big(1 + o(1)\Big).$$

אומדנים לא אסימפטוטיים של פונקציית העצרת נתונים על ידי

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \le n! \le e\left(\frac{n}{e}\right)^n.$$

באמצעות אומדנים אלו נוכל לקבל הערכה גם לפונקציית הבינום מן הצורה

$$\left(\frac{n}{k}\right)^k \le \binom{n}{k} \le \left(\frac{ne}{k}\right)^k.$$

כאשר k גדול יחסית, נוח להשתמש גם בחסם הבא, שנובע ישירות מנוסחת הבינום,

$$\binom{n}{k} \le 2^k.$$

פרק ב

דוגמאות מפתח ושאלות הסתברותית

> תורת ההסתברות היא בסופו של דבר ביטוי של דבר ביטוי של הגיון ישר במונחים של חשבון דיפרנציאלי; היא מאפשרת לנו להעריך בדייקנות את מה שמוחות חריפים חשים במעין אינסטינקט אשר לעיתים קרובות אין ביכולתם להסביר... ...היא מלמדת אותנו כיצד להימנע מאשליות אשר תכופות מוליכות אותנו שולל;... ...אין מדע שמתאים יותר לכלול במערכת החינוך שלנו.
>
> – פְּיֶיר סִימוֹן לַפְּלַס, תיאוריה אנליטית של ההסתברות, 1812.

ב.1 תפקידן של דוגמאות בתורת ההסתברות

תורת ההסתברות מתאפיינת במערכת היחסים הקרובה בין התיאוריה ליישום. קרבה זו גורמת לכך שההפרדה בין אמיתות מתמטיות, הגדרות מדוייקות והוכחות מלאות לבין נימוקים בלתי-פורמלים ומעברים הנשענים רק על השכל-הישר, הפרדה אשר קנה תלמיד המתמטיקה בעמל רב בהכשרתו המוקדמת בלימודי החשבון הדיפרנציאלי והאינטגרלי, נוטה להיטשטש ולהוביל לשגיאות ולבלבול. מצב זה מחמיר לעיתים, כאשר נוסף אליה הרושם שתורת ההסתברות מעמיסה מערכת מסורבלת של הגדרות מופשטות שניתן היה להסתדר בלעדיהן ולהסתפק במספר קטן של הגדרות פשוטות. המצב מוסיף ומחמיר כאשר אותו תלמיד מעיין בפתרונות של תרגילים בהסתברות ומתרשם שרבים מהמעברים נראים כמו טכסיסים ייעודיים ובלתי-מוסברים אשר קשה לנחשם ולהתרגל אליהם.

הפתרון לכל הקשיים הללו נעוץ במיומנות המכונה על ידי המומחים בתחום "חשיבה הסתברותית". בקווים כללים ניתן לתאר יכולת זו כזיהוי מהיר וטבעי של הכלים והשיטות הנחוצות לתיאור מתמטי מדוייק של בעיה הסתברותית ולמציאת פתרון עבורה. רכישת החשיבה ההסתברותית נשענת על שלושה היבטים.

- הבנה טבעית של המושגים השונים ותפקידם ויישומם על דוגמאות פשוטות ומייצגות.
- היכרות עם דוגמאות קצה הנוגדות את השכל-הישר ועם התופעות המחוללות אותן.
- הבנת התרומה של כל רכיב בתורה בפישוט ועידון ניתוחן של דוגמאות מפתח.

לצורך הגשמת ההיבט השלישי, בחרנו מספר דוגמאות מפתח שתלווינה אותנו לאורך הספר. הראשונה בדוגמאות אלו, סוגיה 1, סוגיית המשחק שנקטע, הוצגה בפרק המבוא. בפרק זה נסקור בקצרה חמש הדוגמאות

נוספות מבלי להיכנס לתיאור מתמטי מדוייק וכן נציג את השאלות שנרצה לשאול על כל אחת מהן ואת שימושיהן המעשיים.

ב.2 דוגמאות חשובות

אחד המודלים החשובים והמועילים ביותר בתורת ההסתברות הוא הגרלה של n מספרים בטווח $[m]$ כאשר לכל מספר סיכוי שווה להיבחר ואין קשר בין מספרים שונים. מודל נפוץ אחר הוא הגרלה בזה אחר זה של n מספרים בטווח $[m]$ כאשר כל מספר נבחר בסיכוי שווה מבין המספרים **שטרם נבחרו**. כל הדוגמאות שנראה להלן קשורות בדרך זו או אחרת לאחד משני המודלים הללו. נשים לב שכאשר m גדול בהרבה מ-n שני המודלים דומים מאוד, מפני שבין כה וכה הסיכוי שנבחר את אחד המספרים פעמיים או יותר קלוש. מצד שני, כאשר $m = n$, שני המודלים שונים בתכלית מפני שבמודל של בחירה-ללא-החזרה אנו יודעים בוודאות שנבחר את כל המספרים ורק סדרם אינו ידוע.

ב.2.1 בעיית יום-ההולדת (birthday problem)

סוגיה 2. [בעיית יום-ההולדת] ניח שלכל אדם מקרי סיכוי שווה להיוולד בכל אחד מימות השנה הלועזית בלתי-מעוברת בת $m = 365$ ימים. בכיתה נמצאים n תלמידים מקריים.

מודל זה מעלה באופן טבעי את השאלות הבאות.

- מה הסיכוי שתלמיד מסויים ימצא בכיתה חבר שנולד באותו יום כמותו?
- מה הסיכוי שאין שני אנשים בכיתה שנולדו באותו יום-הולדת?
- כמה אנשים צריכים להיות בכיתה כדי שבסיכוי 99% לפחות יהיו שני תלמידים בעלי אותו יום-הולדת?
- כמה אנשים צריכים להיות בכיתה כדי שבסיכוי 1% לכל היותר יהיו שני תלמידים בעלי אותו יום-הולדת?
- כמה זוגות תלמידים בכיתה בממוצע נולדו באותו יום?
- מה תהיינה התשובות לשאלות אלו כתלות עבור ערכי m גדולים?
- מה הסיכוי, כתלות בm ו-n שיהיו בכיתה k תלמידים בכיתה שנולדו באותו יום?

לבעיית יום-ההולדת חשיבות יישומית רבה. מערכות תקשורת וברשתות מחשבים נדרש כל משתמש למזהה ספרתי. המערכת מעוניינת להימנע ככל האפשר מהתנגשויות בין מזהים. השיטה הפשוטה ביותר לכך היא כמובן להגדיר מסד נתונים כללי ולחלק מזהים למשתמשים השונים, אלא שלעיתים נרצה להמנע מהמורכבות שבניהול מסד נתונים מרכזי. במצב כזה נרצה שכל משתמש יגריל לעצמו מזהה שימלא את תפקיד "יום ההולדת". ניתוח המודל מאפשר לנו לזהות לכל כמות משתמשים m איזה גודל טווח מזהים n עלינו לבחור כדי להמנע מהתנגשויות. יישומים דומים קיימים גם בתורת ההצפנה ובתורת הקודים. עבור

הקורא שמתמצא במדעי-המחשב נציין גם ניתוח מספר ההתנגשויות הנו בעל חשיבות גם בבעיות של קביעת מרחב כתובת למיפוי זיכרון מחשב, או בקביעת גודלן של טבלאות גיבוב.

ב.2.2 בעיית ההתאמה (matching problem)

סוגיה 3. n מכתבים מיועדים ל-n נמענים שונים. כל מכתב הוכנס למעטפה מקרית ונשלח.

מודל זה מעלה באופן טבעי את השאלות הבאות.

- מה ההסתברות שאף מעטפה לא תגיע לנמען המיועד להן?
- כמה מעטפות באופן טיפוסי תגענה לנמען המיועד להן?
- מה ההסתברות שבדיוק k מעטפות תגענה לנמען המיועד להן?

ב.2.3 בעיית האספן (coupon collector problem)

סוגיה 4. ברכישת ביצת הפתעה מקבל הקונה בובה אחת מתוך m סוגי בובות אהובות. קונה רוכש n ביצי הפתעה.

מודל זה מעלה באופן טבעי את השאלות הבאות.

- מה ההסתברות שהקונה הצליח להשיג לפחות בובה אחת מכל סוג?
- מה ההסתברות שהקונה לא הצליח להשיג סוג בובה מסויים?
- מה ההסתברות שבין הבובות שקנה יש בדיוק k סוגים?

נוכל להכליל את המודל ולתאר מצב שבו הקונה מוסיף ורוכש בובות עד שהוא משיג בובה יחידה מכל סוג. כעת נוכל לשאול:

- כמה רכישות דרושות לקונה באופן טיפוסי עד שישיג בובה אחת מכל סוג?
- מה ההסתברות שתדרשנה לו הרבה פחות רכישות? הרבה יותר רכישות?

שימושיה של בעיית האספן נוגעים בעיקר למגוון הדגימות שנוכל לאסוף כאשר נערוך סקירה מקרי. גרסאות ממושקלות של הבעיה, כאשר לבובות שונות סיכויים שונים להופיע, מהוות מודל לכמות הדגימות הביולוגיות שנקבל בסקירת שטח בת n דגימות ומסייעת לנו לדעת כמה דגימות עלינו לעשות כדי לזהות את כל המינים בשטח מסויים או בדגימת מעבדה מסויימת. שימושים נוספים של המודל נוגעים לקצב התמלאותו של זיכרון המטמון ולגודל זיכרון המטמון הדרוש למשתמש או לתוכנה אשר עשויים להשתמש בכל רגע נתון באחת מ-n יחידות מידע.

ב.2.4 זמן-עד-כשלון (time-between-failures)

סוגיה 5. משק-בית קונה נורה חשמלית. בכל יום הנורה עלולה להשרף בהסתברות p.

מודל זה מעלה באופן טבעי את השאלות הבאות.

- כמה זמן תחזיק הנורה בממוצע?
- מה ההסתברות שנורה יחידה תחזיק במשך שנה שלמה?

נניח שכאשר נורה נשרפת – למחרת היום נרכשת נורה זהה ומותקנת במקומה.

- כמה נורות יצרוך משק הבית בממוצע בשנה?
- כמה נורות על משק הבית לרכוש אם ברצונו שהן תספקנה לו לשנה בסיכוי של 99%?

מודל הזמן-עד-כשלון מהווה מודל בסיסי באקטואריה וחשיבותו בכלכלה ובתחום הביטוח רבה מאוד. נוכל לדמיין שהנורה במודל המתואר היא למעשה פוליסת ביטוח ותקלה היא פדיון הפוליסה. במצב כזה לחברת הביטוח פוליסות רבות והיא מבקשת להעריך מה היא ההסתברות שהיא תדרש לשלם מספר רב מדי של פוליסות בזמן קצר באופן שיגרום לה להקלע לקשיים כלכליים. במציאות אירועי פדיון אינם מתרחשים באופן נפרד ובלתי-תלוי באופן כללי, אך לעיתים קרובות הקירוב הבלתי-תלוי נותן אומדנים מהימנים.

ב.2.5 הילוך שיכור פשוט (simple random walk)

סוגיה 6. שיכור מתחיל מראשית ציר המספרים. בכל יחידת זמן הוא נע, בהסתברות p יחידה אחת ימינה ובהסתברות $1-p$ יחידה אחת שמאלה.

מודל זה מעלה באופן טבעי את השאלות הבאות.

- אם נסתכל על השיכור בזמן n באיזה טווח מרחקים יימצא השיכורבהסתברות של 99%?
- מה ההסתברות שבצעד הn ימצא השיכור בראשית הצירים?
- מה ההסתברות שהשיכור ישוב ויבקר אינספור פעמים בראשית הצירים?

מודל זה של תנועה מקרית "חסרת תנע", כלומר שאינה זוכרת את הכיוון שבו התבצעה התנועה בשלב הקודם מוצא שימושים במגוון רחב של תחומים. ראשית, הוא מהווה מודל לסדרה של הימורים חוזרים ומתאר את מצב הונו של המהמר או של המשקיע, ולכן יש לו חשיבות רבה במתמטיקה פיננסית ומודלים שונים המכלילים אותו מתארים את תנודתיות ערכן של מניות ושווין של חברות. שנית, מודל זה מהווה תיאור טוב למדי של תנועתה בציר ה-X של פרודה בודדת בגז או בנוזל ולכן ניתוחו מאפשר להבין את התפשטות החום בחומר כמו גם פעפוע של שני חומרים אחד בתוך השני.

2.6.ב גרף ארדש-רני (Erdős-Rényi random graph)

סוגיה 7. נסתכל על G_n, גרף מקרי על n קודקודים כך שעבור כל קשת אפשרית בין שני קודקודים עורכים הגרלה נפרדת כדי להחליט עם היא תשתתף בגרף או לא, וההסתברות לשתף אותה היא p.

מודל זה מעלה באופן טבעי את השאלות הבאות.

- בהינתן גרף מסויים H, כמה עותקים של H מופיעים באופן טיפוסי כתת-גרפים של G?
- מה ההסתברות ש-G קשיר? מה ההסתברות שהוא חסר מעגלים?

כאשר מכלילים את המודל כך שמובטח שכל הקשתות שתבחרנה תהייה מתוך תת גרף מסויים של $\overline{G} \subset K_n$, מתקבל מודל המכונה מודל החלחול (percolation model). למשל נוכל לבחור את הגרף להיות גרף השכנות בין קודקודים בריבוע $n \times n$ בשריג השלם $\mathbb{Z}^2$ כך שכל קודקוד מחובר לכל קודקוד שנבדל ממנו בקואורדינטה אחת בערך אחד. במקרה כזה נוכל "לשאול מה ההסתברות שצדו הימני של הריבוע מחובר לצדו השמאלי¿'. מודל זה מהווה מודל חשוב במכאניקה סטטיסטית, התורה הפיסיקלת של החומרים, כיוון שהוא מסביר תופעות שונות בתחום החלחול, ההולכה החשמלית, הולכת החום והולכת הקול.

חלק I

הסתברות בדידה

פרק 1

הסתברות בדידה

’אני סבורה שלפחות אחד מעשרים היה שם לב.’
’ואילו לדעתי את מניחה את מה שצריך להוכיח.’ אמר היידוק, ’ואני כבר יכול לראות איך אורב לי אחד התרגילים הללו בהסתברות על שישה ברנשים בעלי כובע לבן ושישה בעלי כובע שחור אשר עליך לעשות את החשבון עד כמה סביר שכובעיהם יתערבבו ובאיזה יחס. אם תתחילי לחשוב על דברים מסוג זה תצאי מדעתך. את זה אני יכול להבטיח לך!’

– אָגַתָה כְּרִיסְטִי, תעלומת הרצח המשולש.

1.1 מרחב מדגם

כפי שתיארנו במבוא, משמעות המושג **הסתברות** (probability), נגזרת ממושג ה**ניסוי** (experiment), כלומר תהליך בדיקה מתוכנן שבסופו מתקבלת תוצאה. ניסוי כזה יכול לתאר תוצאת הגרלת לוטו, בדיקת מין הילוד במהלך הריון או מדידת גובהו של אדם. הצעד הראשון בתיאור ניסוי שכזה במונחי תורת ההסתברות הוא אפיון קבוצת כל התוצאות האפשריות, קבוצה זו מכונה - **מרחב מדגם** (sample space). על פי רוב נסמן את מרחב המדגם ב-Ω ואת איבריו ב-ω. את אוסף התת-קבוצות של Ω נסמן ב-$\mathcal{F} = 2^{|\Omega|}$ ונכנהו בשם **אוסף המאורעות המדידים** (measurable events). השאלות ההסתברותיות שנשאל את עצמנו תהיינה מן הצורה ”מה ההסתברות שהמאורע A מתרחש?“ עבור $A \in \mathcal{F}$, כלומר ”מה ההסתברות שתוצאת הניסוי היא אחת מן התוצאות החברות ב-A?“. את היחס $\omega \in E$ עבור $E \in \mathcal{F}$ נבטא אפוא במילים המאורע E **מתרחש** עבור ω.

לכל תוצאה אפשרית ב-Ω נשייך בהמשך הסתברות – מספר ממשי בקטע $[0, 1]$, כאשר הסתברותו של מאורע $A \in \mathcal{F}$-ב תתקבל כסכום הסתברויות התוצאות החברות ב-A. לפי הגדרותינו אין כל מניעה שחלק מאוספי התוצאות לא יוכלו להתקבל כלל ועל כן הסתברותם תהיה 0. מסיבה זו, לאותו הניסוי עשויים להתאים מספר מרחבי הסתברות שונים. תמיד נוכל להחליף מרחב מדגם מסויים בכל מרחב מדגם אחר שמכיל אותו.

בדוגמאות שלהלן, כאשר נתבקש לתאר מרחב מדגם עבור ניסוי מסויים, נבחר מרחב פשוט לתיאור ולניתוח ככל האפשר. כיוון שבכל הדוגמאות מרחב המאורעות הוא $\mathcal{F} = \{A \ : \ A \subset \Omega\} = 2^{\Omega}$, עד מהרה נחדל לתארו במפורש.

דוגמא 1.1 (התאמת מרחב הסתברות בדיד לניסוי)**.**

(א) **ניסוי:** הטלת מטבע ורישום התוצאה.

מרחב מדגם: קבוצה בת שני איברים: עץ (H) או פלי (T). כלומר - $\Omega = \{H, T\}$.

מרחב מאורעות: כל התת-קבוצות של Ω, כלומר $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$.

(ב) **ניסוי:** הטלת מטבע שלוש פעמים ורישום תוצאות כלל ההטלות.

מרחב מדגם: מכפלה קרטזית של שלושה מרחבי מדגם של הטלת מטבע יחידה. כלומר:

$$\Omega = \{(a_1, a_2, a_3) \ : \ a_i \in \{H, T\}\} = \{H, T\}^3,$$
$$\mathcal{F} = \{A \ : \ A \subset \Omega\}.$$

(ג) **ניסוי:** הטלת שתי קוביות זהות (כלומר כאלה שלא ניתן להבחין ביניהן) ורישום צמד התוצאות כזוג לא סדור.

מרחב מדגם: מרחב המדגם הקטן ביותר האפשרי הוא $\Omega = \{(i, j) \ : \ 1 \le i \le j \le 6\}$, אולם, נוכל להשתמש גם במרחב $\Omega = \{(a_1, a_2) \ : \ a_i \in \{1, 2, 3, 4, 5, 6\}\} = [6]^2$ המתייחס לשתי קוביות נבדלות. מרחב מדגם זה מכיל יותר מידע מן הדרוש לקביעת תוצאת הניסוי, אך בהמשך נגלה כי הוא נוח יותר לעיבוד מתמטי.

(ד) **ניסוי:** הטלת מטבע שוב ושוב, אינספור פעמים ורישום הפעם הראשונה שבה התקבלה תוצאה של פלי.

מרחב מדגם: מספר ההטלות עד לקבלת פלי, כלומר $\Omega = \mathbb{N} \cup \{\infty\}$. בה-במידה יכולנו לבחור את מרחב המדגם $\Omega = \{H, T\}^{\mathbb{N}}$ ולשוות בנפשנו שאנו ממשיכים להטיל את המטבע עד אינסוף, אלא שמרחב זה אינו בן מניה (למעשה עוצמתו היא עוצמת קבוצת החזקה של $\mathbb{N}$), ולכן לא נתקשה לטפל בו בשלב זה.

(ה) **ניסוי:** מטבע מוטל. אם התוצאה היא עץ, מוטלת קוביה ואם פלי – מטבע מוטל שוב ושוב עד לקבלת פלי.

מרחב מדגם:

$$\Omega = \{(H, a) : a \in [6]\} \cup \{(T, a) : a \in \mathbb{N} \cup \{\infty\}\} = (\{H\} \times [6]) \cup (\{T\} \times (\mathbb{N} \cup \{\infty\})).$$

מרחב זה משקף שתי תצורות של תוצאות ולכן הוא מיוצג על ידי איחוד של שתי קבוצות בעלות תיאור שונה זו מזו. גם כאן נוכל לבחור במרחב פשוט יותר בדמות $\{H, T\} \times (\mathbb{N} \cup \{\infty\})$ וזאת מפני שאין כל מניעה שמרחב המדגם יכיל תוצאות שלעולם אינן מתקבלות.

1.2 הסתברות ומרחב הסתברות

יהי Ω מרחב מדגם בדיד המתאר את תוצאתו של ניסוי מסויים. נבקש לתאר מהי הסתברותה של כל תוצאה שעשויה להתקבל בניסוי. נשוב ונזכיר כי מבחינה רעיונית ניתן לחשוב על מספר זה בתור שכיחותה הטיפוסית של כל תוצאה במדגם גדול של ניסויים חוזרים שאין ביניהם קשר.

הגדרה 1.2. ***פונקציית ההסתברות הנקודתית*** (probability mass function) היא פונקציה $p : \Omega \to \mathbb{R}_+$ המקיימת

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

כאשר את הסכום בהגדרה נפרש לאור הגדרה א.13, כך שמובלעת כאן הדרישה כי p תקבל ערכים חיוביים ממש רק על קבוצה בת-מניה של ערכים. נרצה להכליל את פונקציית ההסתברות הנקודתית לקבלת **פונקציית הסתברות** על מרחב המאורעות המדידים $\mathcal{F} = 2^{\Omega}$, כלומר פונקציה $\mathbb{P} : \mathcal{F} \to [0,1]$ אשר תגדיר לכל מאורע את הסתברותו. זאת נעשה באמצעות הכלל המעניק לכל $A \in \mathcal{F}$ את סכום הסתברויות אברי מרחב המדגם המוכלים בה, כלומר:

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \tag{1.1}$$

כדי שהגדרתנו תחול בהמשך גם על מרחבי הסתברות כללים יותר, לא נגדיר את פונקציית ההסתברות ישירות באמצעות בניה זו. תחת זאת נדרוש מ-$\mathbb{P}$ לקיים מספר תכונות ונסיק כי היא מקיימת את (1.1) (ר' טענה 1.5 להלן).

הגדרה 1.3. פונקציית הסתברות היא פונקציה $\mathbb{P} : \mathcal{F} \to \mathbb{R}_+$ המקיימת

(א) $\mathbb{P}(\Omega) = 1$.

(ב) **סכימות בת-מניה (σ-additivity):** לכל אוסף בן-מניה של מאורעות זרים $\{A_i\}_{i\in\mathbb{N}}$ ב-$\mathcal{F}$ מתקיים

$$\mathbb{P}\left(\dot{\bigcup_{i\in\mathbb{N}}} A_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

לאורך כל חלק I של הספר, נניח כי מתקיים גם התנאי

$$\sum_{\omega\in\Omega} \mathbb{P}(\omega) = \mathbb{P}(\omega), \tag{1.2}$$

ובפרט כי הקבוצה $\{\omega : \mathbb{P}(\omega) > 0\}$ היא בת-מניה. כדי לציין זאת נאמר כי פונקציית ההסתברות היא **פונקציית הסתברות בדידה**.

מתכונותיה של פונקציית הסתברות ניתן לגזור מיד מספר תכונות שימושיות.

טענה 1.4 (תכונות של פונקציית הסתברות)**.** יהי Ω מרחב מדגם ותהי $\mathbb{P}$ פונקציית הסתברות. מתקיימות התכונות הבאות:

(א) $\mathbb{P}(\emptyset) = 0$.

(ב) **סכימות (additivity):** לכל אוסף סופי של מאורעות זרים, $\{A_i\}_{i\in[N]}$, מתקיים

$$\mathbb{P}(\dot{\bigcup_{i\in[N]}} A_i) = \sum_{i\in[N]} \mathbb{P}(A_i).$$

(ג) **מונוטוניות (monotonicity):** עבור $A, B \in \mathcal{F}$ אם $A \subset B$ אז $\mathbb{P}(A) \le \mathbb{P}(B)$.

(ד) לכל $A \in \mathcal{F}$ מתקיים $\mathbb{P}(A) \le 1$.

(ה) לכל $A \in \mathcal{F}$ מתקיים $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$, כאשר $A^c := \Omega \setminus A$ הוא המאורע המשלים של A.

הוכחה. **תכונה (א)** נובעת מסכימות בת-מניה עבור $A_i = \emptyset$ לכל $i \in \mathbb{N}$. **תכונה (ב)** נובעת מסכימות בת-מניה ומתכונה (א) כאשר מציבים $A_i = \emptyset$ לכל $i > N$. להוכחת **תכונה (ג)** נשתמש בתכונה (ב) על הקבוצות A, $B \setminus A$ לקבלת $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. מהאי-שליליות של $\mathbb{P}$ נסיק כי $\mathbb{P}(B) \geq \mathbb{P}(A)$ כנדרש.
תכונה (ד) נובעת ישירות מהגדרת $\mathbb{P}$. **תכונה (ה)** נובעת מהפעלת תכונה (ב) על A ו-A^c ומתכונה (1.3א). ■

כפי שנראה מיד, פונקציית ההסתברות הנקודתית קובעות את פונקציית ההתפלגות ביחידות.

טענה 1.5. יהי Ω מרחב מדגם, תהי p פונקציית הסתברות נקודתית ותהי $\mathbb{P}$ הפונקציה המקיימת לכל $A \in \mathcal{F}$

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega).$$

אזי $\mathbb{P}$ הנה פונקציית הסתברות, והיא היחידה המקיימת $\mathbb{P}(\{\omega\}) = p(\omega)$ לכל $\omega \in \Omega$.

הוכחה. כדי להיווכח כי $\mathbb{P}$ מקיימת את **תכונה (1.3א)**, נציב בהגדרתה את התכונה המאפיינת של פונקציית ההסתברות הנקודתית:

$$\mathbb{P}(\Omega) = \sum_{\omega \in \Omega} p(\omega) \overset{\substack{\text{הגדרה} \\ 1.2}}{=} 1.$$

לצורך הוכחת **תכונה (1.3ב)**, תהיינה $\{A_i\}_{i \in \mathbb{N}}$ קבוצות זרות ב-$\mathcal{F} = 2^{\Omega}$. התכונה נובעת מהגדרת פונקציית ההסתברות באמצעות החישוב הבא:

$$\mathbb{P}\left(\dot{\bigcup_{i \in \mathbb{N}}} A_i\right) = \sum_{\omega \in \dot{\bigcup}_{i \in \mathbb{N}} A_i} p(\omega) = \sum_{i \in \mathbb{N}} \sum_{\omega \in A_i} p(\omega) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i),$$

כאשר השוויונים הקיצוניים נובעים מהגדרת הפונקציה $\mathbb{P}$ והשוויון האמצעי נובע מכך שטור של מספרים אי-שליליים מתכנס בהחלט (במובן הרחב) ולכן אינו מושפע משינוי סדר סכימה (טענה א.12). כדי לראות ש-$\mathbb{P}$ היא **היחידה שמקיימת** $\mathbb{P}(\{\omega\}) = p(\omega)$, נשים לב שלכל $\mathbb{P}'$, פונקציית הסתברות אחרת שמקיימת $\mathbb{P}'(\omega) \equiv p(\omega)$ זו ולכל $A \in \mathcal{F}$ מתקיים

$$\mathbb{P}'(A) = \mathbb{P}\left(\dot{\bigcup_{\omega \in A}} \{\omega\}\right) \overset{\substack{\text{טענה} \\ 1.4\text{ב}}}{=} \sum_{\omega \in A} \mathbb{P}'(\{\omega\}) = \sum_{\omega \in A} p(\omega) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}) \overset{\substack{\text{טענה} \\ 1.4\text{ב}}}{=} \mathbb{P}\left(\dot{\bigcup_{\omega \in A}} \{\omega\}\right) = \mathbb{P}(A).$$

■

כדי לצמצם את מספר הסימונים, נימנע מעתה משימוש בפונקציית ההסתברות הנקודתית $p(\omega)$ ותחת זאת, לאור טענה 1.5, נשתמש בפונקציית ההסתברות הכללית $\mathbb{P}(\{\omega\})$ לתיאור הסתברותה של תוצאה במרחב המדגם. כמו כן, נוכח טענה 1.5 נוכל לאפיין את פונקציית ההסתברות על ידי תיאור ההסתברות שהיא מעניקה ליחידונים $\{\omega\}$ תוך שנוודא כי $\sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) = 1$ (כלומר כי פונקציית ההסתברות הנקודתית מקיימת את תנאי הגדרה 1.2).

הערה: על מרחבי מדגם שאינם בני-מניה, קיימות גם פונקציות הסתברות שאינן בדידות. הגדרתן של פונקציות אלה והטיפול בהן חורגים מתחומו של ספר זה מפני שהם דורשים בקיאות בתורת המידה. אף על פי כן, סקירה ראשונית של ההגדרות המכלילות את מרחב ההסתברות מעבר למקרה הבדיד מופיעה בפרק 7.

השלשה (פונקציית הסתברות, אוסף מאורעות מדידים, מרחב מדגם) מכונה - **מרחב הסתברות**.

הגדרה 1.6. יהי Ω מרחב מדגם, תהי $\mathcal{F} = 2^{\Omega}$, קבוצת כל התת-קבוצות של Ω, ותהי $\mathbb{P}$ פונקציית הסתברות בדידה על $\mathcal{F}$. השלשה $(\Omega, \mathcal{F}, \mathbb{P})$ מכונה - **מרחב הסתברות בדידה**.

להלן מספר דוגמאות מפתח למרחבי הסתברות.

דוגמא 1.7 (ניסוי ברנולי)**.** ייצוגו התיאורטי של ניסוי **הטלת מטבע הוגן** מכונה גם **ניסוי ברנולי עם פרמטר** $\frac{1}{2}$. מרחב המדגם של ניסוי זה הוא $\Omega = \{H, T\}$ (ר. דוגמא 1.1א), ולכן $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$. כיוון שהמטבע הוגן, הרי שפונקציית ההסתברות הנקודתית צריכה לתת לכל תוצאה הסתברות שווה, כלומר

$$p(\{H\}) = p(\{T\}) = \frac{1}{2}$$

והיא מוכללת לפונקציית ההסתברות

$$\mathbb{P}(\{H, T\}) = 1, \mathbb{P}(\emptyset) = 0, \mathbb{P}(\{H\}) = \mathbb{P}(\{T\}) = \frac{1}{2}.$$

השלשה $(\Omega, \mathcal{F}, \mathbb{P})$ מגדירה את מרחב ההסתברות של הניסוי. אולם, לא כל המטבעות הוגנים וניתן לשאול מה הוא מרחב כל ניסויי הטלת המטבעות האפשריים. לצורך כך נגדיר את ניסוי **הטלת מטבע מוטה** עם פרמטר $\alpha \in [0, 1]$ המכונה גם **ניסוי ברנולי עם פרמטר** α. שני הניסויים מתרחשים באותו מרחב המדגם, אלא שבהטלת מטבע מוטה פונקציית ההסתברות מעניקה ערך α לתוצאה "עץ", כלומר -

$$\mathbb{P}(\{H\}) = \alpha$$

כיוון שהתכונה היסודית של פונקציית הסתברות האטומית מחייבת אותנו לקיים

$$\mathbb{P}(\{H\}) + \mathbb{P}(\{T\}) = 1,$$

יתר ערכי הפונקציה נקבעים ביחידות:

$$\mathbb{P}(\{T\}) = 1 - \alpha, \mathbb{P}(\{H, T\}) = 1, \mathbb{P}(\emptyset) = 0.$$

יעקב ברנולי (1655-1705), מתמטיקאי שווייצרי שתרם תרומה מכרעת לענפים מגוונים של המתמטיקה. בספרו "אומנות ההשערה" (Ars Conjectandi), שיצא לאור תשע שנים לאחר מותו, טבע את יסודות תורת ההסתברות, וקישר בינה לבין הבנת הטבע על ידי חקר התפלגות תוצאותיו של ניסוי חוזר בעל שתי תוצאות הקרוי על שמו. הראשון לבסס את הקשר בין הסתברות למחקר מדעי באמצעות חוק המספרים הגדולים (ר' פרק 5).

נשים לב שכאשר $\alpha \in \{0,1\}$ הניסוי הופך **דטרמיניסטי**, כלומר רק תוצאה אחת מקבלת הסתברות חיובית. מצב זה משקף ניסוי שאין בו כל אקראיות. מעתה **ניסוי ברנולי**, ניסוי הטלת המטבע (ההוגן או המוטה), יהפוך בלשוננו למושג מתמטי, ובמקום לתאר התרחשות בעולם הפיזי - יתאר את מרחב ההסתברות שתואר לעיל. תהליך זה של העתקת מושגים מהמרחב הגשמי למרחב המתמטי המופשט נקרא מידול והוא מעניינם של מדעי הטבע. כמתמטיקאים, אנו מקבלים את המודל המוגמר ונדרשים לנתח אותו. איננו נדרשים לשאלה האם המודל עצמו משקף את המציאות.

מכאן ואילך לא נתאר מפורשות את $\mathcal{F}$ עבור מרחבי הסתברות בדידים, אלא נסתפק בכתיב $\mathcal{F} = 2^{\Omega}$.

דוגמא 1.8 (הטלת קוביה). ניסוי **הטלת קוביה הוגנת** מיוצג באופן תיאורטי על ידי מרחב המדגם $\Omega = [6]$. הרצון לייצג קובייה הוגנת, כלומר כזו שהסתברות שווה ניתנת לכול תוצאותיה לא מותיר לנו אלא פונקציית הסתברות אטומית יחידה הנתונה על ידי
$$\mathbb{P}(\{n\}) = \frac{1}{|\Omega|} = \frac{1}{6}$$
לכל $n \in [6]$.

דוגמא 1.9 (שליפת כדורים ממוספרים מכד). בניסוי זה שולפים בזה אחר זה k כדורים מתוך כד אטום שבו $k \leq n$ כדורים הממוספרים מ-1 ועד n, ללא החזרה. כל איבר במרחב ההסתברות הוא k-יה סדורה של כדורים, והסתברות כל k-יה זהה. באופן פורמלי:

$$\Omega = \{(a_1,\ldots,a_k) \ : \ a_i \in [n], \forall i,j \in [k], i \neq j \Rightarrow x_i \neq x_j\},$$
$$\mathbb{P}(\{(a_1,\ldots,a_k)\}) = \frac{1}{n\cdot(n-1)\cdots(n-k+1)}.$$

דוגמא 1.10. נסמן $\sum_{n\in\mathbb{N}} \frac{1}{n^2} = a$ (נשים לב כי זהו טור מתכנס, למשל לפי מבחן האינטגרל על הפונקציה $\frac{1}{x^2}$ (טענה א.16)). נתון מרחב הסתברות $\Omega = \mathbb{N}$ ופונקציית הסתברות המוגדרת על ידי

$$\mathbb{P}(\{n\}) = a^{-1}n^{-2}.$$

נבקש לחשב את הסתברותו של המאורע

$$2\mathbb{N} = \{2,4,6,\ldots\}.$$

תשובה: נשים לב כי $\mathbb{P}(\Omega) = a^{-1}\sum_{n\in\mathbb{N}} \frac{1}{n^2} = 1$, ולכן פונקציית ההסתברות מוגדרת כהלכה. כעת נסמן $A = 2\mathbb{N}$ ונחשב לפי סכימות בת-מניה של מרחב הסתברות

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\{2n\}\right) = \sum_{n=1}^{\infty}\frac{1}{a(2n)^2} = \frac{1}{4a}\sum_{n=1}^{\infty}\frac{1}{n^2} = \frac{1}{4}.$$

הערה: מציאת ערכו המדוייק של a בדוגמא לעיל הנה בעיה קלאסית בעלת חשיבות לתורת המספרים. בעיה זו מכונה בעיית באסל (Basel problem) על שם עיר הולדתו של לאונרד אוילר, והיא הוצגה בשנת 1644 על ידי פטרו מנגולי ועמדה פתוחה כ-90 שנה עד אשר נפתרה על ידי אוילר שחישב ומצא כי $a = \frac{\pi^2}{6}$.

לכל בעיה ניתן לבחור מגוון רחב של מרחבי מדגם ומרחבי הסתברות. גם כאשר קיים מרחב הסתברות מצומצם, נעדיף לעיתים לבחור מרחב הסתברות גדול אשר שעליו פונקציית ההסתברות המתאימה לניסוי פשוטה. במיוחד יהיו חביבים עלינו **מרחבי הסתברות אחידים** המוגדרים להלן.

הגדרה 1.11 (מרחב הסתברות אחיד)**.** מרחב הסתברות בדידה $(\Omega,\mathcal{F},\mathbb{P})$ נקרא **אחיד**, אם לכל $\omega_1,\omega_2\in\Omega$ מתקיים

$$\mathbb{P}(\{\omega_1\})=\mathbb{P}(\{\omega_2\})$$

היות ש- $\sum_{\omega\in\Omega}\mathbb{P}(\{\omega\})=1$, הרי שבמרחב אחיד לכל $\omega\in\Omega$ מתקיים $\mathbb{P}(\{\omega\})=\frac{1}{|\Omega|}$. מכאן אנו מסיקים שבמרחב הסתברות אחיד את הסתברותו של כל מאורע $A\in\mathcal{F}$ ניתן לחשב באמצעות חישוב קומבינטורי פשוט:

$$\mathbb{P}(A)=\frac{|A|}{|\Omega|}. \tag{1.3}$$

סימטריה ומרחבים אחידים: במקרים רבים נייחס לבעיה מרחב הסתברות אחיד כאשר תתקיים סימטריה בין כל התוצאות. למשל כאשר נטיל מספר קוביות אשר ניתן להבחין ביניהן, נטען שעל ידי חילוף הכיתוב על פאותיהן נוכל להמיר כל תוצאה בתוצאה אחרת מבלי שנשפיע כל הסתברותה ולכן לכל התוצאות צריכה להיות הסתברות שווה. במקרים אלו נשתמש בביטוי "מטעמי סימטריה".

דוגמא 1.12 (שאלות הסתברותיות על מרחבים בדידים)**.**

(א) שתי קוביות הוגנות מוטלות. מה הסיכוי שסכומן הוא 7?

תשובה: לתיאור מרחב ההסתברות נתאר את זוגות התוצאות של שתי קוביות שניתן להבחין ביניהן. כלומר $\Omega=[6]^2$. המרחב אחיד משיקולי סימטריה ולכן לכל $i\in\Omega$ נגדיר $\mathbb{P}(\{i\})=1/36$. נסמן ב- A את המאורע "תוצאת ההטלה היא 7". ישנם בדיוק שישה צירופי תוצאות הנותנים את הערך 7, כלומר

$$A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\},$$

ומכאן, $\mathbb{P}(A)=|A|/|\Omega|=6/36$. נשים לב שאילו בחרנו שלא להבחין בין הקוביות, ולעבוד עם מרחב ההסתברות המתאים לכל זוגות התוצאה ללא חשיבות לסדר, או אילו בחרנו לעבוד במרחב ההסתברות שבו כל נקודה מתארת את סכום תוצאות הקוביות, היה החישוב הופך מורכב יותר וקשה להצדקה, והיינו מתקשים לפתור בעיות אחרות העוסקות באותו הניסוי. מכאן אנו למדים שמרחב ההסתברות הקטן ביותר, אינו בהכרח גם הטוב ביותר. □

(ב) בכד 11 כדורים, שישה לבנים וחמישה שחורים. שני כדורים נשלפים באקראי. מה ההסתברות לשלוף בדיוק כדור אחד לבן ואחד שחור?

תשובה: נשתמש במרחב ההסתברות של דוגמא 1.9 - שליפת 2 כדורים מכד שבו 11 כדורים ממוספרים. הכדורים הממוספרים במספרים 1,2,3,4,5,6 ייצגו את הכדורים הלבנים והיתר את הכדורים השחורים. כלומר

$$\Omega=\{(a_1,a_2)\ :\ a_1,a_2\in[11],a_1\neq a_2\},$$
$$\mathbb{P}(\{(a_1,a_2)\})=\frac{1}{11\cdot 10}.$$

שוב מדובר במרחב הסתברות אחיד. נסמן את המאורע שאנו מבקשים לחשב את הסתברותו בA כלומר

$$A=\Big(\{1,\ldots,5\}\times\{6,\ldots,11\}\Big)\cup\Big(\{6,\ldots,11\}\times\{1,\ldots,5\}\Big)$$

ישנן חמש אפשרויות לשלוף כדור שחור בשליפה ראשונה, ולכל אחת מהן שש אפשרויות לשלוף כדור לבן בשליפה שניה. לכל אפשרות כזו מתאימה אפשרות אחת של שליפות הכדורים בסדר הפוך ובסך הכל $2\cdot5\cdot6=60$ אפשרויות היות שמדובר במרחב הסתברות אחיד נקבל $\mathbb{P}(A)=\frac{60}{110}=\frac{6}{11}$. □

(ג) חפיסה של 52 קלפי משחק רגילים מחולקת לארבעה שחקנים באקראי, 13 קלפים לכל שחקן, מה ההסתברות שאחד השחקנים יקבל את כל היהלומים?

תשובה: מרחב ההסתברות הטבעי ביותר לבעיה הוא אוסף כל החלוקות של 52 קלפים ל-4 שחקנים. כלומר $\Omega = \left\{f : [52] \to [4] \ : \ \forall i \in [4] \quad f^{-1}(i) = 13\right\}$ חישוב קומבינטורי פשוט נותן כי

$$|\Omega| = \binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{13!\,13!\,13!\,13!}$$

שכן עלינו לבחור שלושה עשר קלפים עבור השחקן הראשון, ולאחר מכן שלושה עשר קלפים עבור השחקן השני וכן הלאה. מפאת הסימטריה בין החלוקות בחלוקה אקראית, זהו מרחב הסתברות אחרי אשר פונקציית ההסתברות האטומית שלו היא $f \in \Omega$

$$\mathbb{P}(\{f\}) = \frac{13!\,13!\,13!\,13!}{52!}.$$

אם נניח ש13 הקלפים המתאימים ליהלומים הם הקלפים הממוספרים מ-1 ועד 13, אז המאורע בו אנו מעוניינים הוא $A = \{f \in \Omega \ : \ \forall i, j \leq 13 \quad f(i) = f(j)\}$. המאורע A מורכב למעשה מארבעה מאורעות זרים ושווי גודל. עבור $k \in [4]$ נגדיר

$$A_k = \{f \in \Omega \ : \ f(i) = k \ \forall i \leq 13\}$$

וניווכח שאומנם

$$A = A_1 \uplus A_2 \uplus A_3 \uplus A_4, \qquad |A_1| = |A_2| = |A_3| = |A_4|,$$

ולכן

$$\mathbb{P}(A) = 4 \cdot \mathbb{P}(A_1)$$

חזרה על אותו חישוב קומבינטורי תגלה לנו כי

$$|A_1| = \binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{39!}{13!\,13!\,13!},$$

ומכאן

$$\mathbb{P}(A) = 4\,\mathbb{P}(A_1) = 4 \cdot \binom{52}{13}^{-1} \approx 6.3 \times 10^{-12}.$$

□

✎ **בעיה 1.1.** להוכיח כי לא קיים מרחב הסתברות אחיד על $\mathbb{Z}$, קבוצת המספרים השלמים.

✎ **בעיה 1.2.** מחלקים כובע שחור או לבן לכל אחד מחמישה ברנשים העומדים בשורה. חלוקת הכובעים נבחרת באופן אחיד מבין כל החלוקות האפשרויות. מה ההסתברות שישנם שני ברנשים סמוכים החובשים כובע שחור?

כבר באמצעות הכלים התיאורטיים הדלים שהצגנו ניתן להוכיח טענות הסתברויות אשר עשויות לסתור את האינטואיציה הטבעית. הדוגמא הבאה היא אחת מן המפורסמות שבטענות אלה אשר הוצגה בסוגיה 2.

✵ **דוגמא 1.13** (פרדוקס יום ההולדת).
נניח שלאדם מקרי סיכוי שווה להיוולד בכל אחד מימות השנה הלועזית הבלתי מעוברת 2001. מה ההסתברות שבין N אנשים שנולדו באותה שנה אין שניים שנולדו באותו תאריך?

תשובה: מרחב ההסתברות הנוח ביותר לפתרון הבעיה הוא מרחב של n הטלות של קובייה בת 365 ימים. כלומר

$$\Omega = [365]^n \qquad \mathbb{P}\Big(\{(x_1, \ldots, x_n)\}\Big) = 365^{-n},$$

לכל $(x_1, \ldots, x_n) \in \Omega$. בחרנו במרחב זה משום שהמרחב המצומצם יותר שבו אין חשיבות לסדר האנשים איננו מרחב הסתברות אחיד, ולכן יקשה על חישובינו.

המאורע שאנו מעוניינים בו הוא המאורע

$$A_N = \Big\{(x_1, \ldots, x_n) \in \Omega \ : \ \forall i, j \in [n], i \neq j \Rightarrow x_i \neq x_j\Big\}$$

כיוון שמדובר במרחב אחיד, די לחשב את היחס $|A_n|/|\Omega|$. כלומר

$$\mathbb{P}(A_n) = \frac{|A_n|}{|\Omega|} = \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

חישוב באמצעות מחשב מגלה כי

$$\mathbb{P}(A_{23}) \approx 0.5 \qquad \mathbb{P}(A_{50}) \approx 0.03 \qquad \mathbb{P}(A_{70}) \approx 1/1000.$$

כלומר די בקבוצה קטנה יחסית של אנשים כדי שבסיכוי לשניים מהם יהיה אותו יום-הולדת. במציאות הסיכוי לקיום זוג כזה גדול יותר מפני שהתפלגות ימי-הולדת לאורך השנה אינה אחידה. □

הסבר, ניתוח והכללות:

רבים מוצאים את העובדה שדי ב-23 אנשים בכדי שבסבירות חצי לפחות לשניים מהם יהיה אותו יום-הולדת, מפתיעה, לפחות במבט ראשון. זאת מפני שאם נשאל מה ההסתברות של כל אחד מאיתנו, כאשר הוא נכנס לחדר שבו 23 אנשים למצא איש נוסף שזהו יום הולדתו, הרי שהסתברות זו היא בדיוק ההסתברות המשלימה למאורע שכל יתר האנשים נולדו בימי הולדת שונים, ולכן ניתן לחשב כי

$$\mathbb{P}(\text{ מישהו בחדר נולד באותו יום כמותי }) = 1 - \left(\frac{365}{366}\right)^2 3 \approx 6\%$$

והלא זו הסתברות נמוכה למדי.

העובדה שהשכל הישר נוטה להתעלם ממנה הוא שמספר **זוגות האנשים** בחדר הוא $\binom{n}{2}$, מספר גדול בהרבה ממספר האנשים, ודי **בזוג** אחד של אנשים שנולדו באותו יום.

לצורך הבנה כללית יותר של התופעה, נכליל את הבעיה לשנה בעלת אורך כללי. כלומר, נסתכל על ניסוי בו מוגרלים n איברים שכל אחד מהם בנפרד מקבל אחד מקרי מבין הערכים 1 עד m עבור $m \in \mathbb{N}$, כך שלכל אחד מהצירופים אותה הסתברות. כלומר

$$\Omega = [m]^n, \mathbb{P}\Big(\{(x_1, \ldots, x_n)\}\Big) = m^{-n},$$

נסמן ב-A_n^m את המאורע

$$A_n^m = \Big\{(x_1,\dots,x_n)\in\Omega \;:\; \forall i,j\in[n], i\neq j \Rightarrow x_i\neq x_j\Big\},$$

ברצוננו לקבל מושג לגבי ההתנהגות האסימפטוטית של הניסוי כתלות בערכי n ו-m. קל לחשב כי

$$\mathbb{P}(A_n^m) = \prod_{i=0}^{n-1}\Big(1-\frac{i}{m}\Big).$$

קשה לנתח את התנהגותן של מכפלות גדולות ולכן נחקור את הלוגריתם של הסתברותו של A_n^m,

$$\log\mathbb{P}(A_n^m) = \sum_{i=0}^{n-1}\log\Big(1-\frac{i}{m}\Big).$$

כדי לחסום את ההסתברות מלמעלה, ניזכר כי $\log(1+x)\leq x$ לכל x עבורו אגף שמאל מוגדר היטב (הקורא מוזמן להיווכח בכך, למשל באמצעות שארית לגרנג' לפולינום טיילור (טענה א.17)). לכן,

$$\log\mathbb{P}(A_N^M) \leq -\sum_{i=0}^{N-1}\frac{i}{M} \overset{\text{סדרה חשבונית}}{=} -\frac{N(N-1)}{2M},$$

$$\mathbb{P}(A_N^M) \leq \exp\Big(-\frac{N(N-1)}{2M}\Big).$$

כדי לחסום את ההסתברות מלמטה נוודא באמצעות שארית לגרנג' כי $x-x^2\leq\log(1+x)$ לכל $x\geq -\frac{1}{4}$. לכן, כל עוד $N\leq M/4$, מתקיים

$$\begin{aligned}\log\mathbb{P}(A_N^M) &\geq -\frac{\sum_{i=0}^{N-1} i}{M} - \frac{\sum_{i=0}^{N-1} i^2}{M^2} = -\frac{N(N-1)}{2M} - \frac{N(N-1)(2N-1)}{6M^2}\\ &= -\frac{N(N-1)}{2M}\Big(1+\frac{2N-1}{3M}\Big).\end{aligned}$$

כאשר המעבר השני הוא חישוב טורים סופיים מוכרים המוצגים בפרק א.4.5. מכאן, שלכל $N\leq M/4$ מתקיים,

$$\exp\Big(-\frac{N(N-1)}{2M}\Big(1+\frac{2N-1}{3M}\Big)\Big) \leq \mathbb{P}(A_N^M) \leq \exp\Big(-\frac{N(N-1)}{2M}\Big).$$

כאשר N קטן ביחס ל-M, החסמים קרובים למדי זה לזה. למשל, עבור $N>M/4$ מתקיים

$$\mathbb{P}(A_N^M) < \mathbb{P}(A_{M/4}^M) \leq \exp\Big(-\frac{M}{32}+\frac{1}{8}\Big),$$

ולכן מובטח לנו שעבור N גדול ההסתברות קטנה להפליא. מכאן שכאשר N קטן יחסית ל-$\sqrt{M}/2$ ההסתברות שהוגרלו זוג איברים זהים אפסית וכאשר N גדול יחסית ל-$\sqrt{M}/2$ הסתברות זו כמעט ודאית. תופעה זו מכונה מעבר חד (sharp transition). בדוגמא זו ראינו לראשון ניתוח מלא ומפורט של התנהגות אסימפטוטית של שאלה הסתברותית. □

1.3 מכפלת מרחבי-הסתברות

בדוגמאות שסקרנו עד כה, הרבנו לעיין בתוצאת ניסויים חוזרים אשר נערכו בתנאים בהם אין כל קשר בין תוצאותיהם. את מרחבי ההסתברות שהתאמנו למקרים כאלו נוכל לכנות בשם **מכפלה קרטזית של מרחבי ההסתברות**. נגדיר פעולה זו באופן מדוייק.

הגדרה 1.14 (מרחב מכפלה)**.** יהיו $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ ו-$(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ מרחבי הסתברות בדידה. **מרחב המכפלה** של מרחבים אלו מוגדר בתור מרחב ההסתברות $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbb{P})$ עבור $\mathbb{P}$ המקיימת

$$\mathbb{P}(\{(\omega_1, \omega_2)\}) = \mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}),$$

לכל $\omega_1 \in \Omega_1$ ו-$\omega_2 \in \Omega_2$.

נשים לב כי מרחב זה מוגדר ביחידות, כיוון שקבענו את פונקציית ההסתברות הנקודתית, וכי הוא מוגדר היטב משום שמתכונותיהם של $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ ו-$(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ נקבל כי:

$$\begin{aligned}\mathbb{P}(\Omega_1 \times \Omega_2) &= \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2}\mathbb{P}(\{(\omega_1, \omega_2)\}) = \sum_{\omega_1\in\Omega_1}\sum_{\omega_2\in\Omega_2}\mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) \\ &= \sum_{\omega_1\in\Omega_1}\left(\mathbb{P}_1(\{\omega_1\})\sum_{\omega_2\in\Omega_2}\mathbb{P}_2(\{\omega_2\})\right) = \sum_{\omega_1\in\Omega_1}\mathbb{P}_1(\{\omega_1\}) = 1.\end{aligned}$$

הגדרה זו של מרחב מכפלה תואמת את המושג האינטואיטיבי של "ניסויים שאין קשר ביניהם", בכך שיחס ההסתברויות לקבל כל שתי תוצאות בניסוי אחד אינו מושפע מתוצאת הניסוי השני. באופן פורמלי - לכל שתי תוצאות אפשריות בניסוי הראשון, $\omega_1, \omega_1' \in \Omega_1$ המקיימות $\mathbb{P}_1(\{\omega_1\}) = a\mathbb{P}_1(\{\omega_1'\})$ ולכל $\omega_2 \in \Omega_2$ מתקיים $\mathbb{P}(\{(\omega_1, \omega_2)\}) = a\mathbb{P}(\{\omega_1', \omega_2\})$.

טענה 1.15 (מאורעות שוליים במרחב מכפלה)**.** יהיו $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ ו-$(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ מרחבי הסתברות בדידים ויהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב המכפלה שלהם. יהיו $A \in \mathcal{F}_1$ ו-$B \in \mathcal{F}_2$. אזי

$$\mathbb{P}(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B).$$

הוכחה. נשוב ונשתמש בסכימות בת מניה לקבלת:

$$\mathbb{P}(A \times B) = \sum_{\omega_1\in A}\sum_{\omega_2\in B}\mathbb{P}_1(\{\omega_1\})\mathbb{P}_2(\{\omega_2\}) = \sum_{\omega_1\in A}P_1(\{\omega_1\})\mathbb{P}_1(B) = \mathbb{P}_1(A)\mathbb{P}_2(B).$$

■

ניתן לכפול גם יותר משני מרחבי הסתברות כפי שמתואר בהגדרה הבאה.

הגדרה 1.16 (מרחב מכפלה של מספר מרחבים בדידים)**.** יהיו $\{(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)\}_{n\in[N]}$ מרחבי הסתברות בדידה. נגדיר את $(\Omega, \mathcal{F}, \mathbb{P})$, מרחב המכפלה של מרחבים אלו באמצעות $\Omega = \times_{n=1}^N \Omega_n$, $\mathcal{F} = \times_{n=1}^N \mathcal{F}_n$ תוך שנגדיר את $\mathbb{P}$ על ידי כך שעבור $\omega = (\omega_1, \ldots, \omega_n) \in \Omega$ נקבע

$$\mathbb{P}(\{(\omega_1, \omega_2, \ldots, \omega_N)\}) = \prod_{n=1}^N \mathbb{P}_n(\{\omega_n\}).$$

כדי לפשט את כתיבתם של מאורעות במרחב מכפלה נשתמש לעיתים בסימן $(\cdot)$ כדי לציין קואורדינטה שאינה משפיעה על שייכות למאורע מסויים. כך למשל במרחב $\Omega = [6]^2$ המתאר הטלת שתי קוביות, נוכל לסמן את המאורע שבקוביה הראשונה התקבלה תוצאה של 3 באמצעות הסימון $\{(3,\cdot)\}$ ואת המאורע שבקוביה השניה התקבלה תוצאה קטנה מ-4 באמצעות הסימון $\{(\cdot,a) \in \Omega \;:\; a<4\}$.

✎ **בעיה 1.3.** להוכיח כי במרחב מכפלה $\{(\Omega_n,\mathcal{F}_n,\mathbb{P}_n)\}_{n\in[N]}$, לכל סדרת מאורעות $\{A_n \in \mathcal{F}_n\}_{n\in[N]}$ מתקיים

$$\mathbb{P}\left(\bigtimes_{n\in N} A_n\right) = \prod_{n\in\mathbb{N}} \mathbb{P}_n(A_n).$$

מעתה באומרנו שאנו מטילים מספר מטבעות, זורקים מספר קוביות או חוזרים על ניסוי מספר פעמים - כוונתנו תהיה שמרחב ההסתברות המתאים לבעיה הוא מרחב מכפלה של מרחבי ההסתברות המתאימים לניסויים אלה.

✎ **בעיה 1.4.** בוחרים באקראי ארבעה ילדים מתוך שמונה בנות ושמונה בנים. מה ההסתברות שיבחרו שני בנים ושתי בנות?

✎ **בעיה 1.5.** שבעה-עשר ברנשים עוטי מצנפות עורכים מסיבה. בסופה, כל אחד מהם לוקח באקראי מצנפת ממתלה הכובעים. מה ההסתברות של כל אחד מהמאורעות הבאים:

1. לפחות אחד הברנשים עוטה את מצנפתו.
2. אריק, זקן הברנשים, עוטה את מצנפתו.
3. בדיוק שלושה ברנשים עוטים את מצנפתם.
4. בדיוק שלושה ברנשים ובהם אריק עוטים את מצנפתם

כתמיד, יש להתחיל באפיון מרחב ההסתברות.

✎ **בעיה 1.6.** קלפים נשלפים מחפיסה תקנית בזה אחר זה. מה ההסתברות שהקלף הארבעה-עשר יהיה אס? מה ההסתברות שהאס הראשון יהיה הקלף הארבעה-עשר? (בחפיסת קלפים תקנית חמישים ושניים קלפים ובהם ארבעה אסים).

1.4 חסם האיחוד

חסם בסיסי ושימושי מאוד בתורת ההסתברות הוא החסם הבא, המכונה **חסם האיחוד**, שקובע כי הסתברות איחוד מאורעות קטנה מסכום הסתברויותיהם.

משפט 1.17 (אי-שווין בול [Boole's Inequality]). יהי $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות. הסתברותם של מאורעות במרחב כזה היא סיגמה תת-חיבורית. כלומר לכל אוסף בן-מניה של מאורעות $\{A_n\}_{n\in\mathbb{N}}$ מתקיים,

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \le \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

הוכחה. נגדיר סדרת מאורעות

$$B_1 = A_1 \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus (A_1 \cup A_2), \quad \ldots, \quad B_n = A_n \setminus \left(\cup_{k=1}^{n-1} A_k\right).$$

כלומר B_n מורכב מכל אברי מרחב המדגם שנמצאים ב-A_n אך לא ב-$A_1, \ldots, A_{n-1}$. לפי הגדרה זו המאורעות B_n הנם זרים, כעת נראה כי איחודם שווה בדיוק לאיחוד המאורעות A_n. ראשית נראה כי

$$\bigcup_{k=1}^{\infty} A_k \subset \dot{\bigcup_{k=1}^{\infty}} B_k. \tag{1.4}$$

יהי $\omega \in \cup_{k=1}^{\infty} A_k$. הקבוצה $\{k \in \mathbb{N} \,:\, \omega \in A_k\}$ אינה ריקה ובפרט קיים בה איבר מינימלי k^*, כלומר $\omega \in A_{k^*}$ אך $\omega \notin A_k$ לכל $k < k^*$. מהגדרת הקבוצות B_n נובע אפוא כי $\omega \in B_{k^*}$. נסיק כי $\omega \in \dot{\bigcup}_{k=1}^{\infty} B_k$. ונקבל את (1.4). את ההכלה בכיוון ההפוך,

$$\bigcup_{k=1}^{\infty} A_k \supset \dot{\bigcup_{k=1}^{\infty}} B_k. \tag{1.5}$$

נסיק מן העובדה ש-$B_n \subset A_n$ (למעשה ההכלה בכיוון זה אינה נחוצה להוכחה ואנו רק מציינים אותה לצורך הבנת הבנייה). מ-(1.4) ו- (1.5) נסיק כי $\bigcup_{k=1}^{\infty} A_k = \dot{\bigcup}_{k=1}^{\infty} B_k$. כעת נשתמש בתכונת המונוטוניות (טענה 1.4ג) ונקבל

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \mathbb{P}\left(\dot{\bigcup_{k=1}^{\infty}} B_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) \le \sum_{k=1}^{\infty} \mathbb{P}(A_k), .$$

■

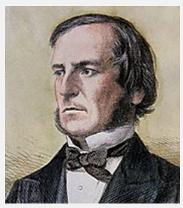
דוגמא 1.18. נתון כי בהסתברות 0.5 לפחות ירד גשם מחר, ובהסתברות 0.7 לפחות טל ילך עבודה. יש להראות כי בהסתברות 0.2 לפחות טל ילך לעבודה למרות שירד גשם.

תשובה: נוכל לנתח את ההסתברות מבלי לטרוח ולהגדיר את מרחב ההסתברות בפרוטרוט. נסמן ב-A את המאורע שירד גשם וב-B את המאורע שאלך לעבודה. נחשב את הסתברות המאורע $A \cap B$ ונשתמש בחישוב הסתברות למאורע משלים, טענה 1.4ה ונקבל.

$$\mathbb{P}(A \cap B) = 1 - \mathbb{P}(A^c \cup B^c)$$

הפעלת חישוב הסתברות משלימה על הנתונים נותנת כי $\mathbb{P}(A^c) \le 0.5$ וכי $\mathbb{P}(B^c) \le 0.3$. מאי-שוויון בול נקבל כי $\mathbb{P}(A^c \cup B^c) \le 0.5 + 0.3 = 0.8$ ולכן $\mathbb{P}(A \cap B) \ge 0.2$, כנדרש. □

נשוב לסוגיה 2 וננסה לנתח אותה ביתר פשטות תוך שימוש באי-שוויון בול.



ג׳ורג׳ בול (1815-1864), מתמטיקאי אנגלי אשר בילה את רוב חייו באירלנד. תרם תרומה מכרעת למדעי המחשב והיה אחד מחלוצי עידן המידע. את **אי-שוויון בול** הקרוי על שמו הציג בספרו ״חוקי המחשבה״ (Laws of Thought) מ-1854 אשר נתחלק בין עיונים באלגברה מעל השדה $\mathbb{F}_2$ אשר מכונה על שמו **בוליאנית** ובין חקר השאלה - ״איזה מידע נוכל להסיק בנוגע להסתברותם של מאורעות מורכבים בהינתן הסתברותם של מאורעות בסיסיים? ״.

דוגמא 1.19.[פרדוקס יום-ההולדת באמצעות אי-שוויון בול] נבקש לקבל חסם עליון טוב ככל הניתן על ההסתברות שבין n אנשים מקריים ימצא זוג אנשים שנולדו באותו היום. ניזכר בסימון

$$A_n = \left\{(x_1, \ldots, x_n) \in \Omega \ : \ \forall i, j \in [n], i \neq j \Rightarrow x_i \neq x_j\right\}.$$

לכל $i, j \in [n]$ נגדיר את $B_{ij} = \{x_i = x_j\}$, המאורע ששני האנשים i ו-j נולדו באותו יום (זו כתיבה מקוצרת של המאורע $B_{ij} = \{x \in \Omega \ : \ x_i = x_j\}$. סגנון זה ילך ויעשה שכיח בכתיבתנו). נשים לב כי לכל $0 \leq i < j \leq n$ מתקיים $|n_{ij}| = 365^{n-1}$ ולכן הסתברותו של כל מאורע כזה היא

$$\mathbb{P}(B_{ij}) = \frac{1}{365}$$

נשתמש באי שוויון בול (משפט 1.17) ונקבל

$$\mathbb{P}(A_n^c) = \mathbb{P}\left(\bigcup_{1 \leq i < j \leq n} B_{ij}\right) \leq \binom{n}{2}\frac{1}{365}.$$

זו הערכה מדוייקת אסימפטוטית, המנבאת למשל כי $\mathbb{P}(A_{18}^c) \leq 0.5$ ו-$\mathbb{P}(A_{23}^c) \leq 0.7$. נוסף על כך ההערכה מבטאת היטב את הרעיון שההסתברות המבוקשת תלויה במספר זוגות האנשים $\binom{n}{2}$ ולא במספר האנשים. □

לקבלת מושג על מידת הדיוק של חסם האיחוד במקרה זה ניתן להשוות את התוצאה לחישובים הנומרים המדוייקים בדוגמא 1.13.

מהטכניקה ששימשה להוכחת אי-שוויון בול נסיק שתי תכונות רציפות של פונקציית ההסתברות.

הגדרה 1.20 (סדרת מאורעות מונוטונית)**.** סדרה של מאורעות $\{A_n\}_{n \in \mathbb{N}}$ במרחב הסתברות נקראת **מונוטונית עולה** אם $A_i \subset A_j$ לכל $i < j$. הסדרה נקראת **מונוטונית יורדת** אם $A_j \subset A_i$ לכל $i < j$.

משפט 1.21 (רציפות פונקציית ההסתברות על מאורעות עולים)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות. ותהי $\{A_n\}_{n \in \mathbb{N}}$ סדרה עולה של מאורעות ב-$\mathcal{F}$. אזי מתקיים,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

הוכחה. נגדיר את סדרת המאורעות $\{B_n\}_{n \in \mathbb{N}}$ הנתונה על ידי

$$B_1 = A_1,$$
$$B_n = A_n \setminus A_{n-1}.$$

אלו מאורעות זרים מפני שעבור $m < n$ מתקיים

$$B_n \cap B_m = (A_n \setminus A_{n-1}) \cap (A_m \setminus A_{m-1}) = (A_n \cap A_m) \setminus (A_{n-1} \cup A_{m-1}) \subseteq A_m \setminus A_{n-1} = \emptyset,$$

כאשר המעבר הימני ביותר משתמש בכך ש-$A_m \subset A_{n-1}$.

הפעם נראה כי לכל $N \in \mathbb{N}$ מתקיים

$$\dot{\bigcup_{k=1}^{n}} B_k = \bigcup_{k=1}^{n} A_k = A_n \tag{1.6}$$

נסתכל על $\omega \in \cup_{k=1}^{n} A_n$ כלשהו. כמקודם, נבחין כי $\{k \in \mathbb{N} \;:\; \omega \in A_k\}$ אינה ריקה ובפרט קיים בה איבר מינימלי $k^* < N$ (איבר זה מתאר את ה-k המינימלי כך ש- $\omega \in A_k$). מהגדרת הקבוצות B_n נובע כי $\omega \in B_{k^*}$ ולפיכך $\omega \in \cup_{k=1}^{n} B_k$. ההכלה בכיוון ההפוך מיידית מכך ש-$B_n \subset A_n$ לכל $n \in \mathbb{N}$ ולכן (1.6) מתקיים. מ-(1.6) נסיק גם כי

$$\dot{\bigcup_{k=1}^{\infty}} B_k = \bigcup_{n=1}^{\infty} A_n \tag{1.7}$$

נשתמש ב-(1.6), ב-(1.7) ובסכימות בת-מניה לקבלת

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \mathbb{P}\left(\dot{\bigcup_{k=1}^{\infty}} B_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}\left(\dot{\bigcup_{k=1}^{n}} B_k\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

כנדרש. ■

מסקנה 1.22 (רציפות פונקציית ההסתברות על מאורעות יורדים)**.** אם $(A_n)_{n\in\mathbb{N}}$ היא סדרת מאורעות יורדים במרחב הסתברות, אזי

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

✎ **בעיה 1.7.** להוכיח את מסקנה 1.22 על סמך טענה 1.21.

(הדרכה: ניתן להראות כי אם סדרת מאורעות יורדת אז סדרת המאורעות המשלימים עולה).

הערה - אי-שוויוני בּוֹנְפֶרוֹנִי: אי-שוויון בול הוא רק אחד משורה של אי-שוויונים אשר ניתן לגזור מעקרון ההכלה וההדרה. להלן הכללה שריכז המתמטיקאי האיטלקי קרלו אמיליו בונפרוני בספרו מ-1936. בהנתן קבוצה של N מאורעות $\{A_i\}_{i\in[N]}$ במרחב הסתברות נסמן

$$S_n = \sum_{1\le i_1<\cdots<i_n\le N} \mathbb{P}\left(\bigcap_{j=1}^{n} A_{i_j}\right)$$

אזי לכל $k \in [N]$ מתקיים

$$\mathbb{P}\left(\bigcup_{n\in[N]} A_n\right) \le \sum_{j\in[2k+1]} (-1)^{j-1} S_j,$$

$$\mathbb{P}\left(\bigcup_{n\in[N]} A_n\right) \ge \sum_{j\in[2k]} (-1)^{j-1} S_j.$$

כתוצאה מכך ניתן לקבל הערכות טובות יותר ויותר על הסתברותו של מאורע על ידי שימוש באי-שוויונים מסדר הולך ועולה.

✎ **בעיה 1.8** (אי-שוויון בונפרוני השני)**.** תהי $(A_n)_{n\in\mathbb{N}}$ סדרת מאורעות במרחב הסתברות. יש להוכיח כי

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} A_n\right) \ge \sum_{i=1}^{N} \mathbb{P}(A_i) - \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{P}(A_i \cap A_j).$$

✎ **בעיה 1.9.** כיצד נשתמש באי-שוויון בונפרוני השני כדי לחסום את $\mathbb{P}(A_N)$ במקרה האסימפטוטי של פרדוקס יום ההולדת (דוגמא 1.13) עבור $N=\frac{\sqrt{M}}{10}$?

1.5 עקרון ההכלה וההדרה

בחישוב הסתברותם של מאורעות מתקיים עקרון שימושי, השאול מתורת הקומבינטוריקה. עקרון זה, המכונה באנגלית inclusion-exclusion principle, זכה למגוון שמות בעברית ובהם "עקרון ההכלה וההפרדה" ו-"עקרון ההכלה וההדחה". ברוח התקופה, אנו נכנה אותו "עקרון ההכלה וההדרה".

טענה 1.23 (הכלה והדרה)**.** יהי $(\Omega,\mathcal{F},\mathbb{P})$ מרחב הסתברות. אזי -

1. לכל שני מאורעות $A,B\in\mathcal{F}$ מתקיים

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

2. לכל שלושה מאורעות $A,B,C\in\mathcal{F}$ מתקיים

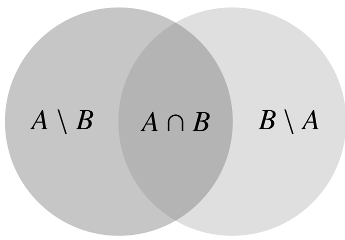
$$\mathbb{P}(A\cup B\cup C)=\mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)-\mathbb{P}(A\cap B)-\mathbb{P}(B\cap C)-\mathbb{P}(A\cap C)+\mathbb{P}(A\cap B\cap C).$$

3. לכל קבוצה של n מאורעות $\{A_i\}_{i\in[n]}$ מתקיים

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i\in[n]}A\right)&=\sum_{1\le i\le n}\mathbb{P}(A_i)-\sum_{1\le i<j\le n}\mathbb{P}(A_i\cap A_j)+\cdots+(-1)^{n+1}\mathbb{P}\left(\bigcap_{i\in[n]}A_i\right)\\&=\sum_{k=1}^{n}(-1)^{k+1}\sum_{\substack{I\subset[n],\\|I|=k}}\mathbb{P}\left(\bigcap_{j\in I}A_j\right)\end{aligned}$$

הוכחה. כדי להראות את סעיף (1), נשים לב כי

$$\begin{aligned}A\cup B&=(A\setminus B)\uplus(B\setminus A)\uplus(A\cap B)\\A&=(A\setminus B)\uplus(A\cap B)\\B&=(B\setminus A)\uplus(A\cap B),\end{aligned}$$



מתכונת הסכימות נובע כי

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)\\ \mathbb{P}(A) &= \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)\\ \mathbb{P}(B) &= \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B).\end{aligned}$$

כל שנותר הוא לחסר את שתי המשוואות האחרונות מהראשונה כדי לקבל את סעיף (1).

כדי להראות את סעיף (2) נשים לב כי לפי סעיף (1) מתקיים כי

$$\begin{aligned}\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A \cup B) + \mathbb{P}(C) - \mathbb{P}((A \cup B) \cap C)\\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) + \mathbb{P}(C) - \mathbb{P}((A \cap C) \cup (B \cap C)),\end{aligned}$$

נפעיל את סעיף (1) בשנית ונקבל את השוויון הנדרש. ■

✎ **בעיה 1.10.** להראות כי אם A ו-B הם שני מאורעות במרחב הסתברות אז הסיכוי שבדיוק אחד מהם ארע הנו

$$\mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(A \cap B).$$

✎ **בעיה 1.11.** להוכיח את סעיף (3) - עקרון ההכלה וההדרה הכללי.

(הדרכה: ניתן להשתמש בהפעלה חוזרת של סעיף (2) עם $A = A_i$ ו-$B = \cup_{i \le j \le n} A_j$).

כעת נראה כיצד בעקרון ההכלה וההדרה מאפשר לנו להציג ניתוח מדוייק של סוגיה 3.

דוגמא 1.24 (בעיית ההתאמה (Matching problem)).

n מכתבים ממוענים מוכנסים באקראי לתוך n מעטפות מבוילות. מהי ההסתברות שאף מכתב לא ישלח למען הנכון? מהי ההסתברות שבדיוק k מכתבים הגיעו ליעדם? יש לחשב הערכות הדוקות אסימפטוטית כאשר n שואף לאינסוף וk קבוע.

תשובה: במקרה זה מוטב לעבוד עם מרחב מדגם שבו שני מכתבים אינם יכולים להיות מותאמים לאותה המעטפה, כלומר מרחב התמורות -

$$\Omega = S_n = \{(x_1, \dots, x_n) \ : \ x_i \neq x_j, \forall i < j\} \qquad \mathbb{P}\Big(\big\{(x_1, \dots, x_n)\big\}\Big) = \frac{1}{n!}.$$

בייצוג זה נפרש את התמורה $(x_1, \dots, x_n)$ בתור התאמה של המכתב שנועד למעטפה ה-i למעטפה x_i. נשים לב כי זהו מרחב הסתברות אחיד.

הגעה של אפס מכתבים לתעודתם. נגדיר מאורע

$$A_n = \{(x_1, \dots, x_n) \in \Omega \ : \ \forall i \neq j \quad x_i \neq x_j\},$$

כלומר המאורע שאף מכתב לא הגיע לייעדו. בכדי לנתח מאורע זה נגדיר את המאורע שהמכתב ה-i הוכנס למעטפה הנכונה על ידי $B_i = \{(x_1, \dots, x_n) \in \Omega \ : \ x_i = i\}$. כעת נוכל לכתוב $(A_n)^c = \bigcup_{m=1}^n B_m$, כלומר

$$A_n = \Omega \setminus \bigcup_{m=1}^{n} B_m.$$

נשתמש בעקרון ההכלה וההדרה ובסימטריה בין המעטפות ונקבל

$$\begin{aligned}\mathbb{P}\Big((A_n)^c\Big) &= \sum_{i=1}^{n}\mathbb{P}(B_i) - \sum_{i<j}\mathbb{P}(B_i\cap B_j) + \sum_{i<j<k}\mathbb{P}(B_i\cap B_j\cap B_k) - \dots \\ &= \binom{n}{1}\mathbb{P}(B_1) - \binom{n}{2}\mathbb{P}(B_1\cap B_2) + \binom{n}{3}\mathbb{P}(B_1\cap B_2\cap B_3) - \dots \\ &\quad + (-1)^{n+1}\binom{n}{n}\mathbb{P}(B_1\cap\dots\cap B_n),\end{aligned}$$

כעת נוכל לחשב באופן קומבינטורי

$$\begin{aligned}\mathbb{P}(B_1) &= \frac{|B_1|}{|\Omega|} = \frac{(n-1)!}{n!} \\ \mathbb{P}(B_1\cap B_2) &= \frac{|B_1\cap B_2|}{|\Omega|} = \frac{(n-2)!}{n!},\end{aligned}$$

וכן הלאה. מכאן נובע כי

$$\begin{aligned}\mathbb{P}\Big((A_n)^c\Big) &= n\,\frac{1}{n} - \binom{n}{2}\frac{(n-2)!}{n!} + \binom{n}{3}\frac{(n-3)!}{n!} - \dots + (-1)^{n+1}\binom{n}{n}\frac{0!}{n!} \\ &= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n+1}\frac{1}{n!} \\ &= \sum_{k=1}^{n}\frac{(-1)^{k+1}}{k!}.\end{aligned}$$

עבור ערכי n גדולים נקבל,

$$\mathbb{P}(A_n) = 1 - \mathbb{P}\Big((A_n)^c\Big) = \sum_{k=0}^{n}\frac{(-1)^k}{k!} \approx e^{-1},$$

כלומר כאשר N שואף לאינסוף, ההסתברות שאף מכתב לא הגיע לתעודתו היא בקירוב 37%. ניתן למצא את העובדה שמספר זה אינו 0 או 1 מפתיעה. □

הואיל ומדובר במרחב אחיד, הרי שמתקיים $\mathbb{P}(A_n) = |A_n|/|\Omega|$. לכן נוכל להסיק מתוך החישוב את הנוסחא הקומבינטורית הבאה למספר התמורות אשר אין להן נקודת שבת,

$$|A_n| = |\Omega|\,\mathbb{P}(A_n) = n!\,\sum_{\ell=0}^{N}\frac{(-1)^\ell}{\ell!}.$$

נוסחא זו תשמשנו בסעיף הבא לניתוח ההסתברות שבדיוק k מכתבים יגיעו לתעודתם.

הגעה של k מכתבים לתעודתם. נגדיר מאורע A_n^k המתאר תמורות שלהן בדיוק k נקודות שבת. תמורות אלו מתאימות למאורע בו המזכירה התאימה בדיוק k מכתבים למעטפה הנכונה. לצורך כך נסמן עבור קבוצה $I \subset [N]$,

$$B_I = \Big(\bigcap_{i\in I} B_i\Big) \cap \Big(\bigcap_{i\in I^c}(B_i)^c\Big).$$

כלומר המאורע שבדיוק המכתבים בקבוצה I הגיעו לתעודתם. כעת נוכל לאפיין את A_n^k על ידי

$$A_n^k = \bigcup_{|I|=k} B_I$$

היות שהמאורעות B_I זרים ושווי הסתברות, נקבל כי

$$\left|A_n^k\right| = \left|B_{\{1,\dots,k\}}\right|\binom{n}{k}$$

והרי $B_{\{1,\dots,k\}}$ הוא המאורע ש-k המכתבים הראשונים הגיעו לתעודתם ואילו אף אחד מהמכתבים הממוענים לנמענים $k+1,\dots,n$ לא הגיעו לתעודתם. נקבל כי $\left|B_{\{1,\dots,k\}}\right| = \left|A_{n-k}\right|$ שכן מספר האפשרויות להתאמה ב-$B_{\{1,\dots,k\}}$ הוא בדיוק מספר האפשרויות להתאים את $n-k$ המכתבים האחרונים לנמענים בטווח $\{k+1,\dots,n\}$ כך שאף אחד ממכתבים אלו לא יגיע לתעודתו. לכן

$$\mathbb{P}(A_n^k) = \frac{\binom{n}{k}(n-k)!\ \sum_{\ell=0}^{n}\frac{(-1)^\ell}{\ell!}}{n!} = \frac{1}{k!}\sum_{\ell=0}^{n-k}\frac{(-1)^\ell}{\ell!}.$$

עבור k קבוע ו-n שואף לאינסוף נקבל

$$\mathbb{P}(A_n^k) \approx \frac{e^{-1}}{k!}.$$

נשוב לתוצאה זו בהמשך כאשר נדון בהתפלגות פואסונית בפרק 3.4. □

בעיות הרחבה והעשרה

✎ **בעיה 1.12** (על פי לואיס קרול). לאיש אחד ישנה אורוות סוסים. 75% מהסוסים באורווה בעלי כתם על המצח, 70% בעלי כתם על הגב, 80% מהם נהדרים ול85% יש שערה של כסף זוהרת בזנב. אם נבחר סוס מקרי, מה הסיכוי הנמוך ביותר שיתכן לכך שיהיה זה סוס נהדר אם כתם על המצח וכתם על הגב ושערה של כסף זוהרת בזנב.

✎ **בעיה 1.13** (הפרדוקס של שֶבַלְיֵיר דֶה-מֵארֶה). מה סביר יותר, תוצאה אחת של 6 לפחות בהטלת 4 קוביות הוגנות, או תוצאה של 12 ב-24 הטלות של זוג קוביות?

(במאה ה-17 הוצג הפתרון כפרדוקס מפני שהוא סתרה את הדעה המקובלת דאז לפיה המאורעות הנם שווי הסתברות).

✎ **בעיה 1.14** (בעיית נְיוּטוֹן-פֶּפִּי). מה סביר יותר, לפחות תוצאה אחת של 6 בהטלת 6 קוביות הוגנות, לפחות שתי תוצאות של 6 בהטלת 12 קוביות הוגנות, או לפחות שלוש תוצאות של 6 בהטלת 18 קוביות הוגנות?

(בעיה זו הייתה מקור לתכתובת בין אַיְיזֶק נְיוּטוֹן לבין סַמוּאֵל פֶּפִּי ב-1693 בנוגע לאינטואיציה הסתברותית).

✎ **בעיה 1.15.** מחפיסה בת N קלפים הממוספרים בסדר עולה שולפים n קלפים. מה הסיכוי שכל קלף יהיה גדול מקודמו?

✎ **בעיה 1.16.** N אבירי השולחן העגול התיישבו בסדר אקראי. מה הסיכוי שבין גלהד ופרסיבל מפרידים בדיוק k אבירים? יש לתת את הדעת למקרה המיוחד $N = 2k + 2$.

✎ **בעיה 1.17 (*).** מוגרלת פונקציה מקרית מ-$[n]$ ל-$[m]$ באופן אחיד. מה הסיכוי שהיא תהיה על?

✎ **בעיה 1.18 (*).** מוגרלת פונקציה מקרית מ-$[n]$ ל-$[m]$ באופן אחיד. מה הסיכוי שהיא תהיה חד-חד-ערכית?

פרק 2

הסתברות מותנית ואי-תלות

"לאחר שהוצאת מכלל חשבון את הבלתי-אפשרי, הרי מה שנותר, בלתי סביר ככל שיהיה, הוא האמת."

– אָרְתוּר קוֹנַאן דוֹיְל, חותם הארבעה.

בפרק זה נערוך היכרות עם שני רעיונות מרכזיים בתורת ההסתברות. הראשון - **הסתברות מותנית**, מעניק לנו כלי פורמלי ראשון לתאר את השינויים שמתרחשים בהערכותינו באשר להסתברותן של התרחשויות לאור מידע חדש שנחשפנו אליו. רעיון זהו הוא מקור לא-אכזב לפרדוקסים ומסקנות לא אינטואיטיביות אך נכונות אשר כמה מן היפים שבהם נכיר בפרק זה. השני - מושג **האי-תלות**, מבטא את הרעיון לפיו ישנן עובדות שידיעתן הנה חסרת כל השפעה על הערכת ההסתברות לתוצאתו של ניסוי מסויים. מושג זה יכליל את מושג מרחב המכפלה וימשיך לשחק תפקיד מרכזי בפרקים הבאים.

2.1 הסתברות מותנית

הסתברות משמשת להערכת שכיחותן של תוצאות של ניסוי שטרם נערך. משעה שנערך הניסוי והתוצאה ידועה - הרי שמנקודת מבט הסתברותית הסתברות התוצאה הזו היא אחת והסתברות כל תוצאה אחרת – אפס : הניסוי הופך דטרמיניסטי. אלא שמעבר זה ממצב של חוסר וודאות למצב של ודאות מוחלטת אינו תמיד כה חד. נסתכל למשל על הדוגמא הבאה.

> מר גלוק מרים את שפופרת הטלפון ומחייג לבתו דליה. בעודו מרים את השפופרת, הוא משער בנפשו כמה סביר שקולה של בת-שיחו אומנם ישמע מעברו השני של הקו. משמשע את צליל החיוג הראשון מתחזק בטחונו שבת-שיחו תענה. הרי הצליל מהווה עדות בטוחה לכך שהתקשורת הצליחה והמכשיר מן העבר השני מצלצל. שניה נוקפת וצלצול שני נשמע. כעת בטחונו בכך שבתו תענה פוחת קמעה, אך אין זה אירוע נדיר שהמענה אינו מגיע בצלצול השני. נוקפים צלצול שלישי ורביעי ובטחונו פוחת והולך ורגע לפני שהחליט לנתק, פתאום נשמעת הקריאה מן העבר השני "האלו?". ומר גלוק משיב, "הו, כבר סבור הייתי שלא תעני...".

בדוגמא זו ביצע מר גלוק הערכה ראשונית להסתברותו של מאורע, במקרה זה ההסתברות לשמוע את קולה של דליה. כעבור רגעים מספר נוסף לדובר מידע - השמיע הטלפון צליל חיוג ראשון. בעקבות המידע אליו נחשף ביצע מר גלוק שערוך של ההסתברות. הוא אמד את **ההסתברות המותנית** של המאורע שדליה תענה **בהנתן** צליל חיוג ראשון. כלומר בהנתן מידע רלוונטי לניסוי שאינו מכריע אותו באופן חד-משמעי. אותה התופעה חזרה על עצמה עם כל צליל חיוג כאשר מר גלוק אמד את **ההסתברות המותנית** של המאורע שדליה תענה **בהנתן** שטרם ענתה לאחר k צלצולים. הסתברות זו הלכה ופחתה כך שלמרות שבתחילה סבור היה שבתו תענה, כשהדבר אומנם קרה לבסוף – מצא עצמו מופתע.

הבה נבחן דוגמא פשוטה יותר לניתוח פורמלי הניתנת לתיאור באמצעות המשוגים שכבר למדנו.

דוגמא 2.1 (סכום קוביות מותנה)**.** שתי קוביות הוגנות מוטלות. מה היא ההסתברות שסכום התוצאות יהיה 8?
אם נודע לנו שתוצאת הקוביה הראשונה היא 3, מה כעת ההסתברות שהסכום יהיה 8?
אם תחת זאת נודע לנו שהתוצאה של לפחות קוביה אחת היא 3, מה ההסתברות שהסכום יהיה 8?

תשובה: לטובת הסעיף הראשון נתאר את מרחב ההסתברות של הניסוי באמצעות חזקת מרחב ההסתברות של הטלת קוביה בודדת. מרחב שתואר בדוגמא (1.8). זהו מרחב הסתברות אחיד עבור $\Omega = [6]^2$. את המאורע "סכום הקוביות יהיה 8" נסמן

$$A = \{(2,6),(3,5),(4,4),(5,3),(6,2)\}\,.$$

לפיכך $\mathbb{P}(A) = |A|/|\Omega| = 5/36$. לצורך החלק השני נסתכל על מרחב הסתברות של הטלת קוביה בודדת ונבקש שתוצאתה תהיה 5 (שהלא תוצאת הקוביה השניה ידוע לנו וערכה 3). זו אפשרות אחת מתוך שש במרחב אחיד ולכן ההסתברות היא $\frac{1}{6}$. אך ישנה גם דרך אחרת להסתכל על בעיה זו. נוכל להסתכל על Ω, מרחב ההסתברות שהגדרנו קודם, ולצמצם את פונקציית ההסתברות שלו למאורעות שעדיין עשויים להתרחש. נסמן $S_1 = \{(3,i) \;:\; i \in [6]\}$ כעת ננסה להגדיר פונקציית הסתברות מתאימה $\mathbb{P}^*$ שתייצג את הסתברותם של המאורעות השונים לאחר שנודע לנו שתוצאת הקוביה הראשונה היא 3. היות שהמידע שברשותנו לא משפיע על יחס ההסתברויות של מאורעות המוכלים ב-S_1, והיות ששומא עלינו להגדיר $\mathbb{P}^*(S_1) = 1$, הרי שלכל $\omega \in S_1$ יהא עלינו להגדיר $\mathbb{P}^*(\{\omega\}) = \mathbb{P}(\{\omega\})/\mathbb{P}(S_1)$. ואילו לכל $\omega \notin S_1$ עלינו להגדיר $\mathbb{P}^*(\{\omega\}) = 0$ (הלא כבר נודע לנו שמאורעות אלו לא התרחשו). כעת נחשב את הסתברות המאורע A לפי $\mathbb{P}^*$ במרחב ההסתברות $(\Omega, 2^\Omega, \mathbb{P}^*)$ והיות שגודלו של המרחב הוא 6 אומנם נקבל $\frac{1}{6}$. אומנם זו דרך סבוכה יותר מזו שהוצגה קודם, אך היא תאפשר לנו להתמודד גם עם הסעיף השלישי. לשם כך נגדיר

$$S_2 = \{(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(1,3),(2,3),(4,3),(5,3),(6,3)\}\,.$$

שוב נגדיר פונקציית הסתברות אחידה על S_2 המעניקה ערך 0 לכל איבר ב-$\Omega \setminus S_2$. אנו מעוניינים במאורע $\{(3,5),(5,3)\}$ והיות ש-$|S_2| = 11$, נסיק שההסתברות המבוקשת הנה 2/11. □

✎ **בעיה 2.1.** נתון כי בהטלת שתי קוביות משחק הוגנות, לפחות אחת הקוביות הראתה תוצאה של 4, מה ההסתברות שסכומן היה 8? מדוע ההסתברות כה נמוכה בהשווה לסעיף (ג) בדוגמא 2.1?

את הדוגמא נכליל באמצעות ההגדרה הבאה.

הגדרה 2.2. יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהי $B \in \mathcal{F}$ מאורע בעל הסתברות חיובית. לכל מאורע $A \in \mathcal{F}$, נגדיר את **ההסתברות המותנית של** A **בהנתן** B על ידי

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

הערה: את מושג ההסתברות המותנית ניתן לפרש בצורות שונות, כתלות בנקודת ההשקפה על תוקפה של תורת-ההסתברות. מנקודת מבט הכרתית היא מייצגת את מידת הוודאות של התרחשות A לאחר שנודע לנו ש-B התרחש, ולכן ההסתברות מכונה לעיתים הסתברות **אפוסטריורי** (לאחר מעשה) להבדיל מהסתברות **אפריורי** (בטרם מעשה). מנקודת מבט שכיחותנית ההסתברות המותנית מתארת את מספר המקרים שבהם ארע מאורע A מתוך כלל המקרים שבהם ארע מאורע B. מנקודת השקפה מהותית הרי היא מיצגת את אחוז העתידים האפשריים שבהם A התרחשה מאותה נקודה בזמן שבה B כבר ארע. כמובן שהטיפול הפורמלי במונח אינו מושפע מההשקפה אותה נבחר לאמץ.

ראשית נראה כי התניה תמיד יוצרת פונקציית הסתברות מוגדרת היטב על Ω, מרחב המדגם של הניסוי.

טענה 2.3 (הסתברות מותנית היא הסתברות)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהי $B \in \mathcal{F}$ מאורע בעל הסתברות חיובית. נגדיר $\mathbb{P}_B(A) = \mathbb{P}(A \mid B)$. אזי $\mathbb{P}_B$ היא פונקציית הסתברות על Ω.

הוכחה. עלינו לוודא כי $\mathbb{P}_B$ מוגדרת היטב ומקיימת את תכונות פונקציית ההסתברות, הגדרה 1.3. מתכונת המונוטוניות (1.4ג) נקבל כי $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ כך ש $\mathbb{P}_B$ היא אכן פונקציה מ-$\mathcal{F}$ לקטע $[0,1]$. תכונה (1.3א) נובעת מיידית מכך ש

$$\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = 1.$$

נותר אפוא להראות סכימות בת מניה (הגדרה 1.3ב). תהי $(A_n)_{n\in\mathbb{N}}$ סדרת מאורעות זרים ב-$\mathcal{F}$. קל לראות שהמאורעות $(A_n \cap B)$ אף הם זרים ולכן, לפי תכונה (1.3ב) של $\mathbb{P}$, נקבל

$$\begin{aligned}\mathbb{P}_B\left(\dot{\bigcup_n} A_n\right) &= \frac{\mathbb{P}((\dot{\bigcup}_n A_n) \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(\dot{\bigcup}_n (A_n \cap B))}{\mathbb{P}(B)} \\ &= \frac{1}{\mathbb{P}(B)} \sum_n \mathbb{P}(A_n \cap B) = \sum_n \mathbb{P}_B(A_n).\end{aligned}$$

■

כאשר נתנה ביותר ממאורע אחד, נשתמש לעיתים קרובות בסימון $\mathbb{P}(D \mid A, B) := \mathbb{P}(D \mid A \cap B)$.

אבחנה 2.4 (תכונות בסיסיות של הסתברות מותנית)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהיו $A, B, D \in \mathcal{F}$ מאורעות בעלי הסתברות חיובית. אזי

(א) אם $B \subset A$ אזי $\mathbb{P}(D \mid A) = \mathbb{P}(D \mid A, B)$.

(ב) **כלל השרשרת (chain-rule):** $\mathbb{P}(A \cap B) = \mathbb{P}(B) \cdot \mathbb{P}(A \mid B)$

(ג) **התניה חוזרת:** $\mathbb{P}_D(A \mid B) = \mathbb{P}(A \mid B \mid D) = \mathbb{P}(A \mid B, D) = \frac{\mathbb{P}(A \cap B \cap D)}{\mathbb{P}(B \cap D)}$

דוגמא 2.5 (הסתברות מותנית)**.**

(א) בכד 5 כדורים צהובים, 10 לבנים ו-10 שחורים. כדור נשלף באקראי, מה ההסתברות שהוא צהוב? מה ההסתברות שהוא צהוב בהינתן שאינו שחור?

תשובה: נשתמש במרחב ההסתברות האחיד על $\Omega = [25]$, כאשר הכדורים הצהובים מיוצגים על ידי המספרים 1-5, הלבנים על ידי 6-15, והשחורים על ידי 16-25. המאורע המבוקש הוא $A = [5]$ ומתוקף אחידות המרחב הסתברותו $5/25 = 1/5$. כעת נתונה במאורע $B = [15]$ ונקבל

$$\mathbb{P}(A \mid B) = \frac{|A \cap B|}{|B|} = \frac{|A|}{|B|} = \frac{5}{15} = \frac{1}{3}.$$

(ב) נשוב לדוגמא המילולית שפתחה את הפרק. נניח כי מר גלוק מתקשר לבתו דליה. ההסתברות שתקלה תמנע תקשורת היא 0.1. אחרת ההסתברות שדליה תענה מיד לאחר הצלצול ה-n נתונה על ידי הטבלה הבאה:

צלצול:	1	2	3	4	לא תענה
הסתברות:	0.05	0.3	0.3	0.05	0.2

חשב את ההסתברות שדליה תענה בהינתן שנשמעו n צלצולים עבור $n = 0, 1, 2, 3, 4$.

תשובה: נגדיר מרחב הסתברות בן חמישה מצבים שיסומנו על ידי

$$\Omega = \{(F, 0), (F, 4), (T, 1), (T, 2), (T, 3), (T, 4)\}$$

האות הראשונה בכל צמד תייצג את אמיתות הטענה "דליה ענתה", ואילו המספר ייצג את מספר הצלצולים שנשמעו. פונקציית ההסתברות המתאימה תהיה זו המקיימת

$$\mathbb{P}(\{(F,0)\}) = 0.1 \quad \mathbb{P}(\{(T,1)\}) = 0.05 \quad \mathbb{P}(\{(T,2)\}) = 0.3$$
$$\mathbb{P}(\{(F,4)\}) = 0.2 \quad \mathbb{P}(\{(T,3)\}) = 0.3 \quad \mathbb{P}(\{(T,4)\}) = 0.05$$

נסמן $A = \{(a, b) \in \Omega \ : \ a = T\}$ וכן $B_n = \{(a, b) \in \Omega \ : \ b \geq n\}$. נציב בנוסחה להסתברות מותנית ונקבל

מאורע	$A \cap B_0$	$A \cap B_1$	$A \cap B_2$	$A \cap B_3$	$A \cap B_4$	B_0	B_1	B_2	B_3	B_4
הסתברות:	0.7	0.7	0.65	0.35	0.05	1	0.9	0.85	0.55	0.25

נחלק ונקבל

מאורע	$A \mid B_0$	$A \mid B_1$	$A \mid B_2$	$A \mid B_3$	$A \mid B_4$
הסתברות:	7/10	7/9	13/17	7/11	1/5

קל לראות כיצד טבלה זו מסבירה את שינויי הערכת ההסתברות לקבלת תשובה לאורך המתנתו של מר גלוק למענה. □

✎ **בעיה 2.2.** יהיה $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהיו $A, B \in \mathcal{F}$ זוג מאורעות בעלי הסתברות חיובית. יש להראות כי אם $\mathbb{P}(A) < \mathbb{P}(A \mid B)$ אז $\mathbb{P}(A) > \mathbb{P}(A \mid B^c)$. מה משמעותה של קביעה זו?

✎ **בעיה 2.3.** יהיו A ו-B שני מאורעות בעלי הסתברות חיובית במרחב הסתברות כלשהו. נאמר כי A **מאשש** את B אם $\mathbb{P}(B \mid A) > P(B)$. יש להראות כי אישוש אינה תכונה עוברת [טרנזיטיבית]. כלומר כי קיימים מאורעות A, B ו-C במרחב הסתברות כך ש-A מאשש את B ו-B מאשש את C אך A אינו מאשש את C.

✎ **בעיה 2.4** (כלל השרשרת - הכללה)**.** לכל שלושה מאורעות A, B ו-C המקיימים $\mathbb{P}(B) > 0$ ו-$\mathbb{P}(B \cap C) > 0$, מתקיים השוויון
$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B, C)\, \mathbb{P}(B \mid C)\, P(C).$$
כיצד נוכיח את נכונותו ונכלילו למספר רב יותר של איברים?

2.2 נוסחת ההסתברות השלמה וכלל בייס

נשווה בנפשנו את המצב הבא:

> בשניית הסיום של משחק כדורסל מכריע בטורניר היורוליג של 2014, כאשר התוצאה 86:87 לטובת ארמאני מילאנו, ביצע אלכס טיוס, שחקן מכבי תל-אביב עברה על קית' לנגפורד, שחקן נבחרת מילאנו. לנגפורד עולה לקו העונשים לשתי קליעות עונשין. אם יקלע את שתיהן - ינצחו האיטלקים, אם יקלע אחת מהן - יגרר המשחק להערכה שבה סיכוי ישראל לנצח הם 60%. אם יחטיא את שתיהן תנצח ישראל. ללנגפורד 75% לקלוע מקו העונשין בכל קליעה באופן בלתי-תלוי.

מה סיכוייה של מכבי לנצח? לשם כך נרצה לבטא את סיכויי המאורע "מכבי ניצחה", באמצעות סיכויי הקליעה של לנגפורד וסיכויי הקבוצה לנצח בהארכה. ביטוי זה יגולם ב**נוסחת ההסתברות השלמה**.

הגדרה 2.6 (חלוקה)**.** יהי Ω מרחב מדגם. אוסף בן-מניה של מאורעות $\mathcal{A}$ יקרא **חלוקה** של Ω אם כל שני מאורעות הנם זרים (כלומר $A_i \cap A_j = \emptyset$ לכל $i, j \in \mathbb{N}$) וכן $\bigcup_{A \in \mathcal{A}} A = \Omega$. במילים אחרות - על כל איבר ב-Ω להמצא בקבוצה אחת ויחידה בחלוקה. כאשר נדבר על חלוקות של מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ - תהא כוונתנו לחלוקות של מרחב המדגם Ω.

טענה 2.7 (נוסחת ההסתברות השלמה)**.** תהי $\mathcal{A}$ חלוקה של מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. לכל מאורע B מתקיים
$$\mathbb{P}(B) = \sum_{A \in \mathcal{A}}^{\infty} \mathbb{P}(B \cap A) = \sum_{A \in \mathcal{A}}^{\infty} \mathbb{P}(B \mid A) \mathbb{P}(A).$$
כאשר אנו מפרשים איברים באגף ימין כשווים ל-0 אם $\mathbb{P}(A_i) = 0$.

הוכחה. כל מאורע B נוכל לפרק לפי
$$B = \dot{\bigcup_{A \in \mathcal{A}}^{\infty}} (B \cap A).$$
ולכן, בזכות תכונת הסכימות בת-המניה (1.3ב) נקבל

$$\mathbb{P}(B)=\sum_{A\in\mathcal{A}}\mathbb{P}(B\cap A)=\sum_{A\in\mathcal{A}}^{\infty}\mathbb{P}(B\,|\,A)\mathbb{P}(A).$$

■

נשים לב שנוסחת ההסתברות השלמה פועלת גם בחלוקות סופיות. כלי זה מאפשר לנו לבצע למידת מקרים פרטיים בהקשר הסתברותי, כלומר לפרק את מרחב ההסתברות למקרים שבהם התרחש כל אחד מקבוצה של מאורעות משלימים.

דוגמא 2.8. מטילים קוביה הוגנת בת שש פאות וקוביה הוגנת נוספת בת שתים-עשרה פאות. מה ההסתברות שהתקבלה בקוביה הראשונה תוצאה גדולה יותר מאשר בקוביה השניה?

תשובה: מרחב ההסתברות הוא המרחב האחיד על $\Omega=[6]\times[12]$. נסמן את המאורע המבוקש

$$B=\{(\omega_1,\omega_2)\in\Omega\ :\ \omega_1>\omega_2\}.$$

נגדיר מאורעות $A_k=\{k\}\times[12]$ המתאימים לתוצאה של k בקוביה הראשונה ונשים לב כי זוהי חלוקה של מרחב ההסתברות. כעת נחשב לפי נוסחת ההסתברות השלמה:

$$\mathbb{P}(B)=\sum_{i\in[6]}\mathbb{P}(A_i)\mathbb{P}(B\,|\,A_i)=\sum_{i\in[6]}\frac{1}{6}\cdot\frac{i-1}{12}=\frac{15}{6\cdot 12}=\frac{5}{24}.$$

□

כמסקנה מנוסחת ההסתברות השלמה, נקבל כי ניתן להגדיר באופן חד-משמעי ניסויים במונחים של הסתברות מותנית. כלומר - נוכל לתאר ניסוי דו-שלבי על ידי כך שנתאר את ההסתברות לכל תוצאה של השלב הראשון של הניסוי, וכן הסתברותה של כל תוצאה של השלב השני כתלות בתוצאת השלב הראשון. ננסח זאת להלן באופן פורמלי, כאשר נתאר את תוצאת השלב הראשון כחלוקה של מרחב ההסתברות ואת תוצאת הניסוי כולו כאיבר במרחב המדגם.

אבחנה 2.9 (ניסוי דו-שלבי). תהי $\mathcal{A}$ חלוקה של מרחב מדגם Ω, ותהיינה $\{\mathbb{P}_A\}_{A\in\mathcal{A}}$ פונקציות הסתברות כך ש-$\mathbb{P}_A(A)=1$, אזי הפונקציה $\mathbb{P}:\mathcal{F}\to\mathbb{R}_+$ אשר לכל $B\subset\Omega$ מקיימת

$$\mathbb{P}(B)=\sum_{\substack{A\in\mathcal{A}\\ \mathbb{P}(A)\neq 0}}^{\infty}\mathbb{P}_A(B)\mathbb{P}(A),$$

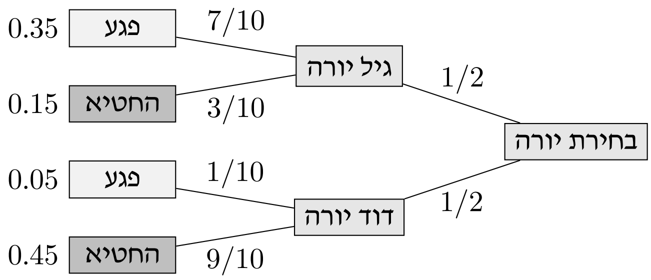
היא פונקציית הסתברות והיא היחידה המקיימת לכל $A\in\mathcal{A}$ ולכל $B\subset\Omega$ מתקיים $\mathbb{P}(B\,|\,A)=\mathbb{P}_A(B)$.

דוגמא 2.10. גיל ידוע כקלע מוכשר הפוגע במטרה בהסתברות 70% בכל ירייה, לעומתו דוד אלרואי קולע רק בהסתברות 10%. בתחרות קליעה בין השניים נבחר הקלע הראשון באקראי באמצעות הטלת מטבע הוגן וירה קליע. בהינתן שהקליע פגע במטרה מה ההסתברות שהוא נורה על ידי גיל?

תשובה: נשתמש במרחב ההסתברות $\Omega=\{\text{גיל},\text{דוד}\}\times\{\text{החטיא},\text{פגע}\}$. המתאר מי ניסה לקלוע ומה הייתה התוצאה. את ההסתברויות המתאימות קיבלנו במונחים של הסתברות מותנית -

$$\mathbb{P}\Big(\{(\text{גיל},\cdot)\}\Big)=0.5 \qquad \mathbb{P}\Big(\{(\text{דוד},\cdot)\}\Big)=0.5$$

$$\mathbb{P}\Big(\{(\cdot,\text{פגע})\}\,\Big|\,\{(\text{גיל},\cdot)\}\Big)=0.7 \qquad \mathbb{P}\Big(\{(\cdot,\text{פגע})\}\,\Big|\,\{(\text{דוד},\cdot)\}\Big)=0.1$$



תרשים 2.1: תיאור גרפי של הניסוי הדו-שלבי המתואר בדוגמא 2.10. על כל קשת כתובה ההסתברות לתוצאה המצויה משמאל בהינתן השלב הכתוב מימין. בשולי התרשים כתובות ההסתברויות הכוללות לכל תוצאה. הסתברויות אלו ניתן לחשב על ידי הכפלת ההסתברויות במורד העץ, כפי שעולה מכלל השרשרת.

ולפי כלל השרשרת, נקבל

$$\mathbb{P}\Big(\{(\text{גיל},\text{פגע})\}\Big) = \mathbb{P}\Big(\{(\cdot,\text{גיל})\}\Big)\mathbb{P}\Big(\{(\text{פגע},\cdot)\}\,\Big|\,\{(\cdot,\text{גיל})\}\Big) = 0.5\cdot 0.7 = 0.35$$

באופן דומה נחשב.

$$\mathbb{P}\Big(\{(\text{גיל},\text{החטיא})\}\Big) = 0.15,\quad \mathbb{P}\Big(\{(\text{דוד},\text{החטיא})\}\Big) = 0.45,\quad \mathbb{P}\Big(\{(\text{דוד},\text{פגע})\}\Big) = 0.05.$$

כעת נשתמש בנוסחה להסתברות מותנית ונקבל

$$\mathbb{P}\Big(\{(\cdot,\text{גיל})\}\,\Big|\,\{(\text{פגע},\cdot)\}\Big) = \frac{\mathbb{P}\Big(\{(\text{גיל},\text{פגע})\}\Big)}{\mathbb{P}\Big(\{(\text{פגע},\cdot)\}\Big)} = \frac{0.35}{0.35+0.5} = 7/8$$

כלומר בהסתברות של שמונים ושבעה אחוזים וחצי – היורה היה גיל. תיאור גרפי של תהליך החישוב שביצענו לפי נוסחת ההסתברות השלמה מפורט בתרשים 2.1. □

נשוב לדוגמא מתחילת הפרק.

דוגמא 2.11 (נס מילאנו - נוסחת ההסתברות השלמה)**.** נגדיר מרחב הסתברות - $\Omega = \{W, L\}\times[2]$ המתאר האם מכבי ניצחה או הפסידה במשחק הכדורסל, וכמה קליעות עונשין קלע לנגפורד. ידוע כי פונקציית ההסתברות על מרחב זה מקיימת:

$$\mathbb{P}(\{(\cdot,0)\}) = \left(\frac{1}{4}\right)^2 \qquad \mathbb{P}(\{(\cdot,1)\}) = 2\left(\frac{3}{4}\right)\left(\frac{1}{4}\right) \qquad \mathbb{P}(\{(\cdot,2)\}) = \left(\frac{3}{4}\right)^2$$

$$\mathbb{P}(\{(W,\cdot)\}\,|\,\{(\cdot,0)\}) = 1 \qquad \mathbb{P}(\{(W,\cdot)\}\,|\,\{(\cdot,1)\}) = \frac{3}{5} \qquad \mathbb{P}(\{(W,\cdot)\}\,|\,\{(\cdot,2)\}) = 0$$

כיצד נחשב את סיכויי של מכבי לנצח, כלומר את $\mathbb{P}(\{(W,\cdot)\}$?

תשובה: נציב בנוסחת ההסתברות השלמה ונקבל

$$\begin{aligned}\mathbb{P}(\{(W,\cdot)\} &= \sum_{k\in\{0,1,2\}} \mathbb{P}(\{(W,\cdot)\}\,|\,\{(\cdot,k)\})\mathbb{P}(\{(\cdot,k)\}) \\ &= \left(\frac{1}{4}\right)^2\cdot 1 + 2\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)\cdot\left(\frac{3}{5}\right) + \left(\frac{3}{4}\right)^2\cdot 0 = \frac{23}{80}.\end{aligned}$$

□

בכדי להוסיף ולפתח את התיאוריה, נעיין בתסריט נוסף הנוגע לאותה דוגמא.

> בדיוק ברגע המכריע השתבש שידור המשחק בישראל והצופה האומלל לא זכה לראות את סופו. למחרת מתברר כי מכבי אומנם ניצחה.

כעת ירצה הצופה לדעת מה הסיכוי שלנגפורד קלע פעם אחת בדיוק והמשחק נגרר להערכה. כדי לענות על שאלה זו - עומד לרשותנו **כלל בייס** המכונה גם **נוסחת ההיפוך**, והוא קושר בין ההסתברות $\mathbb{P}(A \mid B)$ להסתברות $\mathbb{P}(B \mid A)$.

טענה 2.12 (כלל בייס [נוסחת ההיפוך])**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות. ויהיו $A, B \in \mathcal{F}$ שני מאורעות. אזי

$$\mathbb{P}(B)\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A)\mathbb{P}(A).$$

כאשר אנו מפרשים את אגף ימין כשווה ל-0 אם $\mathbb{P}(B) = 0$ ואת אגף שמאל כשווה ל-0 אם $\mathbb{P}(A) = 0$.

הוכחה. הוכחת המשפט מיידית מהגדרת ההסתברות המותנית:

$$\mathbb{P}(B)\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B)\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B \mid A)\mathbb{P}(A).$$

■

כאשר מדובר במאורעות בעלי הסתברות חיובית, נוח לרשום את כלל בייס באופן הבא:

$$\mathbb{P}(A \mid B) = \mathbb{P}(B \mid A)\frac{\mathbb{P}(A)}{\mathbb{P}(B)}.$$

נשים לב שמהנוסחה נובע גם כי לפי אבחנה 2.4ג מתקיים (בכל מקרה בו ההתניות מוגדרות היטב)

$$\mathbb{P}(A \mid B, D) = \mathbb{P}_D(A \mid B) = \mathbb{P}_D(B \mid A)\frac{\mathbb{P}_D(A)}{\mathbb{P}_D(B)} = \mathbb{P}(B \mid A, D)\frac{\mathbb{P}(A \mid D)}{\mathbb{P}(B \mid D)},$$

ניישם נוסחה זו בכדי להכריע מה ההסתברות שלנגפורד קלע קליעה אחת מתוך שתיים בהינתן שמכבי ניצחה. (כל זאת במצב המתואר בדוגמא 2.11).

דוגמא 2.13 (נס מילאנו - נוסחת בייס)**.**

תומָאס בֵּיְיס (1701-1761), מתמטיקאי אנגלי וכומר פרסבטריאני אשר למד את תורת ההסתברות על סמך ספרו של אברהם דה-מואבר "תורת המקריות" שיצא ב-1718. הערותיו לספר שנתפרסמו לאחר מותו תחת הכותרת "מאמר החותר לפתרון בעיה בתורת הסיכויים" (An essay towards solving a problem in the doctrine of chances). פורשו על ידי פייר-סימון לפלס ככלי מתמטי לעדכון סבירותה של תיאוריה נוכח עדויות אמפיריות והוא החל מבסס תורת הסתברות סובייקטיבית. בשנות החמישים של המאה העשרים היוותה הנוסחה כלי מרכזי בפיתוח תורת ההסקה הבייסיאנית המאפשרות להכריע בין השערות על סמך עדויות. כך למשל אם אנו סוברים שאם השערה H נכונה אזי מאורע B מתרחש בהסתברות מסויימת, ואילו אם השערה H' נכונה הרי הוא מתרחש בהסתברות שונה, הרי שאם נראה את B מתרחש נוכל לאמוד את יחס הסבירויות של שתי ההשערות. ספק רב אם בייס עצמו היה שותף לפרשנות זו, שכן אין לה כל ראיות בכתביו.

הפעם נחשב $\mathbb{P}(\{(\cdot,1)\}\,|\,\{(W,\cdot)\})$, כלומר את ההסתברות שלנגפורד קלע בדיוק פעם אחת בהינתן שישראל ניצחה. נציב בנוסחת בייס ונקבל

$$\mathbb{P}(\{(\cdot,1)\}\,|\,\{(W,\cdot)\}) = \mathbb{P}(\{(W,\cdot)\}\,|\,\{(\cdot,1)\})\frac{\mathbb{P}(\{(\cdot,1)\})}{\mathbb{P}(\{(W,\cdot)\})} = \frac{3}{5}\cdot\frac{2\left(\frac{3}{4}\right)\left(\frac{1}{4}\right)}{\frac{23}{80}} = \frac{18}{23} \approx 78\%$$

□

עתה נמחיש בדוגמאות מספר עד כמה נוח לשגות בשימוש בהסתברות מותנית, למרות פשוטתה הפורמלית, ועד כמה שימוש שגוי יכול להוביל לשגיאות מהותיות ולכשלים לוגיים.

דוגמא 2.14 (משפטו של או. ג׳יי. סימפסון)**.**
נוחסאת בייס רלוונטית לעיתים קרובות בעולם המשפט, במיוחד בהקשר של כשל התובע (prosecutor’s fallacy). נסתכל למשל על הדוגמא המפורסמת הבאה ממשפטו של או. ג׳יי. סימפסון ב-1994 בגין רציחתה של אשתו, ניקול בראון. במשפט טענה התביעה שהחשוד המיידי הוא הבעל, מפני שנודע במהלך המשפט שהוא נהג להכותה, והסתברותו של אדם מכה להרוג את אשתו גבוהה יותר. להגנתו טען עורך הדין אלן דרשוביץ כי העובדה הזו כמעט חסרת חשיבות מפני שרק 0.4% מהבעלים המכים רוצחים את נשותיהם.

הבה ננתח שאלה הסתברותית בעלת זיקה למקרה. נסתכל על אדם נשוי מקרי ב-1994. נסמן ב-M את המאורע שאשתו של האיש נרצחה, ב-G את המאורע שהוא רצח את אשתו וב-H את המאורע שהוא נהג להכותה. נשים לב כי $G \subset M$. נתונים סטטיסטים מאותה שנה קובעים בקירוב כי ההסתברות של אישה מוכה להירצח הוא $\mathbb{P}(M\,|\,H) = 1/2250$, ואילו ההסתברות של אישה מוכה להירצח על ידי בעלה היא $\mathbb{P}(G\,|\,H) = 1/2500$, בדיוק כפי שטען מר דרשוביץ.

השאלה שעניינה את המשתתפים במשפט היא הערכת ההסתברות שהאיש אשם ברצח אשתו בהינתן שהיא אישה מוכה שנרצחה, כלומר $\mathbb{P}(G\,|\,H,M)$. נחשב:

$$\mathbb{P}(G\,|\,H,M) \overset{\text{Bayes}}{=} \frac{\mathbb{P}(M\,|\,G,H)\cdot\mathbb{P}(G\,|\,H)}{\mathbb{P}(M\,|\,H)} \overset{G\subset M}{=} \frac{\mathbb{P}(G\,|\,H)}{\mathbb{P}(M\,|\,H)} = \frac{1/2500}{1/2250} = 0.9 \tag{2.1}$$

במילים אחרות - אילו היה מדובר באדם מקרי באותה שנה שכל מה שאנו יודעים עליו הוא שהוא הכה את אשתו והיא נרצחה, אזי ישנם 90 אחוזים שהוא הרוצח. גם התביעה וגם ההגנה במקרה זה התעלמו בנימוקיהם מהנתון הידוע לנו והוא שהאישה אומנם נרצחה. □

מדוגמא זו נסיק כי **בחישוב הסתברותו של מאורע עלינו תמיד להביא בחשבון בתיאור מרחב ההסתברות את כל הידוע לנו**.

דוגמא 2.15 (עשינו עסק [Monty Hall])**.**
בשעשועון הטלוויזיה "עשינו עסק" ששודר בערוץ השני בשנות התשעים היה נהוג התהליך המקרי הבא. מאחורי וילון שנבחר באופן אחיד מבין שלושה וילונות הייתה מוסתרת מכונית. מאחורי שני הוילונות האחרים הוסתר דחליל. המשתתף התבקש לבחור וילון. לאחר מכן היה מנחה השעשועון חושף באקראי את אחד הוילונות שהמשתתף לא בחר בהם מבין אלו שמאחוריהם נמצא דחליל. אז ניתנה למשתתף אפשרות לבחור אם ברצונו להחליף וילון או לדבוק בוילון בו בחר לראשונה. האם החלפת הוילון תעלה את סיכויי המשתתף לזכות במכונית?

תשובה: בלי הגבלת הכלליות, מטעמי סימטריה, נוכל להניח שהמשתתף בוחר בשלב הראשון תמיד את וילון מספר אחד. לתיאור הניסוי ללא החלפה נשתמש במרחב ההסתברות האחיד על $\Omega = [3]$ שיתאר את מספר הוילון הזוכה. במקרה כזה זכיה מתוארת על ידי המאורע $A = \{1\}$ ולכן הסתברותה $(1/3)$.

מה באשר לזכיה לאחר ביצוע החלפה? נשנה את מרחב ההסתברות להיות $\Omega = \{(1,2),(1,3),(2,3),(3,2)\}$ כאשר האיבר הראשון בזוג יתאר את הוילון שמאחוריו המכונית והשני - את הוילון שנפתח. הסתברותם של המאורעות כפי שהיא מתוארת היא

$$\begin{aligned}
&\mathbb{P}\Big(\{(1,2)\}\Big) = \mathbb{P}\Big(\{(1,3)\}\Big) = \mathbb{P}\Big(\{(1,\cdot)\}\Big)\mathbb{P}\Big(\{(1,3)\} \,|\, \{(1,\cdot)\}\Big) = \frac{1}{3}\cdot\frac{1}{2} = \frac{1}{6}\\
&\mathbb{P}\Big(\{(2,3)\}\Big) = \mathbb{P}\Big(\{(3,2)\}\Big) = \mathbb{P}\Big(\{(3,\cdot)\}\Big)\mathbb{P}\Big(\{(3,2)\} \,|\, \{(3,\cdot)\}\Big) = \frac{1}{3}
\end{aligned}$$

כמו כן נשים לב שהחלפה פירושה בחירה בוילון המכונית במאורע $\{(2,3),(3,2)\}$ ולכן הסתברות הזכיה במקרה של החלפה היא $(2/3)$. ואמנם - החלפה נכשלת אך ורק במקרה שבו נבחר בניחוש ההתחלתי את וילון המכונית. נסיק כי סיכוייו של המשתתף לזכות עולים כאשר הוא מחליף את הווילון בו בחר. □

קורא שחושיו מוחים כנגד הפתרון, עשוי להסתייע בכך שידמיין מצב שבו מכונית מסתתרת מאחורי אחד ממאה וילונות, ולאחר שנבחר אחד מהם, המנחה פותח תשעים ושמונה שאינם מכילים מכונית. במקרה כזה קל לראות שקלושים הסיכויים שהוילון שנבחר בתחילה הוא הנכון, ואפוא למנחה אין ברירה אלא להותיר רק את הוילון מאחוריו מסתתרת המכונית.

הערה: כאשר עלה השעשועון בטלויזיה, לא היה ידוע לצופים מה היא מדיניות חשיפת הוילונות של המנחה. כל שהיה ידוע לצופה הוא שלאחר בחירת המשתתף בוילון מסויים, חשף המנחה וילון אחר שמאחוריו התגלה דחליל. כתוצאה מכך, מרחב ההסתברות המתאים לבעיה לא היה מוגדר באופן חד-משמעי ולחישוב שביצענו לא היה תוקף. אם למשל המנחה בחר את הוילון שחשף באקראי (ובאותה מידה היה עשוי לחשוף מכונית) הרי שסיכויי הזכיה לאחר החלפה היו נותרים שליש. אם המנחה היה נוהג לחשוף וילון אחר רק כאשר המשתתף בוחר בוילון שמאחוריו נסתרת המכונית, הרי שאין זה כדאי להחליף כלל וכלל, שכן, בהינתן שהמנחה הציע החלפה - הרי שהמכונית חייבת להיות מאחורי הוילון הראשון שנבחר. מורכבות מסוג זה, הנוגעת לניסוח עמום של בעיה, מבוארת בדוגמא הבאה.

דוגמא 2.16 (בעיית שני הילדים)**.**

הבעיה הבאה הוצגה על ידי מרטין גארדנר (Martin Gardner) ב-1959 כפרדוקס הסתברותי:

- למר ג׳ונס שני ילדים, אשר לפחות אחד מהם הוא בן, מה ההסתברות שגם השני הוא בן?
- למר ג׳ונס שני ילדים, הבכור הוא בן, מה ההסתברות שהצעיר אף הוא בן?

בהנחה שמר ג׳ונס הוא אדם מקרי בעל שני ילדים, הרי שמרחב ההסתברות הוא המרחב האחיד על $\{b,g\}^2$ כאשר b מייצג בן ואילו g מייצג בת, והסדר ביניהם מתאר את סדר לידתם. תחת פרשנות זו המאורע הראשון יכול להתפרש כ- $A = \{(b,b),(g,b),(b,g)\}$ ואילו השני כ- $B = \{(b,b),(b,g)\}$ נחשב ונקבל:

$$\mathbb{P}\Big(\{(b,b)\}\,\Big|\,A\Big) = 1/3, \qquad \mathbb{P}\Big(\{(b,b)\}\,\Big|\,B\Big) = 1/2.$$

אלא שנדמיין את המצב הבא בעולם האמיתי. אנו ניגשים לדלתו של בית ומקישים עליה. פתוח לנו את הדלת נער ואנו שואלים אותו כמה ילדים יש במשפחה. נדמיין שלוש תשובות אפשרויות: "שניים", "שניים, אני הבכור", "שניים, אני הצעיר". לכאורה, לפי הבעיה שפתרנו זה עתה, עלינו להסיק שהסתברות המאורע שיש לו בדיוק אח זכר אחד נוכח התשובה הראשונה היא שליש, ואילו נוכח שתי האחרות - חצי. כעת לכאורה נקבל סתירה לנוסחת

ההסתברות השלמה. נסמן $E =$ {הילד שלפנינו בן}, $F =$ {הילד שלפנינו בכור} ו- $M =$ {הילד האחר הוא בן} ונקבל

$$\frac{1}{3} \overset{?}{=} \mathbb{P}(M \mid E) = \mathbb{P}(M \mid F, E)P(F) + \mathbb{P}(M \mid E, F^c)(1 - \mathbb{P}(F)) \overset{?}{=} \frac{1}{2},$$

היכן הטעות?

תשובה: נוסחת ההסתברות השלמה נכונה, כמובן. מקור הטעות בכך שלא התייחסנו כאן לדרך שבה נבחר הילד המשיב לדלת ויצרנו עמימות. אם בהינתן לפחות בן אחד במשפחה ישיב לדלת בן מקרי, באמת נקבל מקרה התואם את המאורע הראשון של מרטין. במקרה זה $\mathbb{P}\Big(M \mid E\Big) = \mathbb{P}\Big(\{(b,b)\} \,\Big|\, A\Big)$ ו $\mathbb{P}(F \mid A) = 1/2$. נחשב:

$$\begin{aligned}
\frac{1}{3} = \mathbb{P}\Big(\{(b,b)\} \,\Big|\, A\Big) &\overset{\text{הסתברות שלמה}}{=} \mathbb{P}\Big(\{(b,b)\} \,\Big|\, A, F\Big)\mathbb{P}(F \mid A) + \mathbb{P}\Big(\{(b,b)\} \,\Big|\, A, F^c\Big)\mathbb{P}(F^c \mid A) \\
&\overset{\text{כלל בייס}}{=} \mathbb{P}\Big(F \,\Big|\, \{(b,b)\}, A\Big)\mathbb{P}\Big(\{(b,b)\} \,\Big|\, A\Big) + \mathbb{P}\Big(F^c \,\Big|\, \{(b,b)\}, A\Big)\mathbb{P}\Big(\{(b,b)\} \,\Big|\, A\Big) \\
&= (1/3)(1/2) + (1/3)(1/2) = \frac{1}{3}.
\end{aligned}$$

השימוש בכלל בייס במשוואה זו מתאים למקרה המותנה שחושב ב-(2.1)

מצד שני אם הילד המשיב נבחר באקראי נקבל מרחב הסתברות אחיד על $\Omega = \{b,g\}^2 \times \{1,2\}$ כאשר הקואורדינטה הנוספת מייצגת מי מהילדים השיב לדלת. במקרה כזה

$$\mathbb{P}\Big(M \,\Big|\, E\Big) = \mathbb{P}\Big(\{(b,b,\cdot)\} \,\Big|\, \{(x_1,x_2,i) \;:\; x_i = b\}\Big)$$

ונקבל

$$\begin{aligned}
\frac{1}{2} &= \mathbb{P}\Big(\{(b,b,\cdot)\} \,\Big|\, \{(x_1,x_2,i) \;:\; x_i = b\}\Big) \\
&= \mathbb{P}\Big(\{(b,b,1)\} \mid \{(b,\cdot,1)\}\Big)(1/2) + \mathbb{P}\Big(\{(b,b,2)\} \mid \{(\cdot,b,2)\}\Big)(1/2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.
\end{aligned}$$

□

העמימות לגבי הדרך בה מוכרע איזה מן הילדים פותח את הדלת היא שייצרה את הפרדוקס, כיוון שלא היה נהיר לנו מהו בדיוק המאורע בו אנו מתנים. מכאן נסיק כי **ניתוח הסתברותי דורש שתיאור הניסוי יהיה חד-משמעי**.

2.3 הסקה בייסיאנית והפרדת שתי השערות פשוטות

עיסוקה המרכזי של הסטטיסטיקה הוא שימוש בתורת ההסתברות לצורך עיצוב הערכות וניבויים. כלל בייס בשילוב עם נקודת המבט ההכרתית על תורת ההסתברות הולידו פרשנות של תורת ההסתברות אשר מיועדת לתכלית זו. בפרק זה נתאר צורת הסתכלות זו ונראה כיצד נוכל להשתמש בהסתברות מותנית כדי לעדכן את אמונותנו. בכדי להדגיש את ההסתכלות ההכרתית נשתמש במושג ה**סבירות** לתיאור הסתברותו של מאורע, מושג המגלם בתוכו מדד למידת ההפתעה. שאנו חשים נוכח התרחשות מסויימת.

המקרה הפשוט ביותר הוא מצב שבו נאספו נתונים ואנו מנסים להכריע אילו משתי תיאוריות הנוגעות לדרך שבה נתקבלו הנתונים היא הנכונה. במרבית השימושים שתי התיאוריות הללו אינן שוות משקל, אחת מהן היא עמדה נפוצה ואילו השניה נדירה או חדשנית. במצב זה מכנים את ההשערה הראשונה $\mathcal{H}_0$ - השערת האפס ואת רעותה $\mathcal{H}_1$ - ההשערה החלופית. להלן כמה דוגמאות.

- נאספים נתונים רפואיים של אדם במטרה לבדוק אם הוא חולה במחלה מסויימת.
  $\mathcal{H}_0$ - האדם אינו חולה במחלה, $\mathcal{H}_1$ - האדם חולה במחלה.
- נמדדת תעבורת רשת מחשבים כדי לנתר נסיונות פריצה.
  $\mathcal{H}_0$ - הרשת לא נפרצה, $\mathcal{H}_1$ - הרשת נפרצה.
- נבדקים נתוני דואר אלקטרוני כדי לבדוק אם מדובר בדואר-היצף (ספאם) או בהודעה תמימה.
  $\mathcal{H}_0$ - הודעה תמימה, $\mathcal{H}_1$ - הודעת היצף.
- חוקר מודד תוצאה של ניסוי פיסיקלי כדי לבדוק את נכונותה של תיאוריה חדשה.
  $\mathcal{H}_0$ - התיאוריה שגויה, $\mathcal{H}_1$ - התיאוריה נכונה.

הנחתה של תורת ההסקה הבייסיאנית היא שבטרם ערכנו את הניסוי הייתה לנו הערכה לגבי סבירותה של כל אחת מההנחות. כלומר הנחנו ש-$\mathcal{H}_0$ נכונה בהסתברות p_0 ואילו $\mathcal{H}_1$ נכונה בהסתברות $p_1 = 1 - p_0$. מנקודת מבט הסתברותית נראה זאת כניסוי דו-שלבי, כלומר כמרחב הסתברות על מרחב מדגם $\{\mathcal{H}_0, \mathcal{H}_1\} \times \Omega$ כאשר נתונות לנו מראש ההסתברויות $\mathbb{P}(\{\mathcal{H}_0, \omega\} \mid \{\mathcal{H}_0\} \times \Omega), \mathbb{P}(\{\mathcal{H}_1, \omega\} \mid \{\mathcal{H}_0\} \times \Omega)$.

כעת, בהינתן תוצאות הניסוי, נרצה להעריך מחדש את סבירותה של כל אחת מההשערות. כלומר נבקש לחשב את $\mathbb{P}(\mathcal{H}_0 \times \Omega \mid \{\mathcal{H}_0, \mathcal{H}_1\} \times \{\omega\})$.

נמחיש כעת את הגישה ההכרתית לכלל בייס בדוגמא ממחזהו של תום סטופארד "רוזנקרנץ וגילדרשטרן מתים". נקודת מבט זו היא בסיס הגישה הבייסיאנית להסקה סטטיסטית שתידון בהרחבה בפרק הפניה.

דוגמא 2.17 (רוזנקרנץ וגילדרשטרן)**.**
כאשר רוזנקרנץ מוציא מטבע מכיסו, גילדרשטרן מאמין שמדובר במטבע הוגן, אך מוכן לקבל כי בהסתברות של אחד למיליון מדובר במטבע מכושף המוציא תמיד את התוצאה עץ. לאחר מכן רוזנקרנץ מתחיל להטיל את המטבע ופעם אחר פעם התוצאה המתקבלת היא עץ. לאחר שחזה גילדרשטרן בשלושים הטלות מטבע, כיצד עליו לעדכן את הסבירות שהוא מייחס למאורע שמדובר במטבע מכושף?

תשובה: נשתמש במרחב ההסתברות $\Omega = \{\mathcal{H}_0, \mathcal{H}_1\} \times \{H, T\}^{30}$ כאשר $\mathcal{H}_1$ היא ההשערה שהמטבע מכושף ואילו $\mathcal{H}_0$ את ההשערה שהוא הוגן. כמו כן נסמן $A = \{\mathcal{H}_1\} \times \{H, T\}^{30}$ ו-$B = \{\mathcal{H}_0, \mathcal{H}_1\} \times \{H\}^{30}$.

להשקפתו של גילדרשטרן $\mathbb{P}(A) = 1/10^6$ ולכן גם $\mathbb{P}(A^c) = \frac{10^6-1}{10^6}$. כמו כן $\mathbb{P}(B \mid A) = 1$ ומפני שבשלושים הטלות מטבע הוגן יש 2^{30} סדרות תוצאות אפשריות שוות הסתברות, הרי ש-$\mathbb{P}(B \mid A^c) = 2^{-30}$. נחשב לפי נוסחת ההיפוך של בייס ונקבל שההסתברות המבוקשת היא

$$\mathbb{P}(A \mid B) \overset{\text{כלל בייס}}{=} \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} \overset{\text{הסתברות שלמה}}{=} \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c)} = \frac{\frac{1}{10^6}}{\frac{1}{10^6} + \frac{(10^6-1)}{10^6}2^{-30}} \approx 1 - \frac{1}{1073},$$

כלומר, מנקודת מבט הכרתית, לאחר שחזה בתופעה המופלאה של שלושים הטלות עץ רצופות, שינה גילדרשטרן את הערכתו וכעת הסבירות שהמטבע הוגן (A^c) הצטמצמה בעינייו לאחד לאלף. זוהי הדגמה למודל הבייסיאני לתהליך של שינוי האמונה בסברה נוכח מידע סטטיסטי חדש. □

כעת נציג מנקודת מבט שכיחותנית, דוגמא שבה למרות חשיפה נרחבת למדי לראיות נגדיות, לא משתנה הערכה הסבירה. זוהי דוגמא ראשונית ל**מבחן סטטיסטי**, נושא שאף הוא יזכה להרחבה בפרק הפניה.

דוגמא 2.18 (בדיקות סטטיסטיות לתופעה נדירה).

אדם מקרי ניגש לבדוק האם הוא נשא של וירוס מסוכן. ידוע שסיכויו של נבדק מקרי הוא כ-1% לשאת את המחלה. בדיקה חדשה שפותחה מזהה בטעות אדם שאינו נשא בתור נשא בהסתברות α (תוצאה חיובית כוזבת) ומזהה בטעות נשא בתור מי שאינו נשא בהסתברות β (תוצאה שלילית כוזבת). אדם נבדק ונמצא נשא. מה הסיכוי שהוא באמת נושא את הווירוס? יש להעריך בפרט לערכים $\alpha, \beta \in \{0, 0.05\}$.

תשובה: ראשית נגדיר את מרחב המדגם $\Omega = \{\mathcal{H}_0, \mathcal{H}_1\} \times \{+, -\}$ כאשר $\mathcal{H}_0$ מייצגת את המצב בו בדקנו אדם שאינו נשא, ואילו $\mathcal{H}_1$ את - המצב בו בדקנו נשא. נסמן $A = \{\mathcal{H}_1\} \times \{+, -\}$ ו- $B = \{\mathcal{H}_0, \mathcal{H}_1\} \times \{+\}$. נתון לנו כי $\mathbb{P}(B^c \mid A) = \beta$, $\mathbb{P}(B \mid A^c) = \alpha$ ו- $\mathbb{P}(A) = 0.01$. נחשב מנתונים אלו את $\mathbb{P}(B)$ ומתוכו את $\mathbb{P}(A \mid B)$ – הנתון המבוקש. לפי נוסחת ההסתברות השלמה

$$\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(B) = 0.01(1-\beta) + 0.99\alpha$$

ומכאן, לפי כלל בייס, נקבל

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{(1-\beta) \cdot 0.01}{0.01(1-\beta) + 0.99\beta}$$

כמובן שאם נציב $\alpha = \beta = 0$ נקבל $\mathbb{P}(A \mid B) = 1$, כלומר תוצאה אידיאלית. לעומת זאת, בהצבה של הסתברויות $\alpha = \beta = 0.05$ נקבל

$$\mathbb{P}(A \mid B) = 0.0095/0.059 \approx 1/6,$$

כלומר רק שישית מהנבדקים שנמצאו נשאים באמת נושאים את הוירוס!

תופעה דומה מתרחשת עבור $\beta = 0$ ו-$\alpha = 0.05$. ובמקרה זה נקבל $\mathbb{P}(A \mid B) = 0.01/0.0495 \approx 1/5$. לעומת זאת כאשר $\beta = 0$ נקבל $\mathbb{P}(A \mid B) = 0.01/0.01 = 1$. □

מדוגמא זו אנו למדים שבדיקה שמחזירה תוצאה חיובית כוזבת בסיכוי של אחוז אחד אינה מאבחנת ביעילות תופעה ששכיחותה נופלת בהרבה מאחד למאה. כדי להיטיב להבין את הקשר בין שכיחות תופעה וחשיבותן של השגיאות השונות הקורא החרוץ מוזמן לחשב את השפעתן של α ו-β בדוגמא לעיל על ההסתברות שאדם נשא של הווירוס בהינתן שתוצאת הבדיקה הייתה דווקא שלילית.

2.4 אי-תלות

בפרק הקודם עסקנו בהשפעת ההדדית של מאורעות אלה על אלה. ראינו שלעיתים בהינתן מאורע מסויים, הסתברותו של מאורע אחר משתנה והיא עשויה לקטון או לגדול, ואולם לא תמיד זהו המצב. נסתכל על השאלה הבאה:

"גד ויוקי מטילים כל אחד מטבע. בהינתן שליוקי יצא עץ, מה ההסתברות שגם לגד יצא עץ?"

ההגיון הפיסיקלי גורס שכיוון שלא התרחשו כל יחסי-גומלין בין המטבעות, הרי שאין כל חשיבות לתוצאת ההטלה של יוקי ולכן אם נסמן את המאורע - גד הטיל עץ ב-A ואת המאורע יוקי הטילה עץ ב-B, נקבל

$$\mathbb{P}(A \mid B) = \mathbb{P}(A).$$

במקרה כזה נאמר שהמאורעות **בלתי-תלויים**. אולם, הגדרה זו אינה נראית סימטרית, ואילו חוסר יחסי-הגומלין – דווקא כן. נְפַתֵּחַ את ההגדרה ונקבל $\mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A)$. נכפיל ב- $\mathbb{P}(B)$ ונקבל את ההגדרה הבאה.

הגדרה 2.19 (אי-תלות)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהיו $A, B \in \mathcal{F}$ שני מאורעות. נאמר כי A ו-B הנם **בלתי-תלויים (ב"ת)** ונסמן $A \perp\!\!\!\perp B$, אם

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

אחרת המאורעות יקראו **תלויים**.

✎ **בעיה 2.5.** להוכיח כי במרחב מכפלה $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mathbb{P}_1 \times \mathbb{P}_2)$ כל שני מאורעות מהצורה $A \times \Omega_2$ ו-$\Omega_1 \times B$ עבור $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$ הנם בלתי-תלויים. כלומר, **מאורעות שוליים** במרחב מכפלה הנם בלתי-תלויים.

דוגמא 2.20 (ההבדל בין היעדר יחסי-גומלין בין ניסויים לאי-תלות בין מאורעות)**.**

פאוסט בוחר באקראי ובאופן אחיד מספר בין 1 ל-10. נסמן ב-A את המאורע שהמספר קטן מחמש וב-E את המאורע שהוא זוגי. האם A ו-E בלתי-תלויים?

תשובה: גם אם שני המאורעות נוגעים לאותו אובייקט, אין הם בהכרח תלויים. נחשב אפוא. מדובר במרחב האחיד על $\Omega = [10]$.

$$\mathbb{P}(A) = \frac{|[4]|}{|[10]|} = \frac{4}{10} \qquad \mathbb{P}(E) = \frac{|\{2,4,6,8,10\}|}{|[10]|} = \frac{1}{2} \qquad \mathbb{P}(A \cap E) = \frac{|\{2,4\}|}{|[10]|} = \frac{2}{10}$$

ומסתבר שהמאורעות אומנם בלתי-תלויים. □

✎ **בעיה 2.6.** מוטלות שתי קוביות משחק הוגנות בנות שש פאות.

(א) האם המאורעות "בקוביות הראשונה יצא 3" ו-"סכום הקוביות יצא 7" בלתי-תלויים?

(ב) מה בנוגע למאורעות "בקוביות הראשונה יצא 3" ו-"סכום הקוביות יצא 6"?

אבחנה 2.21 (תכונות בסיסיות של אי-תלות)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהיו $A, B \in \mathcal{F}$ מאורעות ב"ת בעלי הסתברות חיובית, אזי

(א) A בלתי-תלוי ב- Ω וב-$\emptyset$.

(ב) $\mathbb{P}(A \mid B) = \mathbb{P}(A), \mathbb{P}(B \mid A) = \mathbb{P}(B)$.

(ג) $\mathbb{P}(A^c \cap B) = \mathbb{P}(A^c)\mathbb{P}(B)$. כלומר גם A^c ו-B ב"ת.

לצורך הוכחת סעיף ג' נשים לב כי $\mathbb{P}(A^c \cap B) = \mathbb{P}(B) - \mathbb{P}(A \cap B) = (1 - \mathbb{P}(A))\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B)$.

את מונח האי-תלות ניתן להכליל גם לקבוצה של מאורעות.

הגדרה 2.22 (אי-תלות של אוסף מאורעות)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות. אוסף מאורעות $\mathcal{A} \subset \mathcal{F}$ יקראו **בלתי-תלויים** אם לכל תת-קבוצה סופית שלהם $\{A_n\}_{n\in[N]}$ מתקיים

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) = \prod_{n\in\mathbb{N}} \mathbb{P}(A_n).$$

אבחנה 2.23. יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהי $\mathcal{A} \cup \{B\} \subset \mathcal{F}$ אוסף מאורעות בלתי-תלויים. אזי $\mathcal{A} \cup \{B^c\} \subset \mathcal{F}$ הוא אוסף מאורעות בלתי-תלויים.

דוגמא 2.24 (אי-תלות של מספר מאורעות)**.**

כל אחד מבין עשרה ילדים מטיל מטבע הוגן. נסמן ב-A_i את המאורע שהילד ה-i קיבל תוצאה של עץ ונסמן ב-B את המאורע שמספר הטלות הפלי בסך הכל היה זוגי. נראה כי $\{A_i \ : \ i \geq k\} \cup \{B\}$ הוא אוסף מאורעות בלתי-תלויים לכל $k > 1$ אך לא עבור $k = 1$.

תשובה: מרחב ההסתברות המתאים הוא חזקה עשירית של מרחב הטלת מטבע הוגן, הלוא הוא המרחב האחיד על $\Omega = \{H, T\}^{10}$. לכן, לכל $I \subset [10]$ לא ריקה מתקיים

$$\mathbb{P}\left(\bigcap_{i\in I} A_i\right) = 2^{-I} = \prod_{i\in I} \mathbb{P}(A_i),$$

$$\mathbb{P}\left(\left\{\omega_1, \ldots, \omega_{10} \ : \ |\{i \in I \ : \ \omega_i = T\}| \mod 2 = 0\right\}\right) = \frac{2^{|I-1|}}{2^{|I|}} = \frac{1}{2},$$

בפרט נסיק כי $\mathbb{P}(B) = \frac{1}{2}$. על סמך חישוב זה נקבל כי

$$\begin{aligned}\mathbb{P}\left(B \cap \bigcap_{i\in I} A_i\right) &= \mathbb{P}\left(\bigcap_{i\in I} A_i \cap \{\omega_1, \ldots, \omega_{10} \ : \ |\{i \notin I \ : \ \omega_i = T\}| \mod 2 = 0\}\right) \\ &= \mathbb{P}\left(\bigcap_{i\in I} A_i\right) \mathbb{P}\left(\{\omega_1, \ldots, \omega_{10} \ : \ |\{i \notin I \ : \ \omega_i = T\}| \mod 2 = 0\}\right),\end{aligned}$$

כאשר השוויון השני נובע מכך שמדובר במאורעות על קואורדינטות שונות במרחב מכפלה. עבור $[10] \setminus I \neq \emptyset$ נסיק כי

$$\mathbb{P}\left(B \cap \bigcap_{i\in I} A_i\right) = \mathbb{P}(B) \prod_{i\in I} \mathbb{P}(A_i)$$

ולכן $\{A_i \ : \ i \geq k\} \cup \{B\}$ הוא אוסף מאורעות בלתי-תלויים לכל $k > 1$. עבור $k = 1$ נשים לב כי

$$\mathbb{P}\left(B \cap \bigcap_{i\in[10]} A_i\right) = 2^{-10} \neq 2^{-11} = \mathbb{P}(B) \prod_{i\in[10]} \mathbb{P}(A_i).$$

ולכן המאורעות תלויים. □

✎ **בעיה 2.7.** להוכיח את אבחנה 2.23.

✎ **בעיה 2.8.** להוכיח כי אם מאורע A בלתי-תלוי בעצמו אז הסתברותו היא 0 או 1.

✎ **בעיה 2.9.** בדוגמא 1.10 נגדיר, לכל k, את המאורע $A_k = k\mathbb{N} = \{k, 2k, 3k, \ldots\}$.

האם המאורעות $\{A_k\}_{k\in\mathbb{N}}$ הם בלתי תלויים?

האם המאורעות $\{A_k\}_{k\in P}$, כאשר P היא קבוצת כל הראשוניים, הם בלתי תלויים?

טענה 2.25. יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהיו B, C ו-$\{A_n\}_{n\in[N]}$ מאורעות בלתי-תלויים. אזי $(B \cap C)$ ו-$\{A_n\}_{n\in[N]}$ הנם בלתי-תלויים וכן $(B \cup C)$ ו-$\{A_n\}_{n\in[N]}$ הנם בלתי-תלויים.

הוכחה. כדי לראות כי $(B \cap C)$ ו-$\{A_n\}_{n\in[N]}$ בלתי-תלויים נשים לב כי היות שהמאורעות B, C ו-$\{A_n\}_{n\in[N]}$ הנם בלתי-תלויים, הרי שמתקיים שלכל תת-קבוצה סופית $\{B_n\}_{n\in[N]}$ של $\{B\} \cup \{C\} \cup \{A_n\}_{n\in[N]}$ מתקיים

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} B_n\right) = \prod_{n\in\mathbb{N}} \mathbb{P}(B_n)$$

והדבר נכון בפרט לכל תת-קבוצה כזו המכילה הן את B והן את C. כדי לראות כי $(B \cup C)$ ו-$\{A_n\}_{n\in[N]}$ בלתי-תלויים נשתמש באבחנה 2.23 בכדי לקבל כי $\{B^c\} \cup \{C\} \cup \{A_n\}_{n\in[N]}$ ו-$\{B^c\} \cup \{C^c\} \cup \{A_n\}_{n\in[N]}$ הם אוספי מאורעות בלתי-תלויים. כעת, לפי אבחנה 2.23 נקבל כי $B^c \cap C^c, \{A_n\}_{n\in[N]}$ בלתי-תלויים וכי $(B^c \cap C^c)^c, \{A_n\}_{n\in[N]}$ הנם בלתי-תלויים. וכיוון שמתקיים $(B^c \cap C^c)^c = (B \cup C)$ נסיק את נכונות הטענה. ■

באופן דומה ניתן להראות כי עבור אוסף מאורעות $\mathcal{A}$ בלתי-תלויים, ועבור תת-קבוצה שלו $\mathcal{B} \subset \mathcal{A}$, מתקיים שכל סדרת פעולות של חיתוך ואיחוד של מאורעות ב-$\mathcal{B}$ תיצור מאורע שיהיה בלתי תלוי בכל המאורעות ב-$\mathcal{A} \setminus \mathcal{B}$. מעתה נרשה לעצמנו להשתמש בעובדה זו כמובנת מאליה.

אי-תלות הנה תכונה של אוסף מאורעות ולא מאפיין של היחסים בין זוגות המאורעות עצמם. כדי לחדד אבחנה זו נגדיר מושג של אי-תלות בזוגות.

הגדרה 2.26 (אי-תלות בזוגות). יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות. אוסף מאורעות $\mathcal{A} \subset \mathcal{F}$ יקראו **בלתי-תלויים בזוגות** אם כל זוג מאורעות $A, B \in \mathcal{A}$ הנם ב"ת.

בדוגמא הבאה נראה כי אי-תלות בזוגות **אינה** שקולה לאי-תלות.

דוגמא 2.27 (ההבדל בין אי-תלות לאי-תלות בזוגות).

פלורין מטיל שני מטבעות הוגנים, אשר תוצאות הטלותיהם בלתי-תלויות. נסמן ב-A את המאורע שיצאה בשניהם אותה התוצאה, ב-B_1 את המאורע שהראשון יצא עץ וב-B_2 את ההמאורע שהשני יצא פלי. האם המאורעות האלו בלתי-תלויים? האם הם בלתי-תלויים בזוגות?

תשובה: זהו מרחב אחיד $\Omega = [H, T]^2$.

$$\mathbb{P}(B_1) = \frac{|\{(H,H),(H,T)\}|}{|\Omega|} = \frac{1}{2} \qquad \mathbb{P}(A) = \frac{|\{(H,H),(T,T)\}|}{|\Omega|} = \frac{1}{2}$$

$$\mathbb{P}(A \cap B_1) = \frac{|\{(H,H)\}|}{|\Omega|} = \frac{1}{4} \qquad \mathbb{P}(B_2) = \frac{|\{(H,T),(T,T)\}|}{|\Omega|} = \frac{1}{2}$$

$$\mathbb{P}(B_1 \cap B_2) = \frac{|\{(H,T)\}|}{|\Omega|} = \frac{1}{4} \qquad \mathbb{P}(A \cap B_2) = \frac{|\{(T,T)\}|}{|\Omega|} = \frac{1}{4}$$

ומסתבר שהמאורעות אומנם בלתי-תלויים בזוגות. ואולם $\mathbb{P}(A \cap B_1 \cap B_2) = \mathbb{P}(\emptyset) = 0$, ולכן הם אינם בלתי-תלויים.

✎ **בעיה 2.10.** האם המאורעות B_{ij} בדוגמא 1.19 בלתי תלויים בזוגות? האם הם בלתי תלויים?

טענה 2.28 (אי-תלות של מאורעות מכפלה במרחב מכפלה). יהיו $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ ו-$(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ מרחבי הסתברות בדידים ויהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב המכפלה שלהם. יהיו $A \in \mathcal{F}_1$ ו-$B \in \mathcal{F}_2$. אזי המאורעות $A \times \Omega_2$ ו-$\Omega_1 \times B$ הנם בלתי-תלויים ב-$(\Omega, \mathcal{F}, \mathbb{P})$.

הוכחה. נחשב לפי תכונת מרחב המכפלה (טענה 1.15).

$$\mathbb{P}(A \times \Omega_2) = \mathbb{P}_1(A) \qquad \mathbb{P}(\Omega_1 \times B) = \mathbb{P}_1(B) \qquad \mathbb{P}(A \times B) = \mathbb{P}_1(A)\mathbb{P}_2(B)$$

כנדרש. ■

הגדרה 2.29 (אי-תלות של שני אוסף מאורעות). יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהיו $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ שני אוספי מאורעות. נאמר כי $\mathcal{A}$ ו-$\mathcal{B}$ **בלתי-תלויים** ונסמן $\mathcal{A} \perp\!\!\!\perp \mathcal{B}$, אם לכל שני מאורעות $A \in \mathcal{A}$ ו-$B \in \mathcal{B}$, מתקיים:

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

✎ **בעיה 2.11.** להראות שאוסף מאורעות $\mathcal{A}$ הנם בלתי-תלויים אם ורק אם לכל $A \in \mathcal{A}$ מתקיים כי A בלתי-תלוי באוסף המאורעות $\mathcal{A} \setminus A$.

✎ **בעיה 2.12.** את כל אחד משני צדדיו של מטבע הוגן החליטו לצבוע בלבן או בשחור (בהסתברות שווה ובאופן בלתי-תלוי). לאחר מכן המטבע הוטל פעמיים ופעמיים יצא צד לבן.

(א) מה ההסתברות ששני צדי המטבע לבנים?

(ב) מה ההסתברות שהמטבע נפל פעמיים על אותו צד?

בעיות הרחבה והעשרה

✎ **בעיה 2.13** (על פי לואיס קרול)**.** בכד אטום נמצא כדור אשר בהסתברות חצי הנו שחור ובהסתברות חצי לבן. הכניסו לכד כדור לבן ולאחר מכן הוציאו ממנו כדור מקרי. מה ההסתברות שהכדור שנותר בפנים לבן?

✎ **בעיה 2.14.** שלושה שופטים מכריעים את גורלו של נאשם על פי דעת רוב. שניים מהשופטים מנוסים ומזהים נכונה את אשמתו של הנאשם בסיכוי 90%. האחרון אינו מנוסה ומזהה נכונה את אשמתו של הנאשם בסיכוי 51%. השווה בין ההסתברויות לזיהוי נכון במקרים הבאים:

(א) החלטת כל שופט בלתי-תלויה

(ב) השופטים המנוסים מחליטים באופן בלתי-תלוי והשופט שאינו מנוסה בוחר אחד מהם באקראי ומצטרף לדעתו

✎ **בעיה 2.15.** על שולחן שלושה כדים אטומים ובכל אחד מהם – כדור אדום ושני כדורים שחורים. מעבירים כדור מקרי מכד א' לכד ב'. לאחר מכן מעבירים כדור מקרי מכד ב' לכד ג' ולאחר מכן מעבירים כדור מקרי מכד ג' לכד א'. לבסוף מוציאים כדור מכד א'. מה ההסתברות שהכדור שהוצא אדום?

✎ **בעיה 2.16** (פרדוקס האסירים)**.** שר האופים, שר המשקים ושר הטבעות יושבים בבית האסורים. אחד הסוהרים מבשר להם כי למחרת היום אחד מהם יוצא להורג והיתר יוצאו לחופשי. נניח כי לכל אחד מהם סיכוי שווה לעלות לגרדום.

(א) באישון לילה שר הטבעות מבקש מהסוהר שיגלה לו את שמו של אחד האסירים האחרים שיצא לחופשי, לטענתו אין בכך מידע לגביו משום שכבר ידוע שלפחות אחד משני האסירים האחרים ישתחרר. הסוהר מסרב מפני שלטענתו לאחר שיגלה את שמו של אסיר כזה - סיכוייו של כל אחד מהנותרים להיהרג יעלה לחצי. מי צודק?

(ב) בסופו של דבר מסכים הסוהר ומודיע לשר הטבעות כי שר המשקים יצא לחופשי. שר הטבעות ממהר ומודיע לשר האופים כי מצבם שניהם בכי רע כי כעת יש לכל אחד מהם סיכוי $\frac{1}{2}$ לקפח את חייהם. האם הוא צודק? אם לא, מי מהם סביר יותר להיות מוצא להורג?

✎ **בעיה 2.17 (*).** [פרדוקס סימפסון] בכל יום גלגמש ונמרוד מתחקים אחר צבי או יעל. הסיכוי שגלגמש יצוד את החיה אחריה התחקה הוא 80% ואילו סיכוייו של נמרוד לעשות כן הם 70%. זאת אף על פי שאם נמרוד מתחקה אחר צבי הוא צד אותו בהסתברות 1 ואילו גלגמש צד צבאים שהתחקה אחר עכבותיהם בהסתברות 7/8, ואילו אם נמרוד מתחקה אחר יעל ההסתברות שיצוד אותה היא 5/8 ואילו הסתברותו של גלגמש לעשות כן היא 1/2. כיצד ניתן להסביר (בכלים הסתברותיים) את העובדה שבכל אחד מסוגי הציד בנפרד מצטיין נמרוד ובכל זאת גלגמש זוכה בצידה רבה יותר?

✎ **בעיה 2.18 (*).** לגד, תלמיד בכיתה א', יש קבוצות חברי משחק שונות בכל יום מימות השבוע. המורה חילקה לכל תלמיד באופן בלתי תלוי ואחיד תג שם שחור או לבן. נסמן ב-A_i את המאורע שביום ה-i בקבוצה של גד יש מספר זוגי של תגים לבנים. האם כל שני מאורעות באוסף בהכרח בלתי-תלויים? מה באשר לכל שלושה? ומה עם כל ארבעה?

פרק 3

משתנים מקריים בדידים

בזמן שכתבתי את הספר [תהליכים סטוכסטים 1953] היה לי ויכוח עם פֶּלֶר. הוא אמר שכולם אומרים **משתנה מקרי** (random variable) ואילו אני אמרתי שכולם אומרים **משתנה אקראי** (chance variable). ברור היה שמוטב שנשתמש באותו שם בשני הספרים שלנו. החלטנו להכריע באופן מקרי. הטלנו מטבע והוא זכה.

– ג'וֹזף דוּב בשיחה עם לָאוּרִי סְנֶל, 1997.

בפרקים הקודמים הנחנו את יסודותיה של תורת ההסתברות על ידי כך שהגדרנו את מרחב ההסתברות וניסחנו את הכללים לחישוב הסתברותם של מאורעות הנוגעים לו. אומנם כלים אלו אפשרו לנו להשיב על שאלות רבות ולאפיין תופעות הסתברותיות שונות, אולם הפתרונות היו לעיתים כרוכים בסרבול רב. אחד מהגורמים לסרבול זה היה הצורך להכריע בין הגדרת מרחב הסתברות מורכב מאוד שיאפשר לנו להתמודד עם מגוון רחב של שאלות בעת ובעונה אחת לבין הגדרת מרחב הסתברות מצומצם ופשוט יותר, המתאים רק לשאלה מסויימת. גם כאשר הגדרנו מרחב מורכב - התקשינו לתאר בתוכו את התנהגותו של היבט מסויים של הניסוי ונדרשנו לסימונים בלתי-פורמליים כגון $(\cdot)$ ולמינים שונים של כתיבה מקוצרת. בפרק זה נציג מונח חדש – **משתנה מקרי**, אשר יאפשר לנו הן להתמודד באופן טוב יותר עם שאלות הסתברותיות כמותיות והן לחקור מגוון שאלות על אותו מרחב הסתברות ולנתח היבטים שונים שלו בנפרד זה מזה.

3.1 משתנה מקרי בדיד

הגדרה 3.1. יהי Ω מרחב מדגם. **משתנה מקרי בדיד (random variable)** על Ω הוא פונקציה $X:\Omega\to\mathbb{R}$.

דוגמא 3.2 (דוגמא למשתנים מקריים על מרחבי הסתברות מוכרים)**.**

(א) סכום שתי קוביות. מרחב המדגם – $\Omega=[6]^2$, המשתנה המקרי – $X(a,b)=a+b$

(ב) מספר העצים בהטלת שלושה מטבעות.
מרחב המדגם – $\Omega=\{H,T\}^3$, המשתנה המקרי – $X(\{a_1,a_2,a_3\})=|\{i\ :\ a_i=H\}|$.

(ג) מספר נקודות השבת בתמורה מקרית על $[N]$ איברים.
מרחב המדגם – $\Omega=S_N$, המשתנה המקרי – $X(\sigma)=|\{i\in[N]\ :\ \sigma(i)=i\}|$.

היות שמשתנה מקרי איננו אלא פונקציה בעלת ערכים ממשיים, הרי שנוכל להגדיר מספר משתנים מקריים על אותו מרחב הסתברות ולבצע עליהם פעולות שונות כאוות נפשנו ובפרט חיבור, כפל, הרכבה של פונקציה ממשית ועוד, כמתואר בדוגמא הבאה.

דוגמא 3.3 (פעולות על משתנים מקריים)**.**
גל שולה מן הים שלושה צדפים, שכל אחד מהם עשוי להיות לבן, שחור או צהוב. נתאר את שללה של גל באמצעות $\Omega = \{\text{לבן}, \text{שחור}, \text{צהוב}\}^3$ ונניח כי שוי כל צדף צהוב הוא שקל, של כל צדף שחור הוא שני שקלים ושל כל צדף לבן – חמישה שקלים. נגדיר שלושה משתנים מקריים שיתארו את מספר הצדפים מכל צבע.

$$\begin{aligned} X((\omega_1,\omega_2,\omega_3)) &= |\{\, i \in [3] \,:\, \omega_i = \text{צהוב}\}| \\ Y((\omega_1,\omega_2,\omega_3)) &= |\{\, i \in [3] \,:\, \omega_i = \text{שחור}\}|, \\ Z((\omega_1,\omega_2,\omega_3)) &= |\{\, i \in [3] \,:\, \omega_i = \text{לבן}\}|. \end{aligned}$$

אזי, הבאים הנם משתנים מקריים:

(א) W_1, מספר הצדפים הצהובים והשחורים: $W_1 = X + Y$.

(ב) W_2, שווים הכולל של הצדפים: $W_2 = X + 2Y + 5Z$.

(ג) W_3, מספר סוגי הצדפים שנשלו: $W_3 = \min(X,1) + \min(Y,1) + \min(Z,1)$.

לעיתים קרובות נרצה להגדיר משתנה מקרי שיהווה אבן בוחן להתרחשותו של מאורע במרחב ההסתברות.

הגדרה 3.4 (משתנה מציין)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות ויהי מאורע $A \in \mathcal{F}$. נכנה בשם **משתנה מציין (indicator) של** A את המשתנה המקרי $X = \mathbb{1}(A) = \mathbb{1}_A$.

כעת נוכל לתאר את הדוגמאות בהן נתקלנו בפרקים הקודמים בשפה של משתנים מקריים.

דוגמא 3.5 (תיאור בעיית ההתאמה (סוגיה 3) במונחי משתנים מקריים)**.**
נתאר ניסוי ובו N מכתבים ממוענים מוכנסים באקראי לתוך N מעטפות מבוילות באמצעות מרחב המדגם $\Omega = S_N$ של כל התמורות על N איברים. נוכל לתאר את המעטפה לתוכה הוכנס המכתב ה-i באמצעות המשתנה המקרי $X_i(\sigma) := \sigma(i)$ לכל $\sigma \in \Omega$. כמו כן נוכל להגדיר משתנה מציין Y_i לכך שהמכתב ה-i הוכנס למעטפה ה-i (המתאימה לו).

$$Y_i := \mathbb{1}(\{\sigma \,:\, \sigma(i) = i\}) = \mathbb{1}(\{\sigma \,:\, X_i(\sigma) = i\}),$$

בנוסף נוכל לתאר את מספר המכתבים שהוכנסו למעטפה המתאימה באמצעות המשתנה המקרי

$$Y := |\{i \,:\, X_i = i\}| = \sum_{i=1}^{n} Y_i.$$

נשים לב כי בדוגמא זו ביטאנו את המשתנה מקרי אשר ביקשנו לנתח כסכום של משתנים מקריים אחרים. בהמשך רעיון זה יאפשר לנו לנתח משתנים אלו בנפרד ולהסיק מסקנות לגבי סכומים.□

3.1.1 התפלגותו של משתנה מקרי

הגדרה 3.6 (פונקציית התפלגות נקודתית של משתנה מקרי בדיד)**.** נגדיר את **פונקציית ההתפלגות הנקודתית (probability mass function)** של משתנה מקרי בדיד X על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$ להיות הפונקציה $p_X : \mathbb{R} \to [0,1]$ הנתונה על ידי

$$p_X(s) = \mathbb{P}\Big(X^{-1}(s)\Big),$$

כאשר $X^{-1}(s) := \{\omega \in \Omega \ : \ X(\omega) = s\}$.
נכנה את הקבוצה $\mathrm{Supp}(X) = \{x \in \mathbb{R} \ : \ p_X(x) > 0\}$ בשם **התומך** של X.

נשים לב שכיוון ש-$\mathbb{P}$, לפי הגדרת מרחב הסתברות בדידה, נתמכת על קבוצה בת-מניה, הרי שהתומך של X אף הוא לכל היותר בן-מניה. נרחיב את ההגדרה הנקודתית להגדרה של פונקציית התפלגות כללית שתפעל על קבוצות ב-$\mathbb{R}$ ולא רק על איברים בודדים.

הגדרה 3.7 (התפלגות משתנה מקרי בדיד)**.** יהי X משתנה מקרי בדיד על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$. נגדיר את $\mathbb{P}_X$, **התפלגותו של** X, **(distribution)** להיות פונקציה $\mathbb{P}_X : 2^{\mathbb{R}} \to [0,1]$ הנתונה על ידי,

$$\mathbb{P}_X(A) := \mathbb{P}(X^{-1}(A)) = \sum_{\omega : X(\omega) \in A} \mathbb{P}(\{\omega\}) = \sum_{s \in A} p_X(s).$$

כך נרשום $X \sim \mathbb{P}_X$ ונאמר כי X מתפלג לפי $\mathbb{P}_X$. כמו כן נסמן $\mathbb{P}(X \in A) = \mathbb{P}_X(A)$ ונאמר כי זו היא **ההסתברות ש-X מקבל ערך ב-A**. נאמר כי X **נתמך** על A אם $\mathbb{P}(X \in A) = 1$.

כך נפרש עבור $A, B \subset \mathbb{R}$ את הכתיב $\mathbb{P}(X \in A \mid X \in B)$, לאור הגדרה 2.2, כ-$\mathbb{P}(X \in A \cap B)/\mathbb{P}(X \in B)$.

כאשר נעסוק במאורעות הנוגעים למשתנים מקריים, נשמיט לעיתים קרובות את סימון הקבוצה. כך למשל

$$\mathbb{P}(X = 5) := \mathbb{P}_X(\{5\}) = \mathbb{P}(X \in \{5\}) = p_X(5),$$
$$\mathbb{P}(X \geq 5) := \mathbb{P}_X([5,\infty)) = \mathbb{P}(X \in [5,\infty)) = \sum_{x \in [5,\infty)} p_X(x)$$

דוגמא 3.8 (סכום הטלת שתי קוביות כמשתנה מקרי)**.**
שחקן מטיל שתי קוביות הוגנות. נסמן ב-X את המשתנה המקרי המתאר את סכום התוצאות. נבנה מרחב הסתברות מתאים, נגדיר את X וחשב את פונקציית ההתפלגות הנקודתית שלו, ואת ההסתברות ש-X בין 4 ל-7.

תשובה: מרחב ההסתברות המתאים לשאלה הוא המרחב האחיד על $[6]^2$. נגדיר $X((a,b)) = a + b$.
נדגים את חישוב $p_x(4)$:

$$p_x(4) = \mathbb{P}(\{\omega \in \Omega \ : \ X(\omega) = 4\}) = \mathbb{P}(\{(1,3),(2,2),(3,1)\}) = \frac{3}{36}$$

באופן דומה נקבל

x	2	3	4	5	6	7	8	9	10	11	12
$p_X(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

כעת נחשב:

$$\mathbb{P}(X \in \{4,5,6,7\}) = \sum_{n=4}^{7} p_X(n) = \frac{3}{36} + \frac{4}{36} + \frac{5}{36} + \frac{6}{36} = \frac{1}{2}.$$

□

3.1.2 התפלגותו של משתנה מקרי כפונקציית הסתברות על $\mathbb{R}$.

נוכל לראות את התפלגותו של משתנה מקרי בתור **דחיפה קדימה** (forward push) של פונקציית ההסתברות שבחרנו אל מרחב המדגם $\mathbb{R}$ באופן שמגדיר עליו מרחב הסתברות בדידה חדש.

אבחנה 3.9 (התפלגותו של משתנה מקרי היא פונקציית הסתברות)**.** יהי X משתנה מקרי על מרחב הסתברות כלשהו. אזי p_X הנה פונקציית הסתברות נקודתית על $\mathbb{R}$ ו-$\mathbb{P}_X$ היא פונקציית ההסתברות המתאימה לה לפי טענה 1.5.

✎ **בעיה 3.1.** להוכיח את אבחנה 3.9.

בראיה זו $\mathbb{P}_X(A)$ מחושבת על ידי משיכה של A למרחב Ω באמצעות ההעתקה X^{-1}. העתקה זו מתאימה לקבוצה A את אוסף כל המקורות שלה לפי X. כעת נקבע כי $\mathbb{P}_X(A)$ היא הסתברותה של קבוצת מקורות זו במרחב ההסתברות המקורי $(\Omega, \mathcal{F}, \mathbb{P})$.

יתר על כן – נוכל לממש כל התפלגות בדידה על $\mathbb{R}$ כמשתנה מקרי.

טענה 3.10 (קיום משתנה מקרי בעל התפלגות נתונה)**.** אם $\mathcal{D}$ פונקציית הסתברות בדידה על $(\mathbb{R}, 2^{\mathbb{R}}, \mathcal{D})$. אזי קיים מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$ ומשתנה מקרי X כך ש-$X \sim \mathcal{D}$.

הוכחה. נבחר $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{F}, \mathcal{D})$ ונגדיר משתנה מקרי $X(\omega) = \omega$. ניווכח כי $\mathbb{P}_X(\{\omega\}) = \mathcal{D}(\{\omega\})$ ולכן, לפי טענה 1.5, $\mathbb{P}_X = \mathcal{D}$. ■

3.1.3 התפלגויות בסיסיות

עד עתה, כאשר רצינו לתאר ניסוי מסויים, תארנו אותו באמצעות מרחב הסתברות. כך עשינו למשל עבור ניסוי ברנולי בדוגמא 1.7. ואולם בהמשך, הניסוי אותו נבקש לתאר יהיה מורכב למדי, ונבקש לטפל בהיבטים שונים שלו בנפרד. לצורך כך נבקש לאפיין את התפלגותם של משתנים מקריים שונים במסגרת אותו ניסוי. כך למשל נוכל לקבוע כי "במסגרת ניסוי של הטלת שתי קוביות הוגנות, זוגיותו של סכום הקוביות תתפלג כמו ניסוי ברנולי עם סיכוי הצלחה חצי", או כי "סכום הקוביות מודולו שש מתפלג כמו הטלת קוביה בודדת". לצורך כך נגדיר התפלגויות שתתאמנה למשתנים מקריים נפוצים כאלה. לאור אבחנה 3.9, נשים לב שהגדרת **התפלגות של משתנה מקרי בדיד** היא למעשה הגדרת **פונקציית הסתברות בדידה** על $\mathbb{R}$.

נפתח בהצגה של משפחת ההתפלגויות הפשוטות ביותר – התפלגויות ברנולי.

הגדרה 3.11 (התפלגות ברנולי)**.** נאמר כי משתנה מקרי X מתפלג לפי **התפלגות ברנולי עם סיכוי הצלחה** p (בקיצור – ברנולי-p) ונכתוב $X \sim \mathrm{Ber}(p)$ אם $p_X(1) = p$ ו-$p_X(0) = 1 - p$.

זוהי משפחה של התפלגויות, כלומר לכל $p \in [0,1]$ נקבל התפלגות אחרת. נשים לב שלמשתנה מציין יש התפלגות ברנולי, וכל משתנה ברנולי X הוא משתנה מציין של הקבוצה $X^{-1}(1)$.

דוגמא 3.12 (מציאת משתנה בעל התפלגות מסויימת במרחב הסתברות)**.** עליסה והמלכה משחקות במשחק הבא: בכד אטום שני כדורים לבנים ושלושה כדורים אדומים. ארבעה כדורים נשלפים בזה אחר זה. אם יצאו יותר כדורים אדומים - המלכה מנצחת ואחרת עליסה מנצחת. נגדיר משתנה מציין למאורע "המלכה ניצחה". נראה שזהו משתנה ברנולי $2/5$.

תשובה: נשתמש במרחב ההסתברות על תמורות בגודל 5 כאשר נתייחס למספרים 1 ו-2 כמייצגים כדורים לבנים וליתר ככדורים אדומים. נשים לב שהמאורע המבוקש ניתן לתיאור באופן הבא:

$$A = \{S \in \Omega \ : \ S_5 \in [2]\}$$

ולפי הגדרה

$$\mathbb{P}(X) = \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{2 \cdot 4!}{5!} = \frac{2}{5}.$$

□

כפי שתיארנו באמצעות התפלגות ברנולי את מספר העצים בניסוי הטלת המטבע, כך נגדיר את משפחת ההתפלגויות האחידות, אשר תתאים לניסויים מספריים המתוארים על ידי מרחב הסתברות אחיד, כמו למשל תוצאת הטלת קוביה.

הגדרה 3.13 (התפלגות אחידה)**.** נאמר שמשתנה מקרי X מתפלג לפי **התפלגות אחידה על קבוצה סופית** $A \subset \mathbb{R}$ **(uniform distribution)** ונכתוב $X \sim \mathrm{Unif}(A)$ אם $\forall i \in A : p_X(i) = \frac{1}{|A|}$.

על פי רוב נתעניין במשתנים בדידים בעלי התפלגות אחידה על $[n]$ ולכן נקצר $\mathrm{Unif}(n) := \mathrm{Unif}([n])$.

✎ **בעיה 3.2.** לחשב את התפלגות סכומן של שלוש קוביות הוגנות.

✎ **בעיה 3.3.** להגדיר על מרחב הסתברות $\Omega = [5]$ שני משתנים מקריים X ו-Y כך שיתקיים:

$$X, Y, X + Y \sim \mathrm{Unif}(\{-2, -1, 0, 1, 2\}).$$

3.2 יחסים בין מספר משתנים מקריים

עיקר התועלת שבשימוש במושג המשתנה המקרי מתבטאת בכך שהוא מאפשר לנו לתאר מערכות יחסים בין משתנים מקריים שונים. בפרק זה נתאר שני סוגים של מערכות יחסים כאלה - מערכות יחסים בין **משתנים מקריים המוגדרים על אותו מרחב הסתברות**, ולעומתן מערכות יחסים בין **התפלגויות של משתנים מקריים**. לצורך כך נתחיל ונשאל למה כוונתנו כאשר אנו אומרים ששני משתנים מקריים שווים.

3.2.1 סוגי שוויון בין משתנים מקריים

מתי שני משתנים מקריים שווים זה לזה? לפי הגדרה 3.1, משתנים מקריים הם פונקציות ממשיות המוגדרות על מרחב מדגם Ω. לפיכך באופן פורמלי עבור שני משתנים מקריים X, Y על אותו מרחב מדגם, השוויון $X \equiv Y$ (קרי: X ו-Y שווים תמיד), הוא בפשטות שוויון בין פונקציות. אלא שעד מהרה נגלה ששוויון זה אינו שימושי לצרכינו. נסתכל על הדוגמא הבאה:

דוגמא 3.14. על קולב תלויות שתי מגבעות שחורות ומגבעת לבנה. בזה אחר זה כל אחד מבין שלושה אדונים חובש מגבעת מקרית מהקולב. נגדיר שני משתנים מקריים: X שיתאר את מספר המגבעות השחורות על ראשם של שני האדונים הראשונים ו-Y שיתאר את מספר המגבעות השחורות על ראשו של האדון השלישי. מובן שנצפה כי סכומם של X ו-Y יהיה 2.

נבחר את מרחב המדגם לתיאור הבעיה להיות $\Omega = \{\text{שחור, לבן}\}^3$ ואת פונקציית ההסתברות הנקודתית להיות מוגדרת באמצעות

$$\mathbb{P}(\{(a_1, a_2, a_3)\}) = \begin{cases} \frac{1}{3} & |\{i \ : \ a_i = \text{לבן}\}| = 1 \\ 0 & \text{אחרת} \end{cases}$$

נגדיר

$$X((a_1, a_2, a_3)) = |\{i \in [2] \ : \ a_i = \text{שחור}\}|$$
$$Z((a_1, a_2, a_3)) = 2 - |\{i \in \{3\} \ : \ a_i = \text{שחור}\}|$$

ונשאל:

(א) האם $X \equiv Z$?

(ב) מה ההסתברות ש-$X = Z$?

תשובה: נשים לב כי $X((1,1,1)) = 2$ ואילו $Z((1,1,1)) = 1$ ולכן $X \not\equiv Z$. מצד שני, נחשב

$$\mathbb{P}(X = Z) = \mathbb{P}\Big(\Big\{\omega \ : \ X(\omega) = Z(\omega)\Big\}\Big) = \mathbb{P}\Big(\Big\{(a_1, a_2, a_3) \ : \ |\{i \ : \ a_i = \text{לבן}\}| = 1\Big\}\Big) = 1$$

כיוון ש-$Z \equiv 2 - Y$ נקבל כי $\{\omega \ : \ X(\omega) + Y(\omega) = 2\}$ הוא מאורע בעל הסתברות 1 כפי שציפינו. □

אבחנה זו מזמינה אותנו להגדיר את מושג השוויון הבא:

הגדרה 3.15 (שוויון כמעט-תמיד)**.** יהיו X, Y שני משתנים מקריים על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נאמר כי X ו-Y שווים **כמעט-תמיד (almost-surely)**, ונסמן $X \overset{\text{a.s.}}{=} Y$ אם מתקיים

$$\mathbb{P}(X = Y) = \mathbb{P}(\{\omega \ : \ X(\omega) = Y(\omega)\}) = 1$$

במילים אחרות X ו-Y שווים כמעט תמיד אם הם שווים כפונקציות על קבוצה שהסתברותה 1, או לחילופין - אם הם שונים זה מזה על קבוצה שהסתברותה 0. בהקשר של מרחב הסתברות, שוויון כמעט-תמיד שימושי בהרבה משוויון תמיד. זאת כיוון שהוספת ערכים למרחב המדגם לצורך פישוט הכתיבה היא פעולה שרירותית והערכים שמשתנה מקרי מתאים לערכי-סרק אלה של מרחב המדגם הנה חסרת חשיבות. באופן דומה נוכל לכתוב $X \overset{\text{a.s.}}{\geq} Y$ אם מתקיים $\mathbb{P}(X \geq Y) = \mathbb{P}(\{\omega \ : \ X(\omega) \geq Y(\omega)\}) = 1$, וכן $X \overset{\text{a.s.}}{>} Y$ אם מתקיים $\mathbb{P}(X > Y) = \mathbb{P}(\{\omega \ : \ X(\omega) > Y(\omega)\}) = 1$. שוויון כמעט-תמיד נשמר תחת הפעלה של פונקציות.

אבחנה 3.16 (שוויון כמעט-תמיד נשמר תחת הפעלת פונקציה)**.** יהיו X, Y משתנים מקריים כך ש-$X \overset{\text{a.s.}}{=} Y$, ותהי $f : \mathbb{R} \to \mathbb{R}$ אזי $f(X) \overset{\text{a.s.}}{=} f(Y)$.

✎ **בעיה 3.4.** להוכיח את אבחנה 3.16.

מובן אחר של שוויון בין משתנים מקריים הוא **שוויון התפלגויות**. מצב זה מתקיים כאשר ההסתברות שמשתנה מקרי יקבל כל ערך מסויים, זהה להסתברות שמשתנה אחר יקבלו. כך למשל, תוצאות ההטלה של שתי קוביות שונות הן שוות התפלגות, אך הן אינן שוות כמעט-תמיד (שכן בהסתברות חיובית מתקבלות תוצאות שונות).

הגדרה 3.17. אם לשני משתנים מקריים שונים X ו-Y (שעשויים להיות מוגדרים על שני מרחבי הסתברות שונים) יש את אותה התפלגות, נאמר כי הם **שווי התפלגות** ונכתוב $X \overset{\text{d}}{=} Y$.

טענה 3.18 (שוויון התפלגויות נשמר תחת הפעלת פונקציה)**.** יהיו X, Y משתנים מקריים שווי התפלגות (לאו דווקא על אותו מרחב הסתברות), ותהי $f : \mathbb{R} \to \mathbb{R}$ אזי $f(X) \overset{\text{d}}{=} f(Y)$.

הוכחה. נבטא את פונקציות ההתפלגות הנקודתיות $p_{f(X)}, p_{f(Y)}$ מבלי להזדקק למרחב ההסתברות.

$$p_{f(X)}(x) = \mathbb{P}(X \in f^{-1}(\{x\})) = \mathbb{P}(Y \in f^{-1}(\{x\})) = p_{f(Y)}(x),$$

כאשר השוויון האמצעי נובע מכך ש-$X \overset{\text{d}}{=} Y$. ■

✎ **בעיה 3.5.** להוכיח כי אם X, Y הנם משתנים מקריים בדידים על $(\Omega, \mathcal{F}, \mathbb{P})$ כך ש- $X \overset{\text{a.s.}}{=} Y$ אז $X \overset{\text{d}}{=} Y$.

✎ **בעיה 3.6** (קוביות סיקרמן Sicherman dice)**.** נתונות זוג קוביות הוגנות בעלות שש פאות. על פאות קוביה אחת כתובים המספרים $\{1, 2, 2, 3, 3, 4\}$ ועל פאות השניה $\{1, 3, 4, 5, 6, 8\}$. יש להראות שהתפלגות סכום שתי הקוביות שווה להתפלגות סכומן של שתי קוביות משחק רגילות.

3.2.2 וקטור מקרי והתפלגות משותפת

דוגמא 3.19. מוטלות שתי קוביות משחק הוגנות. נסמן ב-X_1 את תוצאת הקוביה הראשונה וב-X_2 את תוצאת הקוביה השניה. נסמן $Z = X_1 + X_2$ ו-$Y = X_1 + X_1$. האם אלו משתנה מקריים שווי התפלגות?

תשובה: מבחינה אינטואיטיבית - נהיר לנו ש-Y, שהנו כפליים ערכה של הטלת קוביה, מקבל ערכים זוגיים בלבד, ואילו Z, שהנו סכום הטלות של שתי קוביות - יכול לקבל גם ערכים אי-זוגיים ולכן Y ו-Z אינם שווי-התפלגות. נציג נימוק זה באופן מלא על ידי כך שנראה כי הסתברותו של Y לקבל את הערך 3 היא אפס, בניגוד להסתברות ש-Z יקבל ערך זה. נבחר כמרחב הסתברות את המרחב האחיד על $[6]^2$. נגדיר את X_1 ו-X_2 בהתאם על ידי

$$X_1((a,b)) = a \qquad X_2((a,b)) = b.$$

כעת נחשב לפי נוסחת ההסתברות השלמה

$$\mathbb{P}(Z=j)=\sum_{i=1}^{6}\mathbb{P}(X_1=i)\mathbb{P}(X_2=j-i\mid X_1=i)=\sum_{i=1}^{6}\frac{\mathbb{1}(\{j-i\in[6]\})}{36}=\frac{6-|7-j|}{36}.$$

כך למשל

$$\mathbb{P}(Z=3)=\mathbb{P}(X_1=1)\mathbb{P}(X_2=2)+\mathbb{P}(X_1=2)\mathbb{P}(X_2=1)=\frac{2}{36}\neq 0=\mathbb{P}(Y=3)$$

ולכן $Y\overset{\mathrm{d}}{\neq}Z$. □

בדוגמא 3.19 ראינו כי המידע הנוגע ליחס בין שני משתנים מקריים על אותו מרחב הסתברות אינו מצוי בהתפלגותם. למרות שהתפלגותם של X_1 ושל X_2 זהה, שני הסכומים X_1+X_1 ו-X_1+X_2 אינם שווי-התפלגות. כיוון שאנחנו מעוניינים להימנע מתיאור מפורש של מרחבי הסתברות כאשר אנו עוסקים במשתנים מקריים, נגדיר את מושג ה**וקטור המקרי** שיתאר אוסף סופי של משתנים מקריים ואת התפלגותם המשותפת של משתנים אלה. ההתפלגות המשותפת תתפוס את מלא מערכת היחסים בין המשתנים המקריים.

הגדרה 3.20. אוסף משתנים מקריים בדידים $(X_k)_{k\in[n]}$ על מרחב הסתברות ייקרא **וקטור מקרי בדיד**.

הגדרה 3.21. יהי $\{X_k\}_{k\in[n]}$ וקטור מקרי בדיד על מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$. נכנה בשם **התפלגות המשותפת (joint distribution)** של $\{X_k\}_{k\in[n]}$ את הפונקציה $\mathbb{P}_{X_1,\ldots,X_n}$ מתת-קבוצות של $\mathbb{R}^n$ ל-$[0,1]$ המוגדרת לכל A על ידי

$$\mathbb{P}_{X_1,\ldots,X_n}(A):=\mathbb{P}\Big((X_1,\ldots,X_n)\in A\Big)=\mathbb{P}\left(\{\omega\in\Omega:(X_1(\omega),\ldots,X_n(\omega))\in A\}\right)$$

כדי לתאר התפלגות משותפת של מספר משתנים מקריים בדידים די לתאר את פונקציית ההתפלגות הנקודתית שלהם המוגדרת כדלהלן.

הגדרה 3.22. $\{X_i\}_{i\in[N]}$ וקטור מקרי בדיד על מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$. **פונקציית ההתפלגות המשותפת הנקודתית (joint probability mass function)** של $\{X_k\}_{k\in[n]}$ היא הפונקציה $p_X:\mathbb{R}^n\to[0,1]$ שנתונה על ידי

$$p_{X_1,\ldots X_n}(x_1,\ldots,x_n):=\mathbb{P}(X_1=x_1,\ldots,X_n=x_n)=\mathbb{P}\left(\bigcap_{i=1}^{n}X_k^{-1}(x_k)\right).$$

ואכן, את פונקציית ההתפלגות המשותפת אפשר לבטא בתור

$$\mathbb{P}_X=\mathbb{P}_{X_1,\ldots,X_n}(A)=\mathbb{P}((X_1,\ldots,X_n)\in A)=\sum_{(x_1,\ldots,x_n)\in A}p_{X_1,\ldots,X_n}((x_1,\ldots,x_n)).$$

ההתפלגות המשותפת (הכללית והנקודתית) איננה אלא הכללה של מושג המשתנה המקרי למרחב $\mathbb{R}^N$. כאשר נתונה התפלגות משותפת, נכנה בשם **התפלגות שולית** את התפלגותם בנפרד של כל אחד מהמשתנים. ונוכל לחשב באמצעות נוסחת ההסתברות השלמה

$$\mathbb{P}_{X_k}(A)=\mathbb{P}(X_k\in A)=\sum_{(x_1,\ldots,x_n)\in\mathbb{R}^{k-1}\times A\times\mathbb{R}^{n-k}}p_{X_1,\ldots,X_n}((x_1,\ldots,x_n)).$$

מעתה בכל פעם שנעסוק במספר משתנים מקריים בעלי התפלגות משותפת, תהיה מובלעת בתיאור זה הקביעה כי כולם מוגדרים על אותו מרחב הסתברות.

את מושג שוויון ההתפלגויות נכליל לוקטורים מקריים באופן הבא.

הגדרה 3.23. יהיו X ו-Y שני וקטורים מקריים (שעשויים להיותר מוגדרים על שני מרחבי הסתברות שונים). אם $\mathbb{P}_X = \mathbb{P}_Y$ נאמר כי שני הוקטורים **שווי התפלגות** ונכתוב $X \overset{\mathrm{d}}{=} Y$.

דוגמא 3.24. מוטלות שתי קוביות משחק סטנדרטיות. נסמן ב-X_1, X_2 את תוצאת כל אחת מהקוביות. כמו כן נסמן ב-Y את התוצאה הנמוכה מבין השתיים וב-Z את התוצאה הגבוהה מביניהן. יש לחשב את התפלגותם המשותפת של Y ו-Z ואת ההתפלגויות השוליות.

תשובה: מרחב ההסתברות המתאים לבעיה הוא המרחב האחיד על $\Omega = [6]^2$. נחשב את התפלגותם של X_1 ו-X_2 על ידי מעבר על כל האפשרויות. בכל משבצת יצויין "ערך גבוה-ערך נמוך".

X_2 \ X_1	1	2	3	4	5	6
1	1-1	1-2	1-3	1-4	1-5	1-6
2	1-2	2-2	2-3	2-4	2-5	2-6
3	1-3	2-3	3-3	3-4	3-5	3-6
4	1-4	2-4	3-4	4-4	4-5	4-6
5	1-5	2-5	3-5	4-5	5-5	5-6
6	1-6	2-6	3-6	4-6	5-6	6-6

לכל אחד מהתאים הסתברות של $1/36$. נבחן ונגלה שהתפלגותם המשותפת של Y ו-Z הנה כמתואר להלן

Y \ Z	1	2	3	4	5	6	Y שולי
1	1/36	2/36	2/36	2/36	2/36	2/36	11/36
2	0	1/36	2/36	2/36	2/36	2/36	9/36
3	0	0	1/36	2/36	2/36	2/36	7/36
4	0	0	0	1/36	2/36	2/36	5/36
5	0	0	0	0	1/36	2/36	3/36
6	0	0	0	0	0	1/36	1/36
Z שולי	1/36	3/36	5/36	7/36	9/36	11/36	1

בטבלה מתוארות גם התפלגויותיהם השוליות של Y ו-Z. את התפלגותו השולית של Y למשל, אנו מחשבים לפי נוסחת ההסתברות השלמה תוך סכימת הסתברותן של כל הקונפיגורציות המתאימות לערך Y מסויים (כלומר סכימת השורה המתאימה בטבלה). □

כשם שהתפלגותו של משתנה מקרי בדיד היוותה פונקציית הסתברות בדידה על $\mathbb{R}$, כך התפלגות משותפת של וקטור מקרי בדיד באורך N מהווה פונקציית הסתברות בדידה על $\mathbb{R}^n$.

אבחנה 3.25 (התפלגות משותפת היא פונקציית הסתברות)**.** יהי $(X_1,\dots,X_n)$ וקטור מקרי בדיד על מרחב הסתברות כלשהו. אזי $p_{X_1,\dots,X_n}$ הנה פונקציית הסתברות נקודתית על $\mathbb{R}^n$ ו-$\mathbb{P}_{X_1,\dots,X_n}$ היא פונקציית ההסתברות המתאימה לה לפי טענה 1.5.

✎ **בעיה 3.7.** להוכיח את אבחנה 3.25.

כמו במקרה של משתנה מקרי בודד, נוכל לממש כל פונקציית הסתברות על $\mathbb{R}^N$ כהתפלגות משותפת של N משתנים מקריים.

טענה 3.26 (קיום מערכת משתנים בעלי התפלגות נתונה)**.** לכל פונקציית הסתברות בדידה על $(\mathbb{R}^n, 2^{\mathbb{R}^n}, \mathcal{D})$ קיים מרחב הסתברות בדידה $(\Omega,\mathcal{F},\mathbb{P})$ ווקטור מקרי $(X_1,\dots,X_n)$ שזוהי פונקציית ההתפלגות המשותפת שלו.

✎ **בעיה 3.8.** להוכיח את אבחנה 3.26 בדומה לטענה 3.10.

כמו כן נכליל את מושגי השוויון שהגדרנו.

הגדרה 3.27 (שוויון כמעט-תמיד של וקטורים מקריים)**.** יהיו $\{X_k\}_{k\in n},\{Y_k\}_{k\in n}$ שני וקטורים מקריים על מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$. נאמר כי X ו-Y שווים **כמעט תמיד**, ונסמן $(X_1,\dots,X_n) \overset{\text{a.s.}}{=} (Y_1,\dots,Y_n)$ אם מתקיים

$$\mathbb{P}((X_1,\dots,X_n)=(Y_1,\dots,Y_n)) = \mathbb{P}(\{\omega : (X_1(\omega),\dots,X_n(\omega)) = (Y_1(\omega),\dots,Y_n(\omega))\}) = 1$$

הגדרה 3.28. נאמר כי $X=(X_1,\dots,X_n)$ ו-$Y=(Y_1,\dots,Y_n)$, שני וקטורים מקריים (שעשויים להיותר מוגדרים על שני מרחבי הסתברות שונים), הנם שווי התפלגות ונכתוב $X \overset{\text{d}}{=} Y$, אם התפלגותם המשותפת זהה.

נשים לב שהן שוויון כמעט תמיד והן שוויון התפלגויות תלויים רק בהתפלגות המשותפת של המשתנים המקריים המעורבים, ולא במשתנים עצמם או במרחב ההסתברות עליו הם מוגדרים.

אבחנה 3.29. יהיו $X=(X_1,\dots,X_n)$ ו-$Y=(Y_1,\dots,Y_n)$ שני וקטורים מקריים בדידים על מרחב הסתברות. אזי $X \overset{\text{a.s.}}{=} Y$ אם ורק אם, הקבוצה $A=\{(a_1,\dots,a_n,a_1,\dots,a_n) : a_1,\dots,a_n\in\mathbb{R}^n\}$ מקיימת

$$\mathbb{P}((X_1,\dots,X_n,Y_1,\dots,Y_n)\in A)=1.$$

כמו כן שני סוגי השוויון נשמרים תחת הפעלה של פונקציות.

אבחנה 3.30. יהיו $X=(X_1,\dots,X_n)$ ו-$Y=(Y_1,\dots,Y_n)$ שני וקטורים מקריים ותהי $f:\mathbb{R}^n\to\mathbb{R}^m$ פונקציה כלשהי. אזי

(א) אם $X \overset{\text{a.s.}}{=} Y$ אז $f(X) \overset{\text{a.s.}}{=} f(Y)$.

(ב) אם $X \overset{\text{d}}{=} Y$ אז $f(X) \overset{\text{d}}{=} f(Y)$.

✎ **בעיה 3.9.** להוכיח את אבחנה 3.30.

מאבחנה 3.30 אנו למדים, למשל, שהתפלגותו של סכום משתנים מקריים תלויה רק בהתפלגותם המשותפת.

לסיכום: בזכות אבחנה 3.26 נוכל לתאר את ניסויינו מעתה באמצעות מערכת של משתנים מקריים מבלי לתאר במפורש את מרחב הסתברות עליו הם מוגדרים. בזכות אבחנה 3.30ב, נוכל לחשב את התפלגותם של משתנים מקריים אשר נוצרים על ידי הפעלת פונקציות על משתנים מקריים אלו, וחישובינו יהיו תקפים לכל מרחב הסתברות שהוא.

3.2.3 אי-תלות בין משתנים מקריים

עד עתה שימש אותנו מושג האי-תלות בשתי צורות. האחת - בצורת אי-תלות בין מאורעות שפירושה שהידיעה האם התרחש אחד המאורעות אינה משליכה על ההסתברות שהמאורע השני התרחש. השניה - בצורת מרחבי מכפלה - כאשר כל שני מאורעות המערבים ממדים שונים של מרחב המדגם הנם בלתי-תלויים. כעת נכיר צורה שלישית - **אי-תלות בין משתנים מקריים**. בדומה לאי-התלות שבמרחבי מכפלה, אי-תלות זו תבטא מצב שבו לתוצאה של ניסוי אחד אין כל השפעה על התפלגות תוצאתו של ניסוי אחר. מבחינה טכנית מושג זה מתגלם בכך שהתפלגותם המשותפת הנקודתית של המשתנים שווה למכפלת ההתפלגויות השוליות הנקודתית.

הגדרה 3.31 (אי-תלות של שני משתנים מקריים בדידים)**.** נאמר שמשתנים מקריים בדידים X ו-Y המוגדרים על אותו מרחב הסתברות הנם **בלתי תלויים** ונסמן $X \perp\!\!\!\perp Y$, אם לכל שתי קבוצות $A, B \subset \mathbb{R}$ מתקיים שהמאורעות $\{X \in A\}$ ו-$\{Y \in B\}$ בלתי תלויים, כלומר

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

באופן דומה נדבר על קבוצה של משתנים מקריים בלתי-תלויים.

הגדרה 3.32 (אי-תלות של משתנים מקריים בדידים)**.** יהי $\mathcal{X}$ אוסף של משתנים מקריים בדידים המוגדרים על אותו מרחב הסתברות. נאמר שזהו אוסף משתנים מקריים בלתי-תלויים אם לכל $\{X_k\}_{k\in[n]}$, תת קבוצה סופית של משתנים מקריים ב-$\mathcal{X}$ ולכל $\{A_k\}_{k\in[n]}$, אוסף קבוצות ב-$\mathbb{R}$, מתקיים:

$$\mathbb{P}(\forall k \in [n]\ X_k \in A_k) = \prod_{k\in[n]} \mathbb{P}(X_k \in A_n).$$

✎ **בעיה 3.10.** להוכיח כי משתנה מקרי X מקיים $X \perp\!\!\!\perp X$ אם ורק אם $c \overset{\text{a.s.}}{=} X$ עבור $c \in \mathbb{R}$ כלשהו.

אבחנה 3.33 (אי-תלות של משתנים מקריים בדידים היא תכונה של פונקציות התפלגות נקודתיות)**.** יהיו $X_1, \ldots, X_n$ משתנים מקריים בלתי-תלויים במרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. אזי $X_1, \ldots, X_n$ בלתי-תלויים

אם ורק אם לכל $x_1, \dots, x_n \in \mathbb{R}$ מתקיים

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^{n} \mathbb{P}(X_k = x_k).$$

✎ **בעיה 3.11.** להוכיח את אבחנה 3.33.

✎ **בעיה 3.12.** יהיו X ו-Y משתנים מקריים בדידים בלתי-תלויים. יש להוכיח כי $|X|, |Y|$ בלתי-תלויים.

✎ **בעיה 3.13.** להראות כי Y ו-Z מדוגמא 3.24 הנם תלויים אך $|Y - 3.5|$ ו-$|Z - 3.5|$ הנם בלתי-תלויים.

✎ **בעיה 3.14.** להראות כי אם $X \sim \mathrm{Ber}(p)$ ו-$Y \sim \mathrm{Ber}(q)$ הנם בלתי-תלויים, אז XY מתפלג ברנולי-pq.

✎ **בעיה 3.15** (נוסחת הקונבולוציה)**.** להראות כי אם X ו-Y נתמכים על $\mathbb{Z}$ ובלתי-תלויים, אז $Z = X + Y$ מקיים לכל $n \in \mathbb{N}$

$$p_Z(n) = \sum_{i \in \mathbb{Z}} p_X(i) p_Y(n - i).$$

באמצעות משתנים מקריים נוכל לתאר ולנתח ביתר קלות בעיות אשר קודם לכן יכולנו להתמודד עמן רק באמצעות מרחבי מכפלה. נדגים זאת באמצעות ניתוח סוגיה 4.

דוגמא 3.34 (בעיית האספן [coupon collector's problem])**.**
בכל ביצת הפתעה נמצאת בובה המשתייכת לאחד מבין n סוגי בובות אהובות. הבובות מוגרלות באופן אקראי ובלתי-תלוי. כיצד נחסום מלמעלה את ההסתברות שיהיה צורך ברכישת יותר מ-k ביצים כדי לאסוף את כל סוגי הבובות?

תשובה: מטרתנו היא לחסום את הסתברותו של המאורע שבין k הביצים הראשונות שקנינו לא הופיעו כל סוגי הבובות. בלי הגבלת הכלליות נסמן את סוגי הבובות האפשריים במספרים מ-1 ועד n. נגדיר משתנה מקרי X_i שיתאר את סוג הבובה בביצה ה-i. נסמן ב-A_j^k את המאורע שבין k הביצים שרכשנו אף אחת לא הכילה בובה מסוג j. המאורע שאנו מעוניינים לחשב את הסתברותו הוא $E^k = \bigcup_{j \in [n]} A_j^k$.

ההסתברות שהביצה ה-i אינה מכילה בובה מסוג j הנה $\mathbb{P}(X_i \neq j) = \frac{n-1}{n}$ ולכן, הואיל וסוגי הבובות בביצים שונות בלתי תלויים, הסתברותו של A_j^k היא $\left(\frac{n-1}{n}\right)^k = \prod_{i \in [k]} \mathbb{P}(X_i \neq j) = \mathbb{P}(A_j^k)$. נשתמש בחסם האיחוד (משפט 1.17) ונקבל

$$\mathbb{P}(E^k) \leq \sum_{j \in [n]} \mathbb{P}(A_j^k) = \sum_{j \in [n]} \left(\frac{n-1}{n}\right)^k = n\left(1 - \frac{1}{n}\right)^k \leq n e^{-k/n}.$$

לפיכך אם נקנה $k = c \cdot n \log n$ ביצי הפתעה, תהא ההסתברות של E^k חסומה על ידי n^{1-c}. אם ניקח $c > 1$ כלשהו הרי שההסתברות שנקבל את כל סוגי הבובות תשאף ל-1 כאשר נשאיף את n לאינסוף. □

הערה: לעת עתה הצגנו רק חסם עליון למספר ביצי ההפתעה הדרושות באופן טיפוסי כדי לזכות בכל סוגי הבובות ואילו כחסם תחתון נאלץ להסתפק בחסם המובן מאליו הקובע שדרושות לפחות n ביצים. בפרקים הבאים, נתוודע לכלים חדשים שיאפשרו לנו להראות שהחסם שהצגנו הדוק באופן אסימפטוטי.

בבקשנו להכליל את מושג האי-תלות של זוג משתנים מקריים לאי-תלות קבוצתית – יסתבר לנו שההגדרות וההוכחות מסתבכות והולכות. יתר ההגדרות וההוכחות בפרק זה מציגות את הגישה הישירה להכללה זו. בפרק 3.5, מוצגת גישה חילופית, עמוקה יותר, אשר מפשטת הגדרות אלו במחיר של הפשטה והצגת מושגים חדשים. היכרות עם גישה זו היא הצעד הראשון בדרך לבניין מרחבי הסתברות שאינם בדידים.

הגדרה 3.35 (אי-תלות של שתי קבוצות של משתנים מקריים בדידים)**.** יהיו $\mathcal{X}, \mathcal{Y}$ שני אוספים של משתנים מקריים בדידים (על אותו מרחב הסתברות). נאמר ש**האוספים בלתי-תלויים** ונסמן $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}$, אם לכל $\{X_k\}_{k\in[n]}$, תת קבוצה סופית של משתנים מקריים ב-$\mathcal{X}$ ו-$\{Y_k\}_{k\in[m]}$, תת קבוצה סופית של משתנים מקריים ב-$\mathcal{Y}$ ולכל $\{A_k\}_{k\in[n]},\{B_k\}_{k\in[m]}$ אוספי קבוצות ב-$\mathbb{R}$, מתקיים:

$$\mathbb{P}(\forall k\in[n]\ X_k\in A_k, \forall k\in[m]\ Y_k\in B_k) = \mathbb{P}(\forall k\in[n]\ X_k\in A_k)\mathbb{P}(\forall k\in[m]\ Y_k\in B_k).$$

אי-תלות בין קבוצות משתנים מקריים עוברת בירושה לפונקציות על אוספים אלה.

טענה 3.36 (שימור אי-תלות תחת הפעלות פונקציות)**.** יהיו $X_1,\dots,X_n$ ו-$Y_1,\dots,Y_m$ שתי קבוצות בלתי-תלויות של משתנים מקריים במרחב הסתברות. תהיינה $f:\mathbb{R}^n\to\mathbb{R}$, $g:\mathbb{R}^m\to\mathbb{R}$ פונקציות ממשיות. אזי המשתנים המקריים $f(X_1,\dots,X_n)$ ו-$g(Y_1,\dots,Y_m)$ הנם בלתי-תלויים.

הוכחה. נסמן $X=f(X_1,\dots,X_n), Y=g(Y_1,\dots,Y_m)$. לאור אבחנה 3.33 די להראות כי לכל שני ממשיים $s\in\operatorname{Supp}(X)$ ו-$t\in\operatorname{Supp}(Y)$ מתקיים כי $\mathbb{P}(X=s,Y=t)=\mathbb{P}(X=s)\mathbb{P}(Y=t)$. נחשב,

$$\begin{aligned}
\mathbb{P}(X=s,Y=t) &= \sum_{\substack{\{x_k\}_{k\in[n]}\\ f(x_1,\dots,x_n)=s}}\ \sum_{\substack{\{y_k\}_{k\in[m]}\\ g(y_1,\dots,y_m)=t}} \mathbb{P}\Big(X_1=x_1,\dots,X_n=x_n,Y_1=y_1,\dots,Y_m=y_m\Big)\\
&\overset{\text{אי-תלות}}{=} \sum_{\substack{\{x_k\}_{k\in[n]}\\ f(x_1,\dots,x_n)=s}}\ \sum_{\substack{\{y_k\}_{k\in[m]}\\ g(y_1,\dots,y_m)=t}} \mathbb{P}(\forall k\in[n]:X_k=x_k)\mathbb{P}(\forall k\in[m]:Y_k=y_k)\\
&=\Bigg(\sum_{\substack{\{x_k\}_{k\in[n]}\\ f(x_1,\dots,x_n)=s}} \mathbb{P}(\forall k\in[n]:X_k=x_k)\Bigg)\Bigg(\sum_{\substack{\{y_k\}_{k\in[m]}\\ g(y_1,\dots,y_m)=t}} \mathbb{P}(\forall k\in[m]:Y_k=y_k)\Bigg)\\
&=\mathbb{P}(X=s)\mathbb{P}(Y=t)
\end{aligned}$$

■

כעת נוכל לתאר אי-תלות בין אוסף של משתנים מקריים במונחים של אי-תלות בין אוספי משתנים מקריים.

טענה 3.37 (תנאי שקול לאי-תלות של משתנים מקריים)**.** יהיו $X_1,\dots,X_n$ משתנים מקריים בדידים במרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$. אזי $X_1,\dots,X_n$ בלתי-תלויים אם ורק אם לכל $k\in[n]$ מתקיים כי $\{X_k\}$ בלתי-תלוי ב-$\{X_1,\dots,X_n\}\setminus\{X_k\}$.

הוכחה. **כיוון ראשון:** נניח כי $X_1,\dots,X_n$ בלתי תלויים ונראה כי לכל $k\in[n]$ מתקיים כי $\{X_k\}$ בלתי-תלוי

ב-$\{X_k\} \setminus \{X_1,\dots,X_n\}$. נניח בלי הגבלת הכלליות כי $k=1$. יהיו $A_1 \subset \mathbb{R}$ ו-$A_2 \subset \mathbb{R}^{n-1}$. נחשב,

$$\begin{aligned}\mathbb{P}(X_1 \in A_1, (X_2,\dots,X_n) \in A_2) &= \sum_{(a_2,\dots,a_n)\in A_2} \mathbb{P}(X_1 \in A_1, X_2 = a_2, \dots, X_n = a_n)\\ &\overset{\text{אבחנה } 3.33}{=} \sum_{(a_2,\dots,a_n)\in A_2} \mathbb{P}(X_1 \in A_1) \prod_{j=2}^{n} \mathbb{P}(X_j = a_j)\\ &= \mathbb{P}(X_1 \in A_1)\mathbb{P}((X_2,\dots,X_N) \in A_2),\end{aligned}$$

ולכן $\{X_1\}$ בלתי-תלוי ב-$\{X_1\} \setminus \{X_1,\dots,X_n\}$. **כיוון שני:** נניח כי לכל $k \in [n]$ מתקיים כי $\{X_k\}$ בלתי-תלוי ב-$\{X_n\} \setminus \{X_1,\dots,X_n\}$ ונראה כי $X_1,\dots,X_n$ בלתי תלויים. יהיו $x_1,\dots,x_n \in \mathbb{R}$, נחשב:

$$\begin{aligned}\mathbb{P}(X_1 = x_1,\dots,X_n = x_n) &= \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2,\dots,X_n = x_n)\\ &= \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\mathbb{P}(X_3 = x_3,\dots,X_n = x_n)\\ &= \dots = \prod_{k=1}^{n} \mathbb{P}(X_k = x_k),\end{aligned}$$

כאשר המעבר הראשון נסמך על אי-תלות בין $\{X_1\}$ לבין $\{X_2,\dots,X_n\}$, המעבר השני – על אי-תלות בין $\{X_2\}$ לבין $\{X_3 \dots, X_n\}$ וכן הלאה. נסיק לפי אבחנה 3.33 כי $X_1,\dots,X_n$ בלתי תלויים.

■

✎ **בעיה 3.16** (קוביות אפרון [Efron's dice]**).** להלן המספרים על צדדיהן של ארבע קוביות הוגנות, A,B,C,D המוטלות באופן בלתי-תלוי.

$$A = \{4,4,4,4,0,0\} \qquad B = \{3,3,3,3,3,3\}$$
$$C = \{6,6,2,2,2,2\} \qquad D = \{5,5,5,1,1,1\}$$

יש להשתמש במשתנים מקריים מתאימים ולחשב עבור כל זוג קוביות את ההסתברות שתוצאת קוביה אחת גבוהה מתוצאתה של השניה.

✎ **בעיה 3.17.** למצא שני משתנים מקריים X ו-Y על מרחב ההסתברות האחיד על $\Omega = [5]$ כך שיתקיים $X, Y, X+Y \sim \mathrm{Unif}\{-2,-1,0,1,2\}$. האם יתכן שמשתנים כאלה יהיו בלתי-תלויים?

3.3 התפלגות מותנית במאורע

בפרק 2.1 התוודענו למושג ההסתברות המותנית אשר תאר את השפעתו של מידע על הסתברותו של מאורע. כעת, נכליל מונח זה לתאר כיצד מידע מחולל שינוי בהתפלגותו של משתנה מקרי.

הגדרה 3.38 (התפלגות מותנית)**.** יהי X משתנה מקרי בדיד על מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$ ויהי $A \in \mathcal{F}$ מאורע המקיים $\mathbb{P}(A) > 0$. נסמן ב-$(X \mid A)$ את המשתנה המקרי X על מרחב ההסתברות $(\Omega,\mathcal{F},\mathbb{P}_A)$ כאשר $\mathbb{P}_A$ היא פונקצית ההתפלגות המתקבלת כאשר מתנים ב-A (ראה טענה 2.3).

הסימון בהגדרה 3.38 מטעה במידה מסויימת, משום ש-$(X\,|\,A)$ ו-X הם סימונים שונים לאותה פונקציה. ההבדל הוא בפונקציית ההסתברות במרחב שעליו מוגדר כל אחד מהמשתנים. הבדל זה הוא שגורם להבדל בהתפלגותם של $(X\,|\,A)$ ו-X ולכן עלינו לראות בהם משתנים מקריים על **מרחבי הסתברות שונים**, אשר אין להם התפלגות משותפת.

דוגמא 3.39. נסמן ב-X וב-Y את תוצאות ההטלה של שתי קוביות משחק הוגנות וב-$Z = X + Y$ את סכומן. מהי התפלגות של Z בהינתן שהוא זוגי? בהינתן ש-Z זוגי איזה מאורע סביר יותר - $\{X = 4\}$ או $\{Z = 4\}$?

תשובה: נסמן ב-A את המאורע ש-Z זוגי, כלומר, $A = \{Z \in 2\mathbb{N}\}$. לפי נוסחת ההסתברות השלמה מתקיים:

$$\mathbb{P}(A) = \sum_{i\in[6]} \mathbb{P}(A\,|\,X = k)\mathbb{P}(X = k) = \sum_{k\in[6]} \frac{1}{2}\mathbb{P}(X = k) = \frac{1}{2}.$$

משיקולי ספירת תצריפים נקבל כי

$$\mathbb{P}(Z = 2) = \mathbb{P}(Z = 12) = \frac{1}{36} \qquad \mathbb{P}(Z = 4) = \mathbb{P}(Z = 10) = \frac{3}{36} \qquad \mathbb{P}(Z = 6) = \mathbb{P}(Z = 8) = \frac{5}{36}.$$

נחשב את ההתפלגות המותנית של Z בהנתן A לפי הנוסחא

$$\mathbb{P}(Z = k\,|\,A) = \frac{\mathbb{P}(Z = k \cap A)}{\mathbb{P}(A)} = \frac{P(Z = k)}{\mathbb{P}(A)}$$

לכל i זוגי. נקבל

i	2	4	6	8	10	12
$\mathbb{P}(Z = i\,\vert\,A)$	1/18	1/6	5/18	5/18	1/6	1/18

וחישבנו את התפלגות $(Z\,|\,A)$. כעת נשים לב כי

$$\mathbb{P}(X = i\,|\,A) = \frac{P(X = i)\mathbb{P}(A\,|\,X = i)}{\mathbb{P}(A)} = \frac{(1/6)\cdot(1/2)}{1/2} = \frac{1}{6}.$$

כלומר X בלתי תלוי במאורע A(!) קיבלנו כי $\mathbb{P}(X = 4\,|\,Z \in 2\mathbb{N}) = \mathbb{P}(Z = 4\,|\,Z \in 2\mathbb{N})$. □

מעתה נוכל לתאר את התפלגותו של משתנה מקרי במונחים של התפלגות מותנית בתוצאה של משתנה מקרי אחר, כפי שעשינו עבור מרחבי הסתברות באבחנה 2.9.

דוגמא 3.40. בוחרים באופן אחיד מספר X ב-$[n]$. לאחר מכן בוחרים באופן אחיד מספר בין 1 ל-X. מה ההסתברות שהמספר שנבחר יהיה 1?

תשובה: בדוגמא מתואר ניסוי דו-שלבי. נסמן ב-X את תוצאת ההטלה הראשונה וב-Y את תוצאת ההטלה השניה. X מתפלג אחיד על $[n]$ ואילו בהנתן ש-$X = k$, Y מתפלג אחיד על $[k]$. לכן, ההתפלגות המשותפת של X ו-Y נתונה על ידי $\mathbb{P}(X = k, Y = \ell) = 1/k\ell$ עבור $1 \le \ell \le k \le n$ ואפס עבור ערכים אחרים של k ו-ℓ. אם כך, ההסתברות ש-$Y = 1$ היא

$$\mathbb{P}(Y = 1) = \sum_{k=1}^{n} \mathbb{P}(X = k, Y = 1) = \sum_{k=1}^{n} \frac{1}{kn} \approx \frac{\log(n)}{n}.$$

□

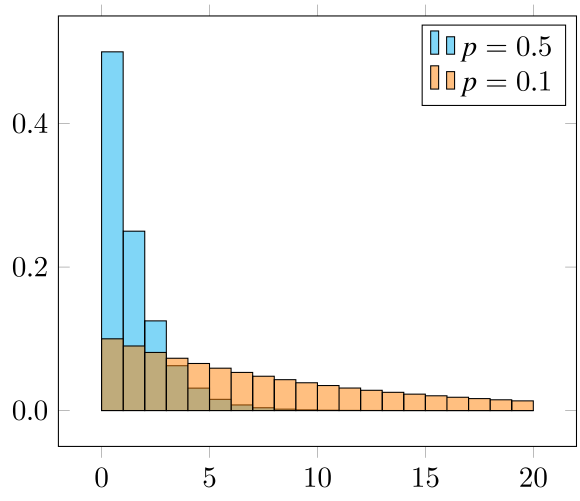
3.4 התפלגויות נפוצות

למרות המגוון הגדול של הניסויים התיאורתיים שאנו יכולים להעלות במחשבתנו, מסתבר שרבים מניסויים אלו טומנים בחובם משתנים מקריים דומים זה לזה. חמש התפלגויות כאלה תקבלנה את מרבית תשומת-ליבנו. שתי הראשונות, התפלגות ברנולי, והתפלגות אחידה כבר הוצגו בהגדרות 3.11 ו-3.13 בהתאמה, וכעת בשלו התנאים לערוך היכרות עם שלוש התפלגויות נוספות.

3.4.1 התפלגות גיאומטרית

נסתכל על ניסוי חוזר בעל הסתברות הצלחה p שאינה תלויה בתוצאות הניסויים הקודמים. אנו מבקשים לדעת מה הוא מספר הניסויים עד להצלחה ראשונה. התפלגות משתנה מקרי זה מכונה **התפלגות גיאומטרית** והיא נתונה להלן.

הגדרה 3.41 (התפלגות גיאומטרית)**.** נאמר שמשתנה מקרי X מתפלג לפי **התפלגות גיאומטרית עם הסתברות** p **להצלחה** ונכתוב $X \sim \mathrm{Geo}(p)$ אם לכל $n \in \mathbb{N}$ מתקיים $p_X(n) = (1-p)^{n-1}p$.



התפלגות גיאומטרית עבור ערכים עד 20 לשני ערכי p.

נוח במיוחד לתאר התפלגות גיאומטרית במונחים של ההסתברות לקבלת ערך גדול ממספר שלם מסוים.

טענה 3.42 (תיאור משתנה גיאומטרי במונחים של התפלגות שיורית)**.** משתנה מקרי שנתמך על השלמים מקיים $X \sim \mathrm{Geo}(p)$ אם ורק אם לכל $r \in \mathbb{N}$ מתקיים $\mathbb{P}(X > r) = (1-p)^r$.

הוכחה. לכל משתנה מקרי X הנתמך על השלמים מתקיים $\mathbb{P}(X = r) = \mathbb{P}(X > r-1) - \mathbb{P}(X > r)$ ולכן אוסף ההסתברויות $\{\mathbb{P}(X > r)\}_{r \in \mathbb{N}}$ קובע את X. די אפוא לחשב עבור X המתפלג גיאומטרית עם סיכוי הצלחה p אם ורק אם

$$\mathbb{P}(X > r) = \sum_{n=r+1}^{\infty} \mathbb{P}(X = n) = \sum_{n=r+1}^{\infty} (1-p)^{n-1}p = (1-p)^r \sum_{n=1}^{\infty} (1-p)^{n-1}p = (1-p)^r.$$

■

כעת ניווכח שמשתנה גיאומטרי מתאר אומנם את מספר הנסיונות על להצלחה בסדרת ניסויים, ונקשר אותו לסוגיה 5 - התפלגות זמן-עד-כשלון.

דוגמא 3.43 (הצדקה לשימוש בהתפלגות גיאומטרית עבור זמן-עד-כשלון)**.**
מבוצעת סדרה של ניסויים חוזרים, כך שכל ניסוי בלתי-תלוי בקודמיו וההסתברות להצלחה בו היא p. נסמן ב-X את מספר הניסויים עד להצלחה הראשונה ונראה כי $X \sim \mathrm{Geo}(p)$.

תשובה: נראה כי תיאור הניסוי מחייב את X לקיים דרישות שמגדירות אותו באופן עקבי ויחיד. את תיאור הניסוי לנו באופן מותנה: ההסתברות שנעצור בניסוי הבא, בהינתן שנכשלנו בכל הקודמים היא p. כלומר לכל $n \in \mathbb{N}$ מתקיים $\mathbb{P}(X = n \mid X > n-1) = p$. נשתמש בעובדה זו להוכיח באינדוקציה כי $\mathbb{P}(X > n+1) = (1-p)^n$ ונסיק, לפי טענה 3.42, כי $X \sim \mathrm{Geo}(p)$. עבור $n = 0$, הטענה מידית. נניח כי הטענה מתקיימת עבור $n-1$ ונוכיח אותה עבור n. נשים לב כי המאורע $\{X > n+1\}$ מוכל במאורע $\{X > n\}$ ונחשב

$$\mathbb{P}(X > n+1) \overset{\text{כלל השרשרת}}{=} \mathbb{P}(X > n+1 \mid X > n)\mathbb{P}(X > n) \overset{\text{הנחת האינדוקציה}}{=} (1-p)^n.$$

□

✎ **בעיה 3.18.** יהיו $X_1, \ldots, X_n$ משתנים מקריים בלתי-תלויים ושווי התפלגות כאשר $X_1 \sim \mathrm{Ber}(p)$, ויהי $Y \sim \mathrm{Geo}(p)$. יש להראות כי לכל $k \le n$ מתקיים $\mathbb{P}(Y = k) = \mathbb{P}(X_1 = 0, \ldots, X_{k-1} = 0, X_k = 1)$.

אם נחשוב על "הצלחה" בתור כשל של פריט ציוד כלשהו, למשל נורה חשמלית (כמו בסוגיה 5), ועל p בתור הסיכוי לכשל כזה בכל יחידת זמן, נוכל להסיק כי הזמן-עד-כשלון מרגע הקניה, או הזמן-בין-כשלונות, אם אנו קונים פריט חדש בכל פעם שבו הקודם כשל, מתפלגים גיאומטרית עם פרמטר p. בפרק 10 נראה הצדקה פורמלית לכך שהתפלגות גיאומטרית מתארת את התפלגות מספר הניסויים עד לכשלון בסדרה של אינספור ניסויים חוזרים בעלי אותה הסתברות הצלחה.

תכונת חוסר הזיכרון לכשלונו מאפיינת באופן ייחודי את התפלגות גיאומטרית כפי שעולה מהטענה הבאה.

טענה 3.44 (אפיון התפלגות גיאומטרית)**.** יהי $X \sim \mathcal{D}$ משתנה מקרי הנתמך על $\mathbb{N}$. שלושת הבאים שקולים:

(א) $\mathcal{D}$ הנה התפלגות גיאומטרית,

(ב) X ו-$(X-1 \mid X > 1)$ שווי התפלגות,

(ג) X ו-$(X-s \mid X > s)$ שווי התפלגות לכל $s \in \mathbb{N}$.

הוכחה. **(א) גורר את (ב).** לכל $n \in \mathbb{N}$ מתקיים

$$\mathbb{P}(X-1 = n \mid X > 1) = \frac{\mathbb{P}(X = n+1)}{\mathbb{P}(X > 1)} = \frac{(1-p)^n p}{1-p} = (1-p)^{n-1}p = \mathbb{P}(X = n).$$

(ב) גורר את (ג). לפי (ב) $(X-1 \mid X>1)$ ו-X שווי התפלגות. נגדיר $Y := (X-(s-1))$, נניח באינדוקציה כי המשתנים $X \overset{\mathrm{d}}{=} (Y \mid Y > 0)$ שווי התפלגות. נגדיר $(X-(s-1)) =$ ונחשב:

$$(Y \mid Y > 0) \overset{\mathrm{d}}{=} X \overset{\mathrm{d}}{=} (X-1 \mid X > 1) \overset{\mathrm{d}}{=} (Y-1 \mid Y > 0, Y > 1) = (Y-1 \mid Y > 1) = (X-s \mid X > s).$$

כאשר השוויון השמאלי ביותר נובע מהנחת האינדוקציה ומהגדרת Y, השני מתכונה (ב), השלישי – שוב משימוש בהנחת האינדוקציה, הרביעי מאבחנה 2.4א לגבי התניה במאורעות המקיימים יחס הכלה והימני ביותר – מהגדרת Y.

(ג) גורר את (א). יהי X משתנה מקרי הנתמך על $\mathbb{N}$, המקיים את תכונה (ג). נרשום לפי כלל השרשרת (טענה 2.4ב) לכל $s \in \mathbb{N}_0$,

$$\mathbb{P}(X > s) = \mathbb{P}(X > s \mid X > s-1)\mathbb{P}(X > s-1 \mid X > s-2)\cdots\mathbb{P}(X > 1) \overset{(ג)}{=} \mathbb{P}(X > 1)^s.$$

משוויון זה נסיק את סעיף (א) באמצעות טענה 3.42. ■

✎ **בעיה 3.19.** יהיו $X_1 \sim \mathrm{Geo}(p)$ ו-$X_2 \sim \mathrm{Geo}(q)$ בלתי-תלויים. יש להראות כי

$$\min(X_1, X_2) \sim \mathrm{Geo}(1-(1-p)(1-q)).$$

✎ **בעיה 3.20.** יהיו $X_1, X_2 \sim \mathrm{Geo}(p)$ בלתי תלויים. יש להראות כי בהינתן המאורע $\{X_1 + X_2 = n\}$ מתקיים כי X_1 מתפלג אחיד על $[n-1]$.

✎ **בעיה 3.21** (התפלגות בינומית שלילית)**.** יהיו $X_1, X_2, \ldots, X_r \sim \mathrm{Geo}(p)$ בלתי תלויים. יש להראות כי

(א) $\mathbb{P}(X_1 + X_2 = 2 + k) = \binom{2+k}{2}(1-p)^2p^k$,

(ב) $\mathbb{P}(\sum_{i=1}^r X_i = r + k) = \binom{k+r}{r}(1-p)^2p^k$.

הערה: התפלגותו של הסכום $X = \left(\sum_{i=1}^r X_i\right) - r$ בבעיה 3.21 מכונה **התפלגות בינומית שלילית** או **התפלגות פסקל** עם r הצלחות ו-p סיכויי הצלחה, ומסמנים $X \sim \mathrm{NB}(r, p)$. התפלגות זו מתאימה למספר הכשלונות עד ל-r הצלחות בסדרה של ניסויים בלתי-תלויים.

✎ **בעיה 3.22.** תורת ההסתברות יכולה גם לשמש להוכחת זהויות קומבינטוריות. יש להראות כיצד ניתן להשתמש בחישוב התפלגותו של סכום של k משתנים גיאומטריים כדי להוכיח את הזהות הבאה:

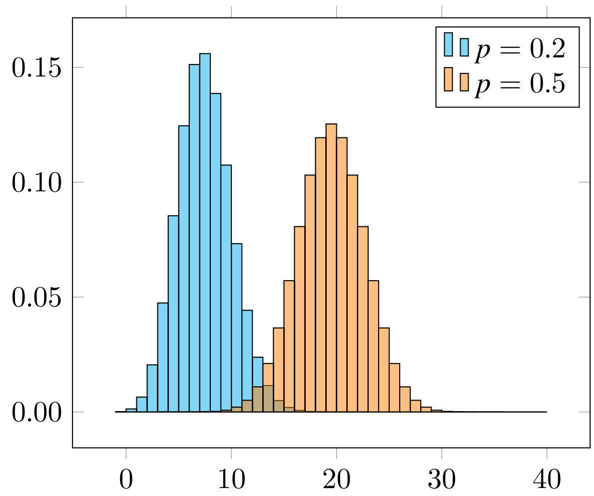
$$\sum_{i=n}^{\infty}\binom{n-1}{k-1}p^k(1-p)^{n-k} = 1.$$

בְּלֵאִיז פָּסְקַל (1623-1662), מתמטיקאי צרפתי ותיאולוג קתולי. בגיל 16 פיתח את הגיאומטריה הפרוייקטיבית ולאורך חייו תרם רבות לפורמליזם המתמטי, ולחשבון הדיפרנציאלי והאינטגרלי. בשנות העשרים לחייו פיתח בהתכתבויות עם פייר פרמה ועם חברו אַנְטוֹאַן גוֹמְבוֹ, שנודע כשבלייר דה מארה, את יסודות תורת ההסתברות ובמיוחד נודע בשל פתרון שלל בעיות בתחום ההימורים. הופעתה של ההתפלגות הבינומית השלילית בפתרונות אלו הביא לכך שנקראה על שמו. בגיל 31 חווה חוויה רוחנית עזה (יתכן שבעקבות תאונת מרכבה בה היה מעורב) והפנה חלק ניכר ממרצו לעיסוקים תיאולוגיים.

3.4.2 התפלגות בינומית

שוב נבחן סדרה של n ניסויים חוזרים, בלתי-תלויים זה בזה, ובעלי הסתברות הצלחה p כך אחד. הפעם נבקש לדעת כמה מהניסויים הצליחו. התפלגות משתנה מקרי זה מכונה **התפלגות בינומית**.

הגדרה 3.45 (התפלגות בינומית)**.** נאמר שמשתנה מקרי X מתפלג לפי **התפלגות בינומית עם n ניסיונות וסיכוי הצלחה** p ונכתוב $X \sim \mathrm{Bin}(n,p)$ אם לכל $k \in \{0,\dots,n\}$ מתקיים $p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



התפלגות בינומית עבור $n = 40$ לשני ערכי p.

נשים לב שהתפלגות בינומית עם $n = 1$ היא לא אחרת מאשר התפלגות ברנולי עם סיכוי הצלחה p, ואומנם היו ברנולי ופסקל הראשונים לחקור התפלגות זו.

תוך כדי כך שנוודא כי התפלגות בינומית הנה אמנם התפלגות מוגדרת היטב, נשים לב לקשר מעניין בינה ובין הבינום של ניוטון אשר ממנו נגזר שמה. נרשום

$$\mathbb{P}(X \in \{0,\dots,n\}) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \big(p + (1-p)\big)^n = 1$$

כאשר השוויון המרכזי הוא בדיוק נוסחת הבינום.

טענה 3.46 (אפיון התפלגות בינומית)**.** יהיו $\{X_i\}_{i\in[n]}$ משתני ברנולי p בלתי-תלויים, אזי

$$\sum_{i\in[n]} X_i \sim \mathrm{Bin}(n,p).$$

הוכחה. יהי $n \in \{0,\dots,N\}$ ונסמן $Y = \sum_{i\in[N]} X_i$ ונחשב

$$\mathbb{P}(Y=n) = \sum_{\substack{(x_1,\dots,x_N)\in\{0,1\}^N \\ \sum_{i\in[N]} x_i = n}} \mathbb{P}(\forall i \in [N],\ X_i = x_i) \overset{\text{אי-תלות}}{=} \sum_{\substack{(x_1,\dots,x_N)\in\{0,1\}^N \\ \sum_{i\in[N]} x_i = n}} \left(\prod_{i\in[N]} \mathbb{P}(X_i = x_i) \right) = \binom{N}{n} p^n (1-p)^{N-n}$$

■

מסקנה 3.47 (חיבור התפלגויות בינומיות)**.** אם $X \sim \mathrm{Bin}(n,p)$, $Y \sim \mathrm{Bin}(m,p)$ בלתי-תלויים, אז $X+Y \sim \mathrm{Bin}(n+m,p)$.

הוכחה. יהיו $X_1,\ldots,X_{n+m}$ משתנים מקריים בלתי-תלויים המתפלגים ברנולי עם הסתברות הצלחה p. לפי טענה 3.46 מתקיים

$$\sum_{k=1}^{n} X_k \sim \mathrm{Bin}(n,p), \qquad \sum_{k=n+1}^{n+m} X_k \sim \mathrm{Bin}(m,p).$$

כיוון שאלו פונקציות של קבוצות משתנים שונות באוסף של משתנים בלתי-תלויים, הרי שלפי טענה 3.36 שני סכומים אלו אף הם בלתי-תלויים ולכן התפלגותם המשותפת זהה לזו של X ו-Y. לכן, כיוון שהתפלגות הסכום תלויה רק בהתפלגויות המשותפת (אבחנה 3.30), הרי שמתקיים:

$$\sum_{i=1}^{n+m} X_i \overset{\mathrm{d}}{=} X+Y.$$

לפי טענה 3.46, התפלגותו של $\sum_{k=1}^{n+m} X_k$ היא בינומית עם $n+m$ נסיונות והסתברות הצלחה p. ■

✎ **בעיה 3.23.** מטילים מטבע הוגן 10 פעמים. לאחר מכן מטילים שוב את כל ההטלות שתוצאתן הייתה עץ. יש להראות שמספר תוצאות הפלי לאחר ההטלות החוזרות מתפלג בינומית ולמצוא את סיכוי ההצלחה.

✎ **בעיה 3.24.** להראות כי אם $X \sim \mathrm{Bin}(n,p)$ אז $(n-X) \sim \mathrm{Bin}(n,1-p)$.

✎ **בעיה 3.25.** בתהליך הייצור של מסך דק כל פִּיקְסֶל עלול להיות פגום בהסתברות של אחד לשני מיליון. אם במסך יש פיקסל פגום הוא נחשב סוג ב', ואם יש בו שני פיקסלים פגומים הוא מושמד. חשב (הנך רשאי להשתמש במכונת חישוב) מה ההסתברות שהמסך יהיה סוג ב' ומה ההסתברות להשמדת המסך ברזולוציות הבאות:

א. (640×480 פיקס') VGA ב. (1280×720 פיקס') 720p ג. (1920×1080 פיקס') 1080p

ד. (2560×1440 פיקס') WQHD ה. (3840×2160 פיקס') UHD-1 ו. (4096×2160 פיקס') 4K

(עד לסדרה של התקדמויות טכנולוגיות סביב שנת 2010, הציבו פיקסלים פגומים מגבלה משמעותית על רזולוציית מסכים.)

✎ **בעיה 3.26.** נניח כי הסתברותו של נדאל לנצח את פדרר במערכה של משחקי טניס היא שני שליש, באופן בלתי תלוי בתוצאת המערכות הקודמות. השניים משחקים עד אשר אחד מהם זוכה בשלוש מערכות בסך הכל. מה ההסתברות שנדאל יזכה בטורניר? מה הקשר להתפלגות בינומית?

כעת נציג חישוב מפורש הנוגע להסתברות בינומית ונבאר את הקשר בינו לבין סוגיה 7 – הילוך שיכור פשוט.

דוגמא 3.48 (חישוב ההסתברות לשוויון במשחק הוגן)**.**
מוטלים $2n$ מטבעות הוגנים. נעריך את ההסתברות שבדיוק ב-n מטבעות תתקבל תוצאה של עץ.

תשובה: לפי טענה 3.46 מספר המטבעות שתוצאת הטלתם היא עץ מתפלג בינומית עם מספר ניסויים $2n$ וסיכוי הצלחה $1/2$. על כן, ההסתברות לשוויון היא בדיוק $\binom{2n}{n}2^{-2n}$, נשתמש בנוסחת סטרלינג האסימפטוטית ונקבל

$$\binom{2n}{n}2^{-2n} = \frac{2^{-2n}\sqrt{4\pi n}(2n/e)^{2n}(1+R(2n))}{\left(\sqrt{2\pi n}(n/e)^n(1+R(n))\right)^2} = \frac{1}{\sqrt{\pi n}} \cdot \frac{1+R(2n)}{(1+R(n))^2}.$$

כאשר $\lim_{n\to\infty} R(n) = 0$. ומכאן שמתקיים: $\lim_{n\to\infty} \sqrt{\pi n}\binom{2n}{n}2^{-2n} = 1$,

כלומר ההסתברות שתוצאת מחצית ההטלות תהיה עץ היא בקירוב $1/\sqrt{\pi n}$. אם ברצוננו חסם כמותי ולא אסימפטוטי, נוכל להשתמש בגרסא המתאימה של נוסחת סטרלינג ולקבל את ההערכה

$$\frac{0.47}{\sqrt{n}} \le \frac{2\sqrt{\pi}}{e^2\sqrt{n}} = \frac{\sqrt{2\pi}(2n)^{2n+1/2}e^{-2n}}{\left(en^{n+1/2}e^{-n}\right)^2 2^{2n}} \le \binom{2n}{n}2^{-2n} \le \frac{e(2n)^{2n+1/2}e^{-2n}}{\left(\sqrt{2\pi}n^{n+1/2}e^{-n}\right)^2 2^{2n}} = \frac{e\sqrt{2}}{2\pi\sqrt{n}} \le \frac{0.62}{\sqrt{n}}.$$

□

דוגמא 3.49 (הסתברות חזרה לראשית של הילוך שיכור מאוזן)**.**
שיכור יוצא מראשית ציר המספרים. בכל יחידת זמן הוא עלול בסיכוי חצי לנוע יחידה אחת ימינה ובסיכוי חצי לנוע יחידה אחת שמאלה באופן בלתי-תלוי ביתר צעדיו. נסמן ב-A_n את המאורע שלאחר n צעדים שב השיכור לראשית. נראה כי $\mathbb{P}(A_{2n}) = \Theta\left(1/\sqrt{n}\right)$ (נשים לב כי השיכור אינו יכול להימצא בראשית לאחר מספר אי-זוגי של צעדים, מפני שבזמנים אלה הוא חייב להימצא באתר אי-זוגי).

תשובה: נסמן ב-$\{X_k\}_{k\in[n]}$ משתני-אינדיקטור לכך שהשיכור נע ימינה בצעד ה-k. נשים לב כי משתנים אלו הנם משתני ברנולי בלתי-תלויים בעלי הסתברות הצלחה $1/2$, ולכן, לפי טענה 3.46 נוכל להסיק כי $\sum_{k\in[n]} X_k \sim \mathrm{Bin}(n,1/2)$. נסמן את מיקומו של השיכור לאחר n צעדים ב-Y_n ונשים לב כי

$$Y_n = \sum_{k\in[n]}\Big(X_k - (1-X_k)\Big) = 2\sum_{k\in[n]} X_k - n$$

מפני שמיקומו הוא מספר הזמנים בהם צעד ימינה פחות מספר הזמנים בהם צעד שמאלה. לכן

$$\mathbb{P}(A_{2n}) = \mathbb{P}(Y_{2n} = \quad = \mathbb{P}\Big(\sum_{k\in[2n]} X_k \sim \mathrm{Bin}(n,1/2) = n\Big)$$

ומכאן שהעובדה $\mathbb{P}(A_{2n}) = \Theta\left(1/\sqrt{n}\right)$ נובעת ישירות מדוגמא 3.48.

✎ **בעיה 3.27** (התפלגות היפר-גיאומטרית)**.** בכד m כדורים שחורים ו-$n-m$ כדורים לבנים. יש לחשב את

(א) התפלגות $X_{n,m,k}$, מספר השליפות של כדור שחור ב-k שליפות של כדורים **עם החזרה**.

(ב) התפלגות $Y_{n,m,k}$, מספר השליפות של כדור שחור ב-k שליפות של כדורים **ללא החזרה**.

הערה: התפלגותו של $Y_{n,m,k}$ בבעיה 3.27 מכונה התפלגות **היפר-גיאומטרית**. הפרמטר n מכונה **גודל אוכלוסיה**, m מכונה **מספר פרטים מבוקשים** ו-k מכונה **מספר שליפות** . התפלגות זו הנה בעלת חשיבות רבה בניתוח משחקי קלפים ובדגימות סטטיסטיות מאוכלוסיה קטנה. זאת משום שבשני המקרי עותק יחיד מכל איבר באוכלוסיה הנדגמת ולא ניתן לשולפו פעמיים באותו מדגם. עבור $k=1$ התפלגות זו מתלכדת עם התפלגות ברנולי עם הסתברות הצלחה $1/n$. כאשר k קבוע ו-n שואף לאינסוף הופכת ההתפלגות דומה יותר ויותר להתפלגות בינומית עם פרמטר n/m. זאת משום שגם אילו היינו שולפים את הכדורים עם החזרה - ההסתברות שאחד הכדורים היה נשלף פעמיים הייתה שואפת לאפס. התפלגות זו חשובה אפוא בעיקר עבור ערכי k גדולים יחסית – או כאשר כל הפרמטרים קטנים.

3.4.3 התפלגות פואסון

ההתפלגות האחרונה שנציג בפרק זה שונה מקודמותיה בכך שהיא אינה מתאימה באופן ישיר למדד כלשהו המתייחס לניסוי חוזר, אלא למעין מקביל רציף שלו. היא גם הייתה האחרונה להתגלות בין ההתפלגויות המתוארות כאן, שכן זכתה לפרסום ברבים לראשונה על ידי סימון דניס פואסון (Siméon Denis Poisson) ב-1837 (אף על פי שהופיעה בחיבור של אברהם דה-מואבר ב-1711 וכנראה שהיה ראוי שתיקרא על שמו). כדי להבין את התועלת שבהתפלגות זו נתאר מחקר שביצע הכלכלן לדיסלאוץ וון בורטקייבייץ' (Ladislaus Von Bortkwicz) ב1898 שבמסגרתו הוא מקדם את השימוש בהתפלגות זו.

> בספרו "חוק המספרים הקטנים", מתאר וון בורטקייביץ' מחקר שערך על מספר החיילים בצבא הפרוסי שנהרגים מבעיטת סוס ידידותי. לכאורה נצפה כי העובדה שחייל מסויים נבעט למוות על ידי סוס היא כמעט ב"ת בגורלם של חיילים אחרים. לכן, טבעי שנשתמש במודל של התפלגות בינומית. ואולם מספר החיילים הוא גדול מאוד והסיכוי של כל אחד מהם להיבעט למוות קטן יחסית (ולראיה רק מספר מצומצם של חיילים אומנם נבעטים למוות – בכל שנה 0.61 חיילים בממוצע בשנה בחטיבה המונה כאלף חיילים). התפלגות בינומית היא מורכבת לניתוח מדוייק. כדי להתגבר על בעיה זו נוכל להשאיף את מספר החיילים בחטיבה לאינסוף ולהשאיף את ההסתברות של כל אחד מהם להיבעט לאפס תוך שימור התכונה שבאופן טיפוסי יבעטו 0.61 חיילים בחטיבה בשנה. ההתפלגות הגבולית במקרה זה היא התפלגות פואסון עם פרמטר 0.61, וון בורטקייביץ' אומנם ממחיש במחקרו כי התפלגות פואסון אומנם מתארת בדיוק מפליא את כמות החיילים שנבעטו למוות.

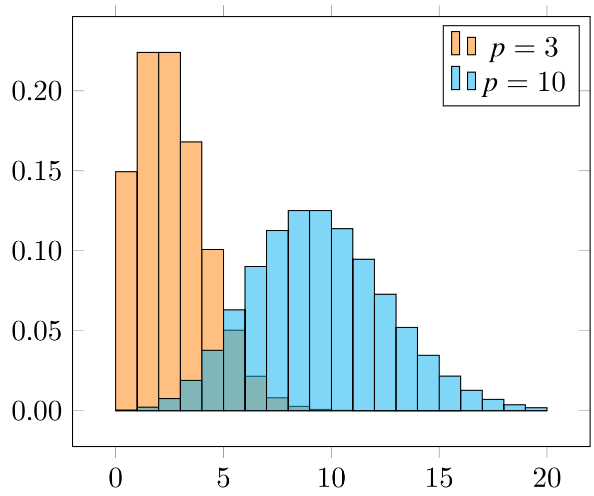
התפלגות פואסון משמשת אפוא לקירוב מספר המאורעות המתרחשים בפועל מתוך מספר רב של מאורעות שעשויים להתרחש באופן בלתי תלוי ובהסתברות נמוכה כל אחד. מלבד הדוגמא הקלאסית שלעיל, היא מתארת היטב את מספר האטומים המתפרקים בפרק זמן נתון בחומר רדיואקטיבי ואת מספר שיחות הטלפון המתבצעות בנקודת זמן מסוימת.

הגדרה 3.50 (התפלגות פואסון)**.** נאמר שמ"מ X מתפלג לפי **התפלגות פואסון (או פואסונית) עם שכיחות** λ ונכתוב $X \sim \mathrm{Po}(\lambda)$ אם לכל $n \in \mathbb{N}_0$ מתקיים

$$p_X(n) = \frac{\lambda^n e^{-\lambda}}{n!}.$$

הברון **סִימוֹן דְּנִיס פּוּאָסוֹן** (1781-1840), מתמטיקאי צרפתי פורה אשר פרסם למעלה משלוש-מאות מאמרים ועבודות, חבר האקדמיה הצרפתית למדעים, זכה להוקרה במשך כל חייו על כישוריו כחוקר וכמרצה. פואסון שהיה תלמידו של לפלאס אשר נהג בו קרבה יתירה, ראה עצמו כממשיך דרכו. תרם תרומה כבדת משקל לאנליזה, למתמטיקה שימושית ולמתמטיקה פיסיקלית (מדינמיקה של נוזלים, דרך תנועת גרמי השמיים ועד תיאור האור כגל). דווקא תרומתו להסתברות הייתה צנועה יחסית, ופרסומה של ההתפלגות הקרויה על שמו, אך אשר נתגלתה על ידי אברהם-דה-מואבר, הופיע בפרסום מתחום המתמטיקה השימושית שעסק במודל הסתברותי להרשעות שווא במערכת המשפט.

נשים לב ש- $\sum_{n\in\mathbb{N}_0} p_X(n) = \sum_{n\in\mathbb{N}_0} \frac{\lambda^n e^{-\lambda}}{n!} = 1$ שכן זהו טור טיילור של הפונקציה e^x בנקודה λ מוכפל ב $e^{-\lambda}$.



התפלגות פואסון עבור ערכים עד 20 לשני ערכי λ.

בפרק 10 נגדיר באיזה אופן בדיוק מתקבלת התפלגות פואסון כגבול של סדרת התפלגויות בינומיות ואולם כבר בשלב זה יוכל הקורא לוודא בעצמו את התכונה הבאה.

✎ **בעיה 3.28** (פואסון כגבול של בינומי במובן נקודתי)**.** תהי $\lambda > 0$ ויהיו $\{X_n\}_{n\in\mathbb{N}}$ משתנים מקריים בלתי-תלויים כך ש- $X_n \sim \mathrm{Bin}(n, \lambda/n)$. אזי לכל $a \in \mathbb{N}$ מתקיים $\lim_{n\to\infty} p_{X_n}(a) = p_Y(a)$ כאשר $Y \sim \mathrm{Po}(\lambda)$.

טענה 3.51 (תכונות התפלגות פואסונית)**.** יהי $X \sim \mathrm{Po}(\lambda)$.

(א) אם $Y \sim \mathrm{Po}(\eta)$ בלתי-תלוי ב-X אז $X + Y \sim \mathrm{Po}(\lambda + \eta)$

(ב) אם לכל $n \in \mathbb{N}_0$ מתקיים $(Y \mid X = n) \sim \mathrm{Bin}(n, p)$ אז $Y \sim \mathrm{Po}(\lambda p)$.

הוכחה. עבור סעיף (א) נחשב לפי נוסחת ההסתברות השלמה עבור $n \in \mathbb{N}_0$,

$$\begin{aligned}\mathbb{P}(X+Y=n) &= \sum_{i\in\mathbb{N}_0} \mathbb{P}(X=i)\mathbb{P}(X+Y=n \mid X=i) \overset{\text{המחוברים } i=n+1 \text{ ואילך הם אפס}}{=} \sum_{i=0}^{n} \frac{\lambda^i e^{-\lambda}}{i!} \frac{\eta^{n-i} e^{-\eta}}{(n-i)!} \\ &= \frac{e^{-\lambda-\eta}}{n!} \sum_{i=0}^{n} \binom{n}{i} \lambda^i \eta^{n-i} \overset{\text{נוסחת הבינום}}{=} \frac{(\lambda+\eta)^n e^{-\lambda-\eta}}{n!}\end{aligned}$$

עבור סעיף (ב) נחשב לכל $n \in \mathbb{N}_0$,

$$\begin{aligned}\mathbb{P}(Y=n) &= \sum_{i=0}^{\infty} \mathbb{P}(X=i)\mathbb{P}(Y=n \mid X=i) \overset{\text{המחוברים } n \text{ הראשונים הם אפס}}{=} \sum_{i=n}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \binom{i}{n} p^n (1-p)^{i-n} \\ &\overset{\text{שינוי אינדקס}}{=} (p\lambda)^n \sum_{i=0}^{\infty} \frac{\lambda^i e^{-\lambda}}{(n+i)!} \binom{n+i}{n} (1-p)^i = \frac{(p\lambda)^n e^{-\lambda}}{n!} \sum_{i=0}^{\infty} \frac{(\lambda - \lambda p)^i}{i!} \overset{\text{טור טיילור}}{=} \frac{(p\lambda)^n e^{-\lambda p}}{n!}\end{aligned}$$

■

לשני הסעיפים בטענה 3.51 ישנו גם הסבר אינטואיטיבי. אשר לסעיף (א), אם נשוב לפרשים הפרוסים הנבעטים על ידי סוסיהם, מובן שנצפה שכאשר נחבר את הנבעטים בשתי חטיבות, קצב ההיבעטות המשותף יהיה בדיוק סכום קצבי ההיבעטות בשתי היחידות. באשר לסעיף (ב), נניח שנסתכל על ניסוי לפיו מקרה של בעיטה למוות מדווח באופן בלתי תלוי בכל מקרה בסיכוי חצי. נוכל אפוא לחשוב על ניסוי כזה כניסוי דו-שלבי שבו קודם מגרילים את מספר הבעיטות לפי התפלגות פואסונית עם פרמטר מסויים, ולאחר מכן מגרילים לכל בעיטה משתנה ברנולי חצי שקובע עם היא מדווחת או לא. מספר הבעיטות המדווחות יהיה סכום של משתני ברנולי אלו ולכן הוא יתפלג בינומית עם מספר נסיונות המתאים למשתנה הפואסוני. כיוון שהיות שהמאורע שפרש ייבעט למוות ומותו ידווח עודנה בלתי תלויה בקירוב עבור פרשים שונים - הרי שהתפלגות מקרי המוות מבעיטת סוסים שדווחו צפויה להיות בקירוב פואסונית. היות שאנו מתוודעים למחצית המקרים בקירוב נצפה שקצב המאורעות המדווחים יהיה מחצית מקצב המאורעות בכלל.

✎ **בעיה 3.29.** להראות כי אם $X \sim \mathrm{Po}(\lambda)$, $Y \sim \mathrm{Po}(\eta)$ בלתי-תלויים, אז לכל $n \in \mathbb{N}$ המשתנה המקרי המותנה $(X \mid X+Y=n)$ מתפלג בינומית עם n נסיונות וסיכוי הצלחה $\lambda/(\lambda+\eta)$. כיצד נסביר תופעה זו?

3.5* החלוקה שיוצר משתנה מקרי

בפרק זה נציג נקודת מבט אחרת על משתנים מקריים כאמצעים לקודד מידע חלקי על תוצאתו של ניסוי. השקפה זו רואה במשתנה מקרי X חלוקה של מרחב ההסתברות למחלקות מן הצורה $X^{-1}(a)$ אשר לא ניתן להבחין ביניהן באמצעות היכרות עם ערכי המשתנה המקרי. בפרק זה נראה שהמבנה של מחלקות אלה מכריע האם משתנים מקריים הנם בלתי-תלויים והאם משתנה אחד יכול להרשם כפונקציה של משתנה אחר. זהו פרק העשרה נפרד שמטרתו לשפר את ההבנה של מושג האי-תלות של משתנים מקריים ולהכין את הקורא ללימודי מושגים שיופיעו בהמשך; הטענות המופיעות בו תוזכרנה בהנמקות והסברים בפרקי העשרה בהמשך, אך לא יעשה בהן שימוש ישיר.

הגדרה 3.52 (חלוקה שנוצרת על ידי משתנה מקרי בדיד)**.** יהי X משתנה מקרי בדיד על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נגדיר את $\pi(X)$, **החלוקה שיוצר** X, על ידי אוסף הקבוצות

$$\pi(X) = \{\{\omega' \in \mathrm{Supp}(X) \ : \ X(\omega) = X(\omega')\}\}_{\omega\in\Omega}.$$

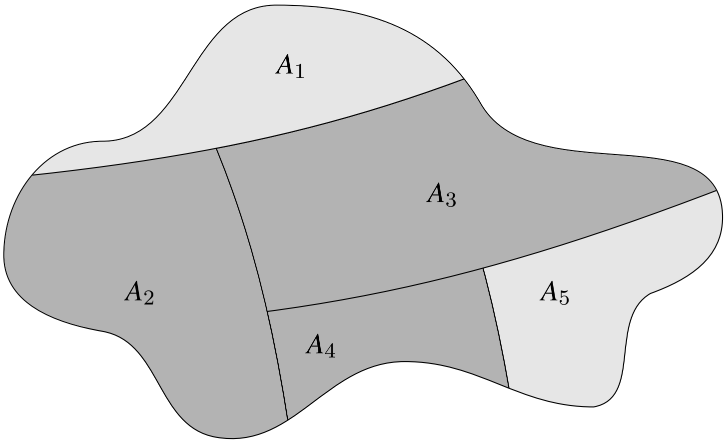
כיוון ש-Ω הוא מרחב בדיד, הרי שזוהי אומנם חלוקה בת-מניה, שכן כל קבוצה שאינה ריקה בה מאופיינת על ידי איבר שונה של Ω. חלוקה זו מתארת למעשה את האיברים במרחב המדגם שהמשתנה X אינו מבדיל ביניהם. כלומר X מאפשר לנו אך ורק להבחין בין מחלקות שונות של $\pi(X)$.

את אי התלות של משתנים מקריים נוכל לתאר במונחים של החלוקה המושרית על-ידיהם.

הגדרה 3.53 (חלוקות בלתי-תלויות)**.** יהי $(\Omega, \mathcal{F}, \mathbb{P})$ מרחב הסתברות בדידה ותהיינה $\pi_1, \pi 2$ שתי חלוקות של Ω. נאמר כי $\pi_1 \perp\!\!\!\perp \pi_2$ אם לכל $A \in \pi_1$, $B \in \pi_2$ מתקיים $A \perp\!\!\!\perp B$.

אבחנה 3.54. $X \perp\!\!\!\perp Y$ אם ורק אם $\pi(X) \perp\!\!\!\perp \pi(Y)$.

אבחנה זו היא מסקנה מידית מהגדרה 3.53 ומאבחנה 3.33.



תרשים 3.1 : תיאור גיאומטרי של חלוקה של מרחב הסתברות באמצעות חלוקה גסה המתוארת על ידי שני גווני אפור, ובאמצעות חלוקה עדינה יותר המתוארת על ידי השטחים $A_1, \ldots, A_5$. נשים לב כי השטח הבהיר ניתן לביטוי כ-$A_1 \cup A_5$ והשטח הכהה כ-$A_2 \cup A_3 \cup A_4$.

באופן דומה נוכל לתאר את החלוקה שמושרית על ידי מספר משתנים מקריים. ניתן לתאר חלוקה זו באופן ישיר, אך נוכל ללמוד יותר מבנייתה ישירות מהחלוקות שהמשתנים המקריים משרים.

הגדרה 3.55 (עדינות וגסות של חלוקות)**.** תהיינה $\mathcal{A}$ ו-$\mathcal{B}$ חלוקות של קבוצה Ω. נאמר כי $\mathcal{A}$ היא עידון $\mathcal{B}$ (או כי $\mathcal{B}$ גסה יותר מ-$\mathcal{A}$) אם לכל $B \in \mathcal{B}$ קיים $I \subset \mathcal{A}$ כך ש-$B = \bigcup_{A \in I} A$.

תיאור גיאומטרי של חלוקה גסה ועידון שלה מתואר בתרשים 3.1.

אבחנה 3.56 (תנאי שקול לעידון)**.** תהיינה $\mathcal{A}$ ו-$\mathcal{B}$ חלוקות של קבוצה Ω. $\mathcal{A}$ מעדנת את $\mathcal{B}$ אם ורק אם כל $\omega_1, \omega_2 \in \Omega$ שנמצאים בשתי מחלקות שונות ב$\mathcal{B}$ נמצאים גם בשתי מחלקות שונות ב-$\mathcal{A}$.

טענה 3.57 (חלוקה נוצרת מחלוקות)**.** יהי $\mathcal{P} = \{\pi_1, \ldots, \pi_n\}$ אוסף של חלוקות של Ω, אזי קיימת חלוקה שמעדנת את כל החלוקות ב-$\mathcal{P}$ שהנה גסה יותר מכל חלוקה אחרת שמעדנת את $\mathcal{P}$. נסמן חלוקה זו ב-$\pi(\mathcal{P})$ ונכנה אותה **החלוקה הנוצרת מ-**$\mathcal{P}$.

הוכחה. נגדיר לכל $A_1 \in \pi_1, \ldots, A_N \in \pi_N$ קבוצה

$$B_{A_1,\ldots,A_N} = \bigcap_{n \in [N]} A_n,$$

ונשים לב כי $\mathcal{B} = \{B_{A_1,\ldots,A_N}\}_{A_i \in \pi_i}$ היא חלוקה בת-מניה של Ω. כדי להיווכח כי B מעדנת את π_i נשים לב כי לכל $A_i \in \pi_i$ מתקיים

$$A_i = \dot{\bigcup_{A_j \in \pi_j, j \neq i}} B_{A_1,\ldots,A_N}$$

כדי להיווכח כי B גסה יותר מכל חלוקה אחרת המעדנת את $\mathcal{P}$, נשתמש באבחנה 3.56 : נשים לב כי כל שני איברים ω_1, ω_2 שאינם באותה מחלקה ב-$\mathcal{B}$, נמצאים בשתי מחלקות שונות לפחות באחת מהחלוקות $\{\pi_i\}_{i \in \mathbb{N}}$, ולכן עליהם להיות בשתי מחלקות שונות של כל חלוקה שמעדנת אותה. ■

הגדרה 3.58 (חלוקה שנוצרת על ידי מספר משתנים מקריים)**.** יהי $(X_1, \ldots, X_N)$ וקטור מקרי בדיד על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נגדיר את $\pi(X_1, \ldots, X_N) := \pi(\pi(X_1), \ldots, \pi(X_N))$, ונכנה אותה **החלוקה שיוצרים** $(X_1, \ldots, X_N)$.

כעת נוכל לתאר אי-תלות בין אוספים של משתנים מקריים במושגים של חלוקות.

טענה 3.59 (אי-תלות של אוסף משתנים מקריים במונחי חלוקות)**.** יהיו $(X_1, \ldots, X_N)$ ו-$(Y_1, \ldots, Y_M)$ וקטורים מקריים בדידים על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. אזי $X_1, \ldots, X_N$ בלתי-תלויים ב-$Y_1, \ldots, Y_M$, אם ורק אם $\pi(X_1, \ldots, X_N) \perp\!\!\!\perp \pi(Y_1, \ldots, Y_M)$.

הוכחה. בכיוון אחד, נניח כי $\pi(X_1, \ldots, X_N) \perp\!\!\!\perp \pi(Y_1, \ldots, Y_M)$ ונוכיח כי המשתנים המקריים $X_1, \ldots, X_N$ בלתי-תלויים ב-$Y_1, \ldots, Y_M$. עבור $A \in \mathbb{R}^N$ ו-$B \in \mathbb{R}^M$ נחשב:

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \sum_{a \in A} \sum_{b \in B} \mathbb{P}(X = a, Y = b) \\ &= \sum_{a \in A} \sum_{b \in B} \mathbb{P}(X = a)\mathbb{P}(Y = b) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B),\end{aligned}$$

כאשר השוויון האמצעי נובע מכך ש $\{\omega \in \mathrm{Supp}(X) : X = a\}$ ו-$\{\omega \in \mathrm{Supp}(Y) : Y = b\}$ הן מחלקות ב-$\pi(X_1, \ldots, X_N)$ וב-$\pi(Y_1, \ldots, Y_M)$ בהתאמה. הכיוון השני נובע באופן מידי מאותה אבחנה עצמה. ■

כעת נראה כי העובדה שמשתנה מקרי יכול להירשם כפונקציה של אוסף משתנים אחרים מקודדת במבנה החלוקות שלהם.

טענה 3.60 (קשר ישיר בין משתנים מקריים במונחי חלוקות)**.** יהי $(X_1, \ldots, X_N)$ וקטור מקרי בדיד על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ ויהי Y משתנה מקרי על מרחב זה. אזי קיימת פונקציה $f : \mathbb{R}^N \to \mathbb{R}$ כך ש-$Y \overset{\text{a.s.}}{=} f(X_1, \ldots, X_n)$ אם ורק אם $\pi(X_1, \ldots, X_N)$ מעדנת את $\pi(Y)$.

הוכחה. **בכיוון אחד** - אם $Y \overset{\text{a.s.}}{=} f(X_1, \ldots, X_n)$ אז

$$\begin{aligned}Y^{-1}(a) &= \{\omega : (X_1(\omega), \ldots, X_N(\omega)) \in f^{-1}(a)\} \\ &= \dot{\bigcup_{(b_1,\ldots,b_N) \in f^{-1}(a)}} \{\omega : X_1(\omega) = b_1, \ldots, X_N(\omega) = b_N\} \\ &= \dot{\bigcup_{(b_1,\ldots,b_N) \in f^{-1}(a)}} \bigcap_{i \in [N]} X_i^{-1}(b_i),\end{aligned}$$

ולכן $\pi(X_1, \ldots, X_N)$ מעדנת את $\pi(Y)$. **בכיוון השני** - אם $\pi(X_1, \ldots, X_N)$ מעדנת את $\pi(Y)$, עבור $x_1, \ldots, x_n$ נסמן $\mathrm{Supp}(X_n) \bigcap_{n \in [N]} X_n^{-1}(x_n) \bigcap_{n \in [N]} = A_{x_1,\ldots,x_n}$ ונסיק כי קיים b כך שלכל $\omega \in A_{x_1,\ldots,x_n}$ מתקיים $Y(\omega) = b$. לכן נוכל להגדיר $f(x_1, \ldots, x_n) = b$ ולקבל כי $Y \overset{\text{a.s.}}{=} f(X_1, \ldots, X_n)$. ■

בעיות הרחבה והעשרה

✎ **בעיה 3.30** (התפלגות פואסון מורכבת). מספר הנכנסים לחנות בגדים בשעה מתפלג פואסונית עם פרמטר 3. כל מבקר רוכש בהסתברות 0.5 פריט אחד ובהסתברות 0.5 שני פריטים. יש לרשום נוסחא מפורשת להתפלגות מספר הפריטים שנרכשו במשך שעתיים.

✎ **בעיה 3.31** (יחס סדר). יהיו X, Y ו-Z שלושה משתנים מקריים שווי התפלגות ובלתי תלויים. נסמן ב-A את המאורע שאין אף זוג משתנים שערכו זהה וב-B את המאורע $X < Y < Z$ יש לחשב את $\mathbb{P}(B \mid A)$ תוך שימוש קפדני בנימוקים פורמליים.

✎ **בעיה 3.32.** עליסה נמצאת בחדר מבואה עם דלת כניסה ושלוש דלתות יציאה. דלת אחת מוליכה לארץ הפלאות, דלת שניה מוליכה לאנגליה והדלת השלישית מוליכה בחזרה לחדר המבואה. בכל ביקור בחדר המבואה עליסה בוחרת בדלת המוליכה לאנגליה בסיכוי p_1 ובדלת המוליכה לארץ הפלאות בסיכוי p_2 עבור $p_1 + p_2 < 1$ נתונים.

(א) מה הסיכוי שעליסה תגיע בסופו של דבר לארץ הפלאות?

(ב) כיצד מתפלג מספר הביקורים של עליסה בחדר המבואה בהינתן שבסופו של דבר תגיע לארץ הפלאות?

✎ **בעיה 3.33 (*).** [דילול משתנה מקרי גיאומטרי] יהי X משתנה מקרי המקיים $(X+1) \sim \mathrm{Geo}(p)$ ויהי Y משתנה מקרי אשר לכל לכל $n \in \mathbb{N}_0$ מקיים $(Y \mid X = n) \sim \mathrm{Bin}(n, q)$. יש להראות כי $(Y+1) \sim \mathrm{Geo}\big(p/(p+q-pq)\big)$. כיצד נסביר תופעה זו?

✎ **בעיה 3.34 (*).** [פואסון כקירוב לבינומי] מכונת צילום שכפלה מאתיים בחינות בהסתברות. בכל בחינה עלולה ליפול טעות שכפול בהסתברות 0.01 באופן בלתי-תלוי. נסתכל על משתנה מקרי X_i המהווה מציין לטעות בבחינה ה-i. ונסמן ב- $\{Y_i\}_{i} \in [200]$ סדרה של משתני פואסון עם פרמטר $\ln 0.99$. כמו כן נרשום $X = \sum_{i=0}^{200} X_i$ ו-$Y = \sum_{i=0}^{200} Y_i$.

(א) יש להראות כי $(Y_1 \mid Y_1 \in \{0,1\})$ ו-X_1 שווי התפלגות.

(ב) מה ההסתברות ש-$Y_i \in \{0,1\}$?

(ג) כיצד נחסום משני הצדדים באמצעות התפלגות פואסון את ההסתברות שתפולנה טעויות שכפול בדיוק ב-k בחינות?

(רמז: ניתן להראות כי $\mathbb{P}(Y = k) - \mathbb{P}(Y \neq X) < \mathbb{P}(X = k) < \mathbb{P}(Y = k) + \mathbb{P}(Y \neq X)$)

✎ **בעיה 3.35 (**).** [פרדוקס הבחירה] יהיו X ו-Y שני משתנים מקריים שווי התפלגות ובלתי תלויים על $\mathbb{N}$ שהתפלגותם אינה קבועה. בשתי מעטפות נפרדות מונחים X שקלים ו-Y שקלים. יש למצא אסטרטגיה (שיכולה להיות הסתברותית ותלויה באקראיות נוספת) לבחור את המעטפה שבה סכום גבוה יותר בהסתברות **גדולה ממש** מחצי בהינתן ש-$X \neq Y$. לאסטרטגיה אסור להתחשב בהתפלגות של X ו-Y (היא אינה ידועה).

פרק 4

התוחלת

ה"יתרון" בתורת הסיכויים הוא מכפלת הסכום שאנו מייחלים לו בהסתברות לזכות בו; זה הוא הסכום שראוי לחלקו כאשר אנחנו לא מעוניינים ליטול את הסיכון אלא לחלק את הסכום המיוחל לפי ההסתברויות. חלוקה זו היא החלוקה ההוגנת היחידה אם מתעלמים מנסיבות משונות; מפני שנתחים שווים של הסתברות מובילים לנתחים שווים מהסכום המיוחל. אנחנו נכנה את המושג הזה בשם "תוחלת מתמטית".

– פְּיֵיר סִימוֹן לָפְּלָס, הצגת מושג התוחלת ב**תיאוריה אנליטית של הסתברות**, 1814.

בפרק זה נגדיר את מושג ה**תוחלת**. מושג מתמטי שיתאר תוצאה ממוצעת צפויה של ניסוי. נשים לב שזו הפעם הראשונה בה נתעניין בניתוח הערכים שמקבל משתנה מקרי, ולא רק במאורעות שמיוצגים על ידם. בפרק הבא נקשור את התוחלת עם הממוצע של תוצאות ניסויים חוזרים בלתי-תלויים, ואולם עוד בטרם נבסס קשר זה, יתגלו לפנינו שימושים נוספים של התוחלת עת נקבל באמצעותה הערכות טובות יותר להתפלגותם של משתנים מקריים. בפרקים הבאים נכליל את התוחלת ואת שימושיה בהציגנו את מושגי ה**שונות** וה**מומנט**.

4.1 תוחלת

את הגדרתה של התוחלת המובאת להלן אפשר לראות כ"ממוצע משוקלל של מרחב המדגם לפי פונקציית ההסתברות הנקודתית". אם נשוב ונדמה את הניסוי ההסתברותי לשליפת כדורים שוויה הסתברות מכד אטום ורשימת המספר המוטבע עליהם, הרי שתוחלת הניסוי מחושבת על ידי הוצאת כל הכדורים מהכד ועריכת ממוצע של ערכיהם. את המושג טבע לראשונה כְּרִיסְטִיאַן הוּיְגֶנְס (Christiaan Huygens) ההולנדי ב-1657, אך טיפול מדוייק בו נאלץ לחכות לעבודתו של פְּיֵיר-סִימוֹן לָפְּלָס (Laplace Pierre-Simon) שיצאה לאור ב-1814.

הגדרה 4.1 (תוחלת)**.** יהי X משתנה מקרי המוגדר על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$.

ה**תוחלת (expectation)** של X, מוגדרת על ידי

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$$

כאשר טור זה מתכנס בהחלט. אחרת נאמר שלמשתנה X אין תוחלת סופית.

את התוחלת נוכל לבטא גם במונחי התפלגות מבלי להידרש למרחב ההסתברות עצמו.

טענה 4.2 (הגדרה שקולה לתוחלת)**.** יהי X משתנה מקרי המוגדר על מרחב הסתברות בדידה. התוחלת של X ניתנת לחישוב על ידי

$$\mathbb{E}(X) = \sum_{s \in \mathbb{R}} s \cdot \mathbb{P}(X = s)$$

כאשר טור זה מתכנס בהחלט. אחרת נאמר כי ל-X אין תוחלת סופית. בפרט, התוחלת של משתנה מקרי תלויה רק בהתפלגותו.

הוכחה.

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{s \in \mathrm{Supp}(X)} \sum_{\omega \in X^{-1}(s)} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{s \in \mathbb{R}} s \cdot \mathbb{P}(X = s)$$

■

מטענה 4.2 אנו למדים כי תוחלת היא תכונה של התפלגותו של משתנה מקרי, שאינה תלויה במרחב ההסתברות שעליו הוא מוגדר. מעתה נעשה שימוש מובלע בטענה 4.2, כמעין הגדרה חלופית של התוחלת, מבלי לצטטה.

דוגמא 4.3 (תוחלת משתנים מקריים מוכרים)**.**

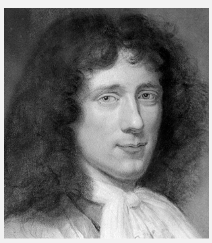
(א) תוחלת משתנה מקרי **ברנולי** $X \sim \mathrm{Ber}(p)$ היא

$$\mathbb{E}(X) = p \cdot 1 + (1 - p) \cdot 0 = \mathbf{p}.$$

לפיכך, התוחלת של $\mathbb{1}(A)$ (האינדיקטור של המאורע A) היא $\mathbb{P}(A)$.

(ב) תוחלת משתנה מקרי **אחיד** על $[n]$ היא

$$\mathbb{E}(X) = \sum_{k \in [n]} k\,\mathbb{P}(X = k) = \frac{n(n+1)}{2n} = \frac{\mathbf{n+1}}{\mathbf{2}}.$$



כְּרִיסְטִיאַן הוּיְגֶנְס (1629-1695), מתמטיקאי, פיסיקאי, מהנדס ואסטרונום הולנדי, ממחוללי המהפיכה המדעית. נודע יותר מכל בשל המצאת שעון המטוטלת. מכונה לעיתים קרובות כאבי הפיסיקה התיאורטית. היה הראשון להציע תיאור של האור במונחי גלים. ספרו "הרציונל במשחקי מזל" (Van Rekeningh in Spelen van Gluck) הרחיב והכליל את תגליותיהם של פסקל ופרמה בחקר סוגיית המשחק שנקטע, וביסס אותן לכדי תיאוריה שעיקר עניינה – ניתוח מושג התוחלת והשימוש בו להגדרה וניתוח של משחקים הוגנים.

(ג) תוחלת משתנה מקרי **בינומי** $X \sim$ [illegible](N, p) היא

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \quad = \sum_{k=0}^{n} k \frac{n!}{(k-n)!k!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \overset{[m=k-1]}{=} \sum_{m=0}^{n-1} \frac{n!}{(n-m-1)!m!} p^{m+1} (1-p)^{n-m-1} \\ &= np \sum_{m=0}^{n-1} \frac{(n-1)!}{(n-1-m)!m!} p^m (1-p)^{n-1-m} \overset{\text{בינום}}{=} np\Big(p + (1-p)\Big)^{n-1} = \mathbf{np}.\end{aligned}$$

(ד) תוחלת משתנה מקרי **פואסוני** $X \sim \mathrm{Po}(\lambda)$ היא

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \overset{[m=n-1]}{=} \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \boldsymbol{\lambda}.$$

נשים לב שזו גם תוחלתו של משתנה $Y \sim \mathrm{Bin}(n, \lambda/n)$ כפי שניתן היה לצפות, לאור בעיה 3.28 (אף-על-פי שנחוצים תנאים מתאימים בכדי שגבול התוחלות יהיה שווה לתוחלת הגבול).

(ה) תוחלת משתנה מקרי **גיאומטרי** $X \sim \mathrm{Geo}(p)$ היא

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} np(1-p)^{n-1}$$

כדי לחשב טור זה נזכר בנוסחה לסכום של טור הנדסי עבור $|x| < 1$:

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

נגזור את שני הצדדים ונקבל:

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

כאשר הגזירה מוצדקת (לפי טענה א.19) כיון שטור הנגזרות הוא טור חזקות המתכנס במידה שווה בסביבת x. נציב כעת $x = 1-p$ ונקבל:

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

$\square$

דוגמא 4.4 (משתנה ללא-תוחלת)**.**

יהי $X \sim (\mathrm{Geo}\, 1/2)$, ונגדיר $Y = 2^X$ ו-$Z = (-2)^X$, $W = (-2)^X/X$. . נראה כי $\mathbb{E}(Y)$,$\mathbb{E}(Z)$ ו-$\mathbb{E}(W)$ אינן מוגדרות.

תשובה: שתי התוחלות $\mathbb{E}(Y)$ ו-$\mathbb{E}(Z)$, אינן מתכנסות לפי חישוב ישיר

$$\mathbb{E}(Y) = \sum_{n=1}^{\infty} \mathbb{P}(Y=n)2^n = \sum_{n=1}^{\infty} 1 = \infty, \qquad \mathbb{E}(Z) = \sum_{n=1}^{\infty} \mathbb{P}(Z=n)(-2)^n = \sum_{n=1}^{\infty} (-1)^n.$$

אשר ל-$\mathbb{E}(W)$ - אומנם

$$\sum_{n=1}^{\infty} \mathbb{P}(Z=n)\frac{(-2)^n}{n} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} = 0.$$

אבל טור זה אינו מתכנס בהחלט - כיוון שמתקיים

$$\sum_{n=1}^{\infty} \mathbb{P}(Z=n)\left|\frac{(-2)^n}{n}\right| = \sum_{n=1}^{\infty} \frac{1}{n} = \infty,$$

ולכן גם תוחלת זו אינה מוגדרת. □

טענה 4.5 (תכונות התוחלת)**.** יהיו X, Y משתנים מקריים בדידים על אותו מרחב הסתברות, המקיימים $\mathbb{E}(Y), \mathbb{E}(X) \in \mathbb{R} \cup \{\infty\}$. אזי

(א) **חיוביות:** אם $X \overset{\text{a.s.}}{\geq} 0$ אז $\mathbb{E}(X) \geq 0$. ואם $X \overset{\text{a.s.}}{>} 0$ אז $\mathbb{E}(X) > 0$.

(ב) **ליניאריות:** $\mathbb{E}(aX+bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ לכל $a, b \in \mathbb{R}$.

(ג) **מונוטוניות:** אם $\mathbb{P}(X \geq Y) = 1$ אז $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

הוכחה. **סעיף א.** לפי הגדרה מתקיים

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$$

והסעיף נובע מכך שסכום גורמים אי-שליליים הנו אי-שלילי וסכום גורמים חיוביים הנו חיובי.

סעיף ב. נחשב לפי הגדרה:

$$\begin{aligned}\mathbb{E}(aX+bY) &= \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega))\mathbb{P}(\{\omega\}) \\ &= a\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) + b\sum_{\omega \in \Omega} Y(\omega)\mathbb{P}(\{\omega\}) = a\mathbb{E}(X) + b\mathbb{E}(Y).\end{aligned}$$

סעיף ג. נרשום $X = (X-Y) + Y$ ונחשב לפי הסעיפים הקודמים:

$$\mathbb{E}(X) \overset{\text{סעיף ב}}{=} \mathbb{E}(X-Y) + \mathbb{E}(Y) \overset{\text{סעיף א}}{\geq} \mathbb{E}(Y).$$

■

נוכל להסיק מליניאריות התוחלת עבור סכום של שני משתנים מקריים, נוכל להסיק באינדוקציה ליניאריות לכל מספר סופי של משתנים $\{X_i\}_{i\in[N]}$ בעלי תוחלת סופית. או בכתיב פורמלי –

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N} E(X_i).$$

דוגמא 4.6 (תוחלת משתנה בינומי באמצעות ליניאריות התוחלת)**.**

יהיו $p \in (0,1)$ ו-$n \in \mathbb{N}$ ויהי $X \sim \text{Bin}(n,p)$. בדוגמא 5.29ג חישבנו את תוחלתו של X באופן ישיר. כעת נחשבה ביתר קלות באמצעות ליניאריות התוחלת. יהיו $\{X_k\}_{k\in[n]}$ משתנים מקריים בלתי-תלויים המתפלגים ברנולי עם סיכוי הצלחה p ונסמן את סכומם ב S. S ו-X הנם שווי התפלגות לפי טענה 3.46 – ולכן הם גם שווי תוחלת. כעת, לפי ליניאריות התוחלת, טענה 4.5ב, מתקיים

$$\mathbb{E}\left(\sum_{k=1}^{n} X_i\right) \overset{\text{ליניאריות}}{=} \sum_{k=1}^{n} \mathbb{E}(X_i) = \sum_{k=1}^{n} p = np.$$

כאשר את השוויון האמצעי קיבלנו בדוגמא 5.29א. □

טענה 4.7 (נוסחא לתוחלת של משתנה מקרי טבעי באמצעות פונקציית ההתפלגות השיורית)**.** יהי X משתנה מקרי המקיים $\text{Supp}(X) \subset \mathbb{N}$. אזי,

$$\mathbb{E}(X) = \sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n)$$

הוכחה. נחשב תוחלת לפי ההגדרה השקולה וננצל את העובדה שהמחוברים בטור אי-שליליים כדי להחליף סדר סכימה באמצעות משפט פוביני (משפט א.15).

$$E(X) = \sum_{n=1}^{\infty} n\mathbb{P}(X=n) = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \mathbb{P}(X=n) = \sum_{\substack{k,n\in\mathbb{N}\\ k\leq n}} \mathbb{P}(X=n) = \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} \mathbb{P}(X=n) = \sum_{n\in\mathbb{N}} \mathbb{P}(X\geq n)$$

■

דוגמא 4.8 (תוחלת משתנה מקרי גיאומטרי באמצעות פונקציית התפלגות שיורית)**.**

יהי $X \sim \text{Geo}(p)$ עבור $p \in (0,1)$. בדוגמא 5.29ה חישבנו את תוחלתו של X תוך שימוש בחשבון אינפיניטיסימלי. כעת נחשבה ביתר קלות באמצעות טענה 4.7. נרשום

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

נשים לבי כי ליניאריות התוחלת מתקיימת גם עבור משתנים מקריים תלויים. בדוגמא שלהלן נראה כיצד נוכל לחשב את התוחלתו של משתנה מקרי מורכב באמצעות פירוקו לסכום של משתנים מקריים פשוטים יותר.

דוגמא 4.9 (תוחלת היפר-גיאומטרית)**.**

מכד אטום שבו a כדורים אדומים ו-b כדורים שחורים נשלפים k כדורים. מהי תוחלת מספר הכדורים האדומים שישלפו?

תשובה: כמו בדוגמא 3.12, נוכל לתאר את שליפה הכדורים באמצעות תמורה על $a+b$ כדורים כאשר הכדורים שנשלף בצעד ה-i יתאים לתמונתו של האיבר i ואנו משווים בנפשנו כאילו המשכנו לשלוף את כל הכדורים בזה אחר זה. משיקולי סימטריה על מרחב זה, התפלגות צבע הכדור ה-i זהה בדיוק להתפלגות צבע הכדור ה-j. נסמן ב-X_i משתנה אינדיקטור למאורע "הכדור ה-i שנשלף היה אדום", וניווכח אפוא כי

אלו משתנים שווי התפלגות (אך לא בלתי-תלויים!). נשים לב שעלינו לחשב את $\mathbb{E}\left(\sum_{i\in[k]} X_i\right)$. נחשב ונקבל

$$\mathbb{E}\left(\sum_{i\in[k]} X_i\right) = \sum_{i\in[k]} \mathbb{E}(X_i) = \sum_{i\in[k]} \mathbb{E}(X_1) = k\mathbb{P}(\{\text{הכדור הראשון אדום}\}) = \frac{ka}{a+b}.$$

נשים לב כי לאור דוגמא 4.17, נוכל להסיק כי תוחלת מספר הכדורים האדומים שישלפו בדוגמא 4.9 אינה משתנה אם נחזיר כל כדור ששלפנו לכד. □

רבים מן המשתנים המקריים שיצרנו בפרק הקודם הוצגו כפונקציה של משתנים מקריים אחרים. כעת נבקש לבטא את תוחלתם של משתנה כזה במונחים של התפלגות המשתנים שיוצרים אותו.

טענה 4.10 (תוחלת של פונקציה של משתנה מקרי). יהי X משתנה מקרי בדיד ותהי $f : \mathbb{R} \to \mathbb{R}$ פונקציה. אזי המ"מ $Y = f(X)$ מקיים

$$\mathbb{E}\big(Y\big) = \sum_{x\in\mathbb{R}} f(x)p_X(x)$$

בתנאי שטור זה מתכנס בהחלט.

הוכחה. נחשב לפי הגדרה.

$$\begin{aligned}\mathbb{E}(Y) &= \sum_{\omega\in\Omega} f(X(\omega))\mathbb{P}(\{\omega\}) = \sum_{x\in\mathbb{R}} \sum_{\omega\in X^{-1}(x)} f(X(\omega))\mathbb{P}(\{\omega\}) \\ &= \sum_{x\in\mathbb{R}} f(x) \sum_{\omega\in X^{-1}(x)} \mathbb{P}(\{\omega\}) = \sum_{x\in\mathbb{R}} f(x)p_X(x).\end{aligned}$$

נשים לב שלפי הגדרה, התוחלת קיימת אם ורק אם הטורים מתכנסים בהחלט ולכן החלפת סדר הסכימה מותרת לפי משפט רימן על שינוי סדר סכימה לטורים מתכנסים בהחלט (טענה א.12). ■

מסקנה 4.11 (תוחלת פונקציה של וקטור מקרי). הי וקטור מקרי $X_1, \ldots, X_k$ בדיד ותהי $f : \mathbb{R}^k \to \mathbb{R}$ פונקציה. המשתנה המקרי $Y = f(X_1, \ldots, X_k)$ מקיים

$$\mathbb{E}\big(Y\big) = \sum_{(x_1,\ldots,x_k)\in\mathbb{R}^k} f(x_1, \ldots, x_k)p_{X_1,\ldots,X_k}(x_1, \ldots, x_k)$$

כאשר $p_{X_1,\ldots,X_k}$ היא פונקציית ההתפלגות המשותפת הנקודתית שהוגדרה בהגדרה 3.22).

✎ **בעיה 4.1.** להוכיח את מסקנה 4.11.

אי-שוויון הבא, אותו גילה המתמטיקאי הדני יוהאן ינסן בשנת 1906, קושר בין תוחלת של פונקציה קמורה של משתנה מקרי לבין הפעלתה של אותה פונקציה על התוחלת.

טענה 4.12 (אי-שוויון ינסן). יהיו $a, b \in \{\pm\infty\} \cup \mathbb{R}$, תהי $f : \mathbb{R} \to \mathbb{R}$ פונקציה קמורה (ר' פרק א.4.2) על (a, b) ויהי X משתנה מקרי בדיד הנתמך על (a, b). אזי מתקיים

$$f(\mathbb{E}(X)) \leq \mathbb{E}(f(X))$$

הוכחה. נסמן $x_0 = \mathbb{E}(X)$. ניזכר כי מכך ש-f קמורה נובע שקיים β כך שלכל $x \in (a, b)$ מתקיים

$$f(x) \geq f(x_0) + \beta(x - x_0).$$

כעת נחשב

$$\begin{aligned}\mathbb{E}(f(X)) &= \sum_{x\in\mathbb{R}} f(x)\mathbb{P}(X = x) \geq \sum_{x\in\mathbb{R}}(f(x_0) + \beta(x - x_0))\mathbb{P}(X = x) \\ &= f(x_0) + \beta\sum_{x\in\mathbb{R}}(x - x_0)\mathbb{P}(X = x) = f(\mathbb{E}(X)) + \beta(\mathbb{E}(X) - x_0) = f(\mathbb{E}(X))\end{aligned}$$

■

מסקנה 4.13. יהיו $a, b \in \{\pm\infty\} \cup \mathbb{R}$, תהי $f : \mathbb{R} \to \mathbb{R}$ פונקציה קעורה (ר' פרק א.4.2) על (a, b) ויהי X משתנה מקרי בדיד הנתמך על (a, b). אזי מתקיים

$$f(\mathbb{E}(X)) \geq \mathbb{E}(f(X))$$

אי-שוויון ינסן מספר הוכחה פשוטה וישירה של יחס הסדר בין ממוצע הארמוני, הנדסי וחשבוני.

דוגמא 4.14 (אי-שוויון הממוצעים).

נשתמש באי-שוויון ינסן בכדי להראות כי לכל סדרה של ערכים חיוביים $\{x_k\}_{k\in[n]}$ מתקיים

$$\min(\{x_k\}_{k\in\mathbb{N}}) \leq \frac{n}{\sum_{k\in[n]} \frac{1}{x_i}} \leq \prod_{i\in[n]} \sqrt[n]{x_i} \leq \frac{\sum_{i\in[n]} x_i}{n} \leq \sqrt{\frac{\sum_{i\in[n]} x_i^2}{n^2}} \leq \max(\{x_k\}_{k\in\mathbb{N}}).$$

כלומר,

מכסימום ≤ ממוצע ריבועי ≤ ממוצע חשבוני ≤ ממוצע הנדסי ≤ ממוצע הארמוני ≤ מינימום

תשובה: נסתכל על משתנה מקרי X המקבל את הערך x_k בהסתברות $1/n$ לכל $k \in [n]$. נשים לב כי

$$\mathbb{P}(X \in \mathbb{R}) = \mathbb{P}(X \in \{x_k\}_{k\in[n]}) = \sum_{k\in[n]} \frac{1}{n} = 1,$$

ולכן זו התפלגות מוגדרת היטב. נשים לב כי הפונקציה $f(x) = x^\alpha$ הנה קמורה בתחום $\mathbb{R}_+$ לכל $\alpha > 0$ וקעורה לכל $\alpha < 0$. נסמן $m_\alpha = \mathbb{E}(X^\alpha)^{\alpha^{-1}}$ ומהפעלת אי-שוויון ינסן על פונקציות אלה נסיק כי לכל $0 \leq \alpha_1 \leq \alpha_2$ מתקיים

$$m_{\alpha_1} \leq m_{\alpha_2} \tag{4.1}$$

נשים לב כי

$$\lim_{\alpha\to-\infty} m_\alpha = \min(\{x_k\}_{k\in\mathbb{N}}) \qquad m_{-1} = \frac{n}{\sum_{k\in[n]} \frac{1}{x_i}}$$

$$\lim_{\alpha\to 0+} m_\alpha = \prod_{i\in[n]} \sqrt[n]{x_i} \qquad m_1 = \frac{\sum_{k\in[n]} x_i}{n}$$

$$m_2 = \sqrt{\frac{\sum_{i\in[n]} x_i^2}{n^2}} \qquad \lim_{\alpha\to\infty} m_\alpha = \max(\{x_k\}_{k\in\mathbb{N}}).$$

כעת נוכל להסיק את הטענה מהפעלת (4.1) על הביטויים השונים. □

עבור משתנים המקריים בלתי-תלויים מקיימת התוחלת גם את תכונת ה**כפליות**.

טענה 4.15 (כפליות התוחלת למשתנים בלתי-תלויים)**.** יהיו X, Y משתנים מקריים בדידים, בלתי-תלויים ובעלי תוחלת סופית, אז התוחלת של XY קיימת ומקיימת,

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

הוכחה.

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{x\in\mathbb{R}}\sum_{y\in\mathbb{R}} xyp_{X,Y}(x,y) = \sum_{x\in\mathbb{R}}\sum_{y\in\mathbb{R}} xyp_X(x)p_Y(y)\\ &= \sum_{x\in\mathbb{R}} xp_X(x)\sum_{y\in\mathbb{R}} yp_Y(y) = \mathbb{E}(X)\mathbb{E}(Y).\end{aligned}$$

■

למעשה נוכל להגדיר אי-תלות במונחים של כפליות התוחלת באמצעות הטענה הבאה.

טענה 4.16 (תוחלת של משתנים מקריים בלתי-תלויים)**.** יהיו X, Y משתני מקריים בדידים על מרחב הסתברות. אזי X ו-Y בלתי-תלויים אם ורק אם לכל שתי פונקציות $f, g : \mathbb{R} \to \mathbb{R}$ מתקיים

$$\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y)).$$

הוכחה. נניח שלכל $f, g : \mathbb{R} \to \mathbb{R}$ מתקיים $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$. אזי לכל $x \in \operatorname{Supp}(X)$ ולכל $y \in \operatorname{Supp}(Y)$ מתקיים

$$\mathbb{P}(X=x, Y=y) = \mathbb{E}(\mathbb{1}_{X=x}\mathbb{1}_{Y=y}) = \mathbb{E}(\mathbb{1}_{X=x})\mathbb{E}(\mathbb{1}_{Y=y}) = \mathbb{P}(X=x)\mathbb{P}(Y=y)$$

ולכן, לפי אבחנה 3.33 המשתנים המקריים בלתי-תלויים. הכיוון השני, לפי טענה 3.36, נובע מטענה 4.15.

■

דוגמא 4.17 (מכפלת שתי קוביות)**.**

מהי תוחלת מכפלת שתי קוביות בלתי תלויות?

תשובה: אם X_1 ו-X_2 הם משתנים מקריים בלתי תלויים המתפלגים אחיד ב-$[6]$ הרי שתוחלת מכפלתם הינה

$$\mathbb{E}(X_1X_2) = \mathbb{E}(X_1)\mathbb{E}(X_2) = \left(\frac{6+1}{2}\right)^2 = 12\frac{1}{4}.$$

4.2 תוחלת תחת התניה

בפרק הקודם הגדרנו את התפלגותו של משתנה מקרי בהנתן מאורע A בעל הסתברות חיובית באמצעות החלפת פונקציית ההסתברות ב-$\mathbb{P}_A$, פונקציית ההסתברות המותנית בA (ר' הגדרה 2.2). כעת נוכל לחשב את תוחלתו של המשתנה המקרי בהינתן מאורע A לפי

$$\mathbb{E}(X \mid A) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}_A(\{\omega\}) = \sum_{s\in\mathrm{Supp}(X)} s\mathbb{P}(X = s \mid A).$$

נוסחא זו מאפשרת לנו להגדיר תכונה מקבילה לנוסחת ההסתברות השלמה בעבור התוחלת:

טענה 4.18 (נוסחאת התוחלת השלמה)**.** תהי $(A_n)_{n\in\mathbb{N}}$ חלוקה של מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. מ"מ X בעל תוחלת סופית מקיים

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{E}(X\mathbb{1}_{A_i}) = \sum_{i=1}^{\infty} \mathbb{E}(X \mid A_i)\mathbb{P}(A_i).$$

כאשר אנו מפרשים איברים באגף ימין כשווים ל-0 אם $\mathbb{P}(A_i) = 0$.

הוכחה. לקבלת השוויון הראשון, נרשום לפי הגדרה:

$$\mathbb{E}(X) = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{i\in\mathbb{N}}\sum_{\omega\in A_i} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{i\in\mathbb{N}} \mathbb{E}(X\mathbb{1}_{A_i}).$$

לקבלת השוויון השני, נפעיל את נוסחת ההסתברות השלמה ונקבל

$$\begin{aligned}\mathbb{E}(X) &= \sum_{\omega\in\Omega}\sum_{i=1}^{\infty} X(\omega)\,\mathbb{P}(\{\omega\}|A_i)\mathbb{P}(A_i) \overset{\text{התכנסות בהחלט}}{=} \sum_{i=1}^{\infty}\mathbb{P}(A_i)\sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\{\omega\} \mid A_i)\\ &= \sum_{i=1}^{\infty}\mathbb{E}(X \mid A_i)\mathbb{P}(A_i)\end{aligned}$$

■

רעיון יסודי זה של ניצול ליניאריות התוחלת כדי לפרקה לתרומה של חלקים שונים של מרחב ההסתברות מהווה טכניקה מרכזית לחישוב תוחלות.

דוגמא 4.19 (תוחלת תחת התניה)**.**

בקופסה שני מטבעות. הראשון נופל על עץ בהסתברות p_1 השני נופל על עץ בהסתברות p_2. בוחרים באקראי מטבע ומטילים אותו n פעמים. מה תוחלת מספר העצים שהתקבלו?

תשובה: בהנתן שנבחר המטבע הראשון התוחלת היא np_1. בהנתן שנבחר המטבע השני התוחלת היא

np_2. התוחלת של מספר העצים היא לפיכך $\frac{1}{2}np_1 + \frac{1}{2}np_2 = n\frac{p_1+p_2}{2}$.

✎ **בעיה 4.2.** יהי X משתנה מקרי אחיד על $[N]$, ויהיו $Y_i \sim \mathrm{Ber}(1/2)$ משתנים מקריים בלתי-תלויים זה בזה וב-X. יש לחשוב את התוחלת

$$e_1 = \mathbb{E}\left(\sum_{i=1}^{X} Y_i \mid Y_1 = Y_2 = \cdots = Y_X = 1\right) \qquad e_2 = \mathbb{E}\left(\sum_{i=1}^{X} Y_i\right).$$

כיצד נסביר את העובדה שעבור ערכי N גדולים e_1 קטנה בהרבה מ-e_2?

✎ **בעיה 4.3.** יהי X משתנה מקרי בעל תוחלת על $[N]$, ויהי Y המקיים $(Y \mid X) \sim \mathrm{Bin}(X, p)$. יש להראות כי $\mathbb{E}(Y) = p\mathbb{E}(X)$.

4.3 חסם התוחלת (אי-שוויון מרקוב)

אבסורד שגור הוא: "זו כיתה כה מוכשרת עד שלכל התלמידים בה ציונים מעל הממוצע". ובאמת - אחת מתכונותיו של הממוצע, ואפילו של ממוצע משוקלל היא שהוא תמיד מהערך המירבי. תופעה זו ניתן להכליל עבור מספרים חיוביים ולקבוע שלא יתכן שיותר ממחצית הילדים בכיתה יקבלו ציון שהוא כפליים מהממוצע. באמצעות אבחנה זו נגיע לאי-שוויון הבא שהוצע בשנת 1853 בידי ז'ול בְּיֶנֶמֶה (Ir□n□e-Jules Bienaym□), הוכח בידי פאפנוטי צ'בישב (Pafnuty Chebyshev) ב-1867, ונקרא על שם תלמידו אנדרי מרקוב (Andrei Markov).

משפט 4.20 (אי-שוויון מרקוב)**.** יהי X משתנה מקרי אי-שלילי. אזי לכל $a > 0$ מתקיים

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

הוכחה. בהנתן X, נגדיר משתנה מקרי חדש $Y = a\mathbb{1}(X \geq a)$ ונשים לב שמתקיים $X \geq Y$. לפי מונוטוניות התוחלת, $\mathbb{E}(X) \geq \mathbb{E}(Y)$. על כן,

$$\mathbb{E}(X) \geq \mathbb{E}(Y) = a\mathbb{P}(X \geq a)$$

ואם נחלק ב-a נקבל את האי-שוויון המבוקש. ■

יש המוצאים את משפט 4.20 אינטואיטיבי יותר לאחר ההצבה $b = \frac{a}{\mathbb{E}(X)}$.

מסקנה 4.21 (א"ש צ'ב - ניסוח אחר)**.** יהי X מ"מ אי-שלילי בעל תוחלת סופית וסטיית תקן σ_X. לכל $b > 0$ מתקיים

$$\mathbb{P}(X \geq b\mathbb{E}(X)) \leq \frac{1}{b}.$$

כלומר כי ערך שגדול מ-b פעמים התוחלת, יכול להתקבל לכל היותר בהסתברות $\frac{1}{b}$.

במבט ראשון אי-שוויון מרקוב נראה כהערכה גרועה למדי של ההסתברות של המאורע $\{X \geq a\}$, ואולם תכונת הליניאריות של התוחלת מאפשרת לנו לחשב באופן מדוייק גם תוחלת של סכום משתנים מקריים תלויים. מן הבחינה הזו תוחלתם של משתנים מקריים נגישה יותר מאשר פונקציית ההתפלגות שלהם ולכן גם אומדן גס שכזה יכול להיות טוב ופשוט יותר לחישוב מאשר נסיונות ישירים לחסום את הסתברות המאורע.

דוגמא 4.22 (התוחלת ופרדוקס יום ההולדת)**.**
נשוב ונתאר את תסריט דוגמא 1.13 באופן מופשט. יהיו $\{X_n\}_{n\in[N]}$ מ"מ ב"ת אחידים על $[M]$. נבקש לדעת מה ההסתברות שלפחות $k>1$ מהמ"מ קיבלו את אותו הערך.

תשובה: נסמן משתנים מציינים $Y_{i_1,\dots,i_k} = \mathbb{1}(\{X_{i_1} = X_{i_2} = \cdots = X_{i_k}\})$ עבור $i_1,\dots,i_k \in [N]$ שונים זה מזה. נסמן את אוסף המשתנים הללו ב-$\mathcal{Y}$. נשים לב שאנו מעוניינים לחסום את המאורע

$$\mathbb{P}(\sum_{Y\in\mathcal{Y}} \geq 1),$$

לכל $i_1,\dots,i_k \in [N]$ המשתנה $Y_{i_1,\dots,i_k}$ הוא משתנה ברנולי עם סיכוי הצלחה M^{1-k} ולכן, כפי שחישבנו בדוגמא 5.29א, $\mathbb{E}(Y_{i_1,\dots,i_k}) = M^{1-k}$. נחשב לפי א"ש מרקוב

$$\mathbb{P}\left(\sum_{Y\in\mathcal{Y}} Y \geq 1\right) \overset{\text{מרקוב}}{\leq} \mathbb{E}\left(\sum_{Y\in\mathcal{Y}} Y\right) \overset{\text{טענה 4.5ב}}{=} \sum_{Y\in\mathcal{Y}} \mathbb{E}(Y) = \binom{N}{k} M^{1-k} \leq \frac{N^k}{M^{k-1}k!}.$$

נשים לב שהחסם שקיבלנו מתלכד עם החסם מדוגמא 1.19 שהושג באמצעות א"ש בול הדבר אינו מקרי כפי שמתגלה בבעיה 9.1 להלן.

דוגמא 4.23 (התוחלת ובעיית האספן)**.**
נבחן תיאור חדש של בעיית האספן (דוגמא 3.34), הפעם מנקודת מבט של אספן האומד את מספר ביצי ההפתעה שעליו לרכוש בטרם יקבל בובה חדשה.

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. נחשב את תוחלת מספר הבובות שעליו לרכוש, ונשתמש בא"ש מרקוב להעריך כמה בובות עליו לרכוש כדי שבהסתברות של לפחות 0.5 יהיה לו עותק יחיד מכל בובה.

תשובה: היות שבשלב ה-i, הסיכוי בכל רכישה שהביצה שנקנתה תכיל בובה חדשה הוא $\frac{n-i+1}{n}$ באופן שאינו תלוי בכמות הבובות שכבר רכשנו באותו שלב, הרי שלאור תכונת חוסר הזיכרון – ניתן לתאר את כמות הביצים שנרכשו בשלב ה-i כמ"מ גיאומטרי $X_i \sim \mathrm{Geo}(\frac{n-i+1}{n})$. אנו מתבקשים אפוא לחשב את $\mathbb{E}\left(\sum_{i=1}^n X_i\right)$. ניעזר בליניאריות התוחלת ובחשבון מדוגמא 5.29א ונקבל

$$\mathbb{E}\left(\sum_{i=1}^n X_i\right) \overset{\text{טענה 4.5ב}}{=} \sum_{i=1}^n \mathbb{E}(X_i) \overset{\text{דוגמא 5.29א}}{=} \sum_{i=1}^n \frac{n}{n-i+1} \overset{[j=n-i+1]}{=} n\sum_{j=1}^n \frac{1}{j} \leq n(\log n+1)$$

ולפי א"ש מרקוב נקבל

$$\mathbb{P}\left(\sum_{i=1}^n X_i \geq 2n(\log n+1)\right) \leq \frac{1}{2}$$

ולכן די ב-$2n(\log n+1)$ ביצים כדי שהסיכוי לא לקבל ביצה אחת מכל סוג יהיה לכל היותר חצי. נשים לב שתוצאה זו מתלכדת עם החסם שקיבלנו בדוגמא 3.34.

✎ **בעיה 4.4.** להוכיח את א"ש בול (משפט 1.17) באמצעות א"ש מרקוב על משתנים מציינים.

✎ **בעיה 4.5.** להוכיח את עקרון ההכלה וההדרה (טענה 1.23) באמצעות תוחלות. הדרכה: יש להגדיר את המשתנה המקרי $X = \prod_{i=1}^n (\mathbb{1}(\cup_{k=1}^n A_k) - \mathbb{1}(A_i))$ ולהוכיח שהוא שווה תמיד לאפס. כעת ניתן לפתוח את הסוגריים ולחשב את תוחלת X בעזרת לינאריות התוחלת.

כשם שלא יתכן שכל התלמידים קיבלו ציון גבוה מהממוצע, כן לא יתכן שכולם קיבלו ציון נמוך ממנו. למשל, לא יתכן שהתלמיד בעל הציון הגבוה ביותר בכיתה עם ממוצע ציונים של 90 קיבל ציון 75.

טענה 4.24. יהי X משתנה מקרי בעל תוחלת סופית.

$$\mathbb{P}\big(X \geq \mathbb{E}(X)\big) > 0$$

הוכחה למקרה הבדיד. נניח בשלילה כי $\mathbb{P}\big(X \geq \mathbb{E}(X)\big) = 0$ ונקבל שמתקיים $\mathbb{P}\big(X < \mathbb{E}(X)\big) = 1$ ולכן מתכונת החיו[illegible] מתקיים [illegible]$) = \mathbb{E}(\mathbb{E}(X)) < \mathbb{E}(X)$ - סתירה. מכאן שמתקיים $\mathbb{P}\big(X \geq \mathbb{E}(X)\big) > 0$. ■

נשים לב שבאופן כללי טענה 4.24 לא נותנת לנו כל הערכה להסתברות המאורע $X \geq \mathbb{E}(X)$. למשל, יתכן שכל התלמידים בכיתה קיבלו 80, למעט תלמיד אחד שקיבל 90. במקרה זה, תוחלת הציון של תלמיד מקרי תהיה גבוהה במעט מ-80 ולרוב הגדול של התלמידים יהיה ציון נמוך יותר.

דוגמא 4.25 (ארדש 1963)**.** תהיינה $\{A_k\}_{k\in[N]}$ קבוצות בנות n איברים עבור $N < 2^{n-1}$. אזי קיימת קבוצה B שאינה זרה לאף קבוצה A_k ואינה מכילה אף קבוצה A_k.

תשובה: נרשום $A = \cup_{k\in[N]} A_k$. נגדיר לכל $a \in A$ משתנה מקרי $X_a \sim \mathrm{Ber}(1/2)$, כך שכל המשתנים בלתי-תלויים זה בזה. נגדיר עבור $i \in \{0,1\}$ את המאורעות: $E_k^i = \{X_a = 0 \ : \ \forall a \in A_k\}$. נשים לב כי היות שכל הקבוצות בנות n איברים, הרי שמתקיים $\mathbb{P}(E_k^0) = \mathbb{P}(E_k^1) = 2^{-n}$. נסמן $E = \cup_{k\in[N], i\in\{0,1\}} E_k^i$ ונחשב

$$\mathbb{E}(\mathbb{1}_E) \leq \mathbb{E}\Big(\sum_{k\in[N],i} E_k^i\Big) = \sum_{k\in[N],i} \mathbb{E}(E_k^i) < 2^n \mathbb{E}(E_1^0) < 1.$$

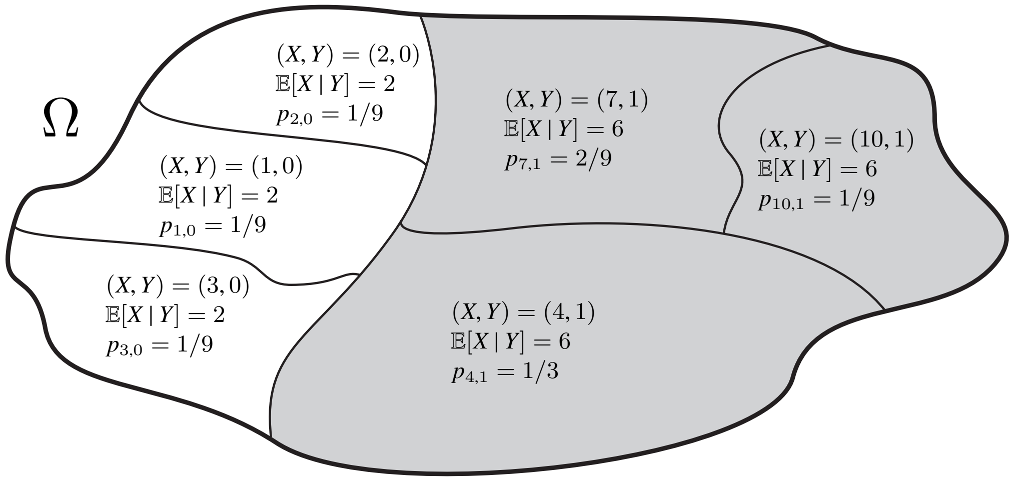
קיבלנו כי $\mathbb{E}(\mathbb{1}_{E^c}) = \mathbb{E}(1 - \mathbb{1}_E) > 0$ ולכן, לפי טענה 4.24, $\mathbb{P}(\mathbb{1}_{E^c} > 0) > 0$, כלומר קיימת בחירה של X_a כך ש-$\{a \ : \ X_a = 1\}$ מקיימת את הטענה.

בחרנו להציג את הדוגמא שלעיל במונחי ליניאריות התוחלת ולא במונחי אי-שווין בול, מתוך כך שאמצעי זה מאפשר להכליל את שיטת הבניה מעבר למקרה הפשוט של משתני אינדיקטור. צורה זו של בניית דוגמאות באמצעות בחירה מקרית פותחה בעיקר על ידי המתמטיקאי ההונגרי היהודי פול ארדש (Paul Erdős) והייתה לה תרומה מכרעת בקומבינטוריקה. לקריאה נוספת ר' פרק הפניה.

*4.4 תוחלת מותנית כמשתנה מקרי

נוכל לשאול "כיצד משתנה התוחלת לאחר שנחשף מידע על המשתנה המקרי". רעיון מתקדם זה מגולם בסוג חדש של משתנה מקרי המכונה **תוחלת מותנית**. אומנם ניתן להבין מושג זה גם מבלי לעסוק להשתמש במושג החלוקה המתואר בפרק 3.5, אך לקורא מומלץ להעמיק בהבנת מושג זה בטרם יקרא את המשך הפרק.

בפרק הקודם ראינו כי בהנתן משתנים מקריים X ו-Y, ניתן לחשב, עבור כל $x \in \mathrm{Supp}(X)$ את התוחלת של Y בהנתן המאורע $\{X = x\}$. נגדיר אפוא



תרשים 4.1 : ייצוג גרפי של $p_{X,Y}$, התפלגות משותפת של וקטור מקרי (X,Y), ושל התוחלת המותנית $\mathbb{E}[Y \mid X]$. גודל השטחים מציג את הסתברותם הנקודתית של איברים במרחב המדגם. נשים לב כי $\mathbb{E}[Y \mid X]$ תלויה רק בערכי של Y וכי $\mathbb{E}(\mathbb{E}[Y \mid X]) = \mathbb{E}(Y) = 6$.

הגדרה 4.26 (תוחלת מותנית)**.** יהיו X, Y משתנים מקריים המוגדרים על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$ כך ש-X בעל תוחלת. נרשום

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

ה**תוחלת המותנית (conditional expectation)** של Y בהינתן X, שנסמנה $\mathbb{E}[Y \mid X]$ היא המשתנה המקרי Z המוגדר על ידי

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

✎ **בעיה 4.6.** להוכיח כי התוחלת המותנית מוגדרת היטב לכל X, Y בדידים בעלי תוחלת.

אם נחשוב על התוחלת $\mathbb{E}(Y)$ בתור התחזית שלנו לערך של Y, הרי שהתוחלת המותנית, $\mathbb{E}[Y \mid X]$ היא התחזית לערך של Y בהנתן המידע הגלום ב-X. במונחי פרק 3.5, נוכל לומר כי מאחר שהתוחלת המותנית הוגדרה כפונקציה של X, הרי שלפי טענה 3.60 החלוקה של התוחלת המותנית גסה יותר מהחלוקה של X. מבחינה טכנית – התוחלת המותנית מתאימה לכל $\omega \in X^{-1}(a)$ את הערך הממוצע של Y על פני מחלקה זו. תיאור גרפי של תוחלת מותנית מוצג בתרשים 4.1.

כעת נוכל לכתוב את טענה 4.18 עבור החלוקה $\{Y^{-1}(x)\}_{x\in \mathrm{Supp}(X)}$ בצורה הנאה

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

דוגמא 4.27 (תוחלת מותנית).

יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים בעלי התפלגות ברנולי עם פרמטר p ונסמן את סכומם $X = \sum_{i=1}^n X_i$. מה התוחלת המותנית של X_1 בהנתן X?

תשובה: נחשב את ההתפלגות המותנית של X_1 בהנתן המאורע $X = k$ עבור $k \in [n]$. לשם כך, עלינו לחשב את ההסתברות של המאורע $X = k$ וגם $X_1 = 0$. למשתנה המקרי $X - X_1$ התפלגות בינומית עם פרמטרים $n-1$ ו-p והוא בלתי תלוי ב-X_1 ולכן

$$\begin{aligned}\mathbb{P}(X_1 = 1 \mid X = k) &= \frac{\mathbb{P}(X_1 = 1\ ,\ X = k)}{\mathbb{P}(X = k)} = \frac{\mathbb{P}(X_1 = 1\ ,\ X - X_1 = k-1)}{\mathbb{P}(X = k)} \\ &= \frac{\mathbb{P}(X_1 = 1)\mathbb{P}(X - X_1 = k)}{\mathbb{P}(X = k)} = \frac{p\binom{n-1}{k-1}(1-p)^{n-k}p^{k-1}}{\binom{n}{k}(1-p)^{n-k}p^k} = \frac{k}{n}\end{aligned}$$

באופן דומה נקבל $\mathbb{P}(X_1 = 0 \mid X = k) = \frac{n-k}{n}$ ולכן $\mathbb{E}(X_1 \mid X = k) = \frac{k}{n}$. נציב את X ונקבל $\mathbb{E}[X_1 \mid X] = \frac{X}{n}$. כעת נוכל לבדוק שאכן מתקיים $\mathbb{E}(\mathbb{E}[X_1 \mid X]) = \mathbb{E}\left(\frac{X}{n}\right) = \frac{np}{n} = p = \mathbb{E}(X_1)$.

✎ **בעיה 4.7.** מה התוחלת המותנית של X בהנתן X_1 בדוגמא 4.27?

טענה 4.28 (תכונות של התוחלת המותנית). יהיו X, Y, Z משתנים מקריים בדידים, תהי $f : \mathbb{R} \to \mathbb{R}$ ונסמן $W = f(Z)$. התכונות הבאות מתקיימות.

(א) **ליניארית:** $\mathbb{E}[X + Y \mid Z] = \mathbb{E}[X \mid Z] + \mathbb{E}[Y \mid Z]$.

(ב) **תכונת ההרכבה:** $\mathbb{E}[\mathbb{E}[X \mid Z] \mid W] = \mathbb{E}[X \mid W]$.

(ג) **מונוטוניות:** אם $X \overset{\text{a.s.}}{\geq} Y$ אז $\mathbb{E}[X \mid Z] \overset{\text{a.s.}}{\geq} \mathbb{E}[Y \mid Z]$.

(ד) **הוצאת החלק הידוע:** $\mathbb{E}[XW \mid Z] \overset{\text{a.s.}}{=} W\,\mathbb{E}[X \mid Z]$.

(ה) **השמטת החלק שאינו תלוי:** אם Z ב"ת ב-X אז $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$ כמעט תמיד.

✎ **בעיה 4.8.** להוכיח את טענה 4.28.

למעשה ניתן לראות כי ערכיו של X כלל לא שיחקו תפקיד בהגדרת $\mathbb{E}(Y \mid X)$ ולכן נוכל לחשוב על התוחלת המותנית במונחים של התניה בחלוקה המושרית על ידי X. צורה אחרת לחשוב על התוחלת המותנית היא בתור המשתנה המקרי המקרב את Y באופן הטוב ביותר מבין כל המשתנים המקריים שמשרים את החלוקה הנתונה על ידי X. נפתח רעיון זה בפרק 5.5.

כשם שהתנינו במשתנה מקרי יחיד - כך נוכל גם להתנות בוקטור מקרי.

הגדרה 4.29 (תוחלת מותנית)**.** יהיו $X_1, \ldots, X_N, Y$ משתנים מקריים המוגדרים על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$. ה**תוחלת המותנית** של Y בהינתן $X_1, \ldots, X_N$, שנסמנה $\mathbb{E}(Y \mid X_1, \ldots, X_N)$ היא המשתנה המקרי Z המוגדר על ידי

$$\mathbb{E}[Y \mid X_1, \ldots, X_N](\omega) = Z(\omega) := \begin{cases} \mathbb{E}[Y \mid X_1 = X_1(\omega), \ldots, X_N = X_N(\omega)] & \mathbb{P}(\omega) > 0 \\ 0 & \mathbb{P}(\omega) = 0 \end{cases}$$

נשים לב שבמונחי חלוקות, זו למעשה "התניה בחלוקה שנוצרת מ-$X_1, \ldots, X_n$".

בעיות הרחבה והעשרה

✎ **בעיה 4.9.** עשרה שופטים בתחרות שחיה צורנית מעניקים לקבוצה הקנדית ציונים מקריים המתפלגים אחיד ב-$[10]$. יש לחשב את תוחלת הציון המינימלי ואת תוחלת הציון המקסימלי שקיבלה הקבוצה.

✎ **בעיה 4.10.** n אנשים מנהלים טורניר של הטלת קוביות. כל זוג אנשים מטילים כל אחד קוביה עד אשר אחד מהם מטיל תוצאה גבוהה יותר. טורניר נקרא טרנזיטיבי אם בכל מקרה שבו שחקן א' ניצח את שחקן ב' ושחקן ב' ניצח את שחקן ג' ניצח שחקן א' את שחקן ג'. יש להשתמש באי-שוויון מרקוב כדי לקבל חסם עליון גס להסתברות שהטורניר שתואר יהיה טרנזיטיבי.

✎ **בעיה 4.11.** מעוניינים לסדר n אנשים במעגל. כל אדם רשאי להציג k דרישות לזוגות אנשים שאינו מעוניין להמצא בדיוק ביניהם. יש להראות שכל עוד $k < \binom{n}{2}$ ישנה דרך אפשרית לסדר את האנשים. (רמז: מושיבים את האנשים באקראי)

✎ **בעיה 4.12** (התפלגות היפר-גיאומטרית)**.** נתון שבהטלת n מטבעות התקבלו m תוצאות של עץ. מה תוחלת מספר העצים ב-k ההטלות הראשונות?

✎ **בעיה 4.13** (הכלה והדרה במונחי תוחלת)**.** יהיו $\{A_n\}_{[N]}$ מאורעות במרחב הסתברות, ויהי $A = \cup_{n\in[N]} A_i$. יש להסיק את טענה 1.23 (עקרון ההכלה וההדרה) מהזהות הבאה, הנתונה במונחי משתנים מקריים,

$$\prod_{n\in[N]} (\mathbb{1}_A - \mathbb{1}_{A_i}) \equiv 0.$$

✎ **בעיה 4.14 (*).** יהי X משתנה מקרי בעל תוחלת הנתמך על $\mathbb{N}$. יש להראות כי $X \sim \mathrm{Geo}(p)$ עבור $p \in (0,1)$ כלשהו אם ורק אם לכל $s \in \mathbb{N}_0$ מתקיים $\mathbb{E}(X \mid X > s) = \mathbb{E}(X) + s$.

✎ **בעיה 4.15 (*).** בכד שלושה כדורים לבנים ושני כדורים שחורים. שחקן שולף כדורים מהכד בזה אחר זה. שליפת כדור לבן גורמת לו להפסיד שקל ושליפת כדור שחור גורמת לו להרוויח שקל. יש להראות כי:

(א) אם השחקן מחליט מראש כמה כדורים לשלוף, אז שום אסטרטגיה לא תעניק לו תוחלת רווח חיובית.

(ב) אם הוא משנה את כמות הכדורים שישלוף לפי תוצאת השליפה, אז ביכולתו להשיג תוחלת רווח חיובית.

✎ **בעיה 4.16 (*).** לתוך שתי מעטפות אטומות מוכנסים סכומי כסף 10^X ו-10^{X+1} 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?

פרק 5

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. גם אם כבר 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. כדי 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**.

5.1 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 5.1 (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 הגדרות השונות נובע מליניאריות התוחלת, טענה 4.5ב, לפי החישוב

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

בעוד שחיוביות השונות מתקבלת מידית מההגדרה הראשונה לפי תכונת החיוביות של התוחלת (טענה 4.5א), הרי שניתן לקבלה גם מההגדרה השניה לפי אי-שוויון ינסן (טענה 4.12) על הפונקציה הקמורה $f(x) = x^2$.

טענה 5.2. 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 5.3 (תכונות השונות)**.** יהי X משתנה מקרי בעל שונות סופית, ויהי $a \in \mathbb{R}$. אזי

(א) $\mathrm{Var}(X) \geq 0$ ושוויון מתקיים רק אם X קבוע.

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

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

(ד) 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.** 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 $\mathbb{P}(\{\omega\}) = 0$.

סעיף ב. לפי ליניאריות התוחלת (טענה 4.5ב) 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 ג. 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 ד. נשתמש בכפליות התוחלת למשתנים בלתי-תלויים (טענה 4.15) ובליניאריות התוחלת ונחשב

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

כאשר השוויון האחרון מתקבל מהחישוב

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

■

דוגמא 5.4 (שונות משתנים מקריים מוכרים)**.**

לאורך הדוגמא נסתמך על חישוב תוחלת מדוגמא 5.29 ועל הגדרת התוחלת.

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

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

(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 $\{Y_n\}_{n\in\mathbb{N}}$ כאשר $Y_n\sim \mathrm{Ber}(p)$ ב"ת. לפי טענה 5.3 וסעיף ב נקבל

$$\mathrm{Var}(X) \overset{\text{טענה 5.3ד}}{=} \sum_{n=1}^{N}\mathrm{Var}(Y_n) \overset{\text{סעיף ב}}{=} \mathbf{Np(1-p)}$$

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

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

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,

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

5.2 שונות ככלי להוכחת ריכוז

בטענה 5.3א ראינו ששונות חיובית מהווה עדות לפיזור כלשהו של המשתנה המקרי (כלומר שהוא אינו קבוע). בפרק זה נכמת את פיזורו של משתנה מקרי במונחים של שונות וסטיית תקן, וזאת באמצעות אי-שוויון מרקוב (משפט 4.20). אומנם, לפי הגדרתה השונות היא **תוחלת** של ריבוע הפרש בין משתנה מקרי לתוחלתו. כתוצאה מכך ניתן להפעיל עליה את אי-שוויון מרקוב ולקבל את התוצאה הבאה באופן מיידי.

משפט 5.5 (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 4.20), 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 5.6 (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 5.7 (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 5.29א ודוגמא 5.4ג מתקיים $\mathbb{E}(X) = 500000$ ו-$\mathrm{Var}(X) = 250000$. נפעיל את א"ש צ'בישב, משפט 5.5 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 כדי לקבוע שתגלית סטטיסטית ראויה לפרסום. רווח בר-סמך משמש גם בהפרדת השערות כדי 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 5.8 (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 4.5ב), מתקיים

$$\mathbb{E}(S_N) \overset{\text{טענה 4.5ב}}{=} \frac{1}{N}\mathbb{E}\left(\sum_{n=1}^{N} X_n\right) \overset{\text{טענה 4.5ב}}{=} \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}(X_n) \overset{\text{שווי התפלגות}}{=} \mathbb{E}(X).$$

כיוון ש-$\{X_n\}_{n\in[N]}$ ב"ת, לפי טענה 5.3 מתקיים

$$\mathrm{Var}(S_N) \overset{\text{טענה 5.3ג}}{=} \frac{1}{N^2}\,\mathrm{Var}\left(\sum_{n=1}^{N} X_n\right) \overset{\text{טענה 5.3ד}}{=} \frac{1}{N^2}\sum_{n=1}^{N}\mathrm{Var}(X_n) \overset{\text{שווי התפלגות}}{=} \frac{\mathrm{Var}(X)}{N}.$$

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.

דוגמא 5.9 (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 5.8, 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 שאביה ניצח בסיבוב ה-i במשחק שמתחיל בתוצאה חמש-שלוש. לפי נוסחת ההסתברות השלמה, טענה 2.7, אם נחלק את המאורע שאביה ינצח, למקרה שבו היא מנצח בסיבוב הראשון, בסיבוב השני לאחר שהובס בראשון או בסיבוב השלישי לאחר שהובס 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 רק $\frac{1}{8}$ ממנו.

דוגמא 5.10 (השונות ובעיית האספן).

נשוב לתיאור בעיית האספן (דוגמא 3.34), לפי דוגמא 4.23. 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 4.23 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 5.5) על מנת למצא טווח שספר הרכישות הנחוץ להשגת עותק של כל בובה ימצא בתוכו בהסתברות של 75%.

תשובה: כמו בדוגמא 4.23 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

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{טענה 5.3}}{=} \sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) \overset{\text{טענה 5.4}}{=} \sum_{i=1}^{n}\frac{\frac{i-1}{n}}{\frac{(n-i+1)^2}{n^2}} = \sum_{i=1}^{n}\frac{n(i+1)}{(n-i+1)^2}$$

נציב $j = n-i+1$ ונקבל

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

כאשר השוויון האחרון הוא פתרון אוילר לבעיית באסל שהוזכר בדוגמא 1.10.

מכאן ש-$\sigma_X \le n\pi/\sqrt{6}$ נפעיל את ניסוחו השני של משפט צ'בישב (מסקנה 5.6), ונקבל

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

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 $n(\log n - 2\pi/\sqrt{6})$ לבין $n(\log n + 2\pi/\sqrt{6} + 1)$. למשל במקרה של 50 סוגי ביצים ידרשו בין 123 ל-317 ביצים.

5.3 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. בהמשך נפתח נוסחא לחישוב שונות של סכום משתנים מקריים תלויים, במונחים של שונויות משותפות.

הגדרה 5.11 (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)**.

אם השונות מודדת את הסטיה הריבועית של משתנה מתוחלתו, הרי השונות המשותפת מודדת את הסטיה של מכפלת X ו-Y מתוחלתם. אלא שמדד זה, שלא כמו השונות, אינו בהכרח חיובי ולכן ניתן לראות אותו כמודד את מידת המתאם בין X ל-Y. נסיבות בהן X ו-Y שניהם יחדיו גדולים מתוחלתם ונסיבות בהן שניהם יחדיו קטנים מתוחלתם מגדילות את השונות המשותפת, ואילו נסיבות בהן האחד גדול מתוחלתו והשני קטן ממנה – מקטינות אותו. כאשר שני אלו מקזזים אלה את אלה, המשתנים בלתי מתואמים. מנקודת מבט זו, טבעי לצפות שכאשר ערכו של Y לא נותן לנו כל מידע לגבי התפלגותו של X – תהיה השונות המשותפת 0, כפי שעולה מן המסקנה הבאה.

מסקנה 5.12 (אי-תלות גוררת חוסר מתואמות)**.** אם X ו-Y בלתי-תלויים אז הם בלתי-מתואמים.

מסקנה זו היא למעשה ניסוח מחודש של טענה 4.15.

נשים לב שהכיוון ההפוך למסקנה 5.12 אינו נכון, כפי שניתן לראות בדוגמא הבאה.

דוגמא 5.13.

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 3.36 המשתנים $(X-Y)$ ו-Z ב"ת. לכן, לפי טענה 4.15 מתקיים

$$\begin{aligned}\mathrm{Cov}(Z,(X-Y)Z) &= \mathbb{E}\left((X-Y)Z^2\right)-\mathbb{E}(Z)\mathbb{E}((X-Y)Z)\\ &\overset{\text{4.15 טענה}}{=} \mathbb{E}\,(X-Y)\,\mathbb{E}\left(Z^2\right)-\mathbb{E}(X-Y)\mathbb{E}(Z)^2\\ &= (\mathbb{E}(X)-\mathbb{E}(Y))\mathbb{E}\left(Z^2\right)-(\mathbb{E}(X)-\mathbb{E}(Y))\mathbb{E}(Z)=0\end{aligned}$$

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. כדי 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 5.14 (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) **סימטריות:** $\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X)$.

(ב) **בי-ליניאריות:** $\mathrm{Cov}(aX+bZ,Y)=a\,\mathrm{Cov}(X,Y)+b\,\mathrm{Cov}(Z,Y)$.

(ג) $\mathrm{Cov}(X,X)=\mathrm{Var}(X).$

✎ **בעיה 5.1.** להוכיח את טענה 5.14.

באמצעות השונות המשותפת נוכל לחשב נוסחא לשונות של סכום שני משתנים מקריים כלליים. נוסחא זו היא למעשה הכללה של טענה 5.3ד.

טענה 5.15 (שונות של סכום מ"מ). יהיו X,Y משתנים מקריים. אזי

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

בכל מקרה בו אגף ימין של המשוואה מוגדר היטב.

הוכחה. נשתמש בהגדרת השונות (הגדרה 6.1) ובליניאריות התוחלת (טענה 4.5ב).

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

■

ניתן להכליל נוסחא זו למספר כלשהו של משתנים מקריים.

טענה 5.16 (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 5.2.** להוכיח את טענה 5.16 תוך שימוש בכך ש- $\operatorname{Var}\left(\sum_{n=1}^N X_n\right) = \operatorname{Cov}\left(\sum_{n=1}^N X_n, \sum_{n=1}^N X_n\right)$ ובביליניאריות השונות המשותפת.

למעשה במסקנה 5.20 להלן נראה שדי בקיום שונות של כל אחד מהמשתנים כדי להבטיח קיומה של שונות משותפת.

דוגמא 5.17 (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 4.5ב):

$$\mathbb{E}(Y) = \sum_{n\in[N-1]} \mathbb{E}(Y_n) = \sum_{n\in[N-1]} \mathbb{E}(X_nX_{n+1}) \overset{\text{טענה } 4.15}{=} \sum_{n\in[N-1]} \mathbb{E}(X_n)\mathbb{E}(X_{n+1}) = (N-1)p^2$$

כמו כן נחשב לפי בעיה 5.2

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

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

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

כאשר $k > n+1$, המשתנים המקריים Y_n ו-Y_k בלתי תלויים לפי טענה 3.36 ולכן $\operatorname{Cov}(Y_n, Y_k) = 0$ במקרה זה. כאשר $k = n+1$, נקבל

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

סך הכל קיבלנו

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

נשים שלמרות ש-Y_n אינם בלתי תלויים, השונות עדיין גדלה לינארית 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.

5.4* קשרים לאלגברה ליניארית

התכונות המתוארות בטענות 5.2, 5.3, 5.14 ו-5.15 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 5.18 (קושי-שוורץ הסתברותי). יהיו X, Y משתנים מקריים בעלי שונויות סופיות. אזי $\mathbb{E}(XY)$ קיימת ומקיימת

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

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

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

ולכן

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

ממונוטוניות ולינאריות התוחלת נובע כי

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

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

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

כמבוקש. ∎

חישוב זה, שהנו הסתברותי גרידא, הוא למעשה הכללה של אי-שוויון קושי-שוורץ מאלגברה ליניארית. נראה כיצד מסיקים שוויון זה מטענה 5.18.

מסקנה 5.19 (קושי-שוורץ)**.** יהיו v, u וקטורים ב-$\mathbb{R}^n$. אזי

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

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

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

וכעת לפי טענה 5.18 מתקיים

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

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

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

מסקנה 5.20 (קיום שונות משותפת)**.** לכל X ו-Y, משתנים מקריים בעלי שונות סופית, $\mathrm{Cov}(X,Y)$ סופי.

הוכחה. לפי טענה 5.18, $\mathbb{E}(XY)$ 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 5.4 להלן נמחיש את משמעות טענה 5.18 ביחס למדד זה.

הגדרה 5.21 (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 5.22 (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 5.3.** להוכיח את טענה 5.22.

✎ **בעיה 5.4.** 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 אשר ניתן לראותו כאנלוג לקוסינוס הזווית בין ווקטורים המשמש כאומדן לתלות הליניארית ביניהם.

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

טענה 5.23 (רגרסיה ליניארית)**.** יהיו X, Y משתנים מקריים בעלי שונויות סופיות. אזי המשתנה המקרי

$$Z = X - \frac{\mathrm{Cov}(X,Y)Y}{\mathrm{Var}(Y)} = X - \sqrt{\frac{\mathrm{Var}(X)}{\mathrm{Var}(Y)}}\,\mathrm{Corr}(X,Y)Y,$$

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 5.14):

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

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 נוסחאת השונות השלמה המוצג בתת-פרק הבא.

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

✎ **בעיה 5.5.** 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 $x=0$)

5.5* שונות מותנית

ראשית נראה כי התוחלת המותנית $\mathbb{E}[X\mid Y]$ בלתי-מתואמת עם המשתנה המקרי $X-\mathbb{E}[X\mid Y]$. תיאור זה עולה בקנה אחד עם הרעיון שהתוחלת המותנית מהווה את הקירוב הטוב ביותר של X שניתן לקבלו כפונקציה של Y (שהרי אילולא הייתה בלתי מתואמת יכולנו לשפר אותה על ידי רגרסיה ליניארית).

טענה 5.24. יהיו X,Y משתנים מקריים על $(\Omega,\mathcal{F},\mathbb{P})$ כך ש-$|\mathbb{E}(X)|<\infty$. אז

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

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

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

■

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

כשם שהגדרנו בפרק 4.4 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 ה**שונות המותנית** כמשתנה מקרי המקבל לכל $x \in \mathbb{R}$ את שונותו של Y בהינתן $X = x$.

הגדרה 5.25 (שונות מותנית)**.** יהיו X, Y משתנים מקריים בדידים על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ כך ש-Y בעל שונות סופית. נרשום

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

ה**שונות המותנית (conditional variance)** של Y בהינתן X, שנסמנה $\mathrm{Var}(Y \mid X)$ 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 5.26 (שונות מותנית)**.** אם למשתנה מקרי X קיימת שונות מותנית, אז מתקיים

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

טענה 5.27. יהי X משתנה מקרי בעל שונות סופית על $(\Omega, \mathcal{F}, \mathbb{P})$ ויהי 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

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

ולכן נוכל לוודא כי

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

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

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

לשונות של X בהינתן Y.

טענה 5.28 (נוסחת השונות השלמה)**.** יהי X מ"מ בעל שונות סופית על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$ ויהי Y משתנה מקרי על אותו מרחב הסתברות. אז $\mathrm{Var}(X) = \mathbb{E}(\mathrm{Var}[X \mid Y]) + \mathrm{Var}(\mathbb{E}[X \mid Y])$. הגורם $\mathrm{Var}(\mathbb{E}[X \mid Y])$ מכונה **השונות המוסברת** (על ידי Y) ואילו הגורם $\mathbb{E}(\mathrm{Var}[X \mid Y])$ מכונה **השונות הבלתי-מוסברת** (על ידי Y)

הוכחה. נחשב:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{Var}(X - \mathbb{E}[X \mid Y] + \mathbb{E}[X \mid Y]) = \\ &= \mathrm{Var}(X - \mathbb{E}[X \mid Y]) + \mathrm{Var}(\mathbb{E}[X \mid Y]) + 2\,\mathrm{Cov}(X - \mathbb{E}[X \mid Y]), \mathbb{E}[X \mid Y]) \\ &\overset{\substack{\text{טענה}\\5.24}}{=} \mathrm{Var}(X - \mathbb{E}[X \mid Y]) + \mathrm{Var}(\mathbb{E}[X \mid Y]) \overset{\substack{\text{טענה}\\5.27}}{=} \mathbb{E}(\mathrm{Var}[X \mid \mathcal{G}]) + \mathrm{Var}(\mathbb{E}[X \mid Y]).\end{aligned}$$

■

המינוחים שונות מוסברת ובלתי-מוסברת מוצדקים על ידי טענה 5.27, אשר מראה שקיימת פונקציה של Y (הלא היא $\mathbb{E}[X \mid Y]$) שהפחתתה מ-X יוצרת משתנה מקרי ששונותו היא השונות הבלתי-מוסברת על ידי Y. כעת נוכל לכנות את $\frac{\mathrm{Var}(\mathbb{E}[X \mid Y])}{\mathrm{Var}(X)}$ בשם **הרכיב היחסי של שונות X המוסבר על ידי Y**. בדומה, את המשלים לאחד של רכיב זה, השווה ל-$\frac{\mathbb{E}(\mathrm{Var}[X \mid Y])}{\mathrm{Var}(X)}$, נכונה בשם **הרכיב היחסי של שונות X שאינו מוסבר על ידי Y**.

הערה: לטענה 5.28 חשיבות מכרעת בסטטיסטיקה, שכן היא מספרת לנו שאת הסטיה הריבועית של נעלם כלשהו נוכל לפרק לסטיה שניתן להסביר על ידי פיזורו של Y ולסטייה שיורית הנובעת מגורמים אחרים. כאשר רכיב השונות המוסבר קרוב ל-1 – נאמר כי Y הוא גורם מרכזי המשפיע על ערכו של X, כאשר רכיב זה קרוב ל-0, נסיק כי ידיעת ערכו של Y אינה מפחיתה באופן משמעותי את פיזורו של X. ניתוח כזה, מכונה ניתוח שונויות (ANOVA בלע"ז) והוא משמש, למשל, כדי לקבוע האם כדאי לפלח את מרחב המדגם לפי ערכי Y כאשר מבצעים ניתוח סטטיסטי. כך למשל נחליט האם בבואנו לנבא את תוצאותיהן של בחירות, כדאי לפלח את המדגם לפי מקום מגורים.

אם הרכיב היחסי של השונות X המוסברת על ידי Y הוא 1, הרי ש-X קבוע בהינתן Y ולכן בהכרח קיימת פונקציה f כך ש- $X \overset{\text{a.s.}}{=} f(Y)$ (זאת בהשוואה לכך שמקדם מתאם בעל ערך מוחלט של 1 מעיד על קיומה של f אפינית בעלת אותה תכונה). מצד שני - עבור משתנים מקריים בלתי-תלויים – רכיב השונות המוסבר הוא 0. ואולם, כמו במקרה של מקדם המתאם, התאפסותו של רכיב השונות המוסבר אינה בהכרח מעידה על אי-תלות, כפי שניתן לראות בדוגמא הבאה.

דוגמא 5.29 (משתנים תלויים בעלי שונות מוסברת אפס)**.**

יהי X משתנה מקרי אחיד על $\{-2, -1, 0, 1, 2\}$ ונסמן $Y = |X|$. נשים לב כי לכל $y \in \{0, 1, 2\}$ מתקיים $\mathbb{E}(X \mid Y = y) = 0$ ולכן $\mathbb{E}[X \mid Y] \overset{\text{a.s.}}{=} 0$ ומכאן שמתקיים $\mathrm{Var}(\mathbb{E}[X \mid Y]) \overset{\text{a.s.}}{=} 0$ זאת בניגוד לכך ש-Y ו-X הנם בברור תלויים.

הטענה הבאה מראה כי $\mathbb{E}[X \mid Y]$ היא אומנם הפונקציה שהפחתתה מביאה למינימום את השונות, ולכן, במובן זה, $\mathbb{E}[X \mid Y]$ היא האומדן הטוב ביותר של X שניתן לחשב מתוך Y.

טענה 5.30. יהי X משתנה מקרי בעל שונות סופית על $(\Omega, \mathcal{F}, \mathbb{P})$ ויהי Y משתנה מקרי על אותו מרחב הסתברות. אזי לכל $f : \mathbb{R} \to \mathbb{R}$ מתקיים

$$\mathbb{E}\left((X - f(Y))^2\right) \geq \mathrm{Var}(\mathbb{E}[X \mid Y]).$$

הוכחה. נסמן $f(Y) = \mathbb{E}[X \mid Y] + g(Y)$ ונחשב

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

תוך שימוש בכך ש-

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

מכך ש-$g(Y)^2$ היא אי-שלילית, מתקבלת הטענה. ■

ריבוע מקדם המתאם הוא מדד חלש יותר לקשר בין שני משתנים מקריים מאשר רכיב השונות המוסבר וזאת מפני שמקדם המתאם מתאר את חלק השונות המוסבר על ידי הקירוב הטוב ביותר של X על ידי פונקציה ליניארית של Y, ואילו השונות המותנית המנורמלת מתארת את חלק השונות המוסבר על ידי הקירוב הטוב ביותר של X על ידי פונקציה כלשהי של Y.

טענה 5.31. יהיו X ו-Y משתנים מקריים בעל שונות סופית על $(\Omega, \mathcal{F}, \mathbb{P})$. אזי

$$\frac{\mathrm{Var}(\mathbb{E}[X \mid Y])}{\mathrm{Var}(X)} \geq \mathrm{Corr}(X, Y)^2$$

הוכחה. נסמן $Z = \frac{\mathrm{Cov}(X,Y)Y}{\mathrm{Var}(Y)}$ ונחשב

$$\mathrm{Var}(X) = \mathrm{Var}(X - Z + Z) \overset{\text{טענה } 5.15}{=} \mathrm{Var}(X - Z) + 2\,\mathrm{Cov}(X - Z, Z) + \mathrm{Var}(Z) \overset{\text{טענה } 5.23}{=} \mathrm{Var}(X - Z) + \mathrm{Var}(Z)$$

$$\mathrm{Var}(X) \overset{\text{טענה } 5.28}{=} \mathbb{E}(\mathrm{Var}[X \mid Y]) + \mathrm{Var}(\mathbb{E}[X \mid Y])$$

נשים לב כי לפי טענה 5.31 מתקיים

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

ולכן

$$\frac{\mathrm{Var}(\mathbb{E}[X \mid Y])}{\mathrm{Var}(X)} \geq \frac{\mathrm{Var}(Z)}{\mathrm{Var}(X)} = \mathrm{Var}\left(\frac{\mathrm{Cov}(X, Y)Y}{\mathrm{Var}(Y)}\right) = \frac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(Y)^2\,\mathrm{Var}(X)^2} = \mathrm{Corr}(X, Y)^2,$$

כנדרש. ■

✎ **בעיה 5.6.** למצא דוגמא למשתנים מקריים X ו-Y בלתי מתואמים כך שרכיב השונות היחסי של X המוסבר על ידי Y יהי 1.

5.6* החוק החלש של המספרים הגדולים

לאמיתו של דבר הדרישה לקיומה של שונות במשפט 5.8 לא הייתה נחוצה כלל ועיקר. נציג את ההכללה הבאה.

משפט 5.32 (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 5.33. 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 5.32. עבור K נגדיר $a_K = \mathbb{E}\left(|X|\mathbb{1}_{|X|\geq K}\right)$. לפי טענה 5.33, 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 5.8, 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 5.8 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 5.33 נקבל כי

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

כנדרש.

■

בעיות הרחבה והעשרה

✎ **בעיה 5.7** (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 5.8.** 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 $\lim_{n\to\infty} np_n = \infty$ אז לכל $\epsilon > 0$ קיים m כך שההסתברות שיש בגן שלושה ילדים שכל שניים מביניהם חברים גדולה מ-$1 - \epsilon$.

✎ **בעיה 5.9** (התפלגות היפר-גיאומטרית)**.** נתון שבהטלת n מטבעות התקבלו m תוצאות של עץ. יש לחשב את שונות מספר העצים בk ההטלות הראשונות.

✎ **בעיה 5.10.** 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 5.11.** 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 ולהסיק כי פעולה זו מגדירה מכפלה פנימית במרחב זה. (מרחב זה שימושי הרבה פחות מאשר המרחב שהוצג בבעיה 5.5)

✎ **בעיה 5.12** (פרדוקס ברטרנד)**.** וינוגרד 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 5.13 (*).** [חוק חלש למשתנים חסרי-תוחלת] עבור פונקציה $f(n) : \mathbb{N} \to [0,1]$ המקיימת

$$\sum_{n=1}^{\infty} f(n) = a \le \infty$$

נגדיר משתנה מקרי X הנתמך על $\mathbb{Z} \setminus \{0\}$ בעל פונקציית התפלגות נקודתית

$$p_z = \frac{f(|z|)}{2a}.$$

לכל $N \in \mathbb{N}$ ו-$\epsilon > 0$ נגדיר

$$p_{N,\epsilon} = \mathbb{P}\left(\left|\frac{\sum_{n=1} NX_n}{N}\right| \le \epsilon\right),$$

עבור $X_1, \ldots, X_N$ בלתי-תלויים ושווי התפלגות ל-X. נאמר כי X **מקיים את החוק החלש** אם לכל $\epsilon > 0$ מתקיים $\lim_{N\to\infty} p_{N,\epsilon} = 1$. יש להראות כי:

(א) עבור $f(n) = 1/(n^2 \log n)$ המשתנה X מקיים את החוק החלש.

(ב) עבור $f(n) = 1/n^2$ המשתנה X לא מקיים את החוק החלש.

פרק 6

מומנטים גבוהים וריכוז מעריכי

...הרמן [רובין] טען שביכולתו להשיג חסם תחתון ביתר קלות. קראתי עליו תגר והוא הוכיח הוכחה בסגנון צ'בישב שהייתה כה פשוטה עד שלא טרחתי לציין את תרומתו.
איזו שגיאה! נראה ששאנון הפעיל את משפט הגבול המרכזי באופן שגוי על הזנב המרוחק של ההתפלגות באחד ממאמריו בתורת האינפורמציה. כששגיאתו נחשפה הוא גילה את החסם התחתון של רובין במאמרי והדבר הציל את תוצאותיו...

– הרמן צ'רנוף על חסם צ'רנוף, זכרונות מחברותי עם הרמן רובין, 2004

בפרק הקודם ראינו כיצד באמצעות מעבר מתוחלת לשונות שיפרנו את חסמי הפיזור שעמדו לרשותנו. כתוצאה מכך עלה בידנו להוכיח כי בתנאים מסויימים מתכנס ממוצע ניסויים חוזרים לתוחלת. עם זאת השוואה של חסמי הפיזור שקיבלנו לחסמים שהתקבלו בחישוב ישיר, למשל בדוגמא 3.34, מעידה שבעוד שהיטבנו להעריך את ערכו הטיפוסי של המשתנה, ההסתברות לסטיה משמעותית מערך הייתה לעיתים קרובות נמוכה בהרבה מהחסמים שהציעו לנו אי-שיוויון מרקוב ואי-שיווין צ'בישב. גישה אחת להכללה ושיפור של אי-שוויונים אלה היא הכללת השונות למושג כללי יותר של **מומנט**.

הגדרה 6.1 (מומנטים פולינומיאלים)**.** יהי X משתנה מקרי. ה**מומנט מסדר** k של X מוגדר בתור

$$m_k(X) = \mathbb{E}\left(X^k\right),$$

כאשר תוחלת זו מוגדרת היטב.

באופן דומה **המומנט המרכזי מסדר** k מוגדר על מ"מ בעל תוחלת סופית בתור

$$\mathbb{E}\left(\left|X - \mathbb{E}(X)\right|^k\right).$$

מנקודת ראות זו השונות היא מומנט מרכזי מסדר 2. ההגיון בבחירת המומנט המרכזי נובע מהטענה הבאה.

טענה 6.2 (השונות היא המומנט השני המוסט המינימלי)**.** יהי X משתנה מקרי בעל מומנט שני סופי, אזי

$$\min\left\{\mathbb{E}\left((X-a)^2\right) \;:\; a \in \mathbb{R}\right\} = \mathrm{Var}(X).$$

הוכחה. נרשום $Y = X - \mathbb{E}(X)$ כך שיתקיים $\mathbb{E}(Y) = 0$. לכל $\Delta \in \mathbb{R}$ נחשב,

$$\mathbb{E}((X - E(X) + \Delta)^2) = \mathbb{E}\left((Y + \Delta)^2\right) \overset{\text{טענה 4.5}}{=} \mathbb{E}(Y^2) + \mathbb{E}(\Delta^2) + 2\mathbb{E}(Y\Delta)$$
$$\overset{\text{טענה 4.15}}{=} \mathrm{Var}(X) + \Delta^2 + 2\mathbb{E}(Y)\mathbb{E}(\Delta) = \mathrm{Var}(X) + \Delta^2$$

ולכן המינימום מתקבל עבור $\Delta = 0$. ■

קיום מומנטים גבוהים היא דרישה מחמירה ביחס להתפלגות. כשם שלא כל התפלגות בעלת תוחלת ניחנה בשונות סופית, כך גם לא כל פונקציה בעלת מומנט מסדר k ניחנה במומנט מסדר $k + 1$.

טענה 6.3. יהי X משתנה מקרי. אם $m_k(X)$ סופי אז גם $m_{k-1}(X)$ סופי.

הוכחה. לפי הגדרה, לכל $k \in \mathbb{N}$ קיומו של $m_k(X) < \infty$ שקול לקיומו של $E(|X|^k)$ נפעיל את אי-שוויון ינסן (טענה 4.12) על הפונקציה הקמורה $f(x) = x^{\frac{k}{k-1}}$ והמשתנה המקרי $|X|^{k-1}$ ונקבל

$$\mathbb{E}(|X^k|) \geq \mathbb{E}(|X^{k-1}|)^{\frac{k}{k-1}}$$

ולכן, אם $m_k(X)$ סופי אז גם $m_{k-1}(X)$ סופי, כנדרש. ■

בעוד שמומנטים גבוהים מרכזיים אינם בהכרח מינימלים בין כל המומנטים המוסטים, עדיין ניתן להשתמש במומנטים אלא כדי לשפר את ההערכות באי-שוויון צ'בישב כפי שניתן לראות בבעיה הבאה.

✎ **בעיה 6.1.** התאם את הוכחת א"ש צ'בישב (משפט 5.5) והוכח כי לכל X מ"מ אי-שלילי בעל תוחלת סופית ומומנט מסדר k. אזי לכל $a > 0$ מתקיים

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\mathbb{E}\left(\left(|X - \mathbb{E}(X)|\right)^k\right)}{a^k}.$$

עם זאת, הכללתו החשובה ביותר של א"ש צ'בישב היא דווקא זו המסתמכת על הפונקציה המעריכית. הכללה זו תעמוד במוקד הפרק הבא.

6.1 פונקציה יוצרת מומנטים ואי שוויון צ'רנוף

נפתח בהיכרות עם המושג **פונקציה יוצרת מומנטים** המכונה לעיתים **המומנט המעריכי של** X.

הגדרה 6.4 (פונקציה יוצרת מומנטים)**.** יהי X משתנה מקרי. הפונקציה הממשית $M_X(t)$ הנתונה על-ידי

$$M_X(t) := \mathbb{E}\left(e^{tX}\right),$$

לכל t עבורו תוחלת זו מוגדרת היטב, מכונה ה**פונקציה יוצרת מומנטים (moment generating function)** של X. משתנה מקרי בעל פונקציה יוצרת מומנטים בסביבה כלשהי של הראשית נקרא בעל **מומנט מעריכי**.

טענה 6.5 (כפליות פונקציה יוצרת מומנטים)**.** יהיו X ו-Y משתנים מקריים בלתי תלויים ויהי $Z = X+Y$. אזי

$$M_Z(t) = M_X(t)M_Y(t).$$

הוכחה. לפי טענה 3.36, המשתנים המקריים e^{tX} ו-e^{tY} בלתי תלויים. נשתמש בכפליות משתנים מקריים בלתי-תלויים (טענה 4.15) ונקבל

$$\mathbb{E}(e^{t(x+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}).$$

■

שימושה החשוב ביותר של פונקציה יוצרת מומנטים, מקורו בשימוש של אי-שיווין מרקוב שהציע הרמן רובין (Herman Rubin) בסביבות 1950 אך נקרא בטעות על שם הרמן צ'ר

משפט 6.6 (אי-שוויון צ'רנוף)**.** יהי X משתנה מקרי בעל מומנט מעריכי. אזי לכל $t > 0$ עבורו $M_X(t)$ מוגדרת ולכל $a \in \mathbb{R}$ מתקיים

$$\mathbb{P}(X \geq a) \leq M_X(t)e^{-ta}$$

הוכחה. נציב בא"ש מרקוב (משפט 4.20) את המ"מ e^{tX} ונקבל

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}} = M_X(t)e^{-ta}.$$

■

נהיר שעבור משתנים מקריים בעלי מומנט מעריכי, אי שיוויון צ'רנוף מציע חסם רב עוצמה על ההסתברות לסטיה משמעותית מהתוחלת. בטרם נראה כי הערכה זו הנה הדוקה עבור משפחה נרחבת של משתנים מקריים, נפתח בחישוב הפונקציה היוצרת מומנטים של מספר משתנים מקריים מוכרים.

דוגמא 6.7 (פונקציה יוצרת מומנטים של משתנים מקריים מוכרים)**.**

(א) פונקציה יוצרת מומנטים של מ"מ **ברנולי** $X \sim \mathrm{Ber}(p)$ היא

$$\mathbb{E}(e^{tX}) = pe^t + 1 - p$$

(ב) פונקציה יוצרת מומנטים של מ"מ **בינומי** $X \sim \mathrm{Bin}(N,p)$ היא

$$\mathbb{E}(e^{tX}) = \sum_{n=0}^{N}\binom{N}{n}e^{nt}p^n(1-p)^{N-n} = (pe^t+1-p)$$

(ג) פונקציה יוצרת מומנטים של מ"מ **פואסוני** $X \sim \mathrm{Po}(\lambda)$ היא

$$\mathbb{E}(e^{tX}) = \sum_{n=0}^{\infty} e^{-\lambda}\frac{(\lambda e^t)^n}{n!} = e^{-\lambda}\sum_{n=0}^{\infty}\frac{(\lambda e^t)^n}{n!} = e^{\lambda(e^t-1)}$$

(ד) פונקציה יוצרת מומנטים של מ"מ **גיאומטרי** $X \sim \mathrm{Geo}(p)$ היא

$$\mathbb{E}(e^{tX}) = \sum_{n=1}^{\infty} e^n p(1-p)^{n-1} = \sum_{n=1}^{\infty} pe^t(e^t - pe^t)^{n-1} = \frac{pe^t}{1-(1-p)e^t}$$

והיא מוגדרת רק עבור $t < -\ln(1-p)$.

✎ **בעיה 6.2** (פונקציה יוצרת מומנטים של התפלגות בינומית שלילית)**.** לחשב את הפונקציה יוצרת מומנטים של Y עבור $Y = \sum_{i=1}^{n} X_i$ כאשר $X_i \sim \mathrm{Geo}(p)$ בלתי-תלויים.

עתה נסתייע בחשבונות אלו ונראה כיצד להוכיח באמצעות אי-שיוויון צ'רנוף ריכוז של ממשתנה מקרי סביב תוחלתו.

דוגמא 6.8 (כדורים בתאים)**.** מחלקים באקראי n כדורים ל-n תאים באופן אחיד ובלתי-תלוי. נחסם מלמעלה את העומס בתא העמוס ביותר. (להוכחה כי חסם זה הנו הדוק אסימפטוטית - ר' הפניה).

תשובה: נחשב לפי חסם האיחוד

$$\mathbb{P}\left(\max_{i\in[n]} S_i < k\right) \leq 1 - \sum_{i\in n} \mathbb{P}(S_i \geq k) \overset{\text{סימטריה}}{=} 1 - n\mathbb{P}(S_1 \geq k)$$

נסמן ב-X_i משתנה אינדיקטור למאורע שהכדור ה-i נפל בתא הראשון. אלו משתני ברנולי בלתי-תלויים ולכן,לפי טענה 3.46 מתקיים $S_1 \sim \mathrm{Bin}(n, 1/n)$. נשתמש בחישוב מדוגמא 6.7ב ונקבל כי

$$M_{S_1}(t) = \left(1 + \frac{e^t - 1}{n}\right)^n \leq \exp\left(\frac{e^t-1}{n} n\right) = \exp\left(e^t - 1\right).$$

נפעיל את אי-שיוויון צ'רנוף ונקבל כל לכל $t \in \mathbb{R}$ מתקיים:

$$\mathbb{P}(S_1 \geq a) \leq e^{-ta} M_{S_1}(t) \leq \exp\left(e^t - 1 - ta\right).$$

נגזור לפי t, נשווה לאפס ונקבל כי $t = \log a$ ממזער את החסם כך שמתקבלת ההערכה

$$\mathbb{P}(S_1 \geq a) \leq \exp\left(a - a\log(a) - 1\right) < \exp\left(-a(\log(a) - 1)\right).$$

כדי להשתמש בחסם האיחוד, נבקש a עבורו $\exp\left(-a(\log(a)-1)\right) = o(n^{-1})$. באופן שקול, נדרוש שיתקיים $a(\log(a)-1) - \log(n) \to \infty$. נחשב ונמצא כי עבור $a = c\log(n)/\log\log(n)$ כאשר $c > 1$ נקבל

$$\begin{aligned} a(\log(a) - 1) - \log(n) &= \frac{c\log(n)}{\log\log(n)}\Big(\log\log(n) + \log(c) - \log\log\log(n) - 1\Big) - \log(n) \\ &= \log(n)\left(c - 1 - \frac{c(\log\log\log(n) + 1 - \log(c))}{\log\log(n)}\right) \to \infty \end{aligned}$$

כמבוקש.

בפרק הבא נכליל את החסם למשפחה רחבה יותר של משתנים מקריים ונמחיש את תועלתו במספר דוגמאות כמותיות.

6.2 אי שיוויון הופדינג

את המסקנות מאי-שווין צ'רנוף למשתנה מקרי בינומי שפגשנו בדוגמא 6.12, הכליל בשנת 1963 המתמטיקאי הפיני וסילי הופדינג (Wassily Hoeffding) לסכום של משתנים מקריים בלתי-תלויים וחסומים. נתחיל בהצגת הלמה של הופדינג העוסקת בחסימת הפונקציה היוצרת מומנטים של משתנה מקרי חסום.

טענה 6.9 (הלמה של הופדינג)**.** יהי X משתנה מקרי המקיים $|X| \leq 1$ וכן $\mathbb{E}(X) = 0$. אזי לכל $t \in \mathbb{R}$ מתקיים

$$\mathbb{E}\left(e^{tX}\right) \leq \exp\left(\frac{t^2}{2}\right).$$

הוכחה. נקבע את t. הפונקציה e^{tx}, כפונקציה של x היא פונקציה בעלת נגזרת שניה חיובית ולכן קמורה. בפרט, אם נסמן ב-$L(x)$ את הפונקציה המתארת את הישר שעובר דרך הנקודות $(-1, e^{-t})$ ו-$(1, e^t)$ אז לכל $x \in [-1, 1]$ מתקיים

$$e^{tx} \leq L(x).$$

משוואת הישר L היא

$$L(x) = \frac{e^t + e^{-t}}{2} + x\frac{e^t - e^{-t}}{2}.$$

ממונוטוניות התוחלת והליניאריות שלה (טענה 4.5) נקבל

$$\mathbb{E}(e^{tX}) \leq \mathbb{E}(L(X)) = \frac{e^t + e^{-t}}{2} + \mathbb{E}(X)\frac{e^t - e^{-t}}{2} = \frac{e^t + e^{-t}}{2}.$$

כעת נותר לבדוק שלכל $t \in \mathbb{R}$ מתקיים

$$\frac{e^t + e^{-t}}{2} \leq e^{t^2/2}.$$

נשווה בין טורי טיילור של שני הצדדים ונקבל

$$\frac{e^t + e^{-t}}{2} = \frac{1}{2}\sum_{n=0}^{\infty}\frac{t^n + (-t)^n}{n!} = \sum_{m=0}^{\infty}\frac{t^{2m}}{(2m)!} \leq \sum_{m=0}^{\infty}\frac{t^{2m}}{2^m m!} = \sum_{m=0}^{\infty}\frac{(t^2/2)^m}{m!} = e^{t^2/2}$$

כנדרש. ■

משפט 6.10 (אי-שוויון הופדינג)**.** יהיו $\{X_i\}_{i\in[N]}$ משתנים מקריים בלתי-תלויים ובעלי תוחלת אפס, אשר מקיימים $|X_i| \overset{\text{a.s.}}{\leq} 1$ לכל $i \in [N]$. אזי לכל $a > 0$ המקיימים

$$\mathbb{P}\left(\sum_{i\in[N]} X_i \geq a\right) \leq \exp\left(-\frac{a^2}{2N}\right).$$

הוכחה. נסמן $X = \sum_{i\in[N]} X_i$.

$$M_X(t) \overset{\substack{\text{טענה}\\6.5}}{=} \prod_{i\in[N]} M_{X_i}(t) \overset{\substack{\text{טענה}\\6.9}}{\le} \prod_{i\in[N]} \exp\left(\frac{t^2}{2}\right) = \exp\left(\frac{Nt^2}{2}\right).$$

מכאן שלפי אי-שיוויון צ'רנוף (משפט 6.6) לכל $t > 0$ מתקיים

$$\mathbb{P}(X \ge a) \le \exp\left(\frac{Nt^2}{2} - ta\right).$$

כדי למצא t שימזער את החסם נגזור את המעריך ונשווה לאפס

$$\frac{d}{dt}\left(\frac{Nt^2}{2} - ta\right) = Nt - a = 0$$

נבחר אפוא $t = a/N$ ונקבל

$$\mathbb{P}(X \ge a) \le \exp\left(\frac{N(a/N)^2}{2} - (a/N)a\right) = \exp\left(-\frac{a^2}{2N}\right),$$

כנדרש. ■

כעת ניווכח בשיפור שמציג חסם זה ביחס לבעיות שעסקנו בהן בפרקים הקודמים. נפתח בבעיה המוצגת בדוגמא 5.7.

דוגמא 6.11 (ריכוז התפלגות בינומית).

מיליון מטבעות הוגנים מוטלים. חסום את הסיכוי שכמות המטבעות שתוצאתם עץ תהייה בין $495{,}000$ ל-$505{,}000$.

תשובה: נסמן ב-X_i מ"מ המקבל את הערך 1 אם המטבע ה-i הראה תוצאה של עץ ו--1 אחרת. כמו כן נסמן $X = \sum_{i=1}^{10^6}(X_i)$. אנו מעוניינים לחסום את ההסתברות ש-$|X|$ קטן מ-10^4. נשים לב כי X_i עומדים בתנאי א"ש הופדינג, משפט 6.10. נפעיל את האי-שוויון ונקבל כי המאורע שאנו מעוניינים בו מקיים

$$\begin{aligned}\mathbb{P}(-10^4 \le X \le 10^4) > 1 - \mathbb{P}(X \ge 10^4) - \mathbb{P}(-X \ge 10^4) \ge 1 - 2\exp\left(-\frac{10^8}{2\cdot 10^6}\right)\\ = 1 - 2\exp(-50) \approx 1 - 10^{-22}.\end{aligned}$$

שיפור בלתי-נתפס בהשוואה לחסם שקיבלנו בדוגמא 5.7.

באופן דומה נוכל להשתמש בא"ש צ'רנוף לקבלת חסם כללי על סטיה בהתפלגות התפלגות בינומית.

דוגמא 6.12 (ריכוז משתנה בינומי).

יהי $X \sim \mathrm{Bin}(N, p)$, נראה כי

$$\mathbb{P}(X - Np \ge a) \le \exp\left(\frac{a^2}{N\max(1-p,p)^2}\right).$$

תשובה: יהיו $Y_i \sim \mathrm{Ber}(p)$ ונשים לב $\mathbb{E}(Y_i - p) = 0$ וכי

$$\left|\frac{Y_i - p}{\max(1-p,p)}\right| \le 1,$$

ולכן $\frac{Y_i-p}{\max(1-p,p)}$ מקיימים את אי-שוויון הופדינג (משפט 6.10). נציב ונקבל:

$$\mathbb{P}(X-Np\geq a)=\mathbb{P}\left(Y\geq\frac{a}{\max(1-p,p)}\right)\leq\exp\left(\frac{a^2}{N\max(1-p,p)^2}\right).$$

6.3* ממומנטים לפונקציה יוצרת מומנטים

השם פונקציה יוצרת מומנטים מקורו בתכונה הבאה:

טענה 6.13. יהי X משתנה מקרי בדיד עבורו M_X מוגדרת היטב בסביבות 0. אז מתקיים

$$M_X^{(k)}(0)=m_k(X)=\mathbb{E}(X^k)$$

כלומר הנגזרת מסדר k של הפונקציה יוצרת מומנטים היא המומנט מסדר k של המשתנה המקרי.

הוכחה. לפי הנחתנו $M_X(s)$ מוגדרת בקטע $(-t_0,t_0)$ עבור $t_0>0$ כלשהו. מפני ש-e^{xt} מונוטונית עולה ב-t עבור x חיובי ומונוטונית יורדת עבור x שלילי, הרי שלכל $|t|<t_0$ מתקיים

$$M_X(t)<\mathbb{E}(e^{xt})+\mathbb{E}(e^{-xt})=M_X(t_0)+M_X(-t_0).$$

נסמן חסם זה ב-$a=M_X(t_0)+M_X(-t_0)$. נזכר כי לכל t מתקיים

$$e^{|t|}<e^{-t}+e^t$$

ומכאן שמתקיים לכל $|t|<t_0$,

$$\sum_{x\in\mathrm{Supp}(X)}e^{|tx|}\mathbb{P}(X=x)\leq\sum_{x\in\mathrm{Supp}(X)}e^{tx}\mathbb{P}(X=x)+\sum_{x\in\mathrm{Supp}(X)}e^{-tx}\mathbb{P}(X=x)<2a$$

נרשום

$$\mathbb{E}\left(e^{tX}\right)=\sum_{x\in\mathrm{Supp}(X)}e^{tx}\,\mathbb{P}(X=x)=\sum_{x\in\mathrm{Supp}(X)}\mathbb{P}(X=x)\sum_{n=0}^{\infty}\frac{x^nt^n}{n!}$$

נשתמש בעובדה ש$\sum_{x\in\mathrm{Supp}(X)}e^{|tx|}\mathbb{P}(X=x)$ סופי כדי להסיק כי הסכום באגף ימין מתכנס בהחלט. לכן נוכל לשנות סדר סכימה ולקבל

$$\mathbb{E}\left(e^{tX}\right)=\sum_{n=0}^{\infty}\sum_{x\in\mathrm{Supp}(X)}\mathbb{P}(X=x)\frac{x^nt^n}{n!}=\sum_{n=0}^{\infty}\frac{\mathbb{E}(X^n)}{n!}t^n$$

זהו טור חזקות המתכנס בהחלט לכל $|t|<t_0$, ולכן בתוך רדיוס ההתכנסות ניתן להחליף נגזרת וסכום. כעת נגזור את הטור ונקבל את המבוקש. ■

בעיות הרחבה והעשרה

✎ **בעיה 6.3.** עבור מ"מ X הנתמך על $\mathbb{N}_0$ נגדיר את ה**פונקציה יוצרת הסתברות** של X בתור

$$G_X(x) = \sum_{i=0}^{\infty} x^i \mathbb{P}(x = i).$$

(א) יש להראות כי $G_X(e^x)$ היא הפונקציה יוצרת מומנטים של X, והסבר כיצד לחשב מומנטים של X מ-$f_X(x)$.

(ב) יש להוכיח כי אם Y בלתי-תלוי ב-X ובעל תומך ב-$\mathbb{N}_0$, ו-$a \in \mathbb{Z}$ אז מתקיים $G_{X+Y} \equiv G_X G_Y$ וכן $G_{aX}(x) = G_X(x^a)$

(ג) יהיו $\{X_n\}_{n\in\mathbb{N}}$ משתנים מקריים בלתי-תלויים הנתמכים על $\mathbb{N}_0$ ושווי התפלגות ויהי Y מ"מ בלתי-תלוי בכולם הנתמך על $\mathbb{N}_0$ יש להראות כי מתקיים

$$G_{\sum_{i=1}^{Y} X_i}(x) = G_Y(G_X(x)).$$

(נשים לב שבשלב זה, אין אנו מכירים מרחב הסתברות שבו קיימים אינסוף משתנים מקריים לא-קבועים ובלתי-תלויים. היכרות עם מרחבים כאלו תערך בפרק 7.

(ד) מספר המכוניות שחולפות בצומת בדקה מתפלג פואסונית עם שכיחות 1. בכל מכונית מספר הנוסעים מתפלג אחיד על $[4]$ באופן בלתי-תלוי. יש לחשב את הפונקציה יוצרת הסתברות של מספר האנשים שחולפים בצומת בדקה. (התפלגות כזו נקראת התפלגות פואסון מורכבת).

✎ **בעיה 6.4.** יהיו $\{X_n\}_{n\in[N+1]}$ משתני ברנולי $1/2$ בלתי-תלויים ונסמן $Y_n = X_n X_{n+1}$. יש לחשב באמצעות אי-שוויון צ'בישב ובאמצעות אי-שוויון הופדינג הערכות להסתברות $\mathbb{P}(\sum_{n\in[N]} Y_n > \frac{aN}{4})$ לכל $a \in \mathbb{R}_+$ ולהשוות ביניהן. (נשים לב ש$\{Y_n\}$ אינם בלתי-תלויים, אך $\{Y_{2n}\}$ דווקא כן).

✎ **בעיה 6.5.** נבחר באופן אחיד גרף על n קודקודים. **המרחק** בין שני קודקודים בגרף הוא אורך המסלול הקצר ביותר המחבר אותם. **קוטר** הגרף הוא המרחק המירבי בין שני קודקודים. יש להראות כי קיימים $c, C > 0$ שאינם תלויים ב-n, כך שההסתברות שקוטר הגרף הוא בדיוק שתיים, גדולה מ-$1 - Ce^{cn}$.

✎ **בעיה 6.6 (*).** להשתמש באי-שיוויון הופדינג ובחסם האיחוד כדי להראות שבהינתן $A_1, \ldots, A_m$ תת-קבוצות של המספרים $1, \ldots, N$, ניתן למצא קבוצה B כך שלכל $i \in [m]$ מתקיים

$$|A_i \cap B| - |A_i \cap B^c| \leq \sqrt{2n \ln(2m)}.$$

הדרכה: יש להראות כי עבור בחירה מקרית ואחידה של B, הסתברותו של המאורע המשלים קטנה מ-1.

חלק II

מרחבי הסתברות כלליים

פרק 7

מבוא למרחבי הסתברות כלליים

> את תורת ההסתברות, מעצם היותה ענף של המתמטיקה, מוטב לבסס על אקסיומות, ממש כמו את הגיאומטריה ואת האלגברה. פירושו של דבר שמהרגע שהגדרנו את הגורמים ואת היחסים הבסיסיים ביניהם והצהרנו על האקסיומות המושלות ביחסים אלו, כל טענה חייבת להיות מבוססת אך ורק על אותן האקסיומות, באופן בלתי-תלוי במשמעות המקובלת המוחשית של אותם גורמים ושל אותם יחסים.
>
> – אנדרי ניקולאביץ' קולמוגורוב, יסודות תורת ההסתברות, 1933

בבואנו להגדיר מרחבי הסתברות בפרק 1, פתחנו בהגדרת **פונקציית ההתפלגות הנקודתית**. בעקבותיה בנינו את **פונקציית ההסתברות**, וכך התאפשרה הגדרתם של מרחבי הסתברות. על מרחבי מדגם בדידים אפשרה גישה זו טיפול בכל סוגיה הסתברותית שעמדה בפנינו, ואולם עד מהרה נגלה שגישה זו מעוררת בעיות קשות, שלא ניתן ליישבן, כאשר פונים למרחבי הסתברות כלליים. נפתח בדוגמא.

נשווה בנפשנו בחירה אקראית של מספר ממשי בין 0 ל-1 המתבצעת על ידי הטלה חוזרת ונשנית של קוביה הוגנת בעלת עשר פיאות. בכל הטלה נקבעת ספרה נוספת של המספר אחרי הנקודה. מרחב המדגם כאן הוא מרחב סדרות הספרות האינסופית, ונשים לב שמשיקולי סימטריה, כל סדרת ספרות הנה בעלת סיכוי שווה להיבחר ולכן אם נצליח להגדיר התפלגות זו הרי שהיא תהיה אחידה על מרחב הסדרות. יתר על כן - התהליך שתואר כאן הוא תהליך שבבירור ניתן לבצעו בפועל (בהנתן פרק זמן בלתי מוגבל) ולכן טבעי לצפות שהתפלגות אחידה על סדרות הספרות אכן תהיה קיימת. בפרק זה נתאר בקווים כללים את הכלים המשמשים להוכחת קיומו של מרחב כזה. בהמשך נגלה כי מרחב זה, המכונה **מרחב ההסתברות התקני** מספיק להגדרת כל מערכת משתנים מקריים שנחפוץ בה.

הקושי בהגדרת מרחב ההסתברות התקני עלה כבר בבעיה 1.1. שם ראינו כי לא ניתן להגדיר פונקציית התפלגות נקודתית אחידה על מרחב אינסופי. הרי מטעמי סימטריה נרצה לייחס לכל סדרה הסתברות שווה, ומכיוון שמספר הסדרות אינסופי – לא יתכן שסכום הסתברויותיהן של כל הסדרות יהיה 1. מתוך מודעות לבעיה זו, הגדרנו בהגדרה 1.3 פונקציית הסתברות באופן שלא נסמך על פונקציית ההתברות נקודתית. אם רק נוכל למצא פונקציית הסתברות שתקיים את הגדרה 1.3, הרי שחלק הארי של התורה שפיתחנו יהא תקף לגביה. כאן מתגלה קושי נוסף ומהותי, אשר דיון מעמיק בו שייך לתחומה של תורת המידה. נתאר את הקושי באמצעות דוגמה. מפונקציית הסתברות אחידה על המרחב $[0,1]$ נבקש כי הסתברותו של כל מקטע $[a,b]$ תהיה

שווה לאורכו. מסתבר שקיומה של פונקציית הסתברות כזו על מרחב המאורעות $2^{\Omega} = \{A \subset [0,1]\}$ איננו עולה בקנה אחד עם האקסיומות המקובלות של המתמטיקה (ובמיוחד עם אקסיומת הבחירה). כדי להתמודד עם בעיה זו ניאלץ לצמצם עצמנו ממרחב המאורעות 2^{Ω} למרחב מאורעות מדידים מצומצם יותר $\mathcal{F}$. על מנת שנוכל להוסיף ולבצע פעולות על מאורעות, נדרוש מ-$\mathcal{F}$ להיות סגור לאוסף טבעי של פעולות. דרישה זו תתבטא בכך שנבקש מ $\mathcal{F}$ להוות σ**-אלגברה**, מושג חדש שיובהר בפרק זה.

מטרתו של פרק זה למסור סקירה שטחית של הפתרונות שמציעה תורת המידה להגדרת מרחבי הסתברות כללים. אי לכך תובאנה בפרק זה מספר טענות בלא הוכחה. לפיתוח מלא ומקיף של תורת ההסתברות למרחבי הסתברות כללים יהא על הקורא לפנות לספרים מתקדמים יותר.

7.1 מרחבי הסתברות כללים

נפתח באפיון הדרישות מאוסף מאורעות מדידים.

הגדרה 7.1 (σ-אלגברה)**.** יהי Ω מרחב מדגם כלשהו. קבוצה $\mathcal{F}$ של תת-קבוצות של Ω נקראת σ**-אלגברה** (קרי: סיגמא-אלגברה), אם היא מקיימת את התכונות הבאות:

(א) $\emptyset \in \mathcal{F}$.

(ב) **(סגירות לאיחוד בן מניה)** $\{U_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ גורר כי $\bigcup_{i\in\mathbb{N}} U_i \in \mathcal{F}$.

(ג) **(סגירות למשלים)** $U \in \mathcal{F}$ גורר כי $\Omega \setminus U \in \mathcal{F}$.

כל מאורע $U \in \mathcal{F}$ יקרא מאורע מדיד לפי $\mathcal{F}$.

✎ **בעיה 7.1** (σ-אלגברה סגורה לחיתוכים בני-מניה)**.** תהי $\mathcal{F}$ σ-אלגברה על מרחב מדגם Ω ותהיינה $\{U_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$. יש להוכיח כי $\bigcap_{i\in\mathbb{N}} U_i \in \mathcal{F}$.

סגירותה של σ-אלגברה לחיתוכים ואיחודים מבטיחה לנו שאם נקח אוסף של מאורעות מדידים לפי $\mathcal{F}$ ונבצע עליהם מספר בן מניה של פעולות בסיסיות (איחוד, חיתוך, משלים, חיסור וכד') התוצאה שתתקבל עדיין תהיה מאורע מדיד לפי $\mathcal{F}$. נשים לב ש-2^{Ω}, כלומר אוסף כל התת-קבוצות של Ω, הוא תמיד σ-אלגברה.

נכליל כעת את הגדרתה של פונקציית ההסתברות הבדידה, על ידי כך שנשמיט את הדרישה המובלעת לקיומה של קבוצה בת-מניה A כך ש-$\mathbb{P}(A) = 1$.

הגדרה 7.2. תהי $\mathcal{F}$ σ-אלגברה על מרחב מדגם Ω. **פונקציית הסתברות** $\mathbb{P} : \mathcal{F} \to \mathbb{R}_+$ על $(\Omega, \mathcal{F})$ היא פונקציה המקיימת את שתי התכונות הבאות.

(א) $\mathbb{P}(\Omega) = 1$.

(ב) **סכימות בת-מניה (σ-additivity):** לכל אוסף בן-מניה של מאורעות זרים $\{A_i\}_{i\in\mathbb{N}}$ ב-$\mathcal{F}$ מתקיים

$$\mathbb{P}\left(\dot{\bigcup_{i\in\mathbb{N}}} A_i\right) = \sum_{i\in\mathbb{N}} \mathbb{P}(A_i).$$

באמצעות מושג ה-σ-אלגברה נוכל להכליל את הגדרה 7.3 ולהגדיר מרחב הסתברות כללי.

הגדרה 7.3 (מרחב הסתברות)**.** תהי $\mathcal{F}$ σ-אלגברה על מרחב מדגם Ω ותהי $\mathcal{P}$ פונקציית התפלגות על $(\Omega, \mathcal{F})$. השלשה $(\Omega, \mathcal{F}, \mathbb{P})$ מכונה - **מרחב הסתברות**.

7.2 אלגברת בורל ומרחב ההסתברות התקני

ה-σ-אלגברה החשובה ביותר לצורכינו היא ה-σ-אלגברה על $\mathbb{R}$ שהוגדרה על ידי אמיל בורל (Émile Borel) ב-1898. σ-אלגברה זו, המכונה σ**-אלגברת בורל** אשר נסמן ב-$\mathcal{F} = \mathbb{B}(\mathbb{R})$, מאופיינת בכך שהיא מכילה את כל הקטעים $[a, b]$ עבור $a \leq b$ ממשיים.

טענה 7.4 (אלגברת בורל)**.** קיימת $\mathbb{B}(\mathbb{R})$, **ה-σ-אלגברה של בורל**, שמכילה את כל הקטעים $[a, b] \subset \mathbb{R}$ והיא מינימלי במובן שהיא מוכלת בכל σ-אלגברה אחרת שמכילה את הקטעים.

לכל קטע $I \subset \mathbb{R}$ (פתוח או סגור, סופי או אינסופי) נגדיר σ-אלגברה של בורל על ידי

$$\mathbb{B}(I) = \{A \cap I \ : \ A \in \mathbb{B}(\mathbb{R})\}.$$

הוכחת טענה 7.4 מובאת בפרק 7.4 להלן.

כעת נבקש להגדיר מרחב הסתברות $([0, 1], \mathbb{B}([0, 1]), \mathbb{P})$ שיתאר התפלגות אחידה על המספרים הממשיים בין 0 ל-1. לשם כך נבקש פונקציית הסתברות על $\mathbb{B}([0, 1])$ שתייצג את ההתפלגות האחידה על הקטע $[0, 1]$. תוצאה מתורת המידה גורסת כי פונקציית הסתברות כזו אומנם קיימת והיא מכונה **מידת לבג** על שם המתמטיקאי הצרפתי אנרי לבג (Henri Lebesgue) שפיתח אותה בסביבות 1902.

טענה 7.5 (מידת לבג)**.** קיימת פונקציית הסתברות על $\mathbb{B}([0, 1])$ כך שלכל $0 \leq a \leq b \leq 1$ מתקיים

$$\mathbb{P}([a, b]) = b - a.$$

כעת נגדיר את מרחב ההסתברות התקני.

הגדרה 7.6. המרחב $([0, 1], \mathbb{B}, \mathbb{P})$ יכונה **מרחב ההסתברות התקני**, או מרחב הסתברות אחיד על $[0, 1]$.

נשים לב שזהו המרחב הראשון שנתעניין בו שאינו מרחב הסתברות בדידה, ואומנם הוא מקיים את התכונה המעניינת $\mathbb{P}(\{a\}) = 0$ לכל $a \in \Omega$. כלומר הסתברותו של כל יחידון היא 0 ולכן גם $\mathbb{P}([a, b]) = \mathbb{P}((a, b))$ לכל $0 \leq a \leq b \leq 1$. נשים לב גם שהסתברותה של כל קבוצה בת-מניה $A \subset [0, 1]$ היא $\mathbb{P}(A) = 0$ וזאת תוך שימוש בסכימות בת-מניה.

מרחב הסתברות כזה, שבו הסתברותו של כל יחידון היא 0, נקרא מרחב **רציף** ואנו נקדיש את פרק 9 לעיסוק במרחבים מסוג זה.

ישנו קשר הדוק בין חישוב הסתברויות במרחב בורל לבין אינטגרל כפי שנובע מהטענה הבאה (אף היא מובאת ללא הוכחה).

טענה 7.7 (חישוב הסתברויות במרחב בורל)**.** תהי $A \in \mathbb{B}([0,1])$ במרחב בורל, ויהי $([0,1],\mathbb{B},\mathbb{P})$ מרחב ההסתברות התקני. אזי

$$\mathbb{P}(A) = \int_0^1 \mathbb{1}(A)dx,$$

כאשר $\mathbb{1}(A)$ אינטגרבילית רימן.

✎ **בעיה 7.2.** נזכר כי כל $\omega \in [0,1]$ ניתן לרשום בתור

$$\sum_{i=1}^{n} \frac{a_i}{3^n} = \omega$$

עבור $a_i \in \{0,1,2\}$. יש להראות כי ההסתברות של קבוצת המספרים שניתן לרשום באופן זה עם $a_i \neq 1$ לכל $i \in \mathbb{N}$, היא אפס.

7.3 משתנים מקריים מעל מרחבי הסתברות כללים

נרחיב כעת את הגדרת המשתנה המקרי הבדיד (הגדרה 3.1) להגדרה של משתנה מקרי כללי.

הגדרה 7.8. יהי Ω מרחב מדגם ותהי $\mathcal{F}$ σ-אלגברה על Ω. **משתנה מקרי (מ"מ) (random variable)** על $(\Omega,\mathcal{F})$ הוא פונקציה $X:\Omega\to\mathbb{R}$ המקיימת $X^{-1}(B)\in\mathcal{F}$ לכל $B\in\mathbb{B}(\mathbb{R})$.

למעשה הגדרה זו אינה שונה בהרבה מההגדרה שכבר הכרנו, אלא שכעת נבקש שלכל קבוצת בורל $A \in \mathbb{B}(\mathbb{R})$ נוכל להגדיר הסתברות למאורע $\{\omega \in \Omega \ : \ X(\omega) \in A\}$ (ולא לכל קבוצה שהיא ב-$2^{\mathbb{R}}$). עתה התפלגותו של משתנה מקרי תוגדר בדומה להגדרה 3.7, אך מבלי להתייחס להתפלגות נקודתית, אשר קיומה אינו מובטח לנו עוד.

הגדרה 7.9 (התפלגותו של משתנה מקרי כללי)**.** יהי X משתנה מקרי על מרחב הסתברות $(\Omega,\mathcal{F},\mathbb{P})$. **התפלגותו** של X תהיה הפונקציה $\mathbb{P}_X:\mathbb{B}(\mathbb{R})\to[0,1]$ הנתונה על ידי:

$$\mathbb{P}_X(A) := \mathbb{P}\Big(X^{-1}(A)\Big).$$

כמקודם, נרשום $X \sim \mathbb{P}_X$ ונסמן $\mathbb{P}(X\in A) = \mathbb{P}_X(A)$ באומרנו כי זו היא ה**הסתברות שערכו של X נמצא ב-A** או שX **מקבל ערך ב-A**.

כשם שאפיינו את התפלגותו של משתנה מקרי בדיד באמצעות פונקציית ההתפלגות הנקודתית, נוכל לאפיין את התפלגותו של משתנה מקרי כללי באמצעות פונקציית התפלגות מצטברת.

הגדרה 7.10 (פונקציית הסתברות מצטברת)**.** יהיה X משתנה על מרחב הסתברות. נגדיר את **פונקציית ההסתברות המצטברת (comulative distribution function)** $F_X:\mathbb{R}\to\mathbb{R}$ באמצעות

$$F_X(s) = \mathbb{P}(X \le s).$$

נביא ללא הוכחה את הטענה לפיה פונקציית ההתפלגות המצטברת מאפיינת את המשתנה המקרי.

טענה 7.11 (התפלגות מצטברת מאפיינת את פונקציית ההתפלגות)**.** יהיו X ו-Y שני משתנים מקריים על מרחב הסתברות. אם $F_X \equiv F_Y$ אז $\mathbb{P}_X = \mathbb{P}_Y$ (ובמילים אחרות $X \overset{\mathrm{d}}{=} Y$).

כעת נגדיר התפלגות בדידה באופן שיכליל את הגדרה 3.7.

הגדרה 7.12 (משתנה מקרי רציף)**.** X משתנה מקרי על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נאמר כי ל-$X$ התפלגות **בדידה** אם קיימת קבוצה A בת-מניה, כך שמתקיים $\mathbb{P}(X \in A) = 1$.

כעת נציג הגדרה משלימה למשתנה מקרי בדיד בדמות **משתנה מקרי רציף**.

הגדרה 7.13 (משתנה מקרי רציף)**.** יהי X משתנה מקרי על מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נאמר כי ל-$X$ התפלגות **רציפה** אם לכל $x \in \mathbb{R}$ מתקיים $\mathbb{P}(X = x) = 0$.

לא לכל המשתנים המקריים התפלגות רציפה או בדידה. משתנה מקרי שאינו רציף או בדיד נקרא משתנה מקרי **מעורב**. נסתכל על הדוגמה הבאה:

דוגמא 7.14 (משתנה מקרי מעורב)**.** יהי $([0,1], \mathbb{B}([0,1]), \mathbb{P})$ מרחב ההסתברות התקני ונגדיר משתנה מקרי

$$X(\omega) = \min(\omega, 1/2).$$

נשים לב כי מחד-גיסא

$$\mathbb{P}(X = 1/2) = \mathbb{P}(\omega \in [1/2, 1]) = 1 - 1/2 = 1/2$$

ולכן X אינו רציף, אך מאידך גיסא לכל $x \neq 1/2$ מתקיים $\mathbb{P}(X = x) = 0$ ולכן לכל קבוצה בת-מניה A מתקיים

$$\mathbb{P}(X \in A) \leq \mathbb{P}(X = 1/2) = 1/2,$$

כך ש-X אינו בדיד.

העיסוק במשתנים מקריים כללים דורש שימוש נרחב בתורת המידה. אנו נגביל עצמנו למשפחה שימושית במיוחד ונגישה יחסית של משתנים מקריים המכונים **משתנים רציפים בהחלט**, הללו יעמדו במוקדו של פרק 9.

7.3.1 תלות של משתנים מקריים כללים

כשם שהגדרנו אי-תלות עבור משתנים מקריים בדידים נוכל להכליל הגדרה זו למשתנים מקריים כללים. ההבדל היחיד בין שתי ההגדרות הוא שכעת אנו מודעים לכך שבכדי ש-$\{X \in A\}$ יהיה מאורע, על A להיות קבוצת בורל.

הגדרה 7.15 (אי-תלות של שני משתנים מקריים)**.** נאמר שמשתנים מקריים X ו-Y (על אותו מרחב הסתברות) הנם **בלתי תלויים (ב"ת)** ונסמן $X \perp\!\!\!\perp Y$, אם לכל שתי קבוצות $A, B \subset \mathbb{B}(\mathbb{R})$ מתקיים שהמאורעות $\{X \in A\}$ ו-$\{Y \in B\}$ בלתי תלויים, כלומר

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

ד

7.4* סיגמא אלגברה של בורל

בפרק זה נתוודע באופן נרחב יותר למושג ה-σ-אלגברה ונוכיח את טענה 7.4.

טענה 7.16 (חיתוך σ-אלגברות)**.** יהי Ω מרחב מדגם כלשהו. ותהי S קבוצה לא ריקה של σ-אלגבראות עליו. אז אם נסמן ב-$\mathcal{F}$ את חיתוך כל ה-σ-אלגבראות ב-S, הרי ש-$\mathcal{F}$ היא σ-אלגברה על Ω.

הוכחה. נבדוק ששלוש התכונות של σ-אלגברה מתקיימות. לפי הגדרה 7.1א כל הקבוצות בS מכילות את הקבוצה הריקה ולכן היא נמצאת גם ב-$\mathcal{F}$. תהיינה $\{U_i\}_{i\in\mathbb{N}}$ קבוצות שנמצאות בכל האיברים של S. לפי הגדרה 7.1ב מתקיים ש-$\bigcup_{i\in\mathbb{N}} U_i$ נמצאת גם בכל איברי S ולכן גם ב-$\mathcal{F}$. תהי U שנמצאת בכל האיברים של S, לפי הגדרה 7.1ג מתקיים ש- $\Omega \setminus U$ נמצאת בכל איברי S ולכן ב-$\mathcal{F}$. ■

בנוסף ניתן לצמצם σ-אלגברה לתת-קבוצה של המרחב.

טענה 7.17 (σ-אלגברה מצומצמת)**.** יהי Ω מרחב מדגם כלשהו, תהי $\mathcal{F}$ σ-אלגברה על Ω ותהי $W \subset \Omega$. אז, הקבוצה $\mathcal{G} = \{F \cap W \ : \ F \in \mathcal{F}\}$ היא σ-אלגברה על W. זו מכונה **הצמצום של** $\mathcal{F}$ **ל-**W.

הוכחה. נבדוק ששלוש התכונות של σ-אלגברה מתקיימות. לפי הגדרה 7.1א הקבוצה הריקה מצויה ב-$\mathcal{F}$ ולכן גם ב-$\mathcal{G}$. תהיינה $\{U_i\}_{i\in\mathbb{N}}$ קבוצות ב-$\mathcal{G}$, אז קיימות $\{U'_i\}_{i\in\mathbb{N}}$ ב-$\mathcal{F}$ כך ש-$U_i = U'_i \cap W$. לפי הגדרה 7.1ב מתקיים ש-$\bigcup_{i\in\mathbb{N}} U'_i \in \mathcal{F}$ ולכן $\bigcup_{i\in\mathbb{N}} U_i \in \mathcal{G}$ $\bigcup_{i\in\mathbb{N}} U'_i \cap W =$. לסיום, תהי $U \in \mathcal{G}$ אז קיימת $U' \in \mathcal{F}$ כך ש-$U' \cap W = U$ ולפי הגדרה 7.1ג מתקיים כי $\Omega \setminus U' \in \mathcal{F}$ ולכן $(\Omega \setminus U) \cap W = W \setminus U \in \mathcal{G}$. ■

באמצעות כלים אלו ניתן ליצור σ-אלגברה מינימלית המכילה אוסף קבוצות.

הגדרה 7.18 (σ-אלגברה שנוצרת על ידי אוסף קבוצות)**.** יהי Ω מרחב מדגם כלשהו. יהי $\mathcal{G}$ אוסף לא ריק של תת-קבוצות של Ω. ונסמן ב-S את אוסף כל ה-σ-אלגבראות על Ω שמכילות את $\mathcal{G}$. נגדיר את $\mathcal{F}$, **ה-σ-אלגברה שנוצרת על ידי** $\mathcal{G}$ בתור חיתוך כל ה-σ-אלגבראות ב-S.

נשים לב שהגדרה 7.18 מוגדרת היטב, שכן S תמיד מכילה את ה-σ-אלגברה שמורכבת מכל התת-קבוצות של Ω ולכן היא אינה ריקה. עוד ניתן דעתנו על כך ש-$\mathcal{F}$ היא מינימלית מן הבחינה שכל σ-אלגברה שמכילה את $\mathcal{G}$, מכילה גם את $\mathcal{F}$.

הערה: הבניה הנתונה בהגדרה 7.18 איננה מפורשת. קיימת גם בניה מפורשת ל-σ-אלגברה שנוצרת על ידי אוסף יוצרים בתור מערכת של איחודים בני מניה ולקיחת משלים של יוצרים אלה. אולם בניית מערכת זו מסובכת ומערבת אינדוקציה טרנספיניטית (השקולה לגרסא מסויימת של אקסיומת הבחירה). הקורא המתעניין מוזמן בסיום קריאת פרק זה להרחיב את ידיעותיו באמצעות ספרים נוספים בנושא "ההיררכיה של בורל" אשר מהווה צוהר לבניה זו.

כעת טענה 7.4 היא לא יותר ממקרה פרטי של הגדרה 7.18.

הגדרה 7.19 (σ-אלגברה של בורל)**. ה-σ-אלגברה של בורל** הנה ה-σ-אלגברה שנוצרת על ידי

$$\mathcal{G} = \{[a,b] \ : \ a < b\}.$$

הערה: אין זה פשוט כלל ועיקר למצא קבוצה שאינה בורל. בין השנים 1902 ו-1904 הכליל אנרי לבג (Henri Lebesgue) את ה-σ-אלגברה של בורל ל-σ-אלגברה עשירה יותר – **σ-אלגברת לבג**. עצם קיומן של קבוצות שאינן מדידות לבג תלוי באקסיומת הבחירה ולכן לעולם לא ניתקל בקבוצת כאלה. ברמת הקורס אנו נתעלם לעיתים קרובות מהשאלה האם קבוצה היא מדידה, ונסמוך על כך שכל הפעולות שאנו משתמשים בהן משמרות מדידות.

פרק 8

סדרות של משתנים מקריים

"...התורה של ניסויים בלתי-תלויים היא בעת ובעונה אחת החלק הפשוט ביותר מבחינה אנליטית והמתקדם ביותר של תורת ההסתברות."

– וויליאם פלר, מבוא לתורת ההסתברות ושימושיה, 1950.

בפרקים הקודמים הגבלנו את עיסוקנו לקבוצות סופיות של משתנים מקריים. אומנם, באמצעות השאפת אורך הסדרה לאינסוף ניסינו לעמוד על ההתנהגות האסימפטוטית של אוספים כאלו כאשר מספר המשתנים הולך וגדל ולהבין אילו מאורעות נעשים סבירים יותר ויותר ואילו הופכים נדירים יותר ויותר. באמצעות הסתכלות זו ניסחנו והוכחנו את החוק החלש של המספרים הגדולים (משפט 5.32), ותיארנו את התנהגותן של בעיות הסתברותיות שונות. לצורך כך היה עלינו להסתכל על כל אוסף סופי של משתנים מקריים כאילו הוא מוגדר במרחב הסתברות משלו.

בפרק זה נבקש להסתכל על סדרה אינסופית של משתנים מקריים בלתי-תלויים המוגדרת על אותו מרחב הסתברות ולראותה כסדרה מקרית של ערכים. כך נוכל למשל לנסח את השאלה "מה היא ההסתברות שסדרה מסויימת מתכנסת? ". לצורך טיפול בשאלות אלו נידרש למרחב הסתברות שעליו מוגדרים אינסוף משתנים מקריים בלתי-תלויים. אין זה קשה לוודא שאף מרחב הסתברות בדידה אינו מקיים תכונה זו. ואולם, את מרחב ההסתברות האחיד על $[0,1]$ נוכל לראות כמרחב של סדרות אינסופיות של ספרות עשרוניות הנבחרות באופן בלתי-תלוי. מסתבר שניתן להכליל הסתכלות זו לקבלת הטענה הבאה.

טענה 8.1 (קיום סדרה אינסופית של משתנים מקריים בלתי-תלויים)**.** יהי X משתנה מקרי. על מרחב ההסתברות האחיד $([0,1],\mathbb{B}([0,1]),\mathbb{P})$ קיימים משתנים מקריים $\{X_i\}_{i\in\mathbb{N}}$, בלתי-תלויים, כולם שווי-התפלגות ל-X.

בנספח הפניה להלן נדון באופן נרחב יותר בהצדקה לטענה זו, אבל הוכחה פורמלית שלה נסמכת על תוצאות בתורת המידה והיא אינה נכללת במסגרת ספר זה. לאור תכונותיה של $\mathbb{B}([0,1])$ כ-σ-אלגברה, נוכל לשייך הסתברות לכל שאלה הנוגעת לסדרת משתנים מקיים בלתי-תלויים $\{X_i\}_{i\in\mathbb{N}}$ אשר ניתן לנסחה במונחים של רצף סופי של פעולות משלים, חיתוך בן-מניה ואיחוד בן-מניה של מאורעות שכל אחד מהם נוגע למשתנה מקרי בודד (ר' הגדרה 7.1). באופן כללי לא נתעכב על נקודה בהמשך הפרק והעיון בשאלה "לאילו מאורעות ניתן לשייך הסתברות?" ישמר לספרים מתקדמים יותר הנסמכים על תורת המידה.

8.1 התכנסות של סדרת מאורעות

בבואנו לדבר על התכנסות של סדרה אינסופית של מאורעות במרחב הסתברות נתעניין בשני מאורעות מרכזיים.

הגדרה 8.2 (התרחשות אסימפטוטית)**.** תהי $(A_n)_{n\in\mathbb{N}}$ סדרת מאורעות במרחב הסתברות. נגדיר את המאורעות הבאים:

(א) $\{A_n \text{ i.o}\} = \{\omega : |\{n : A_n(\omega)\}| = \infty\}$ המאורע לפיו A_n מתרחשים infinitely often, כלומר, **אינסוף פעמים** (או במילים אחרות אינסוף מתוך A_i מתרחשים). מבחינה פורמלית

$$\{A_n \text{ i.o}\} = \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m.$$

(ב) $\{A_n \text{ a.e}\} = \{\omega : |\{n : A_n^c(\omega)\}| < \infty\}$ המאורע לפיו A_n מתרחשים almost everywhere, או eventually כלומר, **החל ממקום מסויים (המ"מ)** או **כמעט בכל מקום (כב"מ)**. מבחינה פורמלית

$$\{A_n \text{ a.e}\} = \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty} A_m.$$

מקובל גם לסמן $\liminf_{i\to\infty} A_n := \{A_n \text{ a.e}\}$, $\limsup_{i\to\infty} A_n := \{A_n \text{ i.o.}\}$.

אבחנה 8.3. לכל סדרת מאורעות A_i במרחב הסתברות מתקיים

$$\{A_i \text{ i.o.}\}^c = \{A_i^c \text{ a.e}\}$$

כדי להיווכח בנכונות האבחנה, נשים לב כי אם A_n המ"מ – בפרט מתקיים כי A_n אינסוף פעמים. כמו כן המאורע A_n^c כמעט בכל מקום הוא משלימו של המאורע A_n אינסוף פעמים, וכי המאורע A_n^c אינסוף פעמים הוא משלימו של המאורע A_n כמעט בכל מקום.

להמחשה נשווה בנפשנו אינסוף הטלות מטבע המתבצעות בזו אחר זו ונסמן ב-A_n את העובדה שבהטלה מסויימת התקבלה תוצאה של עץ. הטענה "כל המטבעות הראו תוצאה של עץ החל ממקום מסויים" (כלומר A_n המ"מ), שקולה לכך שתוצאת פלי התקבלה לכל היותר במספר סופי של מקומות (כלומר לא A_n^c אינסוף פעמים). העובדה שטענה זו לא מתקיימת, כלומר "לא קיים מקום שממנו ואילך כל המטבעות הראו תוצאה של עץ" שקולה לכך שהתקבלה התוצאה פלי באינסוף מקומות (כלומר A_n^c אינסוף פעמים).

לעיתים קרובות נוח לחשוב על A_n כמאורע המתרחשים בזמן n. מנקודת השקפה זו המאורע $\{A_n \text{ a.e}\}$ מכונה גם A_n **אסימפטוטית כמעט תמיד** (asymptotically almost surely).

העובדה שסדרת מאורעות תתקיים החל ממקום מסויים אינה מבטיחה לנו כמה מן המאורעות לא יתרחשו. כך למשל אם ברשותנו אגרטל ואנחנו מטילים בכל יום מטבע הוגן ובתוצאה של עץ מחליטים לשבור את האגרטל, המאורעות "היום האגרטל שבור" יתרחשו המ"מ ואולם האגרטל עשוי לשרוד כל מספר ימים. כך, גם כאשר מובטח כי אינסוף מתוך סדרת מאורעות יתרחשו אין כל ערבון לכך שתדירותם לא תהיה קטנה להדהים.

המתמטיקאי הצרפתי פייר פאטו (Pierre Fatou), בן זמנו של לבג, הראה את החסמים הפשוטים הבאים

להסתברות שסדרת מאורעות תתרחש אינסוף פעמים ולהסתברות שהיא תתרחש כמעט תמיד.

טענה 8.4 (הלמה של פאטו)**.** תהי $\{A_n\}_{n\in\mathbb{N}}$ סדרה של מאורעות. אזי

$$\mathbb{P}(\{A_i \text{ a.e.}\}) = \mathbb{P}(\liminf_{n\to\infty} A_n) \leq \liminf_{n\to\infty} \mathbb{P}(A_n)$$

$$\mathbb{P}(\{A_i \text{ i.o.}\}) = \mathbb{P}(\limsup_{n\to\infty} A_n) \geq \limsup_{n\to\infty} \mathbb{P}(A_n)$$

הוכחה. ראשית נראה כי הטענה השניה נובעת מהפעלת הטענה הראשונה על סדרת המאורעות $\{A_n^c\}$. כלומר -

$$\mathbb{P}(\{A_i \text{ i.o.}\}) \overset{\substack{\text{אבחנה}\\8.3}}{=} 1 - \mathbb{P}(\{A_i^c \text{ a.e.}\}) \overset{\text{חלק א}}{\geq} 1 - \liminf_{n\to\infty} \mathbb{P}(A_n^c) = 1 - \liminf_{n\to\infty}(1 - \mathbb{P}(A_n)) = \limsup_{n\to\infty} \mathbb{P}(A_n).$$

להוכחת הטענה הראשונה, ניזכר ברציפות פונקציית ההסתברות למאורעות יור (משפט 1.21) ,

$$\liminf_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \inf_{i>n} \mathbb{P}(A_i) \geq \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i>n} A_i\right) \overset{\substack{\text{משפט}\\1.21}}{=} \mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \bigcap_{i>n} A_i\right) = \mathbb{P}(\{A_i \text{ a.e.}\})$$

■

המתמטיקאים אמיל בורל (Émile Borel) ופרנצ'סקו קנטלי (Francesco Paolo Cantelli) הבחינו בשנות השלושים של המאה העשרים בקריטריונים הבאים להתרחשות אינסוף מאורעות ולהתרחשותם המ"מ.

משפט 8.5 (הלמה ה-I של בורל-קנטלי)**.** תהי A_i סדרת מאורעות. אם $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$ אז $\mathbb{P}(A_i \text{ i.o.}) = 0$.

הוכחה. נשים לב כי לפי משפט בול,

$$\mathbb{P}(A_i \text{ i.o.}) \overset{\substack{\text{משפט}\\1.21}}{=} \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=n}^{\infty} A_i\right) \overset{\substack{\text{משפט}\\1.17}}{\leq} \lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbb{P}(A_i) = 0$$

כאשר השוויון האחרון נובע מכך ש- $\sum_{i=1}^{\infty} \mathbb{P}(A_i) < \infty$ אם ורק אם $\lim_{n\to\infty} \sum_{i=n}^{\infty} \mathbb{P}(A_i) = 0$. ■

משפט 8.6 (הלמה ה-II של בורל-קנטלי)**.** תהי A_i סדרת מאורעות בלתי-תלויים. אם $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ אז $\mathbb{P}(A_i \text{ i.o.}) = 1$.

הוכחה. נשים לב כי

$$\mathbb{P}(A_i \text{ i.o.}) = 1 - \mathbb{P}(A_i^c \text{ a.e.}) = 1 - \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{i=m}^{\infty} A_i^c\right) \overset{\substack{\text{משפט}\\1.21}}{=} 1 - \lim_{m\to\infty} \mathbb{P}\left(\bigcap_{i=m}^{\infty} A_i^c\right)$$

על כן, די שנראה כי לכל $m \in \mathbb{N}$ מתקיים:

$$\mathbb{P}\left(\bigcap_{i=m}^{\infty} A_i^c\right) = 0.$$

נחשב ונקבל באמצעות אי תלות

$$\mathbb{P}\left(\bigcap_{i=m}^{\infty} A_i^c\right) \overset{\substack{\text{משפט}\\1.21}}{=} \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=m}^{n} A_i^c\right) = \lim_{n\to\infty} \prod_{i=m}^{n}(1-\mathbb{P}(A_i)) \le \lim_{n\to\infty} \exp\left(-\sum_{i=m}^{n} \mathbb{P}(A_i)\right) = 0$$

כאשר האי-שיוויון נובע מכך ש-$1+x \le e^x$ לכל x והשוויון האחרון משתמש בהנחה $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$, ממנה נובע כי $\sum_{i=m}^{\infty} \mathbb{P}(A_i) = \infty$ לכל m. ■

נשים לב שבמקרה של מאורעות בלתי תלויים, שני חלקי הלמה משלימים זה את זה, ובפרט מתקיים שלכל סדרת מאורעות בלתי-תלויים $\mathbb{P}(A_i \text{ i.o}) \in \{0,1\}$. לעומת זאת לא ניתן להכליל את הלמה השניה למאורעות תלויים. כך למשל מפני שאם נטיל מטבע הוגן ונקבע לכל i את A_i להיות המאורע שהתקבלה תוצאה של ראש, אזי נקבל

$$\mathbb{P}(A_i \text{ i.o.}) = \mathbb{P}(A_i \text{ a.e.}) = \frac{1}{2}.$$

דוגמא 8.7. יהיו $X_n \sim \text{Geo}\left(\frac{e-1}{e}\right)$ בלתי-תלויים. נסמן $A_n = \{X_n > \alpha \log n\}$ יש להראות כי $\mathbb{P}(A_n \text{ i.o.}) = 1$ אם ורק אם $\alpha \le 1$.

תשובה: נחשב לפי פונקציית ההתפלגות השיורית של משתנה מקרי גיאומטרי (טענה 3.42).

$$\mathbb{P}(A_n) = \mathbb{P}(X_n > \alpha \log n) = \left(1 - \frac{e-1}{e}\right)^{\lfloor \alpha \log n \rfloor} = e^{-\lfloor \alpha \log n \rfloor}$$

נסיק כי

$$n^{-\alpha} \le e^{-\lfloor \alpha \log n \rfloor} \le \mathbb{P}(A_n) \le e^{-\lfloor \alpha \log n \rfloor + 1} \le e n^{-\alpha}$$

לכן מתקיים

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty \Leftrightarrow \alpha \le 1.$$

מכאן שלפי הלמה הראשונה של בורל-קנטלי (משפט 8.5) לכל $\alpha > 1$ מתקיים $\mathbb{P}(A_n \text{ i.o.}) = 0$ ולפי הלמה השניה של בורל-קנטלי (משפט 8.6) לכל $\alpha \le 1$ מתקיים $\mathbb{P}(A_n \text{ i.o.}) = 1$ כנדרש.

גם כאשר המאורעות אינם בלתי-תלויים באופן מוחלט - נוכל לעיתים להשתמש בלמות של בורל-קנטלי כדי להגיע לאפיון מלא של מרחב הפרמטרים שעליו מתקיימים אינסוף מאורעות מתוך אוסף מסויים. נראה זו בדוגמא הבאה.

דוגמא 8.8. יהיו $\{A_n\}_{n\in\mathbb{N}}$ מאורעות בלתי תלויים בעלי הסתברויות $\mathbb{P}(A_n) = 1/n^{\alpha}$. נגדיר $B_n = A_n \cap A_{n+1}$. עבור אילו α מתקיים $\mathbb{P}(B_n \text{ i.o.}) = 1$?

תשובה:

$$\mathbb{P}(B_n) = (n(n+1))^{-\alpha} \approx n^{-2\alpha}$$

מכאן שלפי הלמה הראשונה של בורל-קנטלי (משפט 8.5) לכל $\alpha > 1/2$ מתקיים $\mathbb{P}(B_n \text{ i.o.}) = 0$.

כאשר $\alpha \le 1/2$ הטור $\sum_{n=1}^{\infty} \mathbb{P}(B_n)$ מתבדר, אולם המאורעות $\{B_n\}_{n\in\mathbb{N}}$ אינם בלתי תלויים ולכן לא ניתן להשתמש בלמה השניה של בורל-קנטלי (משפט 8.6) ישירות. ואולם, נשים לב שהמאורעות $\{B_{2n}\}_{n\in\mathbb{N}}$ הינם בלתי תלויים והטור $\sum_{n=1}^{\infty} \mathbb{P}(B_{2n})$ מתבדר ולפי המה השניה של בורל-קנטלי (משפט 8.6) $\mathbb{P}(B_2 n \text{ i.o.}) = 1$ ולפיכך גם $\mathbb{P}(B_n \text{ i.o.}) = 1$ במקרה זה.

נסיים בדוגמא מורכבת מעט יותר.

דוגמא 8.9 (הילוך שיכור חולף או נשנה)**.** יהיו $\{B_i\}_{i\in\mathbb{N}}$ משתנים מקריים בלתי-תלויים המקבלים את הערך 1 בסיכוי p ואת הערך -1 בסיכוי $1-p$ ונגדיר $X_n = \sum_{i=0}^n B_i$, כאשר אנו משווים בנפשנו את X_n כשיכור המטייל על הציר המספרים ובוחר בזמן n באקראי אם ללכת צעד שלם אחד ימינה או שמאלה. תהליך זה מכונה הילוך שיכור (מוטה). כאשר $p < \frac{1}{2}$ ההילוך מוטה שמאלה, כאשר $p > \frac{1}{2}$ הוא מוטה ימינה וכאשר $p = \frac{1}{2}$ ההילוך מאוזן. נסמן $A_n = \{X_n = 0\}$. אם A_n מתרחשים אינסוף פעמים בהסתברות אחת נאמר שההילוך **נשנה** ואם A_n^c החל ממקום מסויים – נאמר שהוא **חולף**. חקור האם ההילוך חולף, נשנה או לא חולף ולא נשנה כתלות בפרמטר p.

תשובה: עבור $p < \frac{1}{2}$ נתקנן את המשתנים המקריים B_i כך שיתקיימו $\frac{(B_i+1-2p)}{2}$ מקיימים את תנאי אי-שיוויון הופדינג (משפט 6.10), מפני ש-

$$\mathbb{E}(B_i + 1 - 2p) = 1 \cdot p + (-1) \cdot (1-p) + 1 - 2p = 0,$$

וכן $|B_i + 1 - 2p| < 2$. נסמן אפוא

$$Y_n = \frac{\sum_{i=1}^n (B_i + 1 - 2p)}{2} = \frac{X_n}{2} + \frac{n(1-2p)}{2}$$

. לפי אי-שיוויון הופדינג, נקבל

$$\mathbb{P}(A_n) = \mathbb{P}(X_n = 0) \le \mathbb{P}\left(\frac{X_n}{2} \ge 0\right) = \mathbb{P}\left(Y_n \ge \frac{n(1-2p)}{2}\right) \overset{\text{הופדינג}}{\le} \exp\left(\frac{-(1-2p)^2 n}{8}\right)$$

ולכן $\sum_{n=1}^\infty \mathbb{P}(A_n) < \infty$ ולפי הלמה הראשונה של בורל קטנלי (משפט 8.5) נקבל $\mathbb{P}(A_n \text{ i.o.}) = 0$.

עבור $p > \frac{1}{2}$ נשים לב כי $X_i \overset{d}{=} -X_i'$, כאשר X_i' הוא הילוך שיכור מוטה עם פרמטר $q = 1-p$ ומכיוון שהראינו ש-X_i' כזה חולף - הרי שגם X_i חולף.

המקרה $p = \frac{1}{2}$ מורכב יותר, ולא נוכל לטפל בו על ידי שימוש פשוט בלֶמות של בורל-קנטלי. ראשית, נשים לב כי $\sum_{i=1}^n \frac{B_i+1}{2} \sim \mathrm{Bin}(n, 1/2)$, ולכן אם $Y \sim \mathrm{Bin}(n,1/2)$, הרי ש-

$$\mathbb{P}(A_{2n}) = \mathbb{P}(X_{2n} = 0) = \mathbb{P}\left(Y = \frac{n}{2}\right) \overset{\substack{\text{דוגמא}\\ 3.48}}{\ge} \frac{0.4}{\sqrt{n}}$$

כעת נסמן עבור $n \in \mathbb{N}_0$ משתני אינדיקטור לביקור ב-0 בזמן n, על ידי $W_n = \mathbb{1}(A_n)$ ונגדיר $W = \sum_{n=1}^\infty W_n$. נקבל

$$\mathbb{E}(W) \ge \sum_{n=1}^\infty \mathbb{P}(A_{2n}) \ge \sum_{n=1}^\infty \frac{0.4}{\sqrt{n}} = \infty.$$

כעת נראה כי מעובדה זו נובע שההילוך בהכרח נשנה. לצורך כך נציג כמה הגדרות. עבור $a = (a_1, \dots, a_\ell)$ סדרה סופית של $+1$ ו--1, נסמן ב-$\ell(a)$ את אורך הסדרה ונגדיר את A_a להיות המאורע $X_i = a_i$ לכל $1 \le i \le \ell(a)$.

קל לראות כי עבור שתי סדרות סופיות a ו-b, אם a היא רישא של b אז המאורע A_a מכיל את A_b ואילו המאורעות זרים אם ורק אם אף אחת מהסדרות אינה רישא של השניה. עבור S קבוצה של סדרות סופיות

נגדיר את A_S להיות $\cup_{a\in S} A_a$, כלומר המאורע שסדרת ההטלות מתחילה באחת הרישות המוכלות ב-S. נסמן את השרשור של a ו-b שתי סדרות סופיות באמצעות ab. ועבור שתי קבוצות של סדרות סופיות S ו-T נגדיר $ST = \{ab : a \in S, b \in T\}$, הלא היא אוסף כל השרשורים האפשריים של סדרה מ-T לסדרה מ-S. כמו כן ננהיג את הסימון $S^2 = SS$. נאמר כי קבוצת סדרות סופיות S היא **חד-משמעית** אם אף סדרה $a \in S$ בקבוצה אינה רישא של סדרה אחרת $b \in \S$ בקבוצה (כלומר לא קיימת סדרה c כך ש-$b = ac$). המאורעות בקבוצה מובהקת הם זרים ולכן לכל קבוצה חד-משמעית S מתקיים

$$\mathbb{P}(S) = \sum_{a\in S} 2^{-\ell(a)}.$$

נסמן ב-S_{ret} את קבוצה כל הסדרות הסופיות a כך שמתקיים $\sum_{i=1}^{\ell(a)} a_i = 0$ ואילו לכל $1 \leq k < \ell(a)$ מתקיים $\sum_{i=1}^{k} a_i \neq 0$. נשים לב שבקבוצה זו אף סדרה אינה רישא של סדרה אחרת ולכן היא חד-משמעית, ונסמן

$$p_{\text{ret}} := \mathbb{P}(A_{S_{\text{ret}}}) = \sum_{a\in S_{\text{ret}}} \mathbb{P}(A_a) = \sum_{a\in S_{\text{ret}}} 2^{-\ell(a)}.$$

נשים לב כי המאורע $A_{S_{\text{ret}}}$ הוא בדיוק המאורע $W \geq 1$, כלומר שהשיכור חזר לפחות פעם אחת לאפס. באופן דומה נשים לב שגם הקבוצות S^k_{ret} לכל $k \in \mathbb{N}$ הן חד-משמעיות מפני שלכל $a \in S^k_{\text{ret}}$ בדיוק k קואורדינטות ℓ_j המקיימות $\sum_{i=1}^{\ell_j} a_i = 0$. נסיק אפוא כי

$$\mathbb{P}(A_{S^k_{\text{ret}}}) = \sum_{a\in S^k_{\text{ret}}} \mathbb{P}(A_a) = \sum_{a_1,a_2,\ldots,a_k\in S_{\text{ret}}} 2^{\ell(a_1)+\cdots+\ell(a_k)} = \prod_{j\in[k]} \sum_{a_j} 2^{\ell(a_j)} = (p_{\text{ret}})^k.$$

כאשר $\mathbb{P}(A_{S^k_{\text{ret}}})$ הוא המאורע שהשיכור חזר לפחות k פעמים ל-0. נשים לב שמתקיים

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(W = \infty) = \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} \{W \geq n\}\right) \overset{\substack{\text{משפט}\\1.21}}{=} \liminf_{n\to\infty} \mathbb{P}(W \geq n) = \liminf_{n\to\infty} (p_{\text{ret}})^k$$

מצד שני

$$\mathbb{E}(W) \overset{\substack{\text{טענה}\\4.7}}{=} \sum_{n\in\mathbb{N}} \mathbb{P}(W \geq n) = \sum_{n\in\mathbb{N}} \frac{1}{(p_{\text{ret}})^n}$$

ותוחלתו זו סופי אם ורק אם $p_{\text{ret}} < 1$. הואיל וחישבנו כבר ש-$\mathbb{E}(W) = \infty$ נסיק כי $p_{\text{ret}} = 1$ ולכן

$$\mathbb{P}(A_n \text{ i.o.}) = 1$$

וההילוך נשנה.

8.1.1 הכללת הלמה השניה של בורל קנטלי

למעשה ניתן להכליל את הלמה השניה של בורל-קנטלי, למאורעות בלתי-תלויים בזוגות (ר' הגדרה 2.26).

טענה 8.10 (הלמה ה-II של בורל-קנטלי למשתנים בלתי-תלויים בזוגות). תהי A_i סדרת מאורעות. אם $\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$ והמאורעות בלתי-תלויים בזוגות אז $\mathbb{P}(A_i \text{ i.o.}) = 1$.

הוכחה. כמקודם, נראה כי לכל m

$$\mathbb{P}\left(\bigcap_{i=m}^{\infty} A_i^c\right) = 0.$$

נגדיר $X_m^n = \sum_{i=m}^n \mathbb{1}(A_i)$ ונשים לב כי

$$\mathbb{P}\left(\bigcap_{i=m}^{\infty} A_i^c\right) \overset{\substack{\text{משפט}\\ 1.21}}{=} \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{i=m}^{n} A_i^c\right) = \lim_{n\to\infty} \mathbb{P}(X_m^n = 0).$$

נסמן $p_i = \mathbb{P}(A_i)$ ונחשב את תוחלת ושונות X_m^n

$$\mathbb{E}(X_m^n) = \sum_{i=m}^n p_i,$$

$$\operatorname{Var}(X_m^n) = \operatorname{Var}\left(\sum_{i=m}^n \mathbb{1}(A_i)\right) \overset{\substack{\text{טענה}\\ 5.16}}{=} \sum_{i=m}^n \operatorname{Var}(\mathbb{1}(A_i)) = \sum_{i=m}^n p_i(1-p_i) \le \sum_{i=m}^n p_i = \mathbb{E}(X_m^n).$$

כעת נשתמש באי-שיוויון צ'בישב,

$$\mathbb{P}\left(\bigcap_{i=m}^{n} A_i^c\right) = \mathbb{P}(X_m^n = 0) \le \mathbb{P}(|X_m^n - \mathbb{E}(X_m^n)| \ge \mathbb{E}(X_m^n)) \overset{\substack{\text{משפט}\\ 5.5}}{\le} \frac{\operatorname{Var}(X_m^n)}{\mathbb{E}(X_m^n)^2} \le \frac{1}{\mathbb{E}(X_m^n)}.$$

נשים לב כי, היות ש-$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$, הרי שמתקיים $\lim_{n\to\infty} \mathbb{E}(X_m^n) = \sum_{i=m}^{\infty} \mathbb{P}(A_i) = \infty$ ומכאן נסיק כי

$$\lim_{n\to\infty} \mathbb{P}(X_m^n = 0) \le \frac{1}{\lim_{n\to\infty} \mathbb{E}(X_m^n)} = 0,$$

כנדרש. ■

8.2 התכנסות של סדרות משתנים מקריים

נציג כעת שני מובנים של התכנסות הנוגעים לסדרות משתנים מקריים המצויים באותו מרחב הסתברות. כדי להבין את ההבדלים בין שתי צורות התכנסות אלו, נלווה את הצגתן בדוגמא מכוננת.

נדמיין שחובב מפות עולם בוחר באקראי מפה מתוך רשימה של מאה מפות הסטוריות ומחליט להתמחות בהעתקתה. לאחר מכן הוא בודק את יצירתו על ידי כך שהוא מסתכל מה אחוז העולם במפה המכוסה במים. נסמן ב-X_n את שטח הים היחסי במפה ה-n-ית וב-X את שטח הים היחסי במפה המקורית. בשנים הראשונות הדגמים שונים מאוד מהמקור, אך ככל שהמעתיק מתמחה X_n נוטה יותר ויותר להתקרב ל-X ואף מזדהה עמו לעתים קרובות.

נאמר ש-X_n מתכנס כמעט תמיד (או כמעט בוודאות) ל-X אם בהסתברות אחת, לאחר אימון מספק, יהיה שטח הים היחסי בכל המפות המועתקות החל מהעתקה מסויימת ואילך, קרוב כרצוננו לשטח הים היחסי במפה המקורית. התכנסות זו נתונה בהגדרה הבאה.

הגדרה 8.11 (התכנסות כמעט תמיד)**.** תהי $(X_n)_{n\in\mathbb{N}}$ סדרת משתנים מקריים במרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נאמר כי סדרה זו מתכנסת למשתנה המקרי X ***כמעט-תמיד***, ונסמן $X_n \overset{\text{a.s.}}{\to} X$, אם מתקיים

$$\mathbb{P}\left(\lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1.$$

נאמר ש-X_n מתכנס בהסתברות ל-X אם הסיכוי לטעות בכל גודל סופי שטח הים היחסי במפה ה-n שואף לאפס כאשר n שואף לאינסוף. דוגמא למקרה כזה היא כאשר שכיחות הטעויות של המעתיק יורד והולך, אך בכל זאת מפעם בפעם הוא ממשיך ומבצע טעויות העתקה.

הגדרה 8.12 (התכנסות בהסתברות)**.** תהי $(X_n)_{n\in\mathbb{N}}$ סדרת משתנים מקריים במרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. נאמר כי סדרה זו מתכנסת למשתנה המקרי X **בהסתברות**, ונסמן $X_n \xrightarrow{\text{p}} X$, אם לכל $\epsilon > 0$ מתקיים

$$\lim_{n\to\infty} \mathbb{P}(|X_n(\omega) - X(\omega)| \le \epsilon) = 1.$$

בהתקלות ראשונה, שתי ההגדרות נדמות דומות להפליא. כדי לעמוד על ההבדלים שביניהם, נבהיר את האתגר הטכני בהוכחת כל אחד מסוגי ההתכנסות במושגים של חשבון אינפיניטיסימלי.

טענה 8.13. תהי $(X_n)_{n\in\mathbb{N}}$ סדרת משתנים מקריים במרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. אזי

$$X_n \xrightarrow{\text{p}} X \iff \forall\epsilon > 0 \ : \ \limsup_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

$$X_n \xrightarrow{\text{a.s.}} X \iff \forall\epsilon > 0 \ : \ \mathbb{P}(\limsup_{n\to\infty} |X_n - X| > \epsilon) = 0 \iff \forall\epsilon > 0 \ : \ \mathbb{P}(|X_n - X| \le \epsilon \text{ a.e. }) = 1$$

הוכחה. להוכחת השקילות הראשונה, נשים לב כי, לפי הגדרה, $X_n \xrightarrow{\text{p}} X$ שקול לכך ש-

$$\lim_{n\to\infty} \mathbb{P}(|X_n(\omega) - X(\omega)| > \epsilon) = 0,$$

לכל $\epsilon > 0$. כעת מכיוון ש $\mathbb{P}(|X_n(\omega) - X(\omega)| > \epsilon) \ge 0$ טענה זו שקולה לכך ש-

$$\limsup_{n\to\infty} \mathbb{P}(|X_n(\omega) - X(\omega)| > \epsilon) = 0,$$

כנדרש.

להוכחת השקילות השניה, נשים לב כי, לפי הגדרת הגבול, התכנסות כמעט תמיד, $X_n \xrightarrow{\text{a.s.}} X$, מתקיימת אם ורק אם

$$\mathbb{P}\left(\forall\epsilon > 0 \ \exists n_0 \in \mathbb{N} \ \forall n > n_0 \ : \ |X_n - X| \le \epsilon\right) = 1,$$

נסתכל על המאורע המשלים ונקבל שטענו זו שקולה לכך ש-

$$\mathbb{P}\left(\exists k \in \mathbb{N} \forall n_0 \in \mathbb{N} \ \exists n > n_0 \ : \ |X_n - X| > k^{-1}\right) = 0,$$

נשים לב כי

$$\mathbb{P}\left(\exists k \in \mathbb{N} \forall n_0 \in \mathbb{N} \ \exists n > n_0 \ : \ |X_n - X| > k^{-1}\right) \le \sum_{k\in\mathbb{N}} \mathbb{P}\left(\forall n_0 \in \mathbb{N} \ \exists n > n_0 \ : \ |X_n - X| > k^{-1}\right),$$

וכי אגף שמאל שווה ל-0 אם ורק אם כל אברי הסכום מתאפסים. והרי

$$\forall_{k\in\mathbb{N}} \mathbb{P}\left(\forall n_0 \in \mathbb{N} \ \exists n > n_0 \ : \ |X_n - X| > k^{-1}\right) \iff \forall_{\epsilon>0} \mathbb{P}\left(\forall n_0 \in \mathbb{N} \ \exists n > n_0 \ : \ |X_n - X| > \epsilon\right).$$

קיבלנו ש-$X_n \xrightarrow{\text{a.s.}} X$ אם ורק אם,

$$\forall\epsilon > 0 \;:\; \mathbb{P}(\limsup_{n\to\infty} |X_n - X| > \epsilon) = 0,$$

כנדרש. ■

טענה 8.14 (התכנסות כמעט תמיד גוררת התכנסות בהסתברות)**.** אם $X_n \xrightarrow{\text{a.s.}} X$ אז $X_n \xrightarrow{\text{p}} X$.

הוכחה. נסמן $A_n^k = \{\omega : |X_n(\omega) - X(\omega)| \leq 1/k\}$. כעת נוכל לכתוב את המאורע $\{X_n(\omega) \to X(\omega)\}$ בתור

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

כלומר התנאי $X_n \xrightarrow{\text{a.s.}} X$ שקול לכך שלכל k מתקיים

$$\mathbb{P}\left(\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} A_n^k\right) = \mathbb{P}\left(\liminf_{n\to\infty} A_n^k\right) = 1.$$

לעומת זאת, התנאי $X_n \xrightarrow{\text{p}} X$ שקול לכך שלכל k מתקיים

$$\lim_{n\to\infty} \mathbb{P}(A_n^k) = 1.$$

ואולם לפי הלמה של פאטו

$$\mathbb{P}\left(\liminf_{n\to\infty} A_n^k\right) \leq \liminf_{n\to\infty} \mathbb{P}\left(A_n^k\right)$$

ולכן אם צד שמאל שווה ל-1 אז גם צד ימין. ■

הדוגמאות הבאות ממחישות כי התכנסות בהסתברות אינה גוררת התכנסות כמעט-תמיד.

דוגמא 8.15 (התכנסות כמעט תמיד לעומת התכנסות בהסתברות)**.** יהיו $\{A_n\}_{n\in\mathbb{N}}$ מאורעות בלתי תלויים בעלי הסתברויות $p_n = \mathbb{P}(A_n)$ ונגדיר $X_n = \mathbb{1}(A_n)$. מתי $X_n \xrightarrow{\text{p}} 0$ ומתי $X_n \xrightarrow{\text{a.s.}} 0$?

תשובה: אם $p_n \to 0$ אז לכל $\epsilon > 0$ מתקיים $\mathbb{P}(|X_n - 0| > \epsilon) = \mathbb{P}(X_n = 1) = p_n \to 0$ ולכן $X_n \xrightarrow{\text{p}} 0$ במקרה זה.

לעומת זאת, הסדרה $X_n(\omega)$ מקבלת ערכים שהם אפס או אחד בלבד ולכן היא מתכנסת לאפס אם ורק אם היא שווה לאפס החל ממקום מסוים, או במילים אחרות, בדיוק כאשר מתקיים $\{A_n^c \text{ a.e.}\}$. זה שקול למאורע $\{A_n \text{ i.o.}\}^c$ ולפי הלמות של בורל-קנטלי מאורע זה קורה בהסתברות 1 אם ורק אם $\sum_{n\in\mathbb{N}} p_n < \infty$.

דוגמא 8.16. יהי $Y \sim \text{Unif}([0,1])$ ויהיו $X_{m,n} = \mathbb{1}(Y \in [\frac{m-1}{2^n}, \frac{m}{2^n}])$ עבור $m, n \in \mathbb{N}$ המקיימים $m \leq 2^n$. נראה כי

$$X_{1,n} \xrightarrow{\text{a.s.}} 0, \qquad X_{1,n} \xrightarrow{\text{p}} 0, \qquad X_{n,\lceil \log_2 n\rceil} \xrightarrow{\text{p}} 0, \qquad X_{n,\lceil \log_2 n\rceil} \overset{\text{a.s.}}{\not\to} 0.$$

תשובה: נראה כי $X_{1,n} \xrightarrow{\text{a.s.}} 0$. לפי טענה 8.13, עלינו להראות כי לכל $\epsilon > 0$ מתקיים

$$\mathbb{P}\left(\forall n_0 \in \mathbb{N}\ \exists n > n_0\ :\ |X_{1,n}| > \epsilon\right) = 0.$$

נשים לב כי

$$\mathbb{P}\left(\exists n > n_0\ :\ |X_n| > \epsilon\right) \le \sum_{n=n_0+1}^{\infty} \mathbb{P}\left(|X_n| > \epsilon\right) = \sum_{n=n_0+1}^{\infty} \mathbb{P}\left(Y \in \left[0, \frac{1}{2^n}\right]\right) = \sum_{n=n_0+1}^{\infty} 2^{-n} = 2^{-n_0}.$$

ולכן

$$\mathbb{P}\left(\forall n_0 \in \mathbb{N}\ \exists n > n_0\ :\ |X_{1,n}| > \epsilon\right) \overset{\substack{\text{משפט}\\1.21}}{=} \lim_{n_0\to\infty} \mathbb{P}\left(\exists n > n_0\ :\ |X_{1,n}| > \epsilon\right) \le \lim_{n_0\to\infty}\left(2^{-n_0}\right) = 0,$$

כנדרש.

נראה כי $X_{1,n} \xrightarrow{\text{p}} 0$. עלינו להראות כי לכל $\epsilon, \eta > 0$, נוכל לבחור $n_0 \in \mathbb{N}$ כך שלכל $n > n_0$ יתקיים $\mathbb{P}\left(|X_{1,n}| > \epsilon\right) < \eta$. נחשב

$$\mathbb{P}\left(|X_{1,n}| > \epsilon\right) = \mathbb{P}\left(Y \in \left[\frac{n-1}{2^n}, \frac{n}{2^n}\right]\right) = 2^{-n}$$

ולכן עבור $n > -\log_2(\eta)$, נקבל $\mathbb{P}\left(|X_{1,n}| > \epsilon\right) < \eta$ כנדרש.

נראה כי $X_{n,\lceil\log n\rceil} \xrightarrow{\text{p}} 0$. כמקודם ניווכח כי $X_{n,\lceil\log n\rceil} \xrightarrow{\text{p}} 0$. נחשב כמקודם

$$\mathbb{P}\left(|X_{n,\lceil\log n\rceil}| > \epsilon\right) = \mathbb{P}\left(Y \in \left[\frac{n-1}{2^{\lceil\log n\rceil}}, \frac{n}{2^{\lceil\log n\rceil}}\right]\right) = 2^{-\lceil\log n\rceil} \le \frac{1}{n}$$

ולכן עבור $n > \frac{1}{\eta}$, נקבל $\mathbb{P}\left(|X_{n,\lceil\log n\rceil}| > \epsilon\right) < \eta$ כנדרש.

לסיום, **נראה כי** $X_{n,\lceil\log n\rceil} \overset{\text{a.s.}}{\not\to} 0$. לצורך כך עלינו להראות כי

$$\mathbb{P}\left(\forall n_0 \in \mathbb{N}\ \exists n > n_0\ :\ |X_{n,\lceil\log n\rceil}| > \epsilon\right) > 0.$$

נסמן $E_{n_0} = \{\exists n > n_0\ :\ |X_{n,\lceil\log n\rceil}| > \epsilon\}$ ו- $A_{n,m} = \{Y \in \left[\frac{m-1}{2^{\lceil\log n\rceil+1}}, \frac{m}{2^{\lceil\log n\rceil+1}}\right]\}$. יהי $n_0 \in \mathbb{N}$ נשים לב כי

$$\mathbb{P}\left(E_{n_0}\right) = \sum_{m=1}^{2^{\lceil\log n\rceil+1}} \mathbb{P}\left(A_{n,m}\right)\mathbb{P}\left(E_{n_0}\,\middle|\,A_{n,m}\right) \ge \sum_{m=2^{\lceil\log n\rceil}+1}^{2^{\lceil\log n\rceil+1}} \mathbb{P}\left(A_{n,m}\right)\mathbb{P}\left(E_{n_0}\,\middle|\,A_{n,m}\right)$$

והרי עבור $m \in [2^n+1, 2^{n+1}]$ מתקיים

$$\mathbb{P}\left(E_{n_0}\,\middle|\,A_{n,m}\right) \ge \mathbb{P}\left(\left|X_{m,\lceil\log m\rceil}\right| = 1\,\middle|\,A_{n,m}\right) = \mathbb{P}\left(\left|X_{m,\lceil\log n\rceil+1}\right| = 1\,\middle|\,A_{n,m}\right) = 1$$

ולכן

$$\mathbb{P}\left(\mathbb{E}_{n_0}\right) \ge \sum_{m=2^{\lceil\log n\rceil}+1}^{2^{\lceil\log n\rceil+1}} \mathbb{P}\left(Y \in \left[\frac{m-1}{2^{n+1}}, \frac{m}{2^{n+1}}\right]\right) = \frac{1}{2},$$

ולכן $X_{n,\lceil\log n\rceil} \overset{\text{a.s.}}{\not\to} 0$, כנדרש.

את מושגי ההתכנסות שהגדרנו נוכל לנסח גם באמצעות **קריטריון קושי** להתכנסות באופן שלא יחייב אותנו לדעת מראש מהו גבול הסדרה. כך :

סדרת משתנים מקריים $\{X_n\}_{n\in\mathbb{N}}$ תתכנס בהסתברות אם לכל $\epsilon, \eta > 0$ קיים n_0 כך שלכל $n, m > n_0$ מתקיים $\mathbb{P}(|X_n - X_m| > \epsilon) < \eta$.

סדרת משתנים מקריים $\{X_n\}_{n\in\mathbb{N}}$ תתכנס כמעט תמיד אם לכל $\epsilon, \eta > 0$ קיים n_0 כך שניתן לחסום מלמעלה על ידי η את הסתברותו של המאורע "קיימים $n, m > n_0$ המקיימים $|X_n - X_m| > \epsilon$".

8.3 חוקי המספרים הגדולים

את החוק החלש של המספרים הגדולים (משפט 5.8) נוכל לנסח מחדש במונחים של התכנסות בהסתברות.

משפט 8.17 (החוק החלש של המספרים הגדולים)**.** תהי $\{X_n\}_{n\in\mathbb{N}}$ סדרת משתנים מקריים בלתי-תלויים ושווי התפלגות, שתוחלתם μ. אזי

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{p}} \mu$$

כאשר אנו מזהים את μ עם המשתנה המקרי המקבל את הערך μ בהסתברות 1.

כזכור, הוכחנו את החוק תחת התנאי הנוסף שלמשתנים המקריים שונות סופית בטענה 5.8 ולמשתנים בעלי תוחלת בטענה 5.32. כעת נוכל לתאר גם את החוק החזק של המספרים הגדולים. להלן נביא את טענת החוק תחת ההנחה המגבילה שהמשתנים המקריים חסומים, ובפרק 8.4 נציג את החוק בגרסתו הכללית.

משפט 8.18 (החוק החזק של המספרים הגדולים)**.** יהי $M > 0$ תהי $\{X_n\}_{n\in\mathbb{N}}$ סדרת משתנים מקריים בלתי-תלויים ושווי התפלגות כך ש-$|X_i| \overset{\text{a.s.}}{\leq} M$ ו-$\mathbb{E}(X_i) = \mu$ עבור $\mu \in \mathbb{R}$. אזי

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu$$

כאשר אנו מזהים את μ עם המשתנה המקרי המקבל את הערך μ בהסתברות 1.

הוכחה. נגדיר משתנים מקריים חדשים

$$Y_n = \frac{X_n - \mu}{2M}$$

ונשים לב שהם מקיימים את תנאי אי-שיוויון הופדינג (משפט 6.9). לכן, לכל n ולכל $a > 0$ נקבל

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i\right| \geq a\right) \leq 2\exp\left(-\frac{a^2}{2n}\right).$$

עבור $\epsilon > 0$ כלשהו, אם נציב $a = \epsilon n$ נקבל

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2}n\right).$$

נסמן

$$A_n^k = \left\{\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| \geq 1/k\right\}.$$

נשים לב שלכל $k \in \mathbb{N}$ מתקיים $\sum_{i=1}^{n} \mathbb{P}(A_n^k) < \infty$, ולכן לפי הלמה הראשונה של בורל-קנטלי (משפט 8.5), נקבל

$$\mathbb{P}\left(A_n^k \text{ i.o. in } n\right) = 1 - \mathbb{P}\left((A_n^k)^c \text{ a.e. } n\right) = 0$$

ניקח איחוד על כל ערכי k ונקבל .

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_n^k \text{ i.o. in } n\right) = 0$$

או, באופן שקול

$$\mathbb{P}\left(\bigcap_{k=1}^{\infty} (A_n^k)^c \text{ a.e. } n\right) = 1$$

וזו ההגדרה של $\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{\text{a.s.}} 0$ וכמובן ש-$\frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{\text{a.s.}} 0$ אם ורק אם $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu$. ■

8.4* החוק החזק של המספרים הגדולים

בפרק זה נוכיח את גרסתו המלאה של החוק החזק של המספרים הגדולים.

משפט 8.19 (החוק החזק של המספרים הגדולים)**.** יהי 0 תהי $\{X_n\}_{n\in\mathbb{N}}$ סדרת משתנים מקריים בלתי-תלויים ושווי התפלגות. אזי $\mathbb{E}(X_i) = \mu$ עבור $\mu \in \mathbb{R}$ אם ורק אם

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{a.s.}} \mu.$$

לצורך הוכחת המשפט ניעזר בהגדרה הבאה ובשתי טענות הנוגעות אליה.

הגדרה 8.20 (זהות המ"מ)**.** שתי סדרות מ"מ $\{X_n\}_{n\in\mathbb{N}}$ ו-$\{Y_n\}_{n\in\mathbb{N}}$ נקראות זהות המ"מ (Tail Equivalent) אם

$$\mathbb{P}(X_n = Y_n \text{ a.e.}) = 1.$$

טענה 8.21. אם $\{X_n\}_{n\in\mathbb{N}}$ ו-$\{Y_n\}_{n\in\mathbb{N}}$ זהות המ"מ אז לכל סדרה $a_n \to \infty$ ולכל משתנה מקרי Z מתקיים

$$\lim_{n\to\infty}\sum_{m=1}^{n}\frac{X_m}{a_n} \xrightarrow{\text{a.s.}} Z \quad\iff\quad \lim_{n\to\infty}\sum_{m=1}^{n}\frac{Y_m}{a_n} \xrightarrow{\text{a.s.}} Z.$$

הוכחה. לאור הסימטריה בין שני צידי השקילות, די שנניח כי

$$\lim_{n\to\infty}\sum_{m=1}^{n}\frac{X_m}{a_n} \xrightarrow{\text{a.s.}} Z$$

ונוכיח כי

$$\lim_{n\to\infty}\sum_{m=1}^{n}\frac{Y_m}{a_n} \xrightarrow{\text{a.s.}} Z.$$

יהי $n_0 = \min\{n \;:\; \forall m > n\; X_m = Y_m\}$. נרשום

$$\lim_{n\to\infty}\sum_{m=1}^{n}\frac{Y_m}{a_n} \overset{\text{a.s.}}{=} \lim_{n\to\infty}\sum_{m=1}^{n_0-1}\frac{Y_n}{a_n} + \lim_{n\to\infty}\sum_{m=n_0}^{\infty}\frac{X_m}{a_n} = 0 + \lim_{n\to\infty}\sum_{m=n_0}^{\infty}\frac{X_m}{a_n} = \lim_{n\to\infty}\sum_{m=1}^{n}\frac{X_m}{a_n} \overset{\text{a.s.}}{=} Z,$$

כנדרש. ■

טענה 8.22. תהי $\{X_n\}_{n\in\mathbb{N}}$ סדרת משתנים מקריים בלתי-תלויים ותהי $\{x_n\}_{n\in\mathbb{N}}$ סדרה של מספרים ממשיים. אזי מתקיים $\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>x_n)<\infty$ אם ורק אם $\{X_n\}_{n\in\mathbb{N}}$ ו- $\{X_n\cdot \mathbb{1}_{X_n\in[-x_n,x_n]}\}_{n\in\mathbb{N}}$ זהות המ"מ.

הוכחה. המאורעות $A_n=\{|X_n|>x_n\}$ הנם בלתי-תלויים, ולכן לפי שילוב שתי הלמות של בורל-קנטלי (משפטים 8.5 ו-8.6), $\mathbb{P}(\{A_n \text{ i.o.}\})=1$ אם ורק אם $\sum_{n=1}^{\infty}\mathbb{P}(|X_n|>x_n)=\infty$. ולפי הגדרה טענות אלה שקולות לכך ש-$\mathbb{P}(\{A_n^c \text{ a.e.}\})=0$, וזהו בדיוק המאורע ש-$\{X_n\}_{n\in\mathbb{N}}$ ו- $\{X_n\cdot \mathbb{1}_{X_n\in[-x_n,x_n]}\}_{n\in\mathbb{N}}$ זהות המ"מ. ■

הוכחה של משפט 8.19: קיום תוחלת גורר התכנסות כמעט-תמיד של סדרת הממוצעים. נסמן לאורך ההוכחה $S_n=\frac{1}{n}(X_1+\cdots+X_n)$ להיות סדרת הממוצעים האמפיריים (שאת התכנסותה אנו רוצים להוכיח). ההוכחה נחלקת למספר שלבים.

שלב 1: אי-שליליות. ניתן להניח כי $X\geq 0$ כמעט תמיד. זאת מכיוון שבאופן כללי $X=X_+-X_-$ כאשר X_+ ו-X_- הם משתנים מקריים אי-שליליים בעלי תוחלת, לכן מספיק להוכיח את המשפט לכל אחד מהם.

שלב 2: דילול. נשים לב שסדרת הממוצעים האמפריים אינה יכולה להשתנות מהר מדי. באופן פורמלי, עבור אינדקסים $n,m\in\mathbb{N}$ ומספר קטן $\epsilon>0$ כך ש- $n\leq m\leq(1+\epsilon)n$ מתקיים

$$S_m=\frac{1}{m}(X_1+\cdots+X_m)\leq\frac{n}{m}\cdot\frac{1}{n}(X_1+\cdots+X_m+\cdots+X_n)\leq(1+\epsilon)S_n.$$

ככל ש- ϵ קטן יותר (כלומר האינדקסים n ו-m נבדלים בקבוע כפלי קרוב יותר ל-1) הרי שהפער הכפלי בין S_m ל-S_n שואף ל-1. לפיכך, די להוכיח את ההתכנסות של הסדרה $\{S_{n_j}\}_{j\in\mathbb{N}}$ עבור כל $c>1$ וכל סדרת אינדקסים עולה n_j המקיימת $\frac{n_{j+1}}{n_j}\geq c$. סדרת אינדקסים כזו נקראת לקונרית והיא מתאפיינת בכך ש- $\sum_{j=1}^{\infty}\frac{1}{n_j}$.

שלב 3: בורל-קנטלי. לפי למת בורל-קנטלי הראשונה, משפט 8.5, די להראות כי:

$$\sum_{j=1}^{\infty}\mathbb{P}\left(|S_{n_j}-\mathbb{E}X|>\epsilon\right)<\infty, \tag{*}$$

לכל $\epsilon>0$ ולכל סדרה לקונרית $\{n_j\}$. (או אז ינבע כי החל ממקום מסויים לכל j, $|S_{n_j}-\mathbb{E}X|\leq\epsilon$ ולפי שלב 2, ינבע כי החל ממקום מסויים גם לכל m, $|S_m-\mathbb{E}X|\leq\epsilon$).

שלב 4: השוואה למשתנים קטומים זהים כב"מ. לו היה X בעל שונות סופית, אז בעזרת אי-שיוויון צ'בישב (משפט 5.5) היינו מקבלים כי

$$\mathbb{P}\left(|S_{n_j}-\mathbb{E}X|>\epsilon\right)\leq\frac{\mathrm{Var}(S_{n_j})}{\epsilon^2}=\frac{n_j\,\mathrm{Var}(X)}{n_j^2\epsilon^2}=\frac{\mathrm{Var}(X)}{n_j\epsilon^2}.$$

מכיוון ש- n_j סדרה לקונרית, $\sum_{j=1}^{\infty}\frac{1}{n_j}<\infty$, ולכן היינו מקבלים כי (*) מתקיים.

באופן כללי, אי אפשר להניח כי X חסום או בעל שונות סופית, אך ניתן לקרבו על-ידי משתנים כאלה. נשתמש בסימון הבא: לכל מספר $L>0$ נגדיר $X_{\leq L}=X\cdot\mathbb{1}_{X\leq L}$. נשים לב כי מתקיים $\sum_{j=1}^{\infty}n_j\mathbb{P}(X\geq n_j)\leq\mathbb{E}[X]<\infty$. ולכן אם נגדיר $j(i)=\min\{j>i\ :\ n_j\geq i\}$ נקבל לפי טענה 8.22 כי הסדרה $Y_n=X_n\cdot\mathbb{1}_{X_n\leq n_{j(n)}}$

זהה כב"מ לסדרה X_n. נגדיר $\overline{S}_n = \frac{1}{n}(Y_1 + \cdots + Y_n)$ ונקבל, לפי טענה 8.21, כדי להוכיח כי סדרת ממוצעים זו מתכנסת בהסתברות לתוחלת של X. כלומר, לפי בורל-קנטלי, די להראות

$$\sum_{j=1}^{\infty} \mathbb{P}\left(|\overline{S}_{n_j} - \mathbb{E}X| > \epsilon\right) < \infty, \tag{**}$$

קיים j_0 כך ש- $|\mathbb{E}\overline{S}_{n_j} - \mathbb{E}X| = |\mathbb{E}X\mathbb{1}_{X\le n_j}| < \epsilon/2$. לפי צ'בישב נקבל כי לכל $j \ge j_0$,

$$\mathbb{P}\left(|\overline{S}_{n_j} - \mathbb{E}X| > \epsilon\right) \le \mathbb{P}\left(|\overline{S}_{n_j} - \mathbb{E}\overline{S}_{n_j}| > \epsilon/2\right) \le \frac{4\operatorname{Var}[|X_{\le n_j}|^2]}{n_j\epsilon^2} \le \frac{4\mathbb{E}[|X_{\le n_j}|^2]}{n_j\epsilon^2}$$

להוכחת (**), די אפוא אם נוכיח כי $\sum_{j=1}^{\infty} \mathbb{E}[X^2_{\le n_j}/n_j] < \infty$. לפי משפט ההתכנסות המונוטונית ניתן להחליף סכום ואינטגרל בטור החיובי הבא:

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{1}{n_j}\mathbb{E}[X^2_{\le n_j}] &= \sum_{j=1}^{\infty} \int_\Omega X^2(\omega)\frac{1}{n_j}\mathbb{1}\left\{X(\omega) \le n_j\right\} d\mathbb{P}(\omega)\\ &= \int_\Omega X^2(\omega)\sum_{j=1}^{\infty}\frac{1}{n_j}\mathbb{1}\left\{X(\omega) \le n_j\right\} d\mathbb{P}(\omega)\end{aligned}$$

נסדר מחדש את הסכימה ונשתמש כעת בתנאי הלקונריות של הסדרה $\{n_j\}$:

$$\begin{aligned}\sum_{j=1}^{\infty} \frac{1}{n_j}\mathbb{E}[X^2_{\le n_j}] &= \int_\Omega X^2(\omega)\sum_{j=1}^{\infty}\left(\frac{1}{n_j} + \frac{1}{n_{j+1}} + \dots\right)\mathbb{1}\left\{n_{j-1} < X(\omega) \le n_j\right\} d\mathbb{P}(\omega)\\ &\le \int_\Omega X(\omega)\left(1 + c^{-1} + \cdots + c^{-\ell} + \dots\right)\sum_{j=1}^{\infty}\frac{X(\omega)}{n_j}\mathbb{1}\left\{n_{j-1} < X(\omega) \le n_j\right\} d\mathbb{P}(\omega)\\ &\le \frac{c}{c-1}\int_\Omega X(\omega)\sum_{j=1}^{\infty}\mathbb{1}\left\{n_{j-1} \le X(\omega) \le n_j\right\} d\mathbb{P}(\omega) \le \frac{c}{c-1}\mathbb{E}[X] < \infty,\end{aligned}$$

כנדרש. ■

הוכחה. הוכחה של משפט 8.19: התכנסות כ"ת של ממוצעים גוררת קיום תוחלת] נסמן כמקודם $S_n = \frac{1}{n}(X_1 + \cdots + X_n)$. נניח כי קיים $c \in \mathbb{R}$ כך ש- $S_n \xrightarrow{\text{a.s.}} c$. מתקיים

$$\frac{X_n}{n} = \frac{X_1 + \cdots + X_{n-1} + X_n}{n} - \frac{X_1 + \cdots + X_{n-1}}{n-1}\cdot\frac{n-1}{n} = S_n - \frac{n-1}{n}S_{n-1} \xrightarrow{\text{a.s.}} c - 1\cdot c = 0.$$

בפרט $\mathbb{P}\left(\left\{\left|\frac{X_n}{n}\right| \ge 1\right\} \text{ i.o.}\right) = 0$. לפי הלמה השניה של בורל-קנטלי, משפט 8.6, נגיע למסקנה כי

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\left|\frac{X_n}{n}\right| \ge 1\right) = \sum_{n=1}^{\infty}\mathbb{P}\Big(|X| \ge n\Big) < \infty.$$

ואולם, לפי הגדרת התוחלת,

$$\mathbb{E}|X| \leq \sum_{n=1}^{\infty} n\mathbb{P}\Big(n-1 \leq |X| \leq n\Big) = \sum_{k=0}^{\infty} \mathbb{P}\Big(|X| \geq k\Big) < \infty,$$

כלומר X בעל תוחלת כנדרש. ■

בעיות הרחבה והעשרה

✎ **בעיה 8.1.** יהיו $\{A_n\}_n \in \mathbb{N}$, $\{B_n\}_n \in \mathbb{N}$ מאורעות על מרחב הסתברות.

(א) יש להראות כי אם $\mathbb{P}(A_n \text{ a.e.}) = \mathbb{P}(B_n \text{ a.e.}) = 1$ אז $\mathbb{P}(A_n \cap B_n \text{ a.e.}) = 1$.

(ב) יש למצוא דוגמא שבה $\mathbb{P}(A_n \text{ a.e.}) = \mathbb{P}(B_n \text{ a.e.}) = 0$ אך $\mathbb{P}(A_n \cup B_n \text{ a.e.}) = 1$.

✎ **בעיה 8.2.** תהי $\{A_n\}_n \in \mathbb{N}$ סדרת מאורעות על מרחב הסתברות. נשתמש ברישום $A \oplus B = (A \setminus B) \cup (B \setminus A)$. יש להוכיח כי לכל $\epsilon > 0$ ו-$N \in \mathbb{N}$ קיימים $n \geq m > N$ המקיימים

$$\mathbb{P}\left(\{A_n \text{ a.e.}\} \oplus \left(\bigcap_{m}^{n} A_n\right)\right) < \epsilon$$

✎ **בעיה 8.3.** יהיו $\{X_n\}_n \in \mathbb{N}$, X משתנים מקריים על מרחב הסתברות. יש להראות כי אם $X_n \xrightarrow{\text{p}} X$ אז קיימת תת סדרה $\{m_i\}_{i\in\mathbb{N}}$ כך ש-$X_{m_i} \xrightarrow{\text{a.s.}} X$.

✎ **בעיה 8.4.** יהיו $\{X_n\}_n \in \mathbb{N}$, X, $\{Y_n\}_n \in \mathbb{N}$, Y משתנים מקריים על מרחב הסתברות.

(א) יש להראות כי אם $X_n \xrightarrow{\text{a.s.}} X$ ו-$Y_n \xrightarrow{\text{a.s.}} Y$ אז $X_n + Y_n \xrightarrow{\text{a.s.}} X_n + Y_n$.

(ב) יש להראות כי אם $X_n \xrightarrow{\text{p}} X$ ו-$Y_n \xrightarrow{\text{p}} Y$ אז $X_n + Y_n \xrightarrow{\text{p}} X_n + Y_n$.

✎ **בעיה 8.5.** סיזיפוסה, התרנגולת בת-האלמוות, נכנסת ללול בגודל אינסופי. בכל בוקר היא מטילה שלוש ביצים. בכל ערב היא בוחרת באקראי את אחת הביצים השלמות בלול, דוגרת עליה ומבקיעה אותה. מה הסיכוי שבסופו של דבר תוטלנה בלול ביצים שתשארנה שלמות עד קץ-הימים.

✎ **בעיה 8.6 (*).** יהיו $\{X_n\}_{n\in\mathbb{N}}$ משתנים מקריים. יש להראות כי

(א) אם $X_n \xrightarrow{\text{p}} X$ אז קיימת תת-סדרה $\{X_{n_k}\}_{k\in\mathbb{N}}$ כך שמתקיים $X_{n_k} \xrightarrow{\text{a.s.}} X$.

(ב) $X_{n_k} \xrightarrow{\text{p}} X$ אם ורק אם לכל תת-סדרה $\{X_{n_k}\}_{k\in\mathbb{N}}$ קיימת תת-תת-סדרה $\left\{X_{n_{k_j}}\right\}_{j\in\mathbb{N}}$ כך שמתקיים $X_{n_{k_j}} \xrightarrow{\text{a.s.}} X$.

✎ **בעיה 8.7 (*).** נגדיר הילוך שיכור דו-ממדי להיות סדרת וקטורים מקריים (X_n, Y_n) שכל אחד מהם הילוך שיכור מוטה (ר' דוגמא 8.9) עם הסתברויות לנוע ימינה $p_x, p_y \in (0,1)$ בהתאמה. יש להראות כי ההילוך נשנה אם ורק אם $(p_x, p_y) = (1/2, 1/2)$. עוד ניתן להראות כי הילוך שיכור מאוזן בממדים שלוש ומעלה חולף.

✎ **בעיה 8.8 (*).** [הלמה המותנית של בורל-קנטלי] יהיו $\{A_n\}_n \in \mathbb{N}$ מאורעות על מרחב הסתברות ונניח כי לכל $B_1, \ldots, B_n$ כך ש-$B_i \in \{A_i, A_i^c\}$ מתקיים $\mathbb{P}(A_n \mid B_1, \ldots, B_{n-1}) \geq a_n$ עבור $\{a_n\}_n \in \mathbb{N}$ המקיימים $\sum_{n\in\mathbb{N}} a_n = \infty$. אז $\mathbb{P}(A_n \text{ i.o.}) = 1$.

✎ **בעיה 8.9 (*).** יהיו $\{X_n\}$ משתנים מקריים בלתי-תלויים ושווי התפלגות כך ש-$X_n \overset{\text{a.s.}}{\geq} 1$, בעלי תוחלת סופית. יש להראות כי קיים $L \in \mathbb{R}_+$ כך שמתקיים

$$\left(\prod_{i=1}^{n} X_i\right)^{\frac{1}{n}} \xrightarrow{\text{a.s.}} L$$

פרק 9

משתנים מקריים רציפים בהחלט

"אם היה עלי לתאר, במילה אחת, את כוכב הצפון אשר סביבו סובב רקיע המתמטיקה, את הרעיון המרכזי אשר מושל כרוח נעלמה בכל גוף הידע של התורה המתמטית, הייתי בוחר ב**רציפות** כפי שהיא מתמבטאת בהגדרת המרחב שלנו, ואומר - זהו - זה!"
–ג'יימס ג'וזף סילבסטר, נאום לאיגוד המתמטי הבריטי, 1869

עד כה לא הרחבנו את הדיבור על משתנים מקריים מעל מרחבי הסתברות כלליים. בפרק זה נכיר משפחה של משתנים מקריים אשר הנה בה בעת הנגישה ביותר והחשובה ביותר בין כל משפחות המשתנים המקריים שאינם בדידים – **משפחת המשתנים המקריים הרציפים בהחלט**. משתנים אלו מאופיינים בקיומה של פונקציית צפיפות המתארת את "הסבירות ליחידת שטח" שהמשתנה מקבל ערך באיזור מסויים. בפרק זה נכליל את הטכניקות ששימשו אותנו לטיפול במשתנים מקריים בדידים למשתנים ממשפחה זו ונכיר מספר התפלגויות רציפות בהחלט בעלות חשיבות מיוחדת.

הגדרה 9.1 (משתנה מקרי רציף בהחלט)**.** יהיה X משתנה מקרי על מרחב הסתברות. נאמר ש-X **רציף בהחלט** אם קיימת פונקציה אינטגרבילית $f : \mathbb{R} \to \mathbb{R}_+$ כך שלכל $-\infty \le a \le b \le \infty$ מתקיים,

$$\mathbb{P}(X \in [a,b]) = \int_a^b f(x)dx.$$

הפונקציה f נקראת **הצפיפות של** X.

פונקציית הצפיפות של משתנה מקרי רציף בהחלט אינה יחידה, שכן שינוי של מספר בן-מניה של ערכי $f(x)$ לא ישפיע על האינטגרל. עם זאת, שתי פונקציות צפיפות של אותו משתנה מקרי מזדהות כמעט בכל $\mathbb{R}$. את התפלגותו של משתנה מקרי רציף בהחלט נוכל לתאר גם באמצעות פונקציית התפלגות מצטברת.

אבחנה 9.2 (פונקציית הסתברות מצטברת)**.** יהיה X משתנה רציף בהחלט על מרחב הסתברות. אזי

$$F_X(s) = \mathbb{P}(X \le s) = \int_{-\infty}^s f(x)dx.$$

נשים לב כי פונקציית ההתפלגות המצטברת של משתנה מקרי רציף בהחלט הנה רציפה ועולה במובן החלש

מעצם הגדרתה, שכן כל אינטגרל של פונקציה ממשית וחיובית מקיים תכונות אלה. לפי משפט ניוטון-לייבניץ (משפט א.7), ניתן לבחור פונקציית צפיפות של משתנה מקרי כך שהצפיפות תהיה נגזרת פונקציית ההתפלגות המצטברת בכל מקום בו פונקציה זו גזירה.

חישוב הסתברויות עבור מאורעות הנוגעים למשתנה מקרי רציף בהחלט נוכל לחשב גם באמצעות פונקציית ההתפלגות המצטברת.

טענה 9.3 (חישוב הסתברויות - קטעים)**.** יהי X משתנה מקרי רציף בהחלט בעל צפיפות f_X. אזי לכל קטע $[a,b]$ ב-$\mathbb{R}$ מתקיים

$$\mathbb{P}(X \in [a,b]) = F(b) - F(a).$$

בפרט לכל $a \in \mathbb{R}$ מתקיים $\mathbb{P}(X = a) = 0$.

הוכחה. נשים לב שמתוך תכונות האינטגרל (טענה א.8), פונקציית ההתפלגות המצטברת היא רציפה ולכן לכל $a \in \mathbb{R}$ מתקיים

$$\mathbb{P}(X = a) = \mathbb{P}(X \leq a) - \mathbb{P}(X < a) \overset{\substack{\text{משפט}\\1.21}}{=} F(a) - \lim_{x \to a^-} F_X(x) = F(a) - F(a) = 0.$$

כעת נחשב לפי פונקציית ההתפלגות המצטברת

$$\mathbb{P}(X \in [a,b]) = \mathbb{P}(X \in [-\infty, b]) - \mathbb{P}(X \in [-\infty, a]) + \mathbb{P}(X = a) = F(b) - F(a)$$

כנדרש. ■

דוגמא 9.4 (חישוב הסתברות וצפיפות)**.** נתון משתנה מקרי X בעל צפיפות $f_X(x) = \frac{x+1}{2}\mathbb{1}([-1,1])(x)$. נחשב את הסתברות המאורע $\mathbb{P}(X \geq 1/2)$ ומצא את צפיפותם של $Y = 5X$ ושל $Z = X^2$.

תשובה: ראשית נבחין שמדובר אומנם בפונקציית צפיפות, שכן

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_{-1}^{1} \frac{x+1}{2} dx = \left(\frac{x^2+2x}{4}\right)\Bigg|_{-1}^{1} = 1.$$

נחשב לפי טענה 9.3,

$$\mathbb{P}(X \geq 1/2) = \int_{1/2}^{\infty} f_X(x)dx = \int_{1/2}^{1} \frac{x+1}{2} dx = \left(\frac{x^2+2x}{4}\right)\Bigg|_{1/2}^{1} = \frac{3}{4} - \frac{5}{16} = \frac{7}{16}$$

לחישוב צפיפותו של Y, נחשב את פונקציית ההתפלגות המצטברת F_Y. ניתן דעתנו לכך שעבור $y < -5$ מתקיים $\mathbb{P}(Y \leq y) = 0$ ואילו עבור $y > 5$ מתקיים $\mathbb{P}(Y \leq y) = 1$. עבור $y \in [-5, 5]$ נחשב

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(5X \leq y) = \mathbb{P}(X \leq y/5) \\
&= F_X(y/5) = \int_{-\infty}^{y/5} f_X(x)dx = \int_{-1}^{y/5} \frac{x+1}{2} dx = \left(\frac{x^2+2x}{4}\right)\Bigg|_{-1}^{y/5} \\
&= \frac{y^2+10y}{100} + \frac{1}{4}
\end{aligned}$$

נגזור ונקבל $f_Y = (y/50+1/10)\mathbb{1}_{[-5,5]}(x)$. כעת נחשב את צפיפותו של Z, שוב ע"י חישוב תחילה של פונקצית ההתפלגות המצטברת שלו. עבור $z < 0$ מתקיים $\mathbb{P}(Z \leq z) = 0$ ואילו עבור $z > 1$ מתקיים $\mathbb{P}(Z \leq z) = 1$. עבור $z \in [0,1]$ נחשב

$$F_Z(z) = \mathbb{P}(X^2 \leq z) = \mathbb{P}(X \in [-\sqrt{z}, \sqrt{z}]) = \int_{-\sqrt{z}}^{\sqrt{z}} f_X(x)dx = \int_{-\sqrt{z}}^{\sqrt{z}} \frac{x+1}{2}dx$$
$$= \left(\frac{x^2+2x}{4}\right)\Bigg|_{-\sqrt{z}}^{\sqrt{z}} = \sqrt{z}$$

נגזור ונקבל כי הצפיפות של Z היא $f_Z(x) = \dfrac{\mathbb{1}([0,1])}{2\sqrt{z}}$.

כפי שדוגמא 9.4 המחישה, ניתן בקלות יחסית למצוא את השינוי בצפיפות שנגרם מהפעלה של פונקציה מונוטונית על משתנה מקרי בעל צפיפותו ידועה. ננסח אבחנה זו כטענה.

טענה 9.5. יהי X משתנה מקרי בעל צפיפות f_X ותהי $g : \mathbb{R} \to \mathbb{R}$ פונקציה מונוטונית ממש וגזירה. אז $Y = g(X)$ הוא משתנה מקרי בעל צפיפות ולכל $y \in \mathbb{R}$ מתקיים
$$f_Y(y) = \frac{f_X(x)}{|g'(x)|}$$
כאשר $x = g^{-1}(y)$.

הוכחה. נוכיח עבור המקרה של g עולה. המקרה של g יורדת מוכח בצורה דומה. נראה שהפונקציה המוצעת בתור צפיפות של Y אכן מקיימת את ההגדרה. נשים לב כי כיוון ש-g מונוטונית עולה מתקיים

$$g'(x) = g'(g^{-1}(y)) \geq 0$$

לכל $y \in \mathbb{R}$. נשים לב כי לפי נוסחא לנגזרת פונקציה הופכית (טענה א.6) מתקיים:

$$\frac{dg^{-1}(y)}{dy} = \frac{1}{g^{-1}(y)}.$$

ולכן

$$f_Y(y) = \frac{f_X(x)}{|g'(x)|} = f_X(g^{-1}(y))\,(g^{-1})'(y).$$

נציב ונחשב בעזרת החלפת משתנה עבור נקודה a המקיימת $g(a) = b$:

$$\int_{-\infty}^{b} f_Y(y)dy = \int_{-\infty}^{b} f_X(g^{-1}(y))\,(g^{-1})'(y)\,dy \qquad \begin{bmatrix} u &=& g^{-1}(y) \\ du &=& (g^{-1})'(y)dy \end{bmatrix}$$
$$= \int_{-\infty}^{a} f_X(u)\,du = \mathbb{P}\big(X \leq a) = \mathbb{P}(Y \leq b).$$

וקיבלנו כי $f_Y(y)$ היא אכן פונקציית צפיפות של Y, כנדרש.

■

הערה: בהפעלת פונקציה כללית g, לא בהכרח מונטונית, על משתנה מקרי בעל צפיפות X, ניתן לחשב את הצפיפות של $g(X)$ ע"י חלוקה לתחומי מונוטוניות של g וסכימת הצפיפויות המתקבלות מתחומים אלה. דבר זה נעשה בחלק השני של דוגמא 9.4 כאשר חישבנו את הצפיפות של X^2 – למעשה סכמנו את הצפיפות המתקבלת מהפעלת $x \mapsto x^2$ בתחום $(-\infty, 0)$ ובתחום $(0, \infty)$, כאשר בדוגמא הספציפית שלנו התרומות היו שוות בשל הסימטריה של X^2 סביב ראשית הצירים.

✎ **בעיה 9.1.** להוכיח את טענה 9.5 למקרה של פונקציה מונוטונית יורדת.

9.0.1 פונקציית האחוזון של משתנה מקרי רציף בהחלט.

בשימושים סטטיסטים רבים, נעשה שימוש בפונקציה המכונה **פונקציית האחוזון (quantile)**. פונקציה זו מתארת לכל מספר $p \in [0, 1]$, את הערך המינימלי a כך שבסיכוי לפחות p מתקיים המאורע $\{X \le a\}$.

הגדרה 9.6 (אחוזון של משתנה מקרי)**.** יהי X משתנה מקרי בעל פונקציית התפלגות מצטברת F_X. נגדיר את ה**אחוזון** של X כפונקציה $Q : [0, 1] \to \mathbb{R} \cup \{\pm\infty\}$ הנתונה על ידי

$$Q_X(t) = \inf(F_X^{-1}((t, \infty))$$

האחוזון המתאים ל-$Q_X(1/2)$ מכונה **החציון** של X והוא משמש כאומד נוסף להתנהגותו הטיפוסית של המשתנה המקרי. עבור משתנים מקריים רציפים מתקיים כי $F(Q_X(t)) = t$, וזאת מפני שפונקציית ההתפלגות המצטברת של משתנים כאלו הנה רציפה.

✎ **בעיה 9.2.** להראות כי כאשר הפונקציה $F_X(t)$ הפיכה, מתקיים $Q_X(t) = F^{-1}(t)$.

הבחירה להשתמש inf בהגדרת האחוזון הנה שרירותית. הגדרה מקבילה המשתמשת ב-sup תניב אובייקט מתמטי בעל אותן תכונות ואותם שימושים. ההבדל נובע מכך שכאשר ישנו טווח ערכים $I = [a, b]$ כך שההסתברות ש-X קטן מכל ערך ב-I היא p (ולכן ההסתברות ש-$X \in I$ היא אפס), עלינו לבחור באופן שרירותי איזה איבר בטווח יהיה האחוזון ה-p של X. בהגדרה המקובלת אנו בוחרים את הערך הקטן ביותר.

9.0.2 תוחלת של משתנה מקרי רציף

לבטא את תוחלתו של משתנה מקרי רציף בהחלט באמצעות פונקציית הצפיפות באופן אנלוגי לנוסחת התוחלת עבור משתנה מקרי בדיד.

הגדרה 9.7 (תוחלת משתנה מקרי רציף בהחלט)**.** יהי X משתנה מקרי רציף בהחלט בעל פונקציית צפיפות $f_X(x)$. אזי

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

בכל מקרה בו אינטגרל זה מתכנס בהחלט. אחרת אין ל-X תוחלת.

כפי שעשינו עבור משתנה מקרי גיאומטרי, נוכל לפתח נוסחה לתוחלת משתנה מקרי רציף בהחלט במושגים של פונקציית התפלגות מצטברת.

טענה 9.8. לכל משתנה מקרי רציף X מתקיים

$$\mathbb{E}(X) = \int_0^\infty (1 - F(x)) - \int_{-\infty}^0 F(x)$$

הוכחה. נחשב באמצעות משפט פוביני (טענה א.10):

$$\int_0^\infty (1 - F_X(x)) = \int_0^\infty \int_t^\infty f(x)dxdt \overset{\text{פוביני}}{=} \int_t^\infty f(x)dx \int_0^x dt = \int_0^\infty xf(x)dx$$

$$\int_{-\infty}^0 F_X(x) = \int_0^\infty \int_t^\infty f(-x)dxdt \overset{\text{פוביני}}{=} \int_t^\infty f(-x)dx \int_0^x dt = \int_0^\infty xf(-x)dx \quad \begin{bmatrix} y = -x \\ dy = -dx \end{bmatrix}$$

$$= \int_{-\infty}^0 yf(y)dx$$

קיבלנו אפוא ש-

$$\int_{-\infty}^\infty xf(x)dx = \int_{-\infty}^0 yf(y)dx + \int_0^\infty xf(x)dx = \int_0^\infty (1 - F(x)) - \int_{-\infty}^0 F(x),$$

כנדרש. ■

לצורך חישוב תוחלת של פונקציות של משתנים מקריים רציפים נציג ללא הוכחה טענה מקבילה לטענה 4.10.

טענה 9.9 (תוחלת פונקציה של משתנה מקרי). יהי X משתנה מקרי בעל צפיפות f_X ותהי $g : \mathbb{R} \to \mathbb{R}$ פונקציה מדידה.

אז $Y = g(X)$ הוא משתנה מקרי המקיים

$$\mathbb{E}\big(Y\big) = \int_{-\infty}^\infty g(x)f_X(x)dx$$

אם ורק אם האינטגרל מתכנס בהחלט.

נשים לב שלא דרשנו בטענה זו כי המשתנה Y יהיה רציף, ואומנם הטענה נכונה באופן כללי. על סמך טענה 9.9 נוכל לחשב מומנטים של משתנים מקריים רציפים בהחלט. כך למשל שונות של משתנה מקרי רציף X תתקבל מן החשבון

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_{-\infty}^\infty x^2 f_X(x)dx - \left(\int_{-\infty}^\infty xf_X(x)dx\right)^2.$$

התוחלת והשונות של משתנים מקריים רציפים מוסיפות לקיים תכונות התוחלת והשונות שהראינו בטענות 4.5 ו-5.3.

טענה 9.10 (תכונות התוחלת). יהיו X, Y משתנים מקריים רציפים בהחלט על אותו מרחב הסתברות, המקיימים $\mathbb{E}(Y), \mathbb{E}(X) \in \mathbb{R} \cup \{\infty\}$. אזי

(א) **חיוביות:** אם $X \overset{\text{a.s.}}{\geq} 0$ אז $\mathbb{E}(X) \geq 0$, ואם $X \overset{\text{a.s.}}{>} 0$ אז $\mathbb{E}(X) > 0$.

(ב) **ליניאריות:** $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ לכל $a, b \in \mathbb{R}$.

(ג) **מונוטוניות:** אם $\mathbb{P}(X \geq Y) = 1$ אז $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

טענה 9.11 (תכונות השונות)**.** יהי X משתנה מקרי בעל שונות סופית, ויהי $a \in \mathbb{R}$. אזי

(א) $\mathrm{Var}(X) \geq 0$ ושוויון מתקיים רק אם X קבוע.

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

(ג) $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$ (ולכן $\sigma(aX) = |a| \cdot \sigma(X)$).

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

✎ **בעיה 9.3.** להשתמש בתכונות האינטגרל ולהוכיח את טענות 9.10 ו-5.3.

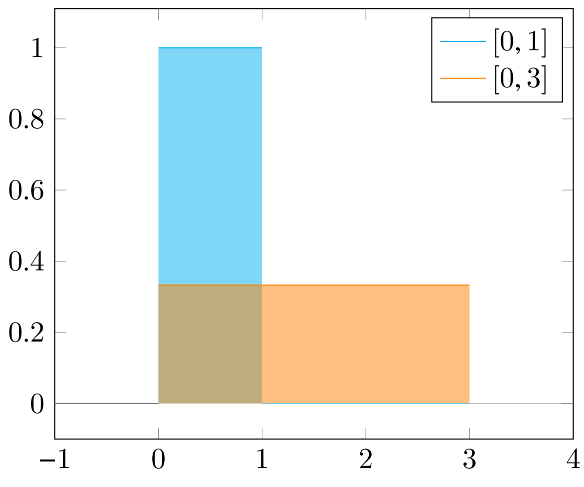
9.1 התפלגויות רציפות חשובות

בפרק זה נכיר את שלוש ההתפלגויות הרציפים החשובות ביותר, אשר תלווינה אותנו לאורך הפרק. בהצגת כל התפלגות נציג פונקציית התפלגות מצטברת, תוחלת, שונות ופונקציה יוצרת מומנטים. הקורא עשוי למצא תועלת בהשוואת חישובי התוחלת השונות והפונקציה יוצרת מומנטים עבור פונקציות אלו, לחישובים עבור התפלגויות בדידות מקבילות בפרק 3.4.

9.1.1 התפלגות אחידה

הגדרה 9.12 (התפלגות אחידה)**.** יהי $[a, b] \subset \mathbb{R}$ קטע. נאמר שלמ"מ מקרי X **התפלגות אחידה על** $[a, b]$ ונכתוב $X \sim \mathrm{Unif}([a, b])$, אם צפיפותו היא

$$f_X(x) = \frac{\mathbb{I}([a, b])(x)}{b - a}.$$



צפיפות התפלגות אחידה עבור שני קטעים ב$\mathbb{R}$.

ניתן לתת כמה דוגמאות להתפלגות אחידה מחיי היום-יום:

(א) הזמן בין הגעתנו לצומת מרומזר בין-עירוני לזמן חילוף הרמזור הבא מתפלג בקירוב אחיד

(ב) הזווית שבה יפול סביבון מתפלגת אחיד על $[0, 2\pi]$.

(ג) בבחירת אדם מקרי באוכלוסיה, אחוז האנשים הגבוהים ממנו מתפלג אחיד על $[0, 1]$

אבחנה 9.13 (תכונות התפלגות אחידה). יהי X מ"מ $X \sim \mathrm{Unif}([a, b])$. אזי מתקיים

$$F_X(t) = \begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \le t \le b \\ 1 & t > b \end{cases} \qquad \mathbb{E}(X) = \frac{a+b}{2}$$

$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)} \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

כמו כן לכל $\alpha > 0, \beta \in \mathbb{R}$ מתקיים $\alpha X + \beta \sim \mathrm{Unif}([\alpha a + \beta, \alpha b + \beta])$.

הוכחה.

$$F_X(t) = \int_{-\infty}^{t} \frac{\mathbb{I}([a, b])ds}{b-a} = \begin{cases} 0 & t < a \\ \frac{t-a}{b-a} & a \le t \le b \\ 1 & t > b \end{cases}.$$

$$\mathbb{E}(X) = \int_a^b \frac{xdx}{b-a} = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}.$$

$$\begin{aligned} \mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \int_a^b \frac{x^2dx}{b-a} - \left(\frac{a+b}{2}\right)^2 = \frac{b^3 - a^3}{3(b-a)} - \frac{(a+b)^2}{4} \\ &= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

$$M_X(t) = \mathbb{E}(e^{Xt}) = \int_a^b \frac{e^{tx}dx}{b-a} = \frac{e^{tb} - e^{ta}}{t(b-a)}.$$

ולפי טענה 9.5 מתקיים

$$f(\alpha X + \beta) = \begin{cases} 0 & t < \alpha a + \beta \\ \frac{1}{\alpha}\frac{1}{b-a} & \alpha a + \beta \le t \le \alpha b + \beta \\ 0 & t > \alpha b + \beta \end{cases}$$

■

טענה 9.14 (אחוזון של התפלגות רציפה בהחלט מתפלג אחיד על $[0, 1]$). יהי X משתנה מקרי רציף בהחלט בעל צפיפות $f_X(x)$, פונקציית התפלגות מצטברת $F_X(x)$ ופונקציית אחוזון $Q_X(x)$. כמו כן, יהי $Y \sim \mathrm{Unif}([0, 1])$ אזי

(א) $F_X(X) \overset{d}{=} Y$
(ב) $Q_X(Y) \overset{d}{=} X$

הוכחה. פי שראינו, F_X רציפה מפני ש-X רציף בהחלט ולכן $F(\inf(F_X^{-1}(t))) = t$. נחשב לכל $0 \le \alpha \le 1$, כי $\mathbb{P}(F_X(X) \le \alpha) = \mathbb{P}(X \le \inf(F_X^{-1}(\alpha)))$. מכך נסיק כי

$$\mathbb{P}(F_X(X) \le \alpha) = \mathbb{P}(X \le \inf(F_X^{-1}(\alpha))) = F_X(\inf(F_X^{-1}(\alpha))) = \alpha.$$

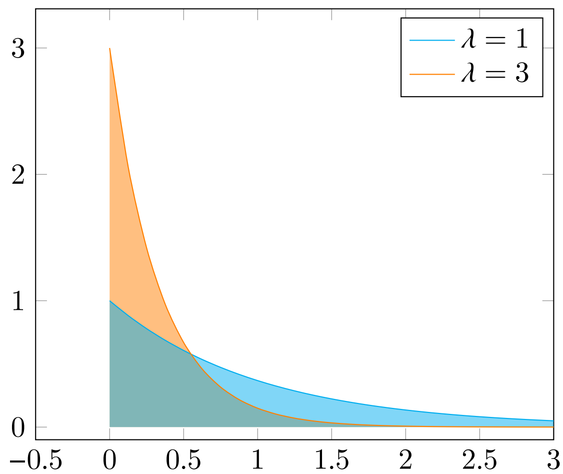
נסיק מטענה 7.11 כי $F_X(X) \sim \mathrm{Unif}([0,1])$, כנדרש.

■

הערה: לפונקציית האחוזון חשיבות רבה בסימולציה של משתנים מקריים שכן היא מבטיחה לנו שאם ביכולתנו לחשב את Q_X ולהגריל U, משתנה מקרי אחיד על $[0,1]$ אז נוכל להגריל משתנה מקרי שווה התפלגות ל-X באמצעות חישוב $Q_X(U)$. למעשה ניתן להשתמש ברעיון זה בכדי להראות שניתן לממש כל משתנה מקרי רציף בהחלט כמשתנה מקרי על מרחב הסתברות תקני.

9.1.2 התפלגות מעריכית

הגדרה 9.15 (התפלגות מעריכית)**.** נאמר שלמ"מ מקרי X **התפלגות מעריכית עם פרמטר** λ ונכתוב $X \sim \mathrm{Exp}(\lambda)$, אם צפיפותו היא
$$f_X(x) = \mathbb{1}([0,\infty))(x)\lambda e^{-\lambda x}.$$



צפיפות התפלגות מעריכית עבור שני פרמטרים שונים.

ההתפלגות המעריכית היא מקבילתה הרציפה של ההתפלגות הגיאומטרית. היא מתארת היטב את **הזמן שחולף עד להתרחשות בתנאי חוסר זיכרון** אך גם התפלגויות פיסיקליות נוספות. דוגמאות למשתנים מקריים אקספוננציאליים הנן:

(א) **הזמן עד לכניסת לקוח לחנות.**

(ב) **הזמן בין לידות של תינוקות במדינה.**

(ג) **הזמן עד להתפרקותו של אטום אורניום.**

(ד) הגובה של מולקולה מקרית באטמוספירה.

(ה) האנרגיה הקינטית של אטום מקרי בגז אידאלי (התפלגות בולצמן).

אבחנה 9.16. יהי $X \sim \mathrm{Exp}(\lambda)$. אזי

$$\mathbb{E}(X) = 1/\lambda \qquad F_X(t) = \max(1 - e^{-\lambda t}, 0)$$

$$\mathrm{Var}(X) = 1/\lambda^2 \qquad M_X(t) = \frac{\lambda}{\lambda - t} \quad \text{עבור } t < \lambda$$

כמו כן לכל $\alpha > 0$ מתקיים $\alpha X \sim \mathrm{Exp}(\lambda/\alpha)$.

הוכחה.

$$F_X(t) = \int_0^t \lambda e^{-\lambda s} ds = \max(1 - e^{-\lambda t}, 0).$$

$$\mathbb{E}(X) = \int_0^\infty \lambda s e^{-\lambda s} ds = \lim_{s\to\infty} \left[-xe^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right]_{x=0}^{s} = \frac{1}{\lambda}$$

$$\begin{bmatrix} f(x) = x^2 & f'(x) = 2x \\ g(x) = e^{-\lambda x} & g'(x) = -\lambda e^{-\lambda x} \end{bmatrix}$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \int_0^\infty \lambda s^2 e^{-\lambda s} ds - \frac{1}{\lambda^2}$$

$$= \lim_{s\to\infty} \left([-x^2 e^{-\lambda x}]|_0^s + 2 \int_0^r xe^{-kx} dx \right)$$

$$= \lim_{s\to\infty} \left[-x^2 e^{-\lambda x} - \frac{2}{\lambda} xe^{-\lambda x} - \frac{2}{\lambda^2} e^{-\lambda} \right]_{x=0}^{s} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

כמו כן, עבור $t < \lambda$ מתקיים,

$$M_X(t) = \int_0^\infty \lambda e^{(t-\lambda)s} ds = \lim_{s\to\infty} \left[\frac{\lambda}{\lambda - t} e^{(t-\lambda)x} \right]_{x=0}^{s} = \frac{\lambda}{\lambda - t},$$

וכן $F_{\alpha X}(t) = F_X(\frac{t}{\alpha}) = \max(1 - e^{-\lambda t/\alpha}, 0)$. ■

נציג כעת תכונת חוסר זיכרון רציפה של משתנה מקרי מעריכי, המעידה על חוסר זיכרון לכשלון בכל פרק זמן ממשי. תכונה זה מקבילה לתכונה הבדידה של משתנה מקרי גיאומטרי.

אבחנה 9.17 (חוסר זיכרון). יהי $X \sim \mathrm{Exp}(\lambda)$. אזי המשתנה המקרי $Y = (X - x_0 \,|\, X > x_0)$ מקיים $Y \sim \mathrm{Exp}(\lambda)$.

הוכחה. נחשב

$$\overline{F}_Y(t) = \mathbb{P}(X - x_0 > t \,|\, X > x_0) = \frac{\overline{F}_X(t + x_0)}{\overline{F}_X(x_0)} = \frac{e^{\lambda(t+x_0)}}{e^{\lambda x_0}} = e^{-\lambda t}$$

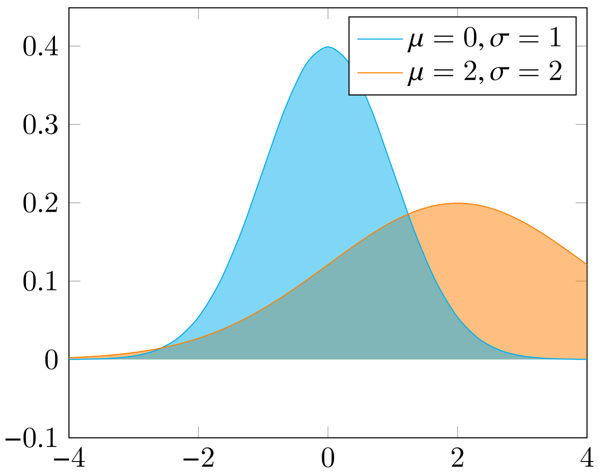
זו פונקציית התפלגות שיורית של משתנה מעריכי עם פרמטר λ והיא קובעת את התפלגותו לפי טענה 7.11 . ■

9.1.3 התפלגות נורמלית

הגדרה 9.18 (התפלגות נורמלית)**.** נאמר שלמ"מ מקרי X **התפלגות נורמלית עם תוחלת** μ **ושונות** σ^2 ונכתוב $X \sim N(\mu, \sigma^2)$, אם צפיפותו היא

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

אם $\mu = 0$ ו-$\sigma = 1$ נאמר ש- X **נורמלי סטנדרטי.**



צפיפות התפלגות נורמלית עבור שני זוגות פרמטרים שונים.

ההתפלגות הנורמלית דומה בצורתה להתפלגות בינומית עבור מספר רב של ניסויים והיא מתארת משתנים מקריים אשר נקבעים כסכום של מספר רב של גורמים שהנן בלתי תלויים בקירוב. דוגמאות למשתנים מקריים המתפלגים בקירוב נורמלית הנן :

(א) גובהו של אדם מקרי באוכלוסיה.

(ב) משך דקה בשעונים אנלוגים שונים

(ג) כמות הגשם במ"מ שירדה בשנה מסויימת

(ד) כמות הנפט שבאר נפט שואבת ביום מסויים

לפונקציית ההתפלגות המצטברת של התפלגות נורמלית סטנדרטית אין נוסחא סגורה. על כן נקבעה ההגדרה הבאה.

הגדרה 9.19 (פונקציית התפלגות מצטברת של התפלגות נורמלית)**.** נסמן מעתה את פונקציית ההתפלגות המצטברת של משתנה מקרי $X \sim N(0,1)$ ב-

$$\Phi(x) = \mathbb{P}(X \le x)$$

אבחנה 9.20. [תכונות משתנה נורמלי] יהי $X \sim N(\mu, \sigma^2)$. אזי

$$F_X(t) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \mathbb{E}(X) = \mu$$
$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \qquad \mathrm{Var}(X) = \sigma^2$$

וכן לכל $\alpha > 0, \beta \in \mathbb{R}$ מתקיים $X + \beta \sim N(\frac{\mu+\beta}{\alpha}, \alpha^2\sigma^2)$.

הוכחה. ראשית נציין כי מדובר בפונקציית צפיפות מוגדרת היטב, כלומר $\int_{-\infty}^{\infty} e^{-x^2} = \sqrt{\pi}$, עובדה אנליטית זו אינה מובנת מאליה והיא מכונה אינטגרל אוילר-פואסון. הוכחתה, החורגת מתחום העניין של קורס זה מובאת להלן בהערה. ראשית נוכיח שלמשתנה נורמלי סטנדרטי יש אומנם תוחלת 0 ושונות 1, כי העתקה ליניארית של מ"מ נורמלי היא נורמלית וכי אם $Y \sim N(0,1)$ ונגדיר $X = \sigma Y + \mu$ אז $X \sim N(\mu, \sigma^2)$. הנוסחא לתוחלת משתנה כללי תנבע אז מליניאריות התוחלת (טענה 4.5), הנוסחא לשונות מתכונות השונות (טענה 5.3), והנוסחא לפונקציית התפלגות מצטברת – מהגדרה 9.19.

בכדי לוודא שהתוחלת מוגדרת היטב, נחשב ונקבל

$$\begin{bmatrix} t = s^2/2 \\ dt = sds \end{bmatrix}$$

$$\begin{aligned}\mathbb{E}(|Y|) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} |s| e^{\frac{-s^2}{2}} ds = 2\int_0^{\infty} \frac{1}{\sqrt{2\pi}} s e^{\frac{-s^2}{2}} ds \\ &= \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t} dt = \frac{1}{\sqrt{2\pi}}\end{aligned}$$

כעת מכיוון שצפיפות Y היא פונקציה זוגית, כלומר $f_Y(t) = f_Y(-t)$, נקבל כי $-\max(Y,0) \overset{d}{=} \min(Y,0)$. ולכן

$$\mathbb{E}(Y) = \mathbb{E}(\min(Y,0)) + \mathbb{E}(\max(Y,0)) = 0.$$

בכדי להראות כי שונות Y היא 1, נחשב

$$\begin{bmatrix} f(x) = x & f'(x) = 1 \\ g(x) = -e^{-\frac{x^2}{2}} & g'(x) = xe^{-\frac{x^2}{2}} \end{bmatrix}$$

$$\begin{aligned}&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} \\ &= \frac{1}{\sqrt{2\pi}} xe^{-\frac{x^2}{2}}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} = 1\end{aligned}$$

עוד ניווכח כי, לפי טענה 9.5,

$$f_{\alpha X+\beta}(y) = \frac{1}{\alpha}\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(\frac{x-\beta}{\alpha}-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\alpha^2\sigma^2}} e^{\frac{-(x-\beta-\mu)^2}{\alpha^2\sigma^2}}$$

ולכן צפיפות $\alpha X+\beta$ היא נורמלית. לסיום ההוכחה נחשב את $M_X(t)$.

$$\begin{aligned}
M_X(t)=\mathbb{E}(e^{Xt})&=\int_{-\infty}^{\infty}e^{xt}\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}dx && \begin{bmatrix} s=(x-\mu)/\sigma \\ ds=\frac{1}{\sigma}dx\end{bmatrix}\\
&=\frac{e^{\mu t}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\sigma st}e^{\frac{-s^2}{2}}ds=\frac{e^{\mu t+\sigma^2t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{s^2}{2}+\sigma ts-\frac{\sigma^2}{2}}ds\\
&=\frac{e^{\mu t+\sigma^2t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}(s-\sigma t)^2}ds && \begin{bmatrix} u=s-\sigma t \\ du=ds\end{bmatrix}\\
&=e^{\mu t+\sigma^2t^2/2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{1}{2}u^2}du=e^{\mu t+\sigma^2t^2/2}.
\end{aligned}$$

■

הערה: החישוב $\sqrt{\pi}=\int_{-\infty}^{\infty}e^{-x^2}dx$, מכונה האינטגרל הגאוסי או אינטגרל אוילר-פואסון. שיטת החישוב פותחה על ידי דה-מואבר, והחישוב עצמו חושב לראשונה על ידי לפלאס ב-1774 אך זכה להתפרסם ב-1809 על ידי גאוס שפיתח אותו באופן בלתי-תלוי. כאן נביא שיטה לחשב אינטגרל זה באמצעות העברת הבעיה למישור ושימוש בקואורדינטות קוטביות. נסמן $I=\int_{-\infty}^{\infty}e^{-x^2}dx$. אזי:

$$\begin{aligned}
I^2&=\left(\int_{-\infty}^{\infty}e^{-x^2}dx\right)^2=\left(\int_{-\infty}^{\infty}e^{-x^2}dx\right)\left(\int_{-\infty}^{\infty}e^{-y^2}dy\right)\\
&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(x^2+y^2)}dx\,dy && \begin{bmatrix} x=r\cos\theta, y=r\sin\theta \\ dx\,dy=r\,dr\,d\theta\end{bmatrix}\\
&=\int_0^{2\pi}\int_0^{\infty}e^{-r^2}r\,dr\,d\theta && \begin{bmatrix} u=r^2, \\ du=2r\,dr\end{bmatrix}\\
&=2\pi\cdot\int_0^{\infty}e^{-u}\frac{1}{2}du=\pi.
\end{aligned}$$

9.2* התפלגות משותפת של מספר משתנים מקריים רציפים

הגדרה 9.21 (צפיפות משותפת של שני משתנים מקריים)**.** נאמר כי לשני משתנים מקריים X,Y מעל מרחב ההסתברות משותף $(\Omega,\mathcal{F},\mathbb{P})$ יש צפיפות משותפת אם קיימת פונקציה אינטגרבילית $f_{X,Y}(x,y):\mathbb{R}^2\to[0,\infty)$ המקיימת לכל קבוצה מהטיפוס $A=(-\infty,a]\times(-\infty,b]$,

$$\mathbb{P}_{X,Y}(A)=\mathbb{P}\Big(X\le a,Y\le b\Big)=\int_{-\infty}^{a}\int_{-\infty}^{b}f_{X,Y}(x,y)\,dx\,dy.$$

הפונקציה $f_{X,Y}(x,y)$ נקראת **צפיפות משותפת** של X ו-Y.

את תפקידה של פונקציית ההתפלגות המצטברת עבור זוג משתנים מקריים בעלי צפיפות משותפת תמלא הפונקציה הבאה

$$F_{X,Y}(a,b)=\mathbb{P}\Big(X\le a,Y\le b\Big)=\int_{-\infty}^{a}\int_{-\infty}^{b}f_{X,Y}(x,y)\,dx\,dy.$$

נשים לב שאם נגזור לפי x ואז לפי y, נקבל כי הנגזרת
$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$
שווה ל-$f_{X,Y}$ מלבד בקבוצה בעלת מידה אפס.

על מנת לחשב את הסתברותם של מאורעות המערבים את X ו-Y, יש צורך לחשב את האינטגרל של הצפיפות על התחום המתאים ב-$\mathbb{R}^2$. ניתן לעשות זאת באופן דומה להגדרת אינטגרל רימן על $\mathbb{R}$. אנו נסמן אינטגרל זה באופן הבא:
$$\mathbb{P}((X,Y) \in A) = \iint_A f(x,y) d(x,y).$$
ניתן לחשוב על אינטגרל זה בתור הנפח מתחת לפונקציה f ומעל התחום A. באופן כללי אנו נתעניין בפונקציות אינטגרביליות, על תחומים ב$\mathbb{R}^2$. כדי לחשב אינטגרל של פונקציה כזו, נשתמש בגרסא כללית של משפט פוביני א.11.

דוגמא 9.22 **(חישוב הסתברויות למשתנים מקריים בעלי צפיפות משותפת).** נתונים משתנים מקריים X ו-Y בעלי צפיפות משותפת $f_{XY}(x,y) = (x+y)\mathbb{1}([0,1])(x)\mathbb{1}([0,1])(y)$. נחשב את הסתברות המאורע $\mathbb{P}(X \geq 1/2, Y \geq 1/2)$ ואת הסתברות המאורע $\mathbb{P}(X+Y \leq 1)$.

תשובה: ראשית נבחין שמדובר אומנם בפונקציית צפיפות, שכן
$$\iint_{\mathbb{R}^2} f(x,y) d(x,y) = \iint_{[0,1]^2} (x+y) d(x,y) = \int_0^1 \left(\int_0^1 (x+y) dy \right) dx = \int_0^1 (xy + \frac{y^2}{2}) \Big|_{y=0}^{y=1} dx$$
$$= \int_0^1 (x + \frac{1}{2}) dx = \frac{x^2+x}{2} \Big|_{x=0}^{x=1} = 1.$$
נחשב לפי טענה א.11,
$$\mathbb{P}(X \geq 1/2, Y \geq 1/2) = \iint_{[1/2,\infty)\times[1/2,\infty)} f(x,y) d(x,y) = \int_{1/2}^{\infty} \left(\int_{1/2}^{\infty} f_{X,Y}(x,y) dy \right) dx$$
$$= \int_{1/2}^{1} \left(\int_{1/2}^{1} (x+y) dy \right) dx = \int_{1/2}^{1} (xy + \frac{y^2}{2}) \Big|_{y=1/2}^{y=1} dx$$
$$= \int_{1/2}^{1} (x + \frac{1}{2}) - (\frac{x}{2} + \frac{1}{8}) dx = \int_{1/2}^{1} \frac{x}{2} + \frac{3}{8}$$
$$= \frac{x^2}{4} + \frac{3x}{8} \Big|_{x=1/2}^{x=1} = (\frac{1}{4} + \frac{3}{8}) - (\frac{1}{16} + \frac{3}{16}) = \frac{3}{8}.$$
אם נסמן ב-A את הקבוצה $\{(x,y) | 0 \leq x, 0 \leq y, x+y \leq 1\}$ הרי שבעזרת טענה א.11 שוב נקבל
$$\mathbb{P}(X+Y \leq 1) = \iint_A (x+y) d(x,y) = \int_0^1 \left(\int_0^{1-x} (x+y) dy \right) dx = \int_0^1 (xy + \frac{y^2}{2}) \Big|_{y=0}^{y=1-x} dx$$
$$= \int_0^1 (x(1-x) + \frac{(1-x)^2}{2}) dx = \int_0^1 \frac{1}{2} - \frac{x^2}{2} dx = \frac{x}{2} - \frac{x^3}{6} \Big|_{x=0}^{x=1} = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}.$$

אבחנה 9.23. יהיו X, Y משתנים מקריים בעלי צפיפות משותפת $f_{X,Y}$ אז X ו-Y רציפים בהחלט ומתקיים כי
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx$$
הן צפיפויות של X ו-Y בהתאמה. התפלגויותיהם של X ו-Y נקראות **ההתפלגויות השוליות של X ו-Y**.

✎ **בעיה 9.4.** להוכיח את אבחנה 9.23.

דוגמא 9.24 (חישוב התפלגות שולית למשתנים מקריים בעלי צפיפות משותפת)**.** נתונים משתנים מקריים X ו-Y בעלי צפיפות משותפת $f_{XY}(x,y) = (x+y)\mathbb{1}([0,1]^2)(x,y)$. חשב את התפלגות X.

תשובה: לפי אבחנה 9.23 עבור $0 \le x \le 1$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy = \int_0^1 (x+y)dy = \left(xy + \frac{y^2}{2}\right)\Bigg|_{y=0}^{y=1} = x + \frac{1}{2}.$$

עבור ערכי x אחרים, $f_X(x) = 0$.

כעת נציג קריטריון טבעי לאי-תלות של משתנים רציפים בהחלט, מקביל לאבחנה 3.33.

אבחנה 9.25. יהיו X, Y משתנים מקריים רציפים בהחלט בעלי צפיפויות f_X ו-f_Y בהתאמה. אזי X ו-Y בלתי-תלויים אם ורק אם קיימת להם צפיפות משותפת המקיימת

$$f_{X,Y} = f_X f_Y.$$

הוכחה. בכיוון אחד נניח כי $f_{X,Y} = f_X f_Y$ ונקבל לפי הגדרה 9.21 כי לכל $A, B \subset \mathbb{R}$ מדידות

$$\begin{aligned}\mathbb{P}(X \in A, Y \in B) &= \int_A \int_B f_{X,Y}(x,y)\,dx\,dy = \int_A \int_B f_X(x) f_Y(y)\,dx\,dy \\ &= \int_A f_X(x)\,dx \int_B f_Y(y)\,dy = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)\end{aligned}$$

ולכן, לפי הגדרה 7.15, X ו-Y בלתי-תלויים. בכיוון השני נניח ש-X ו-Y בלתי-תלויים ונקבל לפי אותה משוואה עצמה כי הפונקציה $f_{X,Y} = f_X f_Y$ היא צפיפות משותפת שלהם. ■

דוגמא 9.26 (אי תלות של משתנים מקריים בעלי צפיפות משותפת)**.** נתונים משתנים מקריים X ו-Y בעלי צפיפות משותפת $f_{XY}(x,y) = (x+y)\mathbb{1}([0,1]^2)(x,y)$. האם X ו-Y בלתי תלויים?

תשובה: בדוגמא 9.24, חישבנו וקיבלנו כי צפיפות X היא $f_X(x) = \frac{2x+1}{2}\mathbb{1}([0,1])(x)$. בצורה דומה, צפיפות Y היא $f_Y(y) = \frac{2y+1}{2}\mathbb{1}([0,1])(y)$ לכן

$$f_{X,Y} = (x+y)\mathbb{1}([0,1]^2)(x,y) \neq (x+1/2)(y+1/2)\mathbb{1}([0,1]^2)(x,y) = f_X(x)f_Y(y)$$

כעת ניתן לבדוק שהפונקציות $f_X(x)f_Y(y)$ ו-$f_{X,Y}$ אינן שוות בכל התחום $x, y < 1/8$, לכן לא יתכן ש-$f_X(x)f_Y(y)$ היא צפיפות משותפת של X ו-Y והמשתנים המקריים הללו תלויים לפי אבחנה 9.25.

דוגמא 9.27 (שימוש באי-תלות של משתנים מקריים)**.** נתונים משתנים מקריים X ו-Y בלתי תלויים המתפלגים מעריכית עם פרמטר 1. נחשב את התפלגות $Z = \min(X, Y)$.

תשובה: הצפיפות המשותפת של X ו-Y היא $f_{X,Y}(x,y) = e^{-x-y}$ עבור x ו-y אי שליליים ואפס אחרת.

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) = 1 - \mathbb{P}(Z > z) = 1 - \mathbb{P}(X > z, Y > z) = 1 - \mathbb{P}(X > z)\mathbb{P}(Y > z) \\ &= 1 - \int_z^\infty e^{-x}dx \int_z^\infty e^{-y}dy = 1 - e^{-2z}. \end{aligned}$$

נגזור ונקבל

$$f_Z(z) = F'_Z(z) = 2e^{-2z}.$$

מכך נסיק כי התפלגות Z היא מעריכית עם פרמטר 2.

באופן כללי חישוב צפיפותו של משתנה המתואר כפונקציה רציפה של שני משתנים מקריים בעלי צפיפות משותפת עשויה להיות משימה מורכבת. ברצף הטענות הבא ננסה להמחיש כיצד לבצע זאת עבור סכום של משתנים.

טענה 9.28. [צפיפות משותפת של סכום] יהיו X ו-Y משתנים מקריים בעלי צפיפות משותפת ויהי סכומם $Z = X + Y$. אזי ל-X ו-Z צפיפות משותפת הנתונה על ידי

$$f_{X,Z}(x,z) = f_{X,Y}(x, z-x).$$

הוכחה.

$$\begin{aligned} \mathbb{P}(X \le a, Z \le b) &= \mathbb{P}(X \le a, X + Y \le b) \\ &= \int_{-\infty}^a \int_{-\infty}^{b-x} f_{X,Y}(x,y)dydx = \int_{-\infty}^a \int_{-\infty}^b f_{X,Y}(x, z-x)dzdx \end{aligned}$$

כנדרש. ■

מטענה זו נוכל להסיק נוסחה לחישוב צפיפותו של סכום משתנים מקריים בלתי-תלויים המכלילה נוסחה מקבילה למשתנים הנתמכים על $\mathbb{Z}$ אשרפיתחנו בבעיה 3.15.

אבחנה 9.29. [נוסחת הקונבולוציה] יהיו X ו-Y משתנים רציפים בהחלט ובלתי-תלויים, אז המשתנה המקרי $Z = X + Y$ הנו רציף בהחלט וצפיפותו היא

$$f_Z(z) = \int_{x \in \mathbb{R}} f_X(x) f_Y(z-x)dx.$$

הוכחה. לפי טענה 9.28 הצפיפות המשותפת של X ו-Z שווה ל-

$$f_{X,Z}(x,z) = f_{X,Y}(x, z-x) = f_X(x) f_Y(z-x).$$

לכן הצפיפות השולית של Z היא

$$f_Z(z) = \int_{x \in \mathbb{R}} f_X(x) f_Y(z-x)dx,$$

כנדרש. ■

דוגמא 9.30 (סכום משתנים מקריים מעריכיים)**.** נתונים משתנים X ו-Y בלתי תלויים המתפלגים מעריכית עם פרמטר 1. נחשב את התפלגותו של $Z = X + Y$.

תשובה: נחשב לפי אבחנה 9.40 עבור $z \geq 0$,

$$f_Z(z) = \int_{x \in \mathbb{R}} f_X(x) f_Y(z-x) dx = \int_0^z e^{-x} e^{x-z} dx = \int_0^z e^{-z} dx = ze^{-z}.$$

ועבור $z \leq 0$ הצפיפות מתאפסת.

כפי שציינו וכפי שנראה בפרק הבא, ההתפלגות הנורמלית מתקבלת לעיתים קרובות כגבול מנורמל של סכום של משתנים מקריים ב"ת. צעד ראשון בדרך לאבחנה זו הוא שימור ההתפלגות הנורמלית תחת סכום, עובדה שנוכיח כעת באמצעות נוסחת הקונבולוציה.

טענה 9.31. [סכום נורמלים הוא נורמלי] יהי X ו-Y משתנים מקריים נורמלים בלתי תלויים בעלי תוחלות μ_X, μ_Y ושונויות σ_X, σ_Y, אזי $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

הוכחה. יהיו $Z_0, Z_1 \sim N(0,1)$ בלתי-תלויים. מאבחנה 9.19 אנו למדים כי

$$(X, Y) \stackrel{d}{=} (\mu_X + \sigma_X Z_0, \mu_Y + \sigma_Y Z_1).$$

להוכחת הטענה נראה כי $\frac{Z_0+Z_1}{\sqrt{2}} \sim N(0,1)$ ונסיק מתכונות התוחלת והשונות (טענות 9.10 ו-5.3) את נכונות הטענה.

כעת נחשב לפי נוסחת הקונבולוציה (אבחנה 9.29),

$$\begin{aligned}
f_{Z_0+Z_1}(z) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx \\
&= \int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(-\frac{(z-x)^2 + x^2}{2}\right) dx \\
&= \int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xz + z^2}{2}\right) dx \\
&= \frac{1}{\sqrt{4\pi}} \int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} \exp\left(-\left(x - \frac{z}{2}\right)^2 + \left(\frac{z}{2}\right)^2\right) dx \\
&= \frac{1}{\sqrt{4\pi}} e^{z^2/4} \int_{\mathbb{R}} \frac{1}{\sqrt{\pi}} \exp\left(-\left(x - \frac{z}{2}\right)^2\right) dx && \left[\begin{array}{rcl} \frac{t}{\sqrt{2}} &=& x - \frac{z}{2} \\ \frac{dt}{\sqrt{2}} &=& dx \end{array}\right] \\
&= \frac{1}{\sqrt{4\pi}} e^{z^2/4} \int_{\mathbb{R}} \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{\sqrt{4\pi}} e^{z^2/4}
\end{aligned}$$

נשתמש בטענה 9.5 ונקבל

$$f_{\frac{Z_0+Z_1}{\sqrt{2}}}(z) = \frac{1}{\sqrt{2\pi}} e^{z^2/2},$$

כנדרש. ■

את הנוסחא לתוחלת של פונקציה של מ"מ (טענה 9.9) נוכל להכליל גם עבור פונקציה של מספר משתנים.

טענה 9.32 (תוחלת של פונקציה של וקטור מקרי)**.** יהיו X, Y משתנים מקריים בעלי צפיפות משותפת $f_{X,Y}$ ותהי $g : \mathbb{R}^2 \to \mathbb{R}$ פונקציה מדידה.

אז $Z = g(X, Y)$ הוא משתנה מקרי המקיים

$$\mathbb{E}(Z) = \iint_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) d(x, y)$$

אם ורק אם האינטגרל מתכנס.

טענה זו תאפשר לנו לחשב את השונות המשותפת של שני משתנים מקריים.

9.3* צפיפות מותנית

נשווה בנפשנו את תיאור הניסוי הבא בגרסא בדידה ובגרסא רציפה:

גרסא בדידה: מספר השנים עד לשריפת יער בקליפורניה מתפלג גיאומטרית עם סיכוי הצלחה של $1/2$. נסמן משתנה זה ב-X'. משעה שמתחוללת השריפה, היא נמשכת פרק זמן של Y' שעות אשר מתפלג בינומית עם $2X$ נסיונות וסיכוי הצלחה $1/2$. חשב את התפלגות משך השריפה.

גרסא רציפה: הזמן בשנים עד לשריפת יער בקליפורניה מתפלג מעריכית עם פרמטר $1/2$. נסמן משתנה זה ב-X. משעה שמתחוללת השריפה, היא נמשכת פרק זמן של Y שעות שמתפלג נורמלית עם תוחלת של X וסטיית תקן של $X/4$. חשב את התפלגות משך השריפה.

קל לראות שהשאלות הללו דומות מאוד וההבדל העיקרי ביניהן הוא ההבדל בין התייחסות לזמן כמשתנה רציף או כמשתנה הנמדד ביחידות בדידות. במענה לשאלה הנוגעת לניסוי בגרסתו הבדידה – איננו נתקלים בכל קושי במענה לשאלה כזו באמצעות נוסחאת ההסתברות השלמה. נוכל לרשום

$$\mathbb{P}(Y = a) = \sum_{x \in \mathbb{N}} \mathbb{P}(X = x)\mathbb{P}(Y = a \mid X = x).$$

ואולם, את הניסוי השני איננו יכולים לתאר בכלים אלא, מפני שהמאורע $X = x$ הנו מאורע בעל הסתברות 0 לכל x ולכן לא ניתן להתנות בו. לבעיה זו יש פתרון בדמות מושג ה**צפיפות מותנית**.

9.3.1 צפיפות מותנית

הגדרה 9.33 (צפיפות מותנית)**.** יהיו X, Y מ"מ רציפים בהחלט בעלי התפלגות משותפת. נסמן את **הצפיפות של X בהינתן Y** על ידי

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

עבור כל $y \in \mathbb{R}$ בו $f_Y(y)$ רציפה אשר מקיים $f_Y(y) \neq 0$.

מהגדרה זו ניתן לראות כי כל פונקציית צפיפות מותנית מגדירה התפלגות משותפת.

דוגמא 9.34 (צפיפות מותנית)**.** נתון (X,Y) וקטור מקרי רציף בהחלט בעל צפיפות

$$f_{X,Y}(x,y) = (x+y)\mathbb{1}([0,1]^2)(x,y).$$

מה הצפיפות המותנית של X בהנתן Y?

תשובה: חישבנו וקיבלנו כי צפיפות Y היא $f_Y(y) = \frac{2y+1}{2}\mathbb{1}([0,1])(y)$. לפי הגדרה, הצפיפות המותנית של X בהנתן Y היא

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{2(x+y)}{2y+1}$$

עבור $0 \le y \le 1$. עבור ערכים אחרים של y הצפיפות המותנית אינה מוגדרת.

טענה 9.35. [הגדרת צפיפות משותפת באמצעות צפיפות מותנית] תהי $g(x,y) : \mathbb{R}^2 \to [0,\infty)$ פונקציה אי-שלילית, אינטגרבילית המקיימת לכל $y \in \mathbb{R}$ את התכונה $\int_{x\in\mathbb{R}} g(x,y)dx = 1$ ותהי $\mathcal{D}$ התפלגות רציפה בהחלט בעלת צפיפות f. אזי קיימים X ו-Y כך ש-$Y \sim \mathcal{D}$ ול-X ול-Y' צפיפות משותפת המקיימת

$$f_{X|Y=y}(x) = g(x,y)$$

הוכחה. נסתכל על הצפיפות

$$f_{X,Y} = f(y)g(x,y)$$

נשים לב כי

$$\int_{\mathbb{R}^2} f_{X,Y}(x,y)dxdy = \int_{\mathbb{R}} f(y)\int_{\mathbb{R}} g(x,y)dxdy = \int_{\mathbb{R}} f(y)dy = 1.$$

ולכן לפי טענה **??** זו צפיפות משותפת של זוג משתנים מקריים. כעת לפי חישוב התפלגות שולית

$$f'_Y(y) = \int_{\mathbb{R}} f_{X,Y'}(x,y)dx = f(y)\int_{\mathbb{R}} g(x,y)dx = f(y).$$

ולפי הגדרת צפיפות מותנית

$$f_{X|Y'=y}(x) = \frac{f_{X,Y'}(x,y)}{f_{Y'}(y)} = \frac{f(y)g(x,y)}{f_{Y'}(y)} = g(x,y).$$

■

דוגמא 9.36 (צפיפות מותנית)**.** יהי X משתנה מקרי המתפלג אחיד בקטע $[0,1]$. יהי Y משתנה מקרי שבהנתן ש-$X = x$ מתפלג אחיד בקטע $[0,x]$. חשב את הצפיפות המשותפת של X ו-Y.

תשובה: הצפיפות המותנית של Y בהנתן X היא

$$f_{Y|X=x}(y) = \frac{1}{x}\mathbb{1}([0,x])(y).$$

לכן הצפיפות המשותפת של X ו-Y היא

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X=x}(y) = \frac{1}{x} \cdot \mathbb{1}(\{(x,y)\ :\ 0 \le y \le x \le 1\})(x,y).$$

באמצעות הצפיפות המותנית נוכל להכליל את נוסחת ההסתברות השלמה (טענה 2.7) ואת כלל בייס (טענה 2.12).

טענה 9.37. [נוסחת הסתברות שלמה רציפה] לכל X, Y בעלי צפיפות משותפת מתקיים

$$f_X(x) = \int_{y\in\mathbb{R}} f_Y(y) f_{X|Y=y}(x) dy$$

כאשר נפרש את הביטוי באינטגרל כשווה לאפס אם $f_Y(y) = 0$.

הוכחה. נרשום את הנוסחא ל-$f_X(x)$ כהתפלגות שולית ונציב את הגדרת הצפיפות המותנית

$$f_X(x) = \int_{y\in\mathbb{R}} f_{X,Y}(x,y)dy = \int_{y\in\mathbb{R}} \frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)dy = \int_{y\in\mathbb{R}} f_Y(y) f_{X|Y=y}(x)dy$$

■

דוגמא 9.38 (צפיפות מותנית)**.** יהי X משתנה מקרי המתפלג אחיד בקטע $[0,1]$. יהי Y משתנה מקרי שבהנתן ש-$X = x$ מתפלג אחיד בקטע $[0,x]$. חשב את הצפיפות של Y.

תשובה: לפי טענה 9.37 נקבל כי לכל $0 < y \le 1$

$$f_Y(y) = \int_{x\in\mathbb{R}} f_X(x) f_{Y|X=x}(y)dx = \int_0^1 \frac{1}{x}\mathbb{1}([0,x])(y)dx = \int_y^1 \frac{1}{x}dx = \log(x)\Big|_{x=y}^{x=1} = -\log(y).$$

אבחנה 9.39. [כלל בייס רציף] לכל X, Y בעלי צפיפות משותפת מתקיים

$$f_{X|Y=y}(x) f_Y(y) = f_{Y|X=x}(y) f_X(x)$$

כאשר נתייחס לאגף שאינו מוגדר מפאת אי קיומה של צפיפיות מותנית כשווה ל-0.

כפי שעשינו עבור משתנים מקריים בדידים, נוכל לרשום קריטריון לאי-תלות בשפה של צפיפות מותנית.

אבחנה 9.40. [תנאי לאי-תלות] מ"מ X ו-Y רציפים בהחלט הנם בלתי תלויים אם ורק אם

$$f_{X|Y=y}(x) = f_X(x)$$

לכל y עבורו $f_Y(y) \neq 0$.

הוכחה. העובדה ש-X ו-Y ב"ת שקולה, לפי טענה 9.25, לקיום פונקציית צפיפות משותפת

$$f_{X,Y}(x,y) = f_X(x) f_Y(y) = f_{X|Y=y}(x) f_Y(y)$$

נצמצם ונקבל $f_{X|Y=y}(x) = f_X(x)$ כנדרש. ■

בעיות הרחבה והעשרה

✎ **בעיה 9.5.** יהי Y משתנה מקרי מעריכי עם פרמטר λ, כיצד מתפלג $\lfloor Y \rfloor$?

✎ **בעיה 9.6.** יהיו X, Y נורמלים סטנדרטים בלתי-תלויים. יש להוכיח כי $\sqrt{X^2+Y^2} \sim \exp(1)$.

✎ **בעיה 9.7.** יהיו $X, Y \sim \exp(1)$ משתנים מקריים מעריכיים ויהי Z משתנה אחיד על $[0,1]$ יש להוכיח כי $Z(X+$ [illegible].

✎ **בעיה 9.8.** יהיו $X, Y, Z \sim \mathrm{Unif}([0,1])$ משתנים מקריים אחידים בלתי-תלויים. יש להוכיח כי $(XY)^Z \sim \mathrm{Unif}([0,1])$

✎ **בעיה 9.9 (*).** יהי Z משתנה מקרי פואסוני סטנדרטי (כלומר עם $\lambda = 1$), יהיו $\{X_n\}_{n\in\mathbb{N}}$ משתנים מקריים אחידים על $[0,1]$, בלתי תלויים ב-Z וזה בזה ויהי W משתנה מעריכי סטנדרטי מותנה במאורע $W \le 1$. הוכח כי $W \overset{d}{=} \min(X_1,\dots,X_Z)$.

פרק 10

סדרות של התפלגויות

כמעט שאינני מכיר דבר-מה אשר מותיר רושם כה עז על הדמיון כמו תבנית הסדר הקוסמי המתבטאת באמצעות חוק שכיחות הסטיות. החוק היה זוכה להאנשה על ידי היוונים אילו הכירוהו. הוא מושל בשלווה ובצניעות מושלמת בלב התוהו הפראי ביותר. ככל שרב האספסוף, ניכרת יותר האנארכיה, כך נשעית שליטתו מושלמת. זהו החוק העליון שבשיגעון.

– סר פרנסיס גלטון על משפט הגבול המרכזי, "תורשה טבעית", 1889

במובן מסויים של התכנסות משתנים מקריים נתקלנו בפרק 3, בדוגמא 3.43. שם רצינו להראות שהתפלגות מספר המטבעות המוטלים לפני תוצאה ראשונה של עץ (או סיום הטלת כל המטבעות), כאשר מטילים N מטבעות, מתכנסת להתפלגותו של משתנה גיאומטרי. קשרים דומים שימשו אותנו לניסוח הקשר בין התפלגות בינומית לפואסונית (בעיה 3.28) ולניסוח החוק החלש של המספרים הגדולים (משפט 5.8). בפרק זה נגדיר מושג זה של התכנסות המכונה **התכנסות התפלגויות** או **התכנסות בהתפלגות** ונחקור תכונות אוניברסליות שלו.

10.1 התכנסות התפלגויות

את מושג ההתכנסות עבור התפלגויות נגדיר במונחים של פונקציות התפלגות מצטברות.

הגדרה 10.1 (התכנסות בהתפלגות)**.** תהי $(X_n)_{n\in\mathbb{N}}$ סדרת משתנים מקריים, לא בהכרח על אותו מרחב הסתברות, ויהי X משתנה מקרי. נאמר כי $(X_n)_{n\in\mathbb{N}}$ מתכנסת **בהתפלגות (in distribution)** ל-X, ונסמן $X_n \xrightarrow{\text{d}} X$ אם לכל a שהיא **נקודת רציפות** של F_X מתקיים

$$\lim_{n\to\infty} F_{X_n}(a) = F_X(a).$$

התכנסות בהתפלגות היא תכונה חלשה יותר מ-"התכנסות של ההסתברות ש-X_n נמצאת ב-I לכל קטע ב-$\mathbb{R}$". בכדי להבין את המוטיבציה להגדרה המורכבת והחלשה יותר, נסתכל על הדוגמא הבאה.

דוגמא 10.2. תהי $X_n \overset{\text{a.s.}}{=} c_n$ סדרת משתנים מקריים קבועים, כאשר $c_n \to 0$ ויהי $X \overset{\text{a.s.}}{=} 0$. פונקציית ההתפלגות

המצטברות של הסדרה הן

$$F_{X_n}(a) = \begin{cases} 0 & a < c_n \\ 1 & a \geq c_n \end{cases}$$

ואומנם $\lim_{n\to\infty} F_{X_n}(a) = F_X(a)$ בכל נקודה מלבד 0, נקודת האי-רציפות היחידה של F_X. ניתן לומר שההגדרה המורכבת שבחרנו נועדה לאפשר גם להתכנסות כזו להיקרא התכנסות בהתפלגות, ואכן לפי הגדרה זו $X_n \xrightarrow{d} X$.

דוגמא 10.3 (דוגמאות פשוטות להתכנסות בהתפלגות).

(א) אם X_n, X כולם משתנים מקריים שווי-התפלגות, אז $X_n \xrightarrow{d} X$.

(ב) אם $\mathbb{P}(X_n = c_n) = 1$ ו- $\mathbb{P}(X = c) = 1$, אז $X_n \xrightarrow{d} X$ אם ורק אם $c_n \to c$.

(ג) אם $X_n \sim \mathrm{Unif}([-\frac{1}{n}, \frac{1}{n}])$ אז $X_n \xrightarrow{d} 0$. כך גם לגבי $X_n \sim \mathrm{Unif}([0, \frac{1}{n}])$ ו- $X_n \sim \mathrm{Unif}([\frac{1}{n}, \frac{2}{n}])$. נשים לב שבדוגמאות אלה הערכים של $F_{X_n}(0)$ שונים ולא חייבים להתכנס ל- $F_X(0) = 1$. כיוון ש-0 היא נקודת אי-רציפות של F_X הדבר עולה בקנה אחד עם הגדרת התכנסות בהתפלגות.

כעת נראה כיצד ההתפלגויות הרציפות בהחלט שלמדנו להכיר בפרק 9, מתקבלות כגבול של התפלגויות בדידות. בכך נצדיק את השימושים שהצגנו עבור התפלגויות אלו.

דוגמא 10.4.

(א) $X_n \sim U(\{\frac{1}{n}, \frac{2}{n}, \ldots, 1\})$, $X \sim U([0,1])$. כל a היא נקודת רציפות של F_X. עבור $a > 1$ או $a < 0$ ההתכנסות טריויאלית ואילו עבור $0 < a < 1$ מתקיים $\mathbb{P}(X_n \leq a) = \frac{\lfloor na \rfloor}{n} \overset{n\to\infty}{\to} a = \mathbb{P}(X \leq a)$, ולכן $X_n \xrightarrow{d} X$.

(ב) $X_n \sim \mathrm{Bin}(n, \frac{1}{n})$, $X \sim \mathrm{Po}(1)$. לכל $k \in \mathbb{N}_0$ נחשב את הגבול:

$$\mathbb{P}(X_n = k) = \binom{n}{k}\frac{1}{n^k}\left(1 - \frac{1}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{n^k}\frac{1}{k!}\left(1 - \frac{1}{n}\right)^{n-k} \overset{n\to\infty}{\to} \frac{e^{-1}}{k!} = \mathbb{P}(X = k).$$

קבוצת נקודות הרציפות של F_X היא $\mathbb{R} \setminus \mathbb{N}$. אם נקח $a \in (k, k+1)$ עבור k טבעי, אז

$$F_{X_n}(a) = \sum_{j=0}^{\lfloor a \rfloor} \mathbb{P}(X_n = j) \overset{n\to\infty}{\to} \sum_{j=0}^{\lfloor a \rfloor} \mathbb{P}(X = j) = F_X(a),$$

ולכן $X_n \xrightarrow{d} X$. בדוגמה זו $F_{X_n}(a) \to F_X(a)$ גם בנקודות האי-רציפות של F_X.

(ג) אם $X_n \sim \mathrm{Geo}(\frac{1}{n})$ ו$X \sim \mathrm{Exp}(1)$ אזי $Y_n := \frac{1}{n}X_n \xrightarrow{d} X$. נקח $a > 0$.

$$\mathbb{P}(Y_n \leq a) = \sum_{k=1}^{\lfloor na \rfloor}\left(1 - \frac{1}{n}\right)^k \frac{1}{n} = \frac{1 - (1 - \frac{1}{n})^{na+1}}{1 - (1 - \frac{1}{n})}\frac{1}{n} \overset{n\to\infty}{\to} 1 - e^{-a} = F_X(a).$$

נפקדת יחידה מרשימת דוגמאות זו היא ההתפלגות הנורמלית. ההתכנסות להתפלגות זו מבוארת בפרק הבא.

נשים לב שהדרישה שהסדרה F_{X_n} תתכנס ל-F_X, פונקציית התפלגות מצטברת של משתנה מקרי, אינה מובנת מאליה. לעיתים תיתכן התכנסות של F_{X_n} לפונקציה שכלל אינה פונקציית התפלגות של אף משתנה. כפי שנראה בדוגמא הבאה, תופעה כזו יכולה להתרחש כאשר X_n נוטים להתפזר יותר ויותר רחוק מן הראשית. התכנסות של F_{X_n} לפונקציה כללית נקראת לפעמים **התכנסות חלשה של התפלגויות**.

דוגמא 10.5.

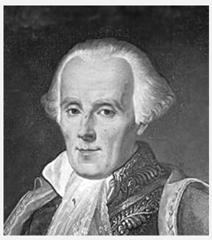
(א) נסתכל על סדרת משתנים מקריים קבועים $X_n \equiv n$. $F_{X_n}(s) = 1\!\!1(s \geq n)$ ולכן $\lim_{n\to\infty} F_{X_n} \equiv 0$. אך 0 אינה פונקציית התפלגות מצטברת של אף משתנה מקרי. מיחידות הגבול נסיק כי X_n אינם מתכנסים בהתפלגות לשום משתנה מקרי.

(ב) נסתכל על סדרת משתנים מקריים אחידים $X_n \sim \mathrm{Unif}([-n,n])$. $F_{X_n}(s) = \min\Big(1, \max\Big(0,$ [illegible] $\Big)\Big)$. לכל a מתקיים $\lim_{n\to\infty} F_{X_n}(a) = \frac{1}{2}$. ולכן $\lim_{n\to\infty} F_{X_n} \equiv \frac{1}{2}$. זו אינה פונקציית התפלגות מצטברת של אף משתנה מקרי. מיחידות הגבול נסיק כי X_n אינם מתכנסים בהתפלגות לשום משתנה מקרי.

10.2 משפט הגבול המרכזי

אחת התוצאות המרשימות ביותר של תורת ההסתברות, אשר תהילתה מובאת בציטוט שבתחילת הפרק, היא **משפט הגבול המרכזי**. משפט זה קובע כי כאשר מספר גדול של גורמים בלתי-תלויים בעלי שונות נסכמים, אז לאחר נרמול בסטיית התקן של הסכום - סטייתו מהתוחלת שואפת תמיד להתפלגות נורמלית. זו דוגמא לתופעה של אוניברסליות, כלומר מצב שבו תנאי ההתחלה המתמטיים אינם משפיעים על התוצאה. תוצאות כאלו זוכות לעיתים קרובות לשימושים רבים בכל תחומי המדע.

גרסא בסיסית של המשפט הייתה ידועה כבר לאברהם דה-מואבר ב-1733, והוא הורחב לתנאים מקלים יותר ויותר, כאשר הגרסא הכללית הראשונה פורסמה על ידי פְּיֵיר-סִימוֹן לָפְּלָס בראשית המאה התשע-עשרה. גרסאותיו המודרניות מאפשרות גם סכימה של משתנים בעלי תלויות ובלבד שהללו תהיינה חלשות דיו. כאן נביא את גרסתו הקלאסית והמוכרת ביותר של המשפט.



פְּיֵיר-סִימוֹן לָפְּלָס (1749-1827), אציל צרפתי, מתמטיקאי ופיסיקאי, שחולל מהפכה בהבנת מדעי הטבע של תקופתו באמצעות אימוץ גישה אנליטית וחקר משוואות דיפרנציאליות. פיתח את תורת הפוטנציאל והיה מהראשונים לשער את קיומם של חורים שחורים. ספרו תיאוריה אנליטית של הסתברות (Théorie analytique des probabilités) היווה ניסיון ראשון לבסס את תורת ההסתברות על יסודות אקסיומטיים והפך להיות ספר ההסתברות החשוב ביותר עד סוף המאה התשע-עשרה. בשני מאמרים מפורסמים מ-1810 ו-1811 ניסח בכלים של אנליזה הארמונית גרסא הכללית הראשונה למשפט הגבול המרכזי.

משפט 10.6 (משפט הגבול המרכזי)**.** תהי $\{X_n\}_{n\in\mathbb{N}}$ סדרת משתנים מקריים בלתי-תלויים ושווי התפלגות, בעלי תוחלת 0 ושונות 1. אזי

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i \xrightarrow{\mathrm{d}} Z$$

כאשר Z הוא משתנה מקרי נורמלי סטנדרטי.

באופן שקול, לכל $a\in\mathbb{R}$ מתקיים

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\le a\right)\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}e^{-x^2/2}dx.$$

גרסא כמותית של המשפט הוכחה ב-1941 על ידי אנדרו ברי (Andrew Berry) וקרל-גוסטב אסין (Carl-Gustav Esseen).

משפט 10.7 (משפט ברי-אסין)**.** תהי $\{X_n\}_{n\in\mathbb{N}}$ סדרת משתנים מקריים בלתי-תלויים ושווי התפלגות, בעלי תוחלת 0, שונות 1 ומומנט מוחלט שלישי $\rho=\mathbb{E}\left[\left|X_1^3\right|\right]<\infty$. אזי לכל $a\in\mathbb{R}$ מתקיים

$$\left|\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\le a\right)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}e^{-x^2/2}dx\right|\le\frac{\rho}{\sqrt{n}}.$$

דוגמא 10.8. תהי $X_n\sim\mathrm{Bin}(n,p)$ סדרת משתנים מקריים בינומים ויהי $X\sim N(0,1)$. נראה כי

$$\frac{X_n-pn}{\sqrt{p(1-p)n}}\xrightarrow{\mathrm{d}}X.$$

תשובה: יהיו $\{Y_n\}_{n\in\mathbb{N}}$ משתנים בלתי-תלויים ושווי התפלגות כך שמתקיים $Y_1\sim\mathrm{Ber}(p)$. נשים לב כי מתקיים $X_n=\sum_{i=1}^{n}Y_i$ ולכן

$$\frac{X_n-pn}{\sqrt{p(1-p)n}}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Y_i-\mathbb{E}(Y_i)}{\sqrt{\mathrm{Var}(Y_i)}}$$

נסמן $Z_i=\frac{Y_i-\mathbb{E}(Y_i)}{\sqrt{\mathrm{Var}(Y_i)}}$ ונשים לב כי אלו משתנים מקריים בלתי-תלויים בעלי תוחלת 0 ושונות 1. נקבל כי לפי משפט הגבול המרכזי

$$\frac{X_n-pn}{\sqrt{p(1-p)n}}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\xrightarrow{\mathrm{d}}X,$$

כנדרש.

10.3 התכנסות בהתפלגות והתכנסות בהסתברות

משפט 10.9 (התכנסות בהסתברות גוררת התכנסות בהתפלגות)**.** תהי X_n סדרת משתנים מקריים ויהי X משתנה מקרי, כולם על אותו מרחב הסתברות. אם $X_n\xrightarrow{\mathrm{p}}X$ אז $X_n\xrightarrow{\mathrm{d}}X$.

הוכחה. תהי a נקודת רציפות של F_X, ויהי $\epsilon > 0$.

$$\begin{aligned}\mathbb{P}(X_n \le a) &= \mathbb{P}(X_n \le a, X \le a+\epsilon) + \mathbb{P}(X_n \le a, X > a+\epsilon) \\ &\le \mathbb{P}(X \le a+\epsilon) + \mathbb{P}(|X - X_n| > \epsilon)\end{aligned}$$

ולכן $F_{X_n}(a) \le F_X(a+\epsilon) + \mathbb{P}(|X - X_n| > \epsilon)$. בצורה דומה

$$\mathbb{P}(X \le a-\epsilon) = \mathbb{P}(X \le a-\epsilon, X_n \le a) + \mathbb{P}(X \le a-\epsilon, X_n > a) \le \mathbb{P}(X_n \le a) + \mathbb{P}(|X - X_n| > \epsilon)$$

ולכן $F_X(a-\epsilon) \le F_{X_n}(a) + \mathbb{P}(|X - X_n| > \epsilon)$. קיבלנו

$$F_X(a-\epsilon) - \mathbb{P}(|X - X_n| > \epsilon) \le F_{X_n}(a) \le F_X(a+\epsilon) + \mathbb{P}(|X - X_n| > \epsilon)$$

נשאיף את n לאינסוף ונקבל מהגדרת ההתכנסות בהסתברות כי לכל $\epsilon > 0$ מתקיים

$$F_X(a-\epsilon) \le \liminf_{n\to\infty} F_{X_n}(a) \le \limsup_{n\to\infty} F_{X_n}(a) \le F_X(a+\epsilon).$$

ומכך ש-a נקודת רציפות נקבל $F_X(a) \le \lim_{n\to\infty} F_{X_n}(a) \le F_X(a)$. ■

ההיפך אינו נכון. למשל, אם $(X_n)_{n\in\mathbb{N}}$ הם משתנים מקריים בלתי תלויים ושווי התפלגות שאינם קבועים, אז $X_n \xrightarrow{\text{d}} X$, כאשר X משתנה מקרי נוסף בעל אותה התפלגות, אך $X_n \overset{\text{p}}{\not\to} X$. עם זאת, מתקיים המשפט הבא.

משפט 10.10 (התכנסות בהתפלגות לקבוע גוררת התכנסות בהסתברות). תהי X_n סדרת משתנים מקריים, יהי X משתנה מקרי ויהי $c \in \mathbb{R}$. אם $X_n \xrightarrow{\text{d}} c$ אז $X_n \xrightarrow{\text{p}} c$.

הוכחה. תהי F_c פונקצית ההתפלגות המצטברת של המשתנה המקרי הקבוע c. לכל $\epsilon > 0$ מתקיים $F_c(c-\epsilon) = 0$ ו-$F_c(c+\epsilon) = 1$. אם $X_n \xrightarrow{\text{d}} c$ אז $F_{X_n}(c-\epsilon) \to 0$ ו-$F_{X_n}(c+\epsilon) \to 1$ ולכן

$$\mathbb{P}(|X_n - c| \le \epsilon) \ge F_{X_n}(c+\epsilon) - F_{X_n}(c-\epsilon) \to 1,$$

כנדרש. ■

10.4* משמעותה של התכנסות התפלגויות

הדרישה בהגדרה של התכנסות בהתפלגות של סדרת משתנים מתייחסת רק להתכנסות של F_{X_n} בנקודות רציפות. ראינו בדוגמא 10.2 שזוהי דרישה משמעותית. ראשית נציג הגדרה חילופית להתכנסות בהתפלגות שאינה מערבת סוגיות רציפות של F_{X_n}.

טענה 10.11. תהי $(X_n)_{n\in\mathbb{N}}$ סדרת משתנים מקריים. X_n מתכנסת בהתפלגות ל-X אם ורק אם לכל $\epsilon > 0$ ולכל $r \in \mathbb{R}$ מתקיים:

$$F_X(r-\epsilon) \le \liminf_{n\to\infty} F_{X_n}(r) \le \limsup_{n\to\infty} F_{X_n}(r) \le F_X(r+\epsilon)$$

הוכחה. קל לראות כי בנקודות רציפות של F_X הטענה שקולה לכך ש-$\lim_{n\to\infty} F_{X_n}(r) = F_X(r)$, ולכן התנאי בטענה 10.11 חזק יותר מאשר התנאי בהגדרה 10.1. בכיוון ההפוך, עלינו להראות כי אם $X_n \xrightarrow{\mathrm{d}} X$, אז בכל נקודת אי-רציפות r מתקיים תנאי הטענה. נשים לב כי $F_X(t)$ הנה פונקציה מונוטונית עולה ולכן יש לה מספר בן-מניה של נקודות אי-רציפות (ר' טענה א.5). יהי $\epsilon > 0$. כיוון שמספר נקודות אי-הרציפות הנו בן מניה, קיימות נקודות $b - \epsilon \le b_- \le b \le b_+ \le b + \epsilon$ שהן נקודות רציפות של F_X. לכן מתקיים לפי תכונת ההתכנסות בהתפלגות

$$F_X(b-\epsilon) \le F_X(b_-) = \lim_{n\to\infty} F_{X_n}(b_-) \le \liminf_{n\to\infty} F_{X_n}(b) \le \limsup_{n\to\infty} F_{X_n}(b) \le F_X(b+\epsilon),$$

כנדרש.

■

✎ **בעיה 10.1.** להכליל את טענה 10.11 ולהראות כי התכנסות X_n מתכנסת בהתפלגות ל-X אם ורק אם לכל $\epsilon > 0$ ולכל $[a,b]$ קיים n_0 כך שלכל $n > n_0$ מתקיים

$$\mathbb{P}(X \in (a+\epsilon, b-\epsilon)) \le \mathbb{P}(X_n \in (a,b)) \le \mathbb{P}(X \in (a-\epsilon, b+\epsilon)).$$

נציג ללא הוכחה את הטענה הבאה המסכמת את מהותה של התכנסות בהתפלגות.

טענה 10.12. תהי $(X_n)_{n\in\mathbb{N}}$ סדרת משתנים מקריים ויהי X משתנה מקרי. אזי $X_n \xrightarrow{\mathrm{d}} X$ אם ורק אם לכל $f : \mathbb{R} \to \mathbb{R}$ רציפה $\mathbb{E}(f(X_n)) \to \mathbb{E}(X)$.

✎ **בעיה 10.2.** להוכיח את טענה 10.11 עבור משתנים מקריים רציפים בהחלט.

Properties [edit]

- Since $F(a) = \Pr(X \le a)$, the convergence in distribution means that the probability for X_n to be in a given range is approximately equal to the probability that the value of X is in that range, provided n is sufficiently large.
- In general, convergence in distribution does not imply that the sequence of corresponding probability density functions will also converge. As an example one may consider random variables with densities $f_n(x) = (1 - \cos(2\pi nx))\mathbf{1}_{(0,1)}$. These random variables converge in distribution to a uniform $U(0, 1)$, whereas their densities do not converge at all.[3]
  - However, according to *Scheffé's theorem*, convergence of the probability density functions implies convergence in distribution.[4]
- The **portmanteau lemma** provides several equivalent definitions of convergence in distribution. Although these definitions are less intuitive, they are used to prove a number of statistical theorems. The lemma states that $\{X_n\}$ converges in distribution to X if and only if any of the following statements are true:[5]
  - $\Pr(X_n \le x) \to \Pr(X \le x)$ for all continuity points of $x \mapsto \Pr(X \le x)$;
  - $\operatorname{E} f(X_n) \to \operatorname{E} f(X)$ for all bounded, continuous functions f (where E denotes the expected value operator);
  - $\operatorname{E} f(X_n) \to \operatorname{E} f(X)$ for all bounded, Lipschitz functions f;
  - $\liminf \operatorname{E} f(X_n) \ge \operatorname{E} f(X)$ for all nonnegative, continuous functions f;
  - $\liminf \Pr(X_n \in G) \ge \Pr(X \in G)$ for every open set G;
  - $\limsup \Pr(X_n \in F) \le \Pr(X \in F)$ for every closed set F;
  - $\Pr(X_n \in B) \to \Pr(X \in B)$ for all continuity sets B of random variable X;
  - $\limsup \operatorname{E} f(X_n) \le \operatorname{E} f(X)$ for every upper semi-continuous function f bounded above;[citation needed]
  - $\liminf \operatorname{E} f(X_n) \ge \operatorname{E} f(X)$ for every lower semi-continuous function f bounded below.[citation needed]

בעיות הרחבה והעשרה

✎ **בעיה 10.3.** יהיו $X_n \sim \mathrm{Unif}[0,1]$ ונסמן $Y_n = \sum_{i=1}^n X_i$ וכן $Z_n = Y_n - \lfloor Y_n \rfloor$. יש להראות כי $Z_n \xrightarrow{\mathrm{d}} Z$ עבור $Z \sim \mathrm{Unif}[0,1]$.

✎ **בעיה 10.4.** לחשב את הגבול

$$\lim_{n\to\infty} e^{-n} \sum_{i=0}^{n} \frac{n^i}{i!}.$$

רמז: לכל n הביטוי הוא הסתברותו של מאורע הקשור למשתנה מקרי מסוים. ניתן להשתמש במשפט הגבול המרכזי בכדי לחשב את הגבול.

חלק III

יישומי תורת ההסתברות

פרק 11

דוגמאות מתקדמות

לעיתים קרובות, כשרעיון מועיל מציג את עצמו – איננו מעריכים אותו מפני שהוא כמעט בלתי-מורגש. יתכן שלמומחה אין רעיונות רבים יותר מאשר לחסר-הניסיון, אבל הוא מעריך במידה רבה יותר את מה שיש לו, ומיטיב להשתמש בו.

–ג'ורג' פוליה, "איך פותרים את זה?", 1945

בפרקים הקודמים נתוודענו ליסודות תורת ההסתברות על הגדרותיה ותוצאותיה. בפרק זה נכיר כמה מודלים מורכבים יותר ונשתמש בשיטות שפיתחנו בפרקים הקודמים בכדי לנתחם. מטרתו של הפרק היא להציג את התורה ההסתברותית במלא הדרה ולהשתמש באופן מתקדם יחסית בכלים שרכשנו. אין לראותו כפרק תרגול של מושגים בסיסיים אלא כפרק המעדן את תפיסת המושגים ומחדד את דקויותיהם.

11.1 כדי פוליה

בשנת 1923 הציע המתמטיקאי ההונגרי היהודי ג'ורג' פוליה (George Pólya) את המודל הבא לאימוץ דעות בקבוצה.

ניסוי 11.1. [כדי פוליה] כד מכיל בזמן 0 שני כדורים - אחד אדום ואחד שחור. בזמן 1 נשלף כדור מקרי מהכד ומוכנס שוב ביחד עם כדור נוסף מאותו הצבע. התהליך חוזר על עצמו עד זמן n, (בו הצטברו בכד $n+2$ כדורים). נסמן ב- R_n את מספר הכדורים האדומים בכד ו-B_n את מספר הכדורים השחורים.

לשיטתו של פוליה צבע הכדורים יכול לייצג למשל את דעותיהם הפוליטיות של חברי הקבוצה (שמאל או ימין) וכל חבר חדש מושפע בדעותיו מאחד מחברי הקבוצה שאותו שאל לדעתו. המודל בנוי כך שהסיכוי להוסיף כדור אדום בצעד ה-$n+1$ הוא בדיוק

$$\frac{R_n}{n+2}.$$

נתאר את הניסוי בצורה פורמלית ובאופן שישמר מידע רב יותר על הכדורים מאשר אנו זקוקים לו. בהמשך נראה כיצד תיאור זה ישרת את צרכינו.

ניסוי 11.2. [כדי פוליה - תיאור פורמלי] תהי $\sigma_0 = (1)$ תמורת הזהות על איבר אחד. נגדיר σ_n תמורה מקרית על $n+1$ איברים באופן אינדוקטיבי על ידי כך שלכל $k \in [n+1]$ נקבע

$$\mathbb{P}\Big(\sigma_n = (s_{n-1}(1), \ldots, s_{n-1}(k-1), n+1, s_{n-1}(k), \ldots, s_{n-1}(n)) \mid \sigma_{n-1} = s_{n-1}, \ldots, \sigma_1 = (1)\Big)$$
$$= \mathbb{P}\Big(\sigma_n = (s_{n-1}(1), \ldots, s_{n-1}(k-1), n+1, s_{n-1}(k), \ldots, s_{n-1}(n)) \mid \sigma_{n-1} = s_{n-1}\Big) = \frac{1}{n+1}$$

כעת נקבע $R_n := \sigma_n^{-1}(1)$ ו-$B_n := n+2-\sigma_n^{-1}(1)$.

תיאור זה של התהליך מתאר את הכדורים בכד כאילו הם יושבים לאורך מערך כאשר במצב ההתחלתי כדור אדום נמצא במקום ה-0 וכדור שחור נמצא במקום ה-1. בכל שלב נבחר כדור מקרי ובאופן שקול ומוכנס כדור לימינו. צבעו של הכדור החדש הופך להיות צבע הכדור שלשמאלו. באופן זה, מספר הכדורים האדומים הוא מספר הכדורים שנמצאים לשמאלו של הכדור השחור הראשון. נשים לב שמספרו של כדור מעיד באיזה תור הוא נכנס.

מעתה עבור תמורה $s \in S_n$ נרשום $s^- \in S_{n-1}$ עבור התמורה על $n-1$ איברים שמתקבלת מ-s על ידי השמטת האיבר ה-n, וכן $s^{-(2)} = s^{--}$, $s^{-(3)} = s^{---}$ וכן הלאה.

טענה 11.3. σ_n היא תמורה מקרית אחידה על $n+1$ איברים.

הוכחה. נוכיח את הטענה באינדוקציה. בסיס האינדוקציה עבור $n=0$ הוא מיידי. נניח נכונות עבור $n-1$ ונוכיח עבור n. נשים לב שלכל $s \in S_{n+1}$, יתכן כי $\sigma_n = s$ רק אם $\sigma_{n-1} = s^-$. לכן,

$$\mathbb{P}(\sigma_n = s) = \mathbb{P}(\sigma_{n-1} = s^-)\mathbb{P}(\sigma_n = s \mid \sigma_{n-1} = s^-) = \frac{1}{n!} \cdot \frac{1}{n+1} = \frac{1}{(n+1)!},$$

כנדרש. ■

טענה 11.4. נראה כי $R_n, B_n \sim \mathrm{Unif}[n+1]$.

הוכחה. נחשב לפי טענה 11.3 לכל $k \in \{0, \ldots, n+1\}$ כי

$$\mathbb{P}(R_n = k) = \mathbb{P}(\sigma_n^{-1}(1) = k+1) = \frac{1}{n+1}$$
$$\mathbb{P}(B_n = k) = \mathbb{P}(R_n = n+2-k) = \frac{1}{n+1}$$

■

נוכח טענה 11.4, נוכל להראות בקלות כי אחוז הכדורים האדומים בכד מתכנס בהתפלגות להתפלגות אחידה.

טענה 11.5. יהי $X \sim \mathrm{Unif}([0,1])$, אזי $\frac{R_n}{n+2} \xrightarrow{\mathrm{d}} X$.

הוכחה. נחשב
$$F_{R_n}(k) = \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{1}(k \le i)$$
ולכן -
$$F_{R_n/(n+2)}(r) = \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbb{1}\left(r \le \frac{i}{n+2}\right)$$
ואכן לכל $r \in [0,1]$ מתקיים
$$\lim_{n\to\infty} F_{\frac{R_n}{n+2}}(s) = \lim_{n\to\infty} \frac{1}{n+1}\sum_{i=1}^{\lfloor (n+2)s \rfloor} 1 = \lim_{n\to\infty} \frac{\lfloor (n+2)s \rfloor}{n+1} = s = F_X(s).$$

■

לאחר שהראינו התכנסות של סדרת ההתפלגויות של $\frac{R_n}{n+2}$ נבקש להראות גם התכנסות כמעט-תמיד. אלא שחישוב התנהגותו של התהליך בעתיד בהינתן מצבו בעבר – קשה ומורכב. מסיבה זו נבצע ראשית את החישוב בכיוון ההפוך. נחשב את ההסתברות לכך ש-R_m מקבל ערך מסויים בהינתן ש-R_n מקבל ערך אחר עבור $n > m$.

טענה 11.6. יהיו $n > m$ ונסמן $b_i = i + 2 - r_i$. אזי
$$\mathbb{P}(R_m = r_m + 1 \mid R_n = r_n + 1) = \frac{\binom{r_n}{r_n - r_m}\binom{b_n}{b_n - b_m}}{\binom{n}{m}} \le \binom{m}{r_m}\left(\frac{r_n}{n}\right)^{r_m}\left(\frac{b_n}{n}\right)^{b_m}\frac{n^m}{n\cdots(n-m+1)}.$$

הוכחה. נסמן $S_{n,m,r_n,r_m} = \{s \in S_n \ : \ (s^{-(n-m)})^{-1}(1) = r_m + 1, s^{-1}(1) = r_n + 1\}$. זוהי בדיוק קבוצת התמורות s כך שאם $\sigma_n = s$ אז $R_n = r_n + 1$ ו- $R_m = r_m + 1$. לכן, כיוון שלפי טענה 11.3 התפלגותה של σ_n אחידה, נוכל לרשום לפי נוסחת ההסתברות השלמה (טענה 2.7)
$$\mathbb{P}(R_m = r_m + 1 \mid R_n = r_n + 1) = \sum_{s \in S_{n,m,r_n,r_m}} \mathbb{P}(R_m = r_m + 1 \mid \sigma_n = s) = \frac{|S_{n,m,r_n,r_m}|}{n!}.$$

נחשב את $|S_{n,m,r_n,r_m}|$ באמצעות חישוב קומבינטורי. עבור $s \in S_{n+1}$ נסמן $t = s^{-1}$ ונשים לב ש-$s \in$ S_{n,m,r_n,r_m} אם ורק אם $|\{i \in \{m+1,\dots,n\} : t(i) < t(1) = r_n + 1\}| = r_n - r_m$. נחשב כמה תמורות $t \in S_n$ כאלו יש על ידי בחירה של $r_n - r_m$ מקומות עבור התמונות של איברים גדולים מ-m בין המספרים 1 עד r_n, בחירה של $b_n - b_m$ של מקומות עבור התמונות של איברים גדולים מ-m בין המספרים r_n עד $n+1$, סידור האיברים גדולים מ-m במקומות אלו וסידור יתר האיברים. סך הכל
$$S = \binom{r_n}{r_n - r_m}\binom{b_n}{b_n - b_m} m!(n-m)!.$$

נציב ונקבל כי
$$\mathbb{P}(R_m = r_m + 1 \mid R_n = r_n + 1) = \frac{\binom{r_n}{r_n - r_m}\binom{b_n}{b_n - b_m}}{\binom{n}{m}}$$

נפתח את המקדמים הבינומיים לקבלת

$$\begin{aligned}\mathbb{P}(R_m = r_m + 1 \mid R_n = r_n + 1) &= \frac{m! r_n! b_n! (n-m)!}{r_m! b_m! (r_n - r_m)! (b_n - b_m)! n!} \\ &= \binom{m}{r_m} \frac{r_n \cdots (r_n - r_m + 1) b_n \cdots (b_n - b_m + 1)}{n \cdots (m+1)} \\ &\le \binom{m}{r_m} \frac{{r_n}^{r_m} {b_n}^{b_m}}{n^m} \frac{n^m}{n \cdots (n-m+1)} \\ &= \binom{m}{r_m} \left(\frac{r_n}{n}\right)^{r_m} \left(\frac{b_n}{n}\right)^{b_m} \frac{n^m}{n \cdots (n-m+1)}\end{aligned}$$

. ■

באמצעות טענה 11.8 ביחד עם כלל בייס (טענה 2.12) נוכל לחשב

אבחנה 11.7. יהיו $m > n$ ונסמן $b_i = i + 2 - r_i$. אזי

$$\begin{aligned}\mathbb{P}(R_n = r_n + 1 \mid R_m = r_m + 1) &= \frac{\mathbb{P}(R_m = r_m + 1 \mid R_n = r_n + 1)\mathbb{P}(R_n = r_n + 1)}{\mathbb{P}(R_m = r_m + 1)} \\ &= \frac{\binom{r_n}{r_n - r_m}\binom{b_n}{b_n - b_m}}{\binom{n}{m}} \frac{m+1}{n+1}.\end{aligned}$$

באמצעות הערכה מדוייקת זו נוכל לאמוד את ההסתברות לסטיה גדולה מהערך הנוכחי בעתיד.

טענה 11.8. יהיו $m_1 < m_2$. אזי

$$\mathbb{P}\left(\left|\frac{R_{m_1}}{m_1} - \frac{R_{m_2}}{m_2}\right| > \epsilon\right) \le 4 \exp\left(-\frac{\epsilon^2 m_1}{1} 6\right).$$

הוכחה. ראשית, נשים לב כי לפי טענה 11.8, לכל $\epsilon > 0$, $n > m, r_n \le n, i \in \{1, 2\}$ מתקיים:

$$\mathbb{P}\left(\left|\frac{R_{m_i}}{m_i} - \frac{r_n + 1}{n}\right| \ge \epsilon \,\middle|\, R_n = r_n + 1\right) \le \sum_{nr_{m_i} \notin [m_i(r_n + 1 - \epsilon), m_i(r_n + 1 + \epsilon)]} \binom{m}{r_m} \left(\frac{r_n}{n}\right)^{r_m} \left(\frac{b_n}{n}\right)^{b_m} \frac{n^m}{n \cdots (n-m+1)},$$

כאשר הסכום באגף ימין נע על פני ערכי r_{m_i} המקיימים את התנאי הנקוב. ניזכר כי את הסכום האחרון חסמנו בדוגמא 6.12, כאשר הראינו באמצעות אי-שוויון הופדינג כי משתנה בינומי $X \sim \mathrm{Bin}\left(m_i, \frac{r_n}{n}\right)$ מקיים

$$\mathbb{P}(|X - m_i(r_n)/n| > \epsilon m_i) = \sum_{nr_{m_i} \notin [m_i(r_n + 1 - \epsilon), m_i(r_n + 1 + \epsilon)]} \binom{m}{r_m} \left(\frac{r_n}{n}\right)^{r_m} \left(\frac{b_n}{n}\right)^{b_m} \le 2 \exp\left(-\frac{\epsilon^2 m_i}{4}\right),$$

נסיק אפוא כי

$$\mathbb{P}\left(\left|\frac{R_{m_i}}{m_i} - \frac{r_n + 1}{n}\right| \ge \epsilon \,\middle|\, R_n = r_n + 1\right) \le 2 \exp\left(-\frac{\epsilon^2 m_i}{4}\right) \frac{n^m}{n \cdots (n-m+1)}$$

וכיוון שהטענה נכונה לכל r_n נקבל כי

$$\mathbb{P}\left(\left|\frac{R_{m_i}}{m_i} - \frac{R_n}{n}\right| \ge \epsilon\right) \le 2 \exp\left(-\frac{\epsilon^2 m_i}{4}\right) \frac{n^m}{n \cdots (n-m+1)}$$

עוד נשים לב כי, לפי אי-שוויון המשולש, המאורע $\{\left|\frac{R_{m_1}}{m_1} - \frac{R_{m_2}}{m_2}\right| \geq \epsilon\}$ מוכל לכל n במאורע

$$\left\{\left|\frac{R_{m_1}}{m_1} - \frac{R_n}{n}\right| \geq \epsilon/2\right\} \cup \left\{\left|\frac{R_{m_2}}{m_2} - \frac{R_n}{n}\right| \geq \epsilon/2\right\}$$

ולכן הסתברותו חסומה על-ידי

$$\mathbb{P}\left(\left|\frac{R_{m_2}}{m_2} - \frac{R_{m_1}}{m_1}\right| \geq \epsilon\right) \leq 4\exp\left(-\frac{(\epsilon/2)^2 m_1}{4}\right)\frac{n^m}{n\cdots(n-m+1)}$$

נשאיף את n לאינסוף ונסיק כי

$$\mathbb{P}\left(\left|\frac{R_{m_2}}{m_2} - \frac{R_{m_1}}{m_1}\right| \geq \epsilon\right) \leq 4\exp\left(-\frac{\epsilon^2 m_1}{1}6\right)$$

■

מטענה 11.8 נוכל להסיק לפי תנאי-קושי להתכנסות בהסתברות כי $X \xrightarrow{\text{p}} \lim_{n\to\infty}\frac{R_n}{n+2}$. מטענה 11.5 נסיק כי X מתפלג אחיד על $[0,1]$.

טענה 11.9. התהליך $X \xrightarrow{\text{a.s.}} \frac{R_n}{n}$ עבור $X \sim \mathrm{Unif}([0,1])$.

הוכחה. נשתמש בתנאי קושי להתכנסות כמעט תמיד. כלומר נראה כי

$$\forall \epsilon, \eta\ \exists m_0 : \mathbb{P}\left(\exists m_1, m_2 > m_0\ :\ \left|\frac{R_{m_2}}{m_2} - \frac{R_{m_1}}{m_1}\right| > \epsilon\right) < \eta.$$

נסתמך על טענה 11.8 בכדי לקבל כי לכל $n > m_0$ מתקיים

$$\mathbb{P}\left(\exists m > m_0\ :\ \left|\frac{R_m}{m} - \frac{R_n}{n}\right| > \epsilon\right) \leq \sum_{m=m_0}^{n}\mathbb{P}\left(\left|\frac{R_m}{m} - \frac{R_n}{n}\right| > \epsilon\right) \leq \sum_{m=m_0}^{n} 2\exp\left(-\frac{\epsilon^2 m}{4}\right) \leq \frac{2\exp\left(-\frac{\epsilon^2 m_0}{4}\right)}{1-e^{-\epsilon^2/4}}.$$

ומסיבה זו מתקיים

$$\mathbb{P}\left(\exists n > m_2 > m_1 > m_0\ :\ \left|\frac{R_{m_1}}{m_1} - \frac{R_{m_2}}{m_2}\right| > \epsilon\right) \leq \mathbb{P}\left(\exists m_1 > m_0\ :\ \left|\frac{R_{m_1}}{m_1} - \frac{R_n}{n}\right| > \epsilon/2\right) \leq \frac{2\exp\left(-\frac{\epsilon^2 m_0}{16}\right)}{1-e^{-\epsilon^2/16}}.$$

כיוון שטענה זו נכונה לכל n, מתקיים תנאי קושי להתכנסות כמעט תמיד, כנדרש. ■

הערה: המתמטיקאי ההונגרי היהודי ג'ורג פוליה פעל בETH ציריך בין 1914 ל1940 ובסטנפורד שבקליפורניה מ1940 ועד מותו ב1985. הוא תרם רבות הן לתורת ההסתברות, והן לספרות העל-מתמטית העוסקת בגישות לפתרון בעיות. ספרו "כיצד לפתור את זה" זכה לתהודה רבה בעולם ההשכלה המתמטית ומהווה עד היום אבן בניין מרכזית בתורת של גישות לפתרון בעיות. את ניסוי כדי פוליה הציג בשנת 1931 במאמר בצרפתית בירחון של המכון על שם אנרי פואנקרה. המאמר זכה להרחבות רבות, בין השאר, מפני שהוא מהווה דוגמא פשוטה יחסית להתכנסות כמעט תמיד למשתנה לא טריוויאלי ושיטות הטיפול בו הורחבו לטיפול בתהליכים מקריים רבים.

הניסוי עצמו הוא חבר מכובד במורשת של ניסויי כדים שראשיתה בספר "אומנות ההשערה" של של יעקב ברנולי מ-1731. מקור מושג הכד האטום המלאה בכדורים ככל הנראה בשיטת הבחירות שהייתה נהוגה בונציה. שם היו מטילים כדורים עם שמות המועמדים לתפקיד הדוג'ה לתוך כד ולאחר מכן מונים אותם.