מבוא להסתברות וסטטיסטיקה
88-165

16 באוקטובר 2017

מהדורה 1.792

סטטיסטיקה היא מערכת של כלים ושיטות הנמצאת בשימוש בכל תחומי החיים: במדע, בכלכלה, במדעי החברה, ועוד. תורת ההסתברות, שפותחה כדי לתאר אירועים ותהליכים אקראיים, היא הבסיס המתמטי של הסטטיסטיקה.

המטרה העיקרית של הקורס שלנו היא לפתח חשיבה הסתברותית, שהיא תנאי חיוני ליכולת להעריך ולאמוד הסתברויות בחיי היום־יום. הלימודים עצמם מתמטיים, עם הגדרות וטכניקות חישוביות ומשפטים והוכחות, אבל המטרה העיקרית היא פיתוח הרגישות למספרים בהקשר ההסתברותי שלהם.

הבנת העקרונות של של תורת ההסתברות והסטטיסטיקה חשובה לכל אדם משכיל. עקרונות אלה מאפשרים לקורא העיתונים הממוצע לקרוא את שטף הנתונים והדיווחים המופיעים בהם מנקודת מבט ביקורתית, ומסייעים לכל מי שמבקש להבין מידע כמותי לא מאורגן.

את מי שאינו מבין מספרים אפשר לרמות באמצעות נתוני סקרים לא רלוונטיים, בהסקות סטטיסטיות שגויות, בהשוואה של תוחלות בלי סטיות תקן. הוא יהמר במקום שאסור, ויחשוש מהימור במקום שזה מתבקש. לדוגמא, התעלמות ממאורעות בעלי הסתברות נמוכה (אבל בשום אופן לא זניחה) היתה הסיבה למפולת הכלכלית הגדולה של 2008.

"37 אחוזים מהדמגוגיה מבוססים על נתונים סטטיסטיים" (זה ציטוט דמגוגי שהמצאתי לצורך המבוא, אבל אתם בוודאי מבינים את ההבדל בינו לבין "37 אחוזים מהנתונים הסטטיסטיים הם דמגוגיה"). בחיים האמיתיים, כשאומרים "A גורם ל־B", מתכוונים "A בדרך כלל גורם ל־B". ההפרכה הכושלת, והשכיחה כל־כך "אבל כלתה של השכנה שלי היא A ואינה B" מבוססת על כך שהשומעים אינם מבינים שקיימות הסתברויות בין אפס לאחד.

ראשיתה של תורת ההסתברות בהערכות של סיכויים, לצורך משחקי קוביה וקלפים, שערכו פייר דה פרמה ובלז פסקל באמצע המאה ה־17. אחריהם תרמו לתאוריה בני משפחת ברנולי, קרל פרידריך גאוס, ורבים אחרים.

הסטטיסטיקה נולדה כמאתיים שנים אחר־כך. פלורנס נייטינגייל, ששירתה במלחמת קרים 1854, זכורה כאחות רחמניה שהכניסה את שיקולי ההגיינה לטיפול הרפואי, מהלך שהוא אולי מאריך החיים הגדול ביותר בהסטוריה. היא לא היתה הראשונה שהעלתה את הרעיון, אבל כדי לקבל הכרה מקצועית מהרופאים ששלטו במקצוע, היא היתה חייבת לאסוף נתונים סטטיסטיים – ואכן, כאות הוקרה על השיטות שפיתחה, נבחרה לחברה המלכותית לסטטיסטיקה ב־1858.

קורס זה נבנה עבור סטודנטים בסמסטר השני של שנת הלימודים הראשונה במתמטיקה. במהלך הקורס נתקל בכמה טכניקות אנליטיות (סיכום טורי חזקות בעזרת גזירה של טורים ידועים, אינטגרציה בחלקים, החלפת משתנים באינטגרציה), אבל נסתפק בהפעלת כלים אלה במקרים הפשוטים והקלים ביותר. במקום אחד נשתמש

באינטגרציה כפולה והחלפת משתנים דו־ממדית, לרבות כלל היעקוביאן. מן האלגברה הליניארית אנו זקוקים לתכונות אלמנטריות של מטריצות, ורק באחד הסעיפים נזכיר מושגים מתקדמים יותר כמו ערכים ווקטורים עצמיים. מלבד אלה, הקורס עומד ברשות עצמו.

תוכן עניינים

1 מבואות **9**

1.1 לשם מה הסתברות 9

1.2 סטטיסטיקה תאורית 11

1.2.1 טיפוסי משתנים 11

1.2.2 אוכלוסיה ומדגם 12

1.2.3 תאור גרפי 12

1.2.4 מדדי מרכז 12

1.2.5 מדדי פיזור 16

1.2.6 מתאם 18

1.3 קומבינטוריקה 19

1.3.1 בחירה עם החזרה ובלי החזרה, עם חשיבות לסדר ובלעדיה 19

1.3.2 עקרון ההכלה וההדחה 23

2 מבוא להסתברות **25**

2.1 מרחבי הסתברות בדידים 25

2.1.1 הסתברות של מאורעות 26

2.1.2 הסתברות מותנית 28

2.1.3 נוסחת ההסתברות השלמה 29

2.1.4 חוק בייס 30

2.1.5 תלות ואי־תלות 32

2.2 משתנים מקריים 33

2.2.1 משתנה יחיד 33

2.2.2 התפלגות משותפת 36

2.2.3 תוחלת של משתנה מקרי בדיד 38

2.2.4 שונות 42

2.2.5 שונות משותפת ומקדם המתאם 44

2.3 התפלגויות בדידות 47

2.3.1 התפלגות אחידה 47

2.3.2 התפלגות ברנולי 48
2.3.3 התפלגות בינומית 49
2.3.4 התפלגות פואסון 52
2.3.5 התפלגות גאומטרית 55
2.3.6 התפלגות בינומית שלילית 60
2.3.7 התפלגות היפרגאומטרית 62
2.3.8 גרפים מקריים 65
2.4 מרחב התפלגות כללי 66
2.4.1 סיכום על קבוצה שאינה בת־מניה 66
2.4.2 אי־קיומן של מידות אינווריאנטיות 66
2.4.3 סיגמא־אלגברות 67
2.4.4 מרחבי הסתברות 68
2.4.5 בעיית ברטרנד 69
2.5 משתנים מקרים רציפים 70
2.5.1 σ-אלגברת בורל 70
2.5.2 פונקציית הצטברות 71
2.5.3 פונקציית צפיפות 72
2.5.4 תוחלת ושונות 72
2.5.5 התפלגות משותפת 74
2.5.6 המקרה המעורב 75
2.5.7 סטטיסטיי הסדר 77
2.5.8 טרנספורמציה של משתנה 77
2.6 התפלגויות רציפות חשובות 78
2.6.1 התפלגות אחידה 78
2.6.2 התפלגות מעריכית 82
2.6.3 התפלגות נורמלית 85
2.6.4 התפלגויות נוספות 87
2.7 חסמים 89
2.7.1 אי־שוויון מרקוב 89
2.7.2 אי־שוויון צ'ביצ'ב 89
2.7.3 אי־שוויון צ'רנוף 90
2.8 פונקציה יוצרת מומנטים 91
2.8.1 מומנטים 91
2.8.2 פונקציה יוצרת מומנטים 92
2.8.3 פונקציות יוצרות אחרות 94
2.9 הלמה של בורל־קנטלי 96
2.10 חוקי המספרים הגדולים 97
2.10.1 החוק החלש של המספרים הגדולים 97

2.10.2 החוק החזק של המספרים הגדולים 98
2.10.3 משפט הגבול המרכזי 99
2.11 שרשראות מרקוב 101
2.11.1 תאור גרפי 102
2.11.2 ההתפלגות הסטציונרית 103
2.11.3 מאורעות שאינם תלויי זמן 104
2.11.4 תוחלת זמן ההגעה 107
2.11.5 הרחבת זכרון 109

3 מבוא לסטטיסטיקה 113
3.1 אמידה 113
3.1.1 אמידה נקודתית 114
3.1.2 רווחי סמך 117
3.1.3 רווחי סמך לפרמטרים של ההתפלגות הנורמלית 118
3.2 בדיקת השערות 122
3.2.1 השערות ושגיאות 123
3.2.2 הליך הבדיקה 124
3.2.3 בדיקת השערות בעזרת רווחי סמך 126

תוכן עניינים

פרק 1

מבואות

1.1 לשם מה הסתברות

תורת ההסתברות מטפלת במצבים של אי־וודאות. לו היינו יודעים לחזות את העתיד על כל פרטיו, לא היה לנו צורך בהסתברויות. מכיוון שאין לנו יכולת כזו (ואם היקום אינו דטרמיניסטי, גם לא תהיה לנו), עלינו לנסות לכמת באופן מספרי את הסבירות של האלטרנטיבות השונות.

אנשים שהחשיבה ההסתברותית שלהם לקויה אומרים לפעמים שאם יש שתי אפשרויות, אז הסיכויים הם 'חצי־חצי'; מן הגישה הרדודה הזו נובע שכל המצבים שבהם יש שתי אפשרויות שקולים זה לזה. באותה נשימה אפשר להזכיר גם את השטות האנטי־אינפורמטיבית 'מי שזה קורה לו, זה קורה לו במאה אחוזים': גישה כזו מרדדת את המציאות ואינה יודעת להבחין בין מידת הזהירות שיש לנקוט בחציית כביש, לבין המאמצים שיש להקדיש להתגוננות מפני מטאוריטים.

אחת המטרות העיקריות של הקורס היא לעזור לכם לפתח תחושה להסתברויות. מכיוון שהאינטואיציה האנושית מוכשרת בטיפול במספרים מתונים (כמו עשירית או שלושת־רבעי) ולא בהסתברויות קיצוניות, נגדיר את **אינדקס הסיכון** בתור מינוס הלוג לפי בסיס 2 של הסיכוי לאירוע. אינדקס 7 (הסתברות של כאחד למאה) מצדיק שינוי של אורח החיים, אינדקס 14 (כאחד לעשרת אלפים) זהירות מסויימת, מאינדקס 21 (כאחד לשני מליון), נאמר, אפשר להתעלם.

דוגמא 1.1.1 להלן הסתברויות לכמה מאורעות, מוכרים או נדירים.

1. יורד גשם ביום שבו החזאי הבטיח עננים ללא גשם: $1/10$, אינדקס 3.
2. גבר צעיר זה יפתח מחלת לב כלילית במהלך חייו: כ־$1/100$, אינדקס 7.
3. ועדת קלפי שוגה ברישום התוצאות בשל החלפה בין שתי מפלגות (על־פי תוצאות הבחירות של *2013*): $3/1000$. אינדקס 8.5.

4. אתה נכנס למעלית והיא נתקעת: $1/1000$. אינדקס 10.
5. הסיכוי שאדם מסויים יתאבד השנה הוא כ־$1/8000$. אינדקס 13.
6. הסיכון של אשה למות בלידה הוא כ־$1/20000$. אינדקס 14.3.
7. הסיכוי של אדם להחנק למוות מדבר מאכל בשנה הקרובה הוא כ־$1/70000$: אינדקס 16.
8. הסיכוי של אדם למות השנה בתאונת אופניים הוא כ־$1/70000$. אינדקס 16.
9. הסיכוי למות השנה מתאונת נשק (בארה"ב) הוא כ־$1/500000$. אינדקס 19.
10. הסיכון שדירה מסויימת תקרוס כתוצאה מדליפת גז במהלך השנה הקרובה הוא, נניח, $1/1000000$. אינדקס 20.
11. הסיכוי של תושב ארה"ב להרג השנה מפגיעת ברק הוא כ־$1/10000000$. אינדקס 23.
12. הסיכוי לזכות בפרס הראשון בלוטו (6 מתוך 49) הוא כ־$1/14000000$. אינדקס 23.7.
13. הסיכוי להרג השנה מפגיעת מטאור הוא כ־$1/75000000$ (אינדקס 26).

דוגמא 1.1.2 תוחלת החיים היא כ־$2^{15} \sim 30,000$ ימים. לכן, אם אבחר לבצע מדי פעילות שאינדקס הסיכון שלה הוא 18, יש סיכוי של כ־99.9% שלא אפגע. לעומת זאת, במדינה שיש בה $8,000,000$ תושבים, מתרחשים מדי שנה כ־$2^{31.4}$ ימי־אדם. מדיניות שתמליץ להסתכן באופן כזה, עד כמה שהיא סבירה לאדם הפרטי, תביא לכ־$10,000$ נפגעים בשנה. לכן אפשר להניח שהרשויות ימליצו על זהירות גם בהתייחס לאירועים שאינדקס הסיכון שלהם הוא כ־30. (הסק, על אחריותך, שהגיוני להתעלם באופן פרטי מהמלצות מסויימות של הרשויות.)

דוגמא 1.1.3 בכל מקרה חשוב לדייק בהגדרת האוכלוסיה. למשל, שיעור המצוננים באוכלוסיה בזמן נתון (בחורף) הוא, נניח, 1%. זה גם שיעור הסכיזופרנים. עם זאת, הסיכוי לחלות בצינון אי־פעם בחיים הוא אולי 98%, והסיכוי לחלות בסכיזופרניה הוא 2%.

אם מדובר באירוע מתמשך, הכפלת משך הזמן מגדילה את הסיכון פי 2, כל עוד שומרים על גבולות הסביר (הסיכון למות בעשרת אלפי השנים הקרובות אינו גדול בהרבה מהסיכוי למות באלף השנים הקרובות).

חשוב לדעת להעריך הסתברויות וכמויות אקראיות הקשורות בהן. במהלך הקורס נתקל בכמה וכמה דוגמאות שבהן חישוב ישיר הוא קשה או בלתי אפשרי, אבל אפשר לערוך אומדן מושכל בקלות יחסית.

תרגיל 1.1.4 השימוש בטלפון סלולרי מגביר את הסיכון לחלות בסרטן מסוג גליומה ב־40%. שיעור החולים בסרטן זה הוא 2/100000. האם תפסיקו להשתמש בטלפון סלולרי?

דוגמא אחרונה לסעיף זה:

ציטוט 1.1.5 "הרופאים דורשים תוספת שכר של 50% לשכר הבסיס ועוד שיפורים בתנאים הנלווים בשיעור של 50%, כך שהתוספת הכוללת שהם דורשים מגיעה ל־100%" (נציגי משרד האוצר, 15.2.2011).

1.2 סטטיסטיקה תאורית

סטטיסטיקה תאורית עוסקת בצמצום של מאגרי נתונים גדולים למספר קטן של נתונים שאפשר להציג באופן מספרי או גרפי. היא נועדה להחליף גודש של מידע בנתונים תמציתיים המייצגים, במובנים מסויימים, את המידע השלם.

תאורים סטטיסטיים חשופים להטעיות ולשגיאות, ולכן חשוב להבין מהי הדרך הנכונה לבצע אותם. אפשר להקדיש לנושא הזה שעורים רבים, ובפקולטות אחרות זה בדיוק מה שעושים. בקורס שלנו הנושא יתפוס כשעור אחד; את כל השאר נשאיר לשכל הישר שלכם. נסתפק בהצגת המושגים המרכזיים, ישמשו אותנו בהמשך, כשנבנה מושגים מקבילים בתורת ההסתברות.

1.2.1 טיפוסי משתנים

כל דבר שמודדים (באופן מספרי) אפשר לקודד למשתנה סטטיסטי.

המיון הבסיסי הוא לפי טיפוס המשתנה: משתנה איכותי (כמו צבע עיניים, מוצא, מין), משתנה אורדינלי (שבו יש משמעות לסדר: השכלה, מידת האהדה לברוקולי), משתנה אינטרוולי (שבו יש משמעות להפרש ולא רק לסדר: שנת לידה, מידת נעליים, נסיון בעבודה), משתנה מנתי (שבו יש גם משמעות ליחס בין ערכים: משכורת, משקל).

המיון אינו מוגדר היטב (אפשר להתווכח האם יש או אין משמעות לכך שאדם פלוני שוקל פי 1.4 מאדם אלמוני, או בעל נסיון גדול פי 5 בעבודה). מטרתו העיקרית להצביע על סוגי הניתוחים הסטטיסטיים שבאים בחשבון. לדוגמא, אין משמעות לשאלה מהו צבע העיניים הממוצע של התלמידים בכתה.

את המשתנים האינטרוואליים והמנתיים אפשר למיין הלאה לשני סוגים: משתנים בדידים (אלו המקבלים ערכים בקבוצה בדידה: סופית או בת־מניה כמו המספרים הטבעיים), ומשתנים רציפים (אלו שאלמלא מגבלות הפיזיקה וההנדסה היו מקבלים ערכים בקטע רצוף). בדרך כלל, אם הערכים בדידים בשל מגבלות דיוק או רזולוציה אבל רבים מאד (גובה משכורת, שיעור המדד) מתייחסים אליהם כאילו הם רציפים. כפי שנראה באריכות בהמשך הקורס, הניתוח המתמטי שונה עבור משתנים בדידים ורציפים, למרות שהאידיאולוגיה דומה.

1.2.2 אוכלוסיה ומדגם

'אוכלוסיה' היא מכלול הישויות הכפופות לניתוח הסטטיסטי שבו אנו מעוניינים. למשל, כשמדברים על מספר ימי המחלה שלוקח עובד ישראלי בשנה, האוכלוסיה כוללת את כל העובדים בישראל. אם נרצה להציג לשר הבריאות את הנתונים על האוכלוסיה כולה, נצטרך לקבוע ישיבת עבודה ארוכה ביותר. גרוע מזה, איסוף הנתונים דורש כח אדם משמעותי וכסף רב.

לכן מסתפקים ב'מדגם': קבוצה חלקית לאוכלוסיה, שנאספה באופן שיאפשר הסקת מסקנות ממנה על האוכלוסיה כולה.

תאור האוכלוסיה ותכונותיה הם תחום העיסוק של הסטטיסטיקה התאורית. הדרכים שבהן אפשר להסיק מן המדגם על האוכלוסיה הן ליבה של הסטטיסטיקה, ובכך נעסוק בשעורים האחרונים של הקורס.

1.2.3 תאור גרפי

מי שקרא עיתון או הפעיל פעם את הפונקציות הגרפיות של Excel אינו זקוק להסברים לגבי תאורים גרפיים. ובכל זאת, בכמה מלים.

משתנה סטטיסטי אפשר להציג בהיסטוגרמה (מתאימה להצגת מספר הנבדקים בעלי תכונה מסויימת; כללי אתיקה: העמודות צריכות להתחיל באפס!), דיאגרמת עוגה (מתאימה להצגת פרופורציות באוכלוסיה), או באלף דרכים אחרות.

כאשר מדובר בזוג משתנים (או יותר), כמו שכר לפני מס ואחרי מס, ציונים בקורס א' לעומת קורס ב', צבע עיניים וגובה - יש הרבה מאד אפשרויות. החשובה ביותר: דיאגרמת פיזור (נקודה פיזיקלית לכל ערך (x,y)).

נסו להציע דרכים לתאר מידע ממימד גבוה יותר (יש שיטות סטטיסטיות שעוסקות בזה).

1.2.4 מדדי מרכז

תפקידו של מדד מרכזי הוא לתת ערך יחיד המתאר את הנתנים במקורב. 'מצאתי את עצמי במסיבה משעממת'. מה היה הגיל הממוצע של שאר המשתתפים? '14' (או '55'). כמובן שלערך יחיד יש מגבלות. הבדיחות ידועות: 'סטטיסטיקאי טבע באגם שעומקו הממוצע חצי מטר'. 'בחדר יש תשע נשים: אחת לקראת לידה ושמונה רווקות. הרופא חושב שהן בממוצע בהריון בחודש הראשון'. הבדיחה כאן היא על חשבון מי שלא מבין סטטיסטיקה: אפשר לטבוע באגם גם אם עומקו הממוצע סנטימטר אחד. אין שום הגיון בהכנסת מי שאינה בהריון לחישוב זמן ההריון הממוצע, כפי שלא מכלילים בחישוב את מנורת השולחן (גם את זמן ההריון הממוצע של נשים בהריון יש טעם לחשב רק במקרים מוגבלים ביותר).

אינטואיטיבית, מדד מרכזי הוא מספר התלוי ברשימה של n נתונים, ו"מייצג" את המרכז שלהם. להלן ארבע דוגמאות חשובות:

1. **הממוצע** של מדגם $x_1,\dots,x_n$ מוגדר לפי $\bar{x}=\frac{x_1+\cdots+x_n}{n}$.

2. **החציון** הוא הערך האמצעי ברשימה המתקבלת לאחר מיון הנתונים, או כל מספר בין שני האמצעיים אם מספר הערכים זוגי.

3. **אמצע הטווח**, $\frac{1}{2}(\max x_i+\min x_i)$.

4. **השכיח** הוא הערך המופיע מספר רב ביותר של פעמים (או אחד מהם, אם יש כמה כאלה).

הערה 1.2.1 יש מדדים שאינם מתאימים לכל טיפוסי המשתנים. למשל, בחציון אפשר להשתמש אם מדובר במשתנה אורדינלי (אבל לא איכותי), ובממוצע אפשר להשתמש אם המשתנה אינטרוולי (אבל לא אורדינלי).

כל אחד מהמדדים האלו מקיים את התכונות הבאות:

(1) סימטריות: $f(x_1,\dots,x_n)=f(x_{\sigma 1},\dots,x_{\sigma n})$ לכל תמורה $\sigma\in S_n$.

(2) הומוגניות: $f(cx_1,\dots,cx_n)=cf(x_1,\dots,x_n)$.

(3) שקיפות להזזה: $f(a+x_1,\dots,a+x_n)=a+f(x_1,\dots,x_n)$.

נסמן את המדד המרכזי של המדגם $x_1,\dots,x_n$ ב־$f_n(x_1,\dots,x_n)$; אפשר לצפות שהוספת ערך מרכזי למדגם לא תשנה את המרכז שלו. אכן, כל המדדים שלנו מקיימים תכונה נוספת:

(4) עקביות: $f_{n+1}(x_1,\dots,x_n,f_n(x_1,\dots,x_n))=f_n(x_1,\dots,x_n)$.

נעיר שהמינימום $\min\{x_1,\dots,x_n\}$ והמקסימום $\max\{x_1,\dots,x_n\}$ הומוגניים רק ביחס לקבועים חיוביים.

תרגיל 1.2.2 הראה שעבור $n=1,2$ הפונקציה היחידה המקיימת את התכונות (1)-(4) לעיל (סימטריות, הומוגניות, שקיפות ועקביות) הן $f_1(x)=x$ ו־$f_2(x,y)=\frac{x+y}{2}$. עבור $n=3$, הראה ש־$f_3(x,y,z)=x+(y-x)g(\frac{z-x}{y-x})$ (עם המוסכמה $0\cdot g(\frac{a}{0})=a\cdot g(0)$ המטפלת במקרה $y=x$) מקיימת את ארבע התכונות אם ורק אם $g:\mathbb{R}\to\mathbb{R}$ היא פונקציה המקיימת $g(\alpha^{-1})=\alpha^{-1}g(\alpha)$ ו־$g(1-\alpha)=1-g(\alpha)$. בפרט אפשר לבחור $g:[0,1/2]\to\mathbb{R}$ כלשהי (עם $g(1/2)=1/2$), ולהשלים לפונקציה על כל הישר לפי המשוואות. מצא את הפונקציה המתאימה עבור הממוצע, החציון, אמצע הטווח והשכיח.

תרגיל 1.2.3 נניח ש־$d(x_1,\dots,x_n;t)$ היא 'פונקציית שגיאה', שעבור מדגם $x_1,\dots,x_n$ מודדת עד כמה הוא מיוצג על־ידי הערך t. הנסיון למזער פונקציות שגיאה טבעיות, כלומר, לחשב $t = g(x_1,\dots,x_n)$ שיביא את השגיאה למינימום, מובילות לכמה ממדדי המרכז העיקריים.

בדוק את הטענות הבאות.

1. הממוצע ממזער את הפונקציה $d_2(x_1,\dots,x_n;t) = \sum_i (x_i - t)^2$. כלומר, הממוצע הוא המספר שסכום ריבועי המרחקים שלו מן הנתונים הוא הקטן ביותר.
2. החציון ממזער את $d_1(x_1,\dots,x_n;t) = \sum |x_i - t|$.
3. אמצע הטווח ממזער את $d_\infty(x_1,\dots,x_n;t) = \max |x_i - t|$.
4. השכיח ממזער את מספר השגיאות האבסולוטיות, $d' = \sum_{x_i \neq t} 1$.

אפשר להכליל ולהגדיר את המדד t_p, שהוא הערך הממזער את $d_p(x_1,\dots,x_n;t) = \sum |x_i - t|^p$, כאשר $p \geq 1$; אלא שבדרך כלל אין נוסחה אנליטית לערך t_p.

למרות שהממוצע הוא מדד טבעי ופשוט, ואולי דווקא בשל כך, הוא עשוי לבלבל כשמשתמשים בו שלא כהלכה. מי שמציג את הממוצע צריך לבחור אילו נתונים נכללים בחישוב, ואילו מושמטים ממנו. כאשר ממצעים נתונים של קבוצה, מדוע דווקא הקבוצה הזו ולא קבוצה רחבה או צרה יותר?

דוגמא 1.2.4 משווים את ממוצע הקליעות לסל של שני שחקנים. מתברר שאחוז הקליעות של A בעונה הראשונה גבוה משל B, וגם אחוז הקליעות שלו בעונה השניה גבוה יותר. היתכן שאחוז הקליעות המצטבר שלו נמוך יותר?

בהחלט. למשל, אם A קלע 60 מתוך 100 בעונה הראשונה לעומת 29 מתוך 50 קליעות של שחקן B; ובעונה השניה A קלע 35 מתוך 50, בעוד ש־B קלע 69 מתוך 100. אחוז הקליעה המצטבר של A הוא $95/150$, לעומת $98/150$ אצל B. (אנו חוזרים לדוגמא זו בהערה 2.1.30).

ציטוט 1.2.5 האחראי לנושא המשכורות וכוח האדם במשרד מבקר המדינה הוא המשנה למנכ"ל, שמואל יונס. לדברי בכיר במשרד המבקר, המניע להעברת עובדי מינהלה ומזכירות לדרגת "מבקר" ולא להסתפק בשכר אקדמאי בדרגת מינהלן, הוא כפול: מחד, לו היו נשארים עובדים אלה בדרגת "עובדי מינהלה" אך שכרם היה עולה - הדבר היה מעלה את ממוצע שכרם של עובדי המינהלה וגורר ביקורת ציבורית. מאידך - צירופם של עובדים אלה לדרג עובדי הביקורת מוריד את ממוצע השכר שם, וממתן את הביקורת הציבורית גם מכיוון זה. (דה־מרקר, "התחקיר שמבקר המדינה לא יכתוב", *27/2/2013*, `http://www.themarker.com/news/1.1938799`.)

תרגיל 1.2.6 "מאחר שמטעמי ביטחון שדה לא ניתן להציג את ממוצע מקבלי הגמלה, מציין האוצר כי 25% מגמלאי מערכת הביטחון קיבלו בדצמבר האחרון פחות מ־$8,299$ שקל, 50% קיבלו יותר מ־$11,752$ שקל ועוד 25% קיבלו יותר מ־$16,255$ שקל. 44.4% מגמלאי מערכת הביטחון קיבלו בדצמבר שכר שבין $10,000$ ל־20 אלף שקל. 29.8% קיבל עד $10,000$ שקל ו־18.5% גמלה של 20-30 אלף שקל. בסך הכל, שולם לגמלאי מערכת הביטחון בדצמבר 564 מיליון שקל, 6.8 מיליארד שקל במונחים שנתיים." (http://www.haaretz.co.il/tm/1.2362545 ,30/6/2014)
הערך: כמה גמלאים יש למערכת הבטחון?

הנה דוגמא לחיוניותו של השכיח.

תרגיל 1.2.7 קבעת עם חבר טוב להפגש ביום רביעי הבא בירושלים. אין לך שום אמצעי ליצור איתו קשר עד אז (הוא במחתרת מסיבות שאתה מקווה לגלות בפגישה). היכן ומתי תחפש אותו? השווה את תשובתך לתשובותיהם של סטודנטים אחרים ובדוק כמה מכם היו מצליחים להפגש בסופו של דבר.

דוגמא 1.2.8 ה**חציון** שימושי בקבלת הכרעות כאשר מצביעים על רצף של אפשרויות. נאמר למשל שהחברים במועדון חלוקים בדעתם לגבי מספר שעות ההתנדבות שחבר חדש במועדון צריך להתחייב עליהן. יש הטוענים שחובה להתנדב 40 שעות, ויש שיסתפקו ב־10. יש גם כאלו הדורשים 12 שעות. הצבעת רוב על שלוש אפשרויות כאלה עלולה לפצל את התומכים במספר קרוב ל־10, ולתת ל־40 לזכות מן ההפקר. היא גם לא מאפשרת למצביעים התומכים ב־35 או 32 לבוא לידי ביטוי. השיטה הנכונה להכריע במקרה כזה היא כדלקמן:

1. הנהלת המועדון מנסחת את השאלה העומדת להכרעה, ודואגת שהתשובה תהיה ערך של משתנה אורדינלי (כגון "מספר השעות הדרוש"). עליה לקבוע מהו **כיוון** ההצבעה: אם מישהו בוחר בערך α, האם חזקה עליו שהוא תומך בכל ערך גדול יותר, או בכל ערך קטן יותר? בדוגמא שלנו הכיוון הוא כלפי מעלה: מי שדורש 15 שעות התנדבות, בוודאי יסתפק ב־18; אבל הוא יתנגד ל־14, משום שאחרת היה בוחר בערך הזה (כאן רואים מדוע חשוב לתת לכל מצביע לבטא את דעתו באופן מלא).

2. כל מצביע מטיל לקלפי את עמדתו בסוגיה, שהיא ערך אחד בסולם הערכים האפשרי. "הנחת הרציונליות" במקרה כזה מחייבת את המצביע לאמץ עמדה מונוטונית. למשל, הוא אינו יכול לדרוש שמספר שעות ההתנדבות יהיה ראשוני דווקא.

3. בתנאים אלו, החציון הוא הערך הקטן ביותר שיש עליו הסכמת רוב, ולכן זו האפשרות שצריכה להבחר.

4. גם בלי להסביר את פרטי המנגנון, ההנהלה יכולה לפרסם הוכחה מסודרת ומשכנעת לכך שהרוב תומך בתוצאה שהתקבלה.

תרגיל. נניח שהתקנון קובע שכדי לשנות סעיף בתקנון, ההחלטה צריכה להתקבל ברוב מיוחס של 60%. הראה שהאחוזון ה־60% של ההצבעות צריך להבחר (בכיוון ההצבעה), משום שזה הערך הקטן ביותר שיש עליו הסכמת רוב מיוחס. **תרגיל.** המועדון מצביע על הסעיף הדורש הכרעה ברוב מיוחס. כל מי שתומך בדרישה לרוב של α יסתפק בוודאי גם ברוב גדול יותר, אבל יתנגד לרוב קטן יותר (אחרת היה בוחר בזה). נניח שמבין 50 המצביעים, יש 20 התומכים בהשארת הרוב המיוחס על 60%, 26 שהיו מסתפקים ברוב רגיל (של 50%), ועוד מצביע אחד לכל אחת מהאפשרויות 52%, 54%, 56%, 58%. הוכח לחברי המועדון שההצבעה הובילה לביטול הרוב המיוחס, למרות שרק לדרישה רוב של 58% ומעלה יש רוב תומך של 60%.

תרגיל 1.2.9 במשחק **נחש וזכה** משחקים על פרסים ששווים המוסכם מראש הוא $a_1,\dots,a_n$. מנהל המשחק שם בקופה אחד מהפרסים. השחקן צריך לנחש איזה פרס נבחר; אם צדק, הוא זוכה בפרס. נסמן $C = (\sum a_i^{-1})^{-1}$. הראה שהאסטרטגיה הטובה ביותר, גם למנהל וגם לשחקן, היא לבחור את הפרס ה־i בהסתברות Ca_i^{-1}. הסק ש־C הוא המחיר ההוגן לכרטיס השתתפות במשחק.

1.2.5 מדדי פיזור

אם נחזור למי שטבע באגם שעומקו הממוצע חצי מטר, הבעיה היא שממוצע (או שכיח, או כל מדד מרכזי אחר) נותן רק תאור של האמצע. הוא לא מספר לנו עד כמה הנתונים מפוזרים סביב האמצע. נחוץ לנו מדד שיבדיל בין מפעל שבו 29 עובדים בשכר 4000 ש"ח ומנהל שיווק בשכר 40000 ש"ח (ממוצע 5200) לבין מקום עבודה שבו 15 עובדים מנוסים מקבלים 5400 ש"ח, ו־15 הפחות מנוסים מקבלים 5000 ש"ח (אותו ממוצע).

מדד פיזור אמור לייצג את מידת הפיזור של הערכים סביב ערך מרכזי. בהמשך נציג כמה דוגמאות, המקיימות כולן את התכונות הבאות (השווה לתכונות המקבילות של מדדי המרכז):

1. סימטריות: $f_n(x_1,\dots,x_n) = f_n(x_{\sigma 1},\dots,x_{\sigma n})$ לכל תמורה $\sigma \in S_n$.
2. הומוגניות חיובית: $f_n(cx_1,\dots,cx_n) = |c|\, f_n(x_1,\dots,x_n)$.
3. אדישות להזזה: $f_n(a+x_1,\dots,a+x_n) = f_n(x_1,\dots,x_n)$.

מדד הפיזור החשוב ביותר הוא סטיית התקן:

***הגדרה 1.2.10** סטיית התקן של סדרת מספרים $x_1, \dots, x_n$ היא השורש s של השונות*

$$s^2 = \frac{1}{n} \sum (x_i - \bar{x})^2.$$

תרגיל 1.2.11 לכל $x_1, \dots, x_n$ מתקיים

$$\frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} (\sum x_i^2) - \bar{x}^2.$$

לאחר הוצאת השורש, המדד הוא הומוגני ממעלה ראשונה. בפרט, יחידת המידה של סטיית התקן היא זו של הנתונים עצמם: סטיית התקן יכולה להיות 15 סנטימטר, או 400 שקל.

יש מדדי פיזור נוספים:

1. **הטווח** $\max x_i - \min x_i$.
2. **טווח בין־רבעוני:** $Q_3 - Q_1$, כאשר $Q_1, M = Q_2, Q_3$ הם ערכים המחלקים את המדגם לארבעה רבעים שווים בגודלם (בקירוב, אם גודל המדגם אינו מתחלק ב־4).

בדוגמת המפעל, סטיית התקן של המפעל עם מנהל השיווק היא כ־6500, בעוד שסטיית התקן במקרה השני היא 203. כדי לפתח הבנה אינטואיטיבית למשמעות המספרית של סטיית התקן, יש להכיר דוגמאות נוספות. עשו לכם מנהג לחשב את סטיית התקן של נתוני מדגמים שאתם נתקלים בהם. במקרה הטיפוסי, חלק נכבד מן הנתונים נמצא במרחק של שתי סטיות תקן לכל היותר מן הממוצע.

כהכללה למושג הרבעון, מגדירים (עבור $0 < \alpha < 1$) את ה"α־יון" P_α להיות הערך ש־α מהנתונים קטנים ממנו והשאר גדולים ממנו. כך החציון הוא $M = P_{0.5}$, הרבעונים הם $Q_1 = P_{0.25}$ ו־$Q_3 = P_{0.75}$; העשירונים $D_1, \dots, D_9$ מוגדרים לפי $D_n = P_{n/10}$, והאחוזונים המוגדרים לפי $C_n = P_{n/100}$ מקיימים $C_1 < \dots < C_{99}$.

ציטוט 1.2.12 "גודלו של העשירון התחתון הוכפל תחת משטר תאצ'ר" (פיליפ בלונד, הוגה דעות פוליטי בריטי, הראלד טריביון, *26.1.2008*).

(אוכלוסיית בריטניה גדלה בתקופה זו ב־1.5%.)

ציטוט 1.2.13 "הבהרה: 'אחוזון עליון' משמעו כי מתוך שנתון של 100,000 תלמידים, אחוז אחד הם בעלי מנת משכל גבוהה מ־135." (אתר האגף לתלמידים מחוננים ומצטיינים, משרד החינוך, *2012*.)

1.2.6 מתאם

החשיבות העיקרית בנתונים סטטיסטיים עשויה להיות השוואתית. בפרט, אנחנו עשויים להיות מעוניינים בקשר בין שני משתנים. הקשר יכול להיות חזק (גובה ומשקל אצל ילדים בני שנתיים) או חלש (גובה של ילד בן שנתיים והקומה שבה הוא גר), בעל משמעות או חסר משמעות. לפעמים הקשר בין שני משתנים חזק, אבל שניהם מוסברים על־ידי משתנה שלישי (גובה וידיעת לוח הכפל בבית־הספר היסודי - שניהם מושפעים מן הגיל).

מקדם המתאם נועד לבדוק את המקרה המיוחד שבו משתנה אחד הוא, בקירוב, פונקציה ליניארית של משתנה אחר.

מגדירים את מקדם המתאם $\rho = \frac{\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y})}{s_x s_y}$, כאשר s_x, s_y הן סטיות התקן המתאימות. נפגוש בנושא זה שוב לקראת סוף הקורס. **הערה.** $\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y}) = \frac{1}{n}\sum x_i y_i - \bar{x}\bar{y}$.

טענה 1.2.14 תמיד מתקיים $-1 \le \rho \le 1$.

הוכחה. נתבונן ב־$\mathbb{R}^n$ כמרחב מכפלה פנימית, ביחס למכפלה הפנימית הסטנדרטית $(x_1,\dots,x_n)\cdot(y_1,\dots,y_n) = \sum x_i y_i$. לפי אי־שוויון קושי־שוורץ,

$$\begin{aligned}\left|\sum(x_i-\bar{x})(y_i-\bar{y})\right| &= |(x_i-\bar{x})_i\cdot(y_i-\bar{y})_i| \\ &\le \|(x_i-\bar{x})_i\|\cdot\|(y_i-\bar{y})_i\| \\ &= \sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2} = n s_x s_y.\end{aligned}$$

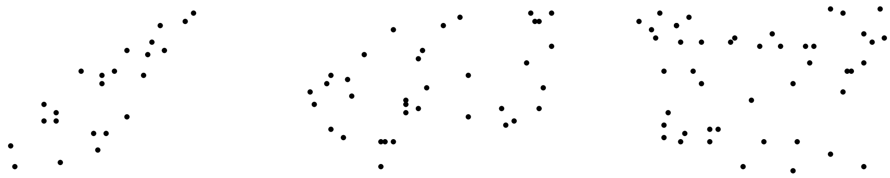
$\square$

אם אין קשר סטטיסטי בין המשתנים, הערך צריך להיות קרוב לאפס. לעומת זאת אם $y_i = ax_i + b$ (עבור ערכים a, b כלשהם), אז $\rho = \pm 1$.

איור 1.1 מדגים נקודות אקראיות עם כמה מתאמים אפשריים.

מקדם המתאם מודד קשר ליניארי בין שני משתנים. הוא מזהה קשר לא ליניארי רק באופן חלקי, ועשוי להיות רחוק מאפס גם כשאין קשר סיבתי בין שני הגורמים.

רעיון 1.2.15 יש מתאם חיובי בין הנטיה לאכול אוכל בריא, לבין הנטיה לחלות בסרטן. מדוע? משום שהרגלי האכילה הבריאה מפחיתים (נכון יותר לומר: דוחים) מחלות לב. בשנים הנוספות יש סיכוי לחלות במחלות חדשות.



איור 1.1: מתאמים (מימין לשמאל): 0.04, 0.41 ו־0.83.

1.3 קומבינטוריקה

אם ישנם n תרחישים אפשריים, שווי הסתברות, ומהם k תרחישים מוצלחים, אז הסיכוי שיתרחש משהו מוצלח הוא $\frac{k}{n}$. מן ההבחנה הפשוטה הזו נובע שכדי לחשב הסתברויות, עלינו לדעת לספור מצבים מוצלחים (תהי ההגדרה אשר תהיה), ומצבים בכלל. בפרק זה ניגע בנושא הקומבינטוריקה - תורת הספירה - על קצה המזלג.

1.3.1 בחירה עם החזרה ובלי החזרה, עם חשיבות לסדר ובלעדיה

האובייקטים הבסיסיים בקומבינטוריקה הם **וקטורים**, כלומר, הופעה סדורה של רכיבים יסודיים, ו**תת־קבוצות**, שבהן מופיעים אותם רכיבים באופן לא סדור. בכל בחירה יש להבדיל בין בחירה עם חזרות (בכל פעם בוחרים מאותו מרחב) וללא חזרות (המרחב משתנה תוך כדי הבחירה בהתאם לערכים שנבחרו).

דוגמא 1.3.1 (בחירה עם חזרות כשיש חשיבות לסדר) נתונה קבוצה X. בחירה של k רכיבים מן הקבוצה הזו, עם חשיבות לסדר הרכיבים, אפשר לקודד לווקטור באורך k, כלומר איבר של הקבוצה X^k. לכן מספר הדרכים לבחור k ערכים מקבוצה כזו, עם חזרות, הוא $\left|X^k\right| = |X|^k$. למשל, בכתה של 30 תלמידים מבקשים להעניק תעודות מתאימות לתלמיד שמקום מגוריו צפוני ביותר, לזה שחולצתו הכהה ביותר, ולזה שציונו הנמוך ביותר: יש 30^3 תוצאות אפשריות.

דוגמא זו מיישמת את "**עקרון המכפלה**" בגרסתו הפשוטה ביותר: מספר הדרכים לבחור איבר $a \in A$ ואיבר $b \in B$, כאשר אין קשר בין הבחירות, הוא $|A \times B| = |A| \cdot |B|$ (מפעילים את העקרון באינדוקציה על k).

דוגמא 1.3.2 לקבוצה A יש בדיוק $2^{|A|}$ תת־קבוצות (לרבות הקבוצה הריקה $\emptyset$ והקבוצה A עצמה). זוהי חזרה על דוגמא 1.3.1, משום שיש התאמה בין תת־קבוצות לפונקציות $A \to \{0,1\}$, ויש התאמה בין פונקציות אלו לבחירת $|A|$ ערכים (עם חזרות) מן הקבוצה $\{0,1\}$.

תרגיל 1.3.3 נתונה קבוצה A_0 בגודל n. כמה דרכים יש לבחור ממנה ועדה ובתוכה תת־ועדה? הראה שמספר השרשראות $A_k \subseteq \cdots \subseteq A_1 \subseteq A_0$ הוא $(k+1)^n$.

דוגמא 1.3.4 ***תמורה*** היא פונקציה חד־חד־ערכית ועל מן הקבוצה $\{1,\dots,n\}$ לעצמה. אנו מסמנים ב־S_n את קבוצת התמורות על n עצמים. יש $n! = n(n-1)(n-2)\cdots 2\cdot 1$ דרכים לסדר n עצמים. (אפשר להוכיח באינדוקציה).

מספר התמורות ב־S_n שווה גם למספר הפונקציות החד־חד־ערכיות $f\colon X\to Y$ בין כל שתי קבוצות X,Y בגודל n. למשל, זהו מספר ה"סידורים" של X, שהם פונקציות $f\colon\{1,\dots,n\}\to X$; וגם מספר התמורות של X (פונקציות החד־חד־ערכיות ועל $\sigma\colon X\to X$).

דוגמא 1.3.5 (בחירה ללא חזרות כשיש חשיבות לסדר) סופרים וקטורים שרכיביהם בקבוצה X בגודל n, עם המגבלה שכל שני רכיבים שונים זה מזה; יש n אפשרויות לרכיב הראשון, $n-1$ לשני, וכן הלאה. מספר הווקטורים באורך n הוא

$$n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

באופן שקול לזה, אנו סופרים פונקציות חד־חד־ערכיות מקבוצה בגודל k לקבוצה בגודל n.

באותה כתה, $30\cdot 29\cdot 28$ הוא מספר האפשרויות לבחור ועד בן שלושה אנשים, שאחד מהם עומד בראשו ואחד הוא סגן.

במקרה המיוחד $k=n$ סופרת דוגמא 1.3.5 תמורות על n עצמים. אם $k>n$ הדרישה היא למצוא פונקציה חד־חד־ערכית מקבוצה גדולה לקבוצה קטנה ממנה, ואין פונקציות כאלה.

"**עקרון הסכום**" הוא העובדה הפשוטה שאם הקבוצות A,B זרות, מספר הדרכים לבחור איבר מ־A או מ־B הוא $|A\cup B| = |A|+|B|$. את העקרון הזה אפשר להכליל באופן הבא:

טענה 1.3.6 (עקרון הסכום המוכלל) תהי $f\colon A\to B$ פונקציה בין קבוצות סופיות. אז $|A| = \sum_{b\in B}|f^{-1}(b)|$.

הוכחה. לפי ההגדרה הפורמלית, f היא הקבוצה $f = |\{(a,f(a))\colon\ a\in A\}|$; התאמה חד־חד־ערכית מראה ש־$|A| = |f|$; $f = \bigcup_{b\in B}\{(a,b)\colon\ f(a)=b\}$ הוא איחוד זר, ולכל b מתקיים $|\{(a,b)\colon\ f(a)=b\}| = |f^{-1}(b)|$. לכן

$$|A| = |f| = \sum_{b\in B}|\{(a,b)\colon\ f(a)=b\}| = \sum_{b\in B}\left|f^{-1}(b)\right|.$$

□

מה הקשר לנוסחה שבדוגמא 1.3.5? אנו סופרים וקטורים באורך k, ללא חזרות, שרכיביהם שייכים לקבוצה X. נסמן ב־$X^{[k]}$ את קבוצת הווקטורים ללא חזרות באורך k, וב־$X^{[k-1]}$ את קבוצת הווקטורים ללא חזרות מאורך $k-1$. הפונקציה $f : X^{[k]} \to X^{[k-1]}$ מוגדרת כך ש־$f(v)$ מתקבל מן הווקטור v על־ידי מחיקת הרכיב הראשון. לפי עקרון הסכום המוכלל,

$$\left|X^{[k]}\right| = \sum_{x \in X} \left|(X - \{x\})^{[k-1]}\right| = \sum_{x \in X} \frac{(n-1)!}{(n-k)!} = \frac{n!}{(n-k)!}.$$

מעקרון המכפלה נובע מיד גם

מסקנה 1.3.7 (עקרון המנה) תהי $f : A \to B$ פונקציה על. אם $m = |f^{-1}(b)|$ לכל $b \in B$, אז $|B| = \frac{1}{m}|A|$.

דרך נוספת לספור וקטורים ללא חזרות היא להתבונן בהטלה $S_n \to X^{[k]}$, המוגדרת על־ידי הקיצוץ $\sigma \mapsto (\sigma(1), \sigma(2), \ldots, \sigma(k))$. כל וקטור $(x_1, \ldots, x_k)$ עם ערכים שונים אפשר להשלים לתמורה בדיוק ב־$(n-k)!$ דרכים, ולפי עקרון המנה $\left|X^{[k]}\right| = \frac{n!}{(n-k)!}$.

דוגמא 1.3.8 (בחירה ללא חזרות כשאין חשיבות לסדר) כאן סופרים תת־קבוצות בגודל קבוע. מספר תת־הקבוצות בגודל k של קבוצה X בגודל n הוא $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. אכן, מספר הדרכים לבחור ועדה שוויונית בת 3 אנשים היא $\binom{30}{3}$.

הערה 1.3.9 המקדמים הבינומיים $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ מקיימים זהויות רבות, ובראשן

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

מצא לזהות זו הסבר קומבינטורי פשוט. העזר בזהות כדי להוכיח ש־$\sum_{n=0}^{N-1} \binom{n+k}{k} = \binom{N+k}{k}$. ראה גם תרגיל 2.3.68.

דוגמא 1.3.10 (בחירה עם חזרות כשאין חשיבות לסדר) גם כאן נתונה קבוצה X בגודל n, ובוחרים מתוכה k ערכים; אלא שהפעם אין חשיבות לסדר הערכים שהתקבלו. תוצאה כזו אפשר לתאר באמצעות המספרים $a_1, \ldots, a_n$, כאשר a_i סופר כמה פעמים נבחר הערך i מהקבוצה. אם כך, אפשר לקודד את הבחירה בווקטור $(a_1, \ldots, a_n)$, המוגבל על־ידי שני אילוצים: $a_i \geq 0$, ו־$a_1 + \cdots + a_n = k$. כפי שנוכיח מיד, מספר הפתרונות הוא $\binom{n+k-1}{k}$. מספר הדרכים לחלק שלושה פרסים זהים לתלמידי הכתה הוא $\binom{30+2}{3}$.

נוסחת הספירה בדוגמא 1.3.10 נובעת מן ההתאמה המסקרנת הבאה:

טענה 1.3.11 מספר הפתרונות $(x_1,\dots,x_n)$ למשוואה $x_1+\dots+x_n=k$, תחת האילוצים $x_i\geq 0$, הוא $\binom{n+k-1}{k}$.

הוכחה. נבנה התאמה חד־חד־ערכית ועל בין קבוצת הפתרונות $(x_1,\dots,x_n)$ לבין קבוצת הווקטורים באורך $n+k-1$ שיש בהם k סימני "$\times$" ו־$(n-1)$ סימני פסיק: הווקטור $(x_1,\dots,x_n)$ מתאים לרשימה רצופה של סימנים המורכבת מ־x_1 סימני $\times$ ואז סימן פסיק; ואז x_2 סימני $\times$ וסימן פסיק; ואז x_3 סימני $\times$ וסימן פסיק; וכן הלאה, עד ל־x_k סימני $\times$.

מספר הווקטורים באורך $n+k-1$ שיש בהם k סימני "$\times$" ו־$(n-1)$ סימני פסיק הוא $\binom{n+k-1}{k}$, משום שיש לבחור היכן ממוקמים סימני הפסיק בקבוצת הרכיבים של הווקטור. □

אלו הן ארבע דוגמאות חשובות לבעיות בחירה, אבל יש בעיות רבות אחרות, בדרך כלל קשות יותר. למשל, כמה וקטורים מסודרים יש באורך קבוע ובעלי סכום קבוע n? ואם אין חשיבות לאורך? ואיך גדל המספר הזה, כפונקציה של n?

ספירה של קבוצות היא בעיה חשובה, שההכללה שלה מתבקשת. אם $|A|=m$, אז מספר תת־הקבוצות בגודל n הוא כאמור $\binom{m}{n}$. כל תת־קבוצה כזו היא למעשה חלוקה של A לאיחוד זר $B\cup B^c$. באופן כללי יותר, מספר האפשרויות לפרק את A לאיחוד זר של תת־קבוצות $B_1,\dots,B_t$, בגדלים שנקבעו מראש $|B_i|=n_i$ כך ש־$n_1+\dots+n_t=m$, הוא

$$\binom{m}{n_1}\binom{m-n_1}{n_2}\cdots\binom{m-n_1-\dots-n_{t-1}}{n_t}=\binom{m}{n_1\ \cdots\ n_t}.$$

דוגמא 1.3.12 בכתה יש 30 תלמידים. רוצים לחלק להם כובעים: 7 כובעי ליצן, 18 מצנפות שינה, ו־5 צילינדרים, כך שכל תלמיד יחבוש בדיוק כובע אחד. מספר הדרכים לעשות זאת הוא $\binom{30}{7\ 18\ 5}=\frac{30!}{7!18!5!}$.

תרגיל 1.3.13 כמה דרכים יש להושיב שבעה אנשים על ספסל? וסביב שולחן עגול? כמה דרכים יש לחלק אותם למנהיג ושני צוותי עבודה שווים בגודלם, שיש להם אותה משימה?

תרגיל 1.3.14 נאמר שקבוצת אנשים מסודרת במבנה $[\ell_1\dots\ell_t]$ אם הם יושבים סביב שולחנות עגולים באורכים $\ell_1,\dots,\ell_t$. לדוגמא, המבנים האפשריים לסידור ארבעה אנשים הם $[1111],[112],[22],[13],[4]$. כמה דרכים יש להושיב ששה אנשים בכל אחד מאחד־עשר המבנים האפשריים? (סכום התשובות הוא $6!$).

תרגיל 1.3.15 1. בטיול על המרחב $\mathbb{Z}^2$ מותר לזוז בכל צעד רק באלכסון: ימינה ולמעלה או ימינה ולמטה. נניח ש־$a>b$ שלמים. כמה מסלולים מגיעים מנקודת הראשית לנקודה $(a+b,a-b)$?

2. ["בעיית ספירת הקולות", William Allen Whitworth (1840–1905), 1878; הפתרון המתואר כאן הוא של Désiré André (1840–1917/18)] בבחירות יש a תומכים למועמד אחד ו־b למועמד השני, כאשר $a > b$. סופרים את הקולות בזה אחר זה. מהם הסיכויים לכך שהמועמד שינצח בסופו של דבר יוביל לאורך הספירה כולה? **הדרכה.** לכל קבוצת מסלולים X, נסמן ב־X' את קבוצת המסלולים ב־X שבהם המנצח מוביל לאורך הספירה כולה, וב־X'' את המסלולים ב־X שבהם מתרחש תיקו לפחות פעם אחת במהלך הספירה. נסמן ב־A את קבוצת המסלולים שהצעד הראשון שלהם הוא כלפי מעלה (תמיכה במועמד שינצח בסופו של דבר) וב־B את המשלים, היינו המסלולים שהצעד הראשון שלהם כלפי מטה (תמיכה במועמד המפסיד). כמובן, $B' = \emptyset$. על־ידי שיקוף החלק של המסלול עד לתיקו הראשון, הראה ש־$|A''| = |B''|$. הסק שההסתברות המבוקשת היא $\frac{a-b}{a+b} = \frac{|A|-|B|}{|A|+|B|} = \frac{|A'|}{|A \cup B|}$.

1.3.2 עקרון ההכלה וההדחה

כידוע, $|A \cup B| = |A| + |B| - |A \cap B|$. בדומה לזה אפשר לחשב

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

זוהי נוסחה שימושית, משום שבדרך כלל קל יותר לחשב כמה איברים מקיימים תכונה מסויימת או זוג תכונות מסויימות או שלוש תכונות, מאשר לחשב כמה אברים מקיימים **לפחות** תכונה אחת.

ההכללה מתבקשת:

משפט 1.3.16 (עקרון ההכלה וההדחה) אם $A_1, \dots, A_t$ קבוצות, אז

$$|A_1 \cup \cdots \cup A_t| = \sum_{i=1}^{t} (-1)^{i-1} \sum_{I \subseteq \{1,\dots,t\},\, |I|=i} \left| \bigcap_{i \in I} A_i \right|.$$

הוכחה. אם נסכים שהחיתוך על־פני אוסף ריק של קבוצות שווה לאיחוד $A_1 \cup \cdots \cup A_t$, נוכל לכתוב את הטענה בצורה הדוקה יותר:

$$\sum_{i=0}^{t} (-1)^{i} \sum_{I \subseteq \{1,\dots,t\},\, |I|=i} \left| \bigcap_{i \in I} A_i \right| = 0,$$

ואפילו

$$\sum_{I \subseteq \{1,\dots,t\}} (-1)^{|I|} \left| \bigcap_{i \in I} A_i \right| = 0.$$

כדי להוכיח את הטענה, יהי $x \in \bigcup A_i$. נסמן $J = \{i \in \{1,\dots,t\}:\ x \in A_i\}$, ו־$t' = |J|$. אז x מופיע באגף ימין בדיוק בחיתוכים $\cap_{i\in I} A_i$ שעבורם $I \subseteq J$, ובסך הכל

$$\sum_{i=1}^{t}(-1)^{i-1}\sum_{I\subseteq J,\,|I|=i} 1 = \sum_{i=1}^{t'}(-1)^{i-1}\binom{t'}{i} = 1-\sum_{i=0}^{t'}(-1)^{i}\binom{t'}{i} = 1-(1-1)^{t'} = 1$$

פעמים. $\square$

נאמר שמערכת הקבוצות $A_1,\dots,A_t$ היא **סימטרית** אם גודל החיתוך $|\cap_{i\in I} A_i|$ תלוי רק ב־$|I|$. עבור מערכת סימטרית, נוסחת ההכלה וההדחה מקבלת צורה נוחה במיוחד:

$$|A_1 \cup \dots \cup A_t| = \sum_{i=1}^{t}(-1)^{i-1}\binom{t}{i}|A_1 \cap \dots \cap A_i|.$$

דוגמא 1.3.17 נספור כמה פונקציות $X \to Y$ הן על. נסמן $n = |X|$ ו־$\ell = |Y|$. לכל $y \in Y$ נסמן $A_y = \{f : X \to Y:\ y \notin \mathrm{Im}(f)\}$. הפונקציות שאינן על נמצאות ב־$\bigcup_{y\in Y} A_y$. לכל $I \subseteq Y$ מתקיים $|\cap_{y\in I} A_y| = (\ell - |I|)^n$. לכן

$$|A_1 \cup \dots \cup A_t| = \sum_{i=1}^{t}(-1)^{i-1}\binom{t}{i}|A_1 \cap \dots \cap A_i| = \sum_{i=1}^{t}(-1)^{i-1}\binom{t}{i}(\ell - i)^n.$$

דוגמא 1.3.18 כמה פתרונות יש למשוואה $x_1 + \dots + x_t = n$, במספרים טבעיים, בכפוף לאילוצים $0 \le x_1,\dots,x_t < m$?

נתבונן בקבוצת הפתרונות של שעבורם $0 \le x_1,\dots,x_t$, ללא האילוצים $x_i \le m$; ספרנו את אלה בטענה 1.3.11. נסמן ב־A_i את קבוצת הפתרונות שעבורם $m \le x_i$. אנו מעוניינים בפתרונות שאינם באף A_i, כלומר במשלים $(A_1 \cup \dots \cup A_t)^{\mathrm{c}}$. המערכת $A_1,\dots,A_t$ סימטרית. כדי לספור את הפתרונות ב־$A_1 \cap \dots \cap A_j$, נציב $x_i = x_i' + m$ עבור $1 \le i \le j$, ונקבל את המשוואה $x_1' + \dots + x_j' + x_{j+1} + \dots + x_t = n - jm$, עם האילוצים הסטנדרטיים $x_1',\dots,x_j',x_{j+1},\dots,x_t \ge 0$, שמספר הפתרונות שלה הוא $\binom{t+n-jm-1}{n-jm}$. לפי נוסחת ההכלה וההדחה, מספר הפתרונות המבוקש הוא

$$|(A_1 \cup \dots \cup A_t)^{\mathrm{c}}| = \sum_{j=0}^{t}(-1)^{j}\binom{t}{j}\binom{t+n-jm-1}{n-jm}.$$

פרק 2

מבוא להסתברות

2.1 מרחבי הסתברות בדידים

יש קושי פילוסופי בהגדרת מושג ההסתברות (אם העולם דטרמיניסטי, כל דבר או שיקרה בוודאות, או שלא יקרה בוודאות). כדי לחסוך בזמן, ננקוט בגישה המתמטית (שהיא: ליצור תאוריה מדוייקת, ולהשאיר את כאב הראש לפילוסופים).

המתמטיקה מסובכת יותר עבור משתנים רציפים, ולכן נתחיל בפיתוח התאוריה עבור משתנים בדידים. מעתה ועד פרק 2.5, תהי Ω קבוצה סופית או בת־מניה.

הערה 2.1.1 אם טור מתכנס בהחלט, סכומו אינו תלוי בסדר האברים.

הגדרה 2.1.2 (מרחב הסתברות בדיד) מרחב הסתברות בדיד הוא זוג סדור (Ω, P) שבו Ω קבוצה סופית או בת מניה, ו־$P:\Omega\to\mathbb{R}^{+}$ פונקציה חיובית המקיימת $\sum_{x\in\Omega} P(x) = 1$.

לפי הערה 2.1.1, אם הטור מתכנס לסידור מסויים של הנקודות במרחב, אז הסכום אינו תלוי בסדר.

ההגדרה הזו אינה מתאימה למרחבים שאינם בני מניה (משום שאז הסכום אינו מוגדר). בהמשך הקורס נתמודד עם המקרה הכללי, ונראה שהוא מכליל את ההגדרות שנפגוש בפרק הזה.

דוגמא 2.1.3 נניח ש־Ω מרחב סופי, $|\Omega| = n$. הפונקציה $P(x) = \frac{1}{n}$ (לכל $x \in \Omega$) הופכת את Ω למרחב הסתברות, הנקרא **מרחב ההסתברות האחיד** (בגודל n).

מצבים טבעיים רבים מוליכים למרחבי הסתברות אחידים: זריקת מטבע או קוביה, זוג קוביות (בצבעים שונים), וכדומה. במקרים רבים המרחב האחיד הוא גדול מכדי להיות מעניין, ומחליפים אותו במרחב מנה קטן, הנוגע באופן ישיר יותר לשאלה שרוצים לחקור.

דוגמא 2.1.4 זורקים מטבע שש פעמים, ושואלים מתי התקבל 1 בפעם הראשונה (אם בכלל). נוכחים כאן שני מרחבי הסתברות: המרחב האחיד Ω על 2^6 הסדרות האפשריות של הטלות מטבע, המתאר את **מרחב התצפיות**, והמרחב $\Omega' = \{1,2,3,4,5,6,-\}$ של תוצאות הניסוי. יש פונקציה $\Omega \to \Omega'$ המתאימה לכל תצפית את התוצאה המתאימה לה (למשל, כל הסדרות $001{*}{*}{*}$ מתאימות לערך $3 \in \Omega'$).

ההתאמה הזו מגדירה פונקציית הסתברות על המרחב Ω', לפי $P(\omega') = \frac{|f^{-1}(\omega')|}{|\Omega|}$.

דוגמא 2.1.5 קוביה אינה חייבת להיות הוגנת: לכל n מספרים ממשיים $p_1,\dots,p_n \geq 0$ שסכומם $p_1+\dots+p_n = 1$, אפשר לקחת $P(x_i) = p_i$.

דוגמא 2.1.6 נגדיר פונקציה $P:\mathbb{N}\to\mathbb{R}$ לפי $P(n) = 2^{-n}$. כך $P(n)$ הוא הסיכוי שנצטרך להטיל מטבע n פעמים עד להצלחה הראשונה (לקבל 'עץ', למשל). זוהי פונקציית הסתברות, משום ש־$\sum_{n=1}^{\infty} 2^{-n} = 1$.

הערה 2.1.7 יכולנו להמשיך בדוגמא הקודמת לפונקציה $P:\mathbb{N}\cup\{\infty\}\to\mathbb{R}$ על־ידי $P(\infty) = 0$, משום שככלות הכל יתכן שסדרת הכשלונות תמשך לאינסוף, גם אם ההסתברות לכך היא 0 (לא 'קטנה' או 'זניחה' ־ אפס).

נקודות כאלה אפשר לזרוק מן המרחב: קיומן או העדרן אינו משפיע כלל על החישובים.

2.1.1 הסתברות של מאורעות

כל תת־קבוצה $A \subseteq \Omega$ נקראת **מאורע**; היא מייצגת דבר שיכול לקרות, ושאפשר לדבר על ההסתברות שלו. נסמן ב־$\mathcal{P}(\Omega)$ את קבוצת החזקה של Ω (לא $P(\Omega)$, משום שאנו זקוקים לאות P כדי לציין הסתברויות). אכן, אנו מרחיבים את ההגדרה מהפונקציה $P:\Omega\to\mathbb{R}$ לפונקציה $P:\mathcal{P}(\Omega)\to\mathbb{R}$ על־ידי

$$P(A) = \sum_{x\in A} P(x) \tag{2.1}$$

(בכל סידור של A הסכומים החלקיים מהווים סדרה עולה וחסומה, ולכן מתכנסת; ומכיוון שהטור מתכנס בהחלט, הסדר אינו חשוב).

טענה 2.1.8 הפונקציה החדשה מקיימת שני תנאים חשובים:

1. $P(\Omega) = 1$;
2. לכל סדרה $A_1, A_2, \dots$ של מאורעות זרים בזוגות, $P(\cup A_n) = \sum_n P(A_n)$.

הוכחה. התכונה הראשונה היא אקסיומה מפורשת; לתכונה השניה, הסכום על $\cup A_n$ הוא סידור מחדש של סכום הסכומים על ה־A_n. $\square$

הערה 2.1.9 בתכונה 2 לעיל אפשר לקחת $A_3 = A_4 = \cdots = \emptyset$, ואז מתקבל עבור כל שתי קבוצות זרות A_1, A_2 הכלל $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

תרגיל 2.1.10 אם $A \subseteq B$ אז $P(B{-}A) = P(B) - P(A)$.

מאורעות A, B הם **זרים** אם $A \cap B = \emptyset$; לפעמים מרשים גם $P(A \cap B) = 0$. אם A, B זרים, אז $P(A \cup B) = P(A) + P(B)$.

לכל שני מאורעות מתקיים $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. עבור קבוצות או יותר, אפשר לנסח את הגרסה ההסתברותית של נוסחת ההכלה וההדחה (עם אותה הוכחה):

$$P(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}).$$

דוגמא 2.1.11 בעיית המזכירה המבולבלת: מזכירה הוציאה n מכתבים מן המעטפות שלהם, והחזירה אותם למקומות באקראי. מה הסיכוי לכך שאף מכתב לא יחזור למקומו?

נסמן ב־A_i את המאורע "המכתב ה־i חזר למקומו". מבקשים לחשב את הסיכוי של המשלים לאיחוד, $P((A_1 \cup \cdots \cup A_n)^{\mathrm{c}})$, בעוד שהסיכוי לכל חיתוך $A_{i_1} \cap \cdots \cap A_{i_k}$ $(i_1 < \cdots < i_k)$ קל לחישוב. לפי נוסחת ההכלה וההדחה,

$$\begin{aligned} P((A_1 \cup \cdots \cup A_n)^{\mathrm{c}}) &= \sum_{k=0}^{n} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \\ &= \sum_{k=0}^{n} (-1)^k \sum_{1 \leq i_1 < \cdots < i_k \leq n} \frac{(n-k)!}{n!} \\ &= \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{(n-k)!}{n!} \\ &= \sum_{k=0}^{n} (-1)^k \frac{1}{k!}, \end{aligned}$$

שהוא קירוב מצויין ל־e^{-1}.

דוגמא 2.1.12 בישראל 7000000 אזרחים. מדי שנה נהרגים כ־500 בתאונות דרכים. פירושו של דבר הוא שהסיכוי של אדם למות השנה בתאונת דרכים הוא 1/14000 (אינדקס סיכון: 16). הסיכוי של מי שיוצא בקביעות לשתות עם חברים למות

בתאונה הוא גבוה בהרבה (נסו להעריך אותו!). הסיכוי של נהגים זהירים נמוך יותר; הסיכוי של מי שאינו יוצא מהבית נמוך עוד יותר.

במוסד גדול לומדים 14000 תלמידים. מה הסיכוי שאחד מתלמידי המוסד יהרג בתאונה השנה? מה דעתכם על הסיכוי שהשנה יהרג חבר־כנסת בתאונת דרכים? (ח"כ מיכה רייסר נהרג בתאונה בשנת 1988; עד היום היו בישראל כ־$7800 = 120 \cdot 65$ שנות ח"כים).

ציטוט 2.1.13 "בעלי התקציב הלחוץ ישמחו לדעת שהביטוי 'יש לה לוק מיליון דולר', מוטעה מבסיסו. האמת היא שאין שום קשר בין כסף לבין סטייל. עזבו את העובדה ששום סכום שבעולם לא יקנה טעם. לעיתים, כיסים עמוקים גורמים להפך הגמור. זה עניין של סבירות פשוטה: אם אתה יכול להרשות לעצמך לרכוש ממבחר גדול יותר של פריטים, הסיכוי לבחור את הלוק הלא נכון גבוה יותר." (הבלוג של רוני בר, "הארץ", http://blogs.haaretz.co.il/ronibar/78).

2.1.2 הסתברות מותנית

כשמחשבים הסתברות, חשוב מאד להגדיר 'מה מתוך מה'.

ציטוט 2.1.14 "57% מן המטבעות של שקל אחד הולכים לאיבוד" (כתבה לכבוד השקת המטבע של שני שקלים, *2009*).

יהי (Ω, P) מרחב התפלגות בדיד. לעתים קרובות, מרחב כזה מקודד את מידת אי־הוודאות שלנו לגבי התרחשות עתידית. מידע נוסף משפר את ההערכה שלנו לגבי העומד להתרחש, דרך שינוי ההסתברויות שאנו צריכים לייחס לכל אפשרות. באופן פורמלי, נניח ש־$B \subseteq \Omega$ מאורע בעל הסתברות חיובית. כדי להניח שמרחב האפשרויות הצטמצם מ־Ω ל־B, אנו מגדירים פונקציית הסתברות חדשה על Ω:

$$P(x|B) = \begin{cases} \frac{P(x)}{P(B)}, & x \in B; \\ 0, & x \notin B. \end{cases}$$

טענה 2.1.15 $(\Omega, P(\cdot|B))$ הוא מרחב הסתברות בדיד.

לכל קבוצה $A \subseteq \Omega$, מתקיים לפי ההגדרה $P(A|B) = \sum_{x\in A} P(x|B) = \frac{P(A\cap B)}{P(B)}$.

על ההתניה אפשר לחזור. כפי שהגדרנו את ההסתברות המותנית ביחס ל־B, אפשר להגדיר גם $P(\cdot|B,C) = P(\cdot|B \cap C)$, וכן הלאה.

תרגיל 2.1.16 אחוז החולים במחלה מסויימת הוא 1%. יש בדיקה לאיתור המחלה, שאחוז הדיוק שלה 90%. מישהו נבדק ונמצא בריא. מה הסיכוי לכך שהוא באמת בריא? (דון בשתי הפרשנויות למלה "דיוק" - הסיכוי שבדיקה של אדם אקראי תחזיר תשובה נכונה, והסיכוי שהבדיקה תחזיר תשובה נכונה בין אם האדם חולה או בריא).

תרגיל 2.1.17 בחפיסה שלושה קלפים: אחד ששני צידיו אדומים, אחד ששני צידיו שחורים, ואחד שיש לו צד אחד אדום וצד שני שחור. בוחרים קלף באקראי ומניחים אותו על השולחן באקראי: הצד החשוף הוא אדום. מה הסיכוי לכך שהצד השני שחור? (מדוע, לדעתכם, התשובה השכיחה לשאלה זו היא 'חצי'? ומדוע תשובה זו שגויה?)

דיון 2.1.18 בכד מונחים n כרטיסי־גירוד, שאחד מהם זוכה בפרס. הילדים מסכימים שאם יחלקו כרטיס אחד לכל ילד, ואז יבדוק כל אחד את הכרטיס שברשותם, אז הסיכוי של ראובן לזכות הוא $\frac{1}{n}$. שכנעו את ראובן שאם מחלקים את הכרטיסים בזה אחר זה, וכל ילד יבדוק מיד את הכרטיס שלו, אין זה משנה היכן הוא עומד בתור לקבלת הכרטיסים.

תרגיל 2.1.19 העד המומחה טוען שרק במקרים נדירים ביותר אורך האמה עולה על היקף פרק כף היד. שלושה אנשים בקהל מדגימים שאורך האמה שלהם עולה על היקף פרק כף היד. האם תופתע לגלות שעורך הדין מטעם ההגנה זימן אותם לשם בכוונה?

2.1.3 נוסחת ההסתברות השלמה

מתקיים $P(A\cap B)=P(A|B)P(B)$. מכאן נובעת **נוסחת ההסתברות השלמה** לשתי קבוצות:

$$P(A)=P(A|B)P(B)+P(A|B^c)P(B^c).$$

דוגמא 2.1.20 הסיכוי שבן אוהב ברוקולי הוא 7%, הסיכוי שבת אוהבת ברוקולי הוא 31%. מבין הקונים בחנות, 70% בנות. הרגע נכנס לקוח. מה הסיכוי שהוא אוהב ברוקולי?

טענה 2.1.21 נוסחת ההסתברות השלמה הכללית: אם Ω מפורק לאיחוד זר $\bigcup B_i$ (סופי או אינסופי), אז

$$P(A)=\sum P(A|B_i)P(B_i). \tag{2.2}$$

(אם $P(B_i)=0$ מפרשים את המכפלה $P(A|B_i)P(B_i)$ כאפס, למרות שהגורם הראשון בה אינו מוגדר.)

תרגיל 2.1.22 מטילים מטבעות בזה אחר זה עד להופעת ה"עץ" הראשון, ואז בוחרים מטבע באקראי (ובסיכויים שווים) מבין אלה שהוטלו. הראה שהסיכוי לכך שהמטבע שנבחר יציג "עץ" הוא בדיוק $\log(2)$.

בני משפחת ברנולי תרמו רבות לתורת ההסתברות. הבעיה הבאה, שהוצגה ליעקב לברנולי, נחשבת לאחד הגורמים שהתניעו את ההתפתחות הזו.

תרגיל 2.1.23 משחק "הראשון לשש" שבו צריך להגיע לתוצאה 6, בין שני שחקנים שקולים, נקטע בתוצאה $5:3$. איך לחלק את הימור, ביחס של $3:1$ (יחס הנקודות החסרות), $3:5$ (יחס הנקודות שכבר הושגו), או אולי באופן אחר? כתוב עץ אפשרויות מתאים, וקבע מה היה צריך ברנולי להשיב.

2.1.4 חוק בייס

נוסחת ההיפוך מאפשרת לעבור מההסתברות של A בהנתן B להסתברות של B בהנתן A:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \tag{2.3}$$

דוגמא 2.1.24 25% מהרופאים מרוויחים מתחת לשכר הממוצע. 2% מבין המרוויחים מתחת לממוצע הם רופאים. אחוז המרוויחים מתחת לממוצע באוכלוסיה הוא 65%. מה אחוז הרופאים באוכלוסיה?

ציטוט 2.1.25 בקוסטה־ריקה עבר ב־*2011* חוק המאפשר לגזור שנתיים מאסר על גבר שהעליב את אשתו. יוזמי החוק הסבירו ש־70% ממקרי האלימות התחילו בהעלבה. מה דעתכם על החוק, אם רק 0.22% ממקרי ההעלבה מסתיימים באלימות?

תרגיל 2.1.26 שליש מזוגות התאומים הם תאומים זהים. באולטרהסאונד מגלים ששני התאומים הם מאותו מין. מה הסיכוי שהם זהים?

טענה 2.1.27 לכל חלוקה של המרחב $\Omega = \cup B_k$ לאיחוד זר, ולכל מאורע A, מתקיים **חוק בייס:** $P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_i P(A|B_i)P(B_i)}$.

תרגיל 2.1.28 התמנון שנודע בתקופת תחרויות הכדורגל המונדיאל של 2010 בשם "פול", הוא, קרוב לוודאי, תמנון פשוט שאין כל קשר בין התחזיות שלו לתוצאות המשחק; יש רק סיכוי קלוש (נאמר, 0.001%) שהוא נביאו של נפטון החוזה בדייקנות כל תוצאה. לאחר שפול קלע אל המטרה 7 פעמים רצופות, מה הסיכוי שהוא יחזה נכונה את תוצאת המשחק השמיני?

איך היית מגיב אם פול היה חוזה בדייקנות 40 משחקים רצופים?

דוגמא 2.1.29 באמצע שנות השמונים התמודדה אוניברסיטה גדולה בקליפורניה עם טענות על אפליה נגד נשים (המספרים כאן מומצאים): מבין הפונים ללימודי דוקטורט, התקבלו 44% מהבנים, ורק 37% מהבנות.

כששלטונות האוניברסיטה ניגשו לבדוק איזו פקולטה אשמה בפער, התברר שבכולן אחוז הנשים המתקבלות ללימודים גבוה מזה של הבנים. במדעי הטבע התקבלו 50% מהבנים ו־57% מהבנות, ובמדעי החברה התקבלו 28% מהבנים ו־33% מהבנות. ראו איזה פלא: למרות שבכל פקולטה הופלו הבנות לטובה, האוניברסיטה ככלל הפלתה אותן לרעה.

האם אפשר לקבוע מה אחוז הבנות מן הפונים לאוניברסיטה?

הערה 2.1.30 המצב שבו סיכום שגוי של שני זוגות שברים יביא להיפוך הסדר ביניהם נקרא **פרדוקס סימפסון**:

$$\frac{a}{x} > \frac{a'}{x'}, \qquad \frac{b}{y} > \frac{b'}{y'}, \qquad \frac{a+b}{x+y} < \frac{a'+b'}{x'+y'};$$

זהו המצב בדוגמא 2.1.29, כאשר x, x' הם מספרי הבנות והבנים שפנו לפקולטה למדעי הטבע, y, y' הם מספרי הבנות והבנים שפנו לפקולטה למדעים מדוייקים, ואילו a, a', b, b' הם מספרי המתקבלים בכל קבוצה. הפרדוקס הוא שלמרות שאחוז אחוז הבנות שהתקבלו בכל פקולטה גדול יותר מאחוז הבנים, אחוז הבנות שהתקבלו בסך־הכל קטן יותר. דוגמא פשוטה:

$$\frac{1}{1} > \frac{5}{6}, \qquad \frac{1}{6} > \frac{0}{1}, \qquad \frac{1+1}{1+6} < \frac{5+0}{6+1}.$$

השוו גם לדוגמא 1.2.4.

דוגמא 2.1.31 (בעיה מ'תחרות הערים' 32#, סתיו תשע"א, גרסת תרגול לכתות ט'־י') בבית ספר מסויים יותר מ־90% מהתלמידים יודעים גם אנגלית וגם גרמנית, ויותר מ־90% יודעים גם אנגלית וגם צרפתית. הוכח שמבין התלמידים שיודעים גם גרמנית וגם צרפתית, לפחות 90% יודעים אנגלית.

תרגיל 2.1.32 נוסעים עולים למטוס בזה אחר זה, לפי סדר. כל נוסע ניגש לכסא שלו: אם הכסא פנוי, הוא מתיישב שם. אחרת הוא בוחר כסא פנוי אחר, באקראי. הנוסע הראשון בחר בטעות בכסא אקראי. כל שאר הנוסעים התנהגו לפי הכללים. מהם הסיכויים לכך שהנוסע האחרון ישב בסופו של דבר בכסא שלו? **הדרכה.** אפשר לחשב את התוצאה באינדוקציה. כדי לקבל פתרון פשוט יותר, הראה שהאדם הראשון שתופס את הכסא הראשון או האחרון קובע את גורל המשחק; התשובה היא לכן חצי (הוכח שאם נוסע i מוצא את מקומו תפוס, זה מפני שכבר ישבו בכסאות $2, \dots, i-1$, ורק בהם).

2.1.5 תלות ואי־תלות

הגדרה 2.1.33 *מאורעות* A, B *הם* ***בלתי תלויים*** *אם*

$$P(A \cap B) = P(A)P(B).$$

ציטוט 2.1.34 "מוחמד לי. מוחמד הוא השם הפרטי הנפוץ ביותר; לי - שם המשפחה הנפוץ ביותר. מכיוון שלא ידעתי את התשובה, חשבתי שכך אשיג יתרון מתמטי" (שלדון, "המפץ הגדול", מנסה לנחש מיהו המומחה שיטוס לתחנת החלל). על אלו הנחות מבסס שלדון את הניחוש שלו?

תרגיל 2.1.35 נניח ש־$P(B) = 0$ או $P(B) = 1$, אז B בלתי תלויה בכל מאורע A.

טענה 2.1.36 נניח $0 < P(A), P(B) < 1$. התנאים הבאים שקולים:

1. $P(A|B) = P(A)$;
2. $P(A|B) = P(A|B^c)$;
3. $P(A \cap B) = P(A) \cdot P(B)$.
4. $P(B|A) = P(B)$;

בפרט, A ו־B בלתי תלויים אם הסיכוי ל־A בהנתן B שווה לסיכוי ל־A ללא נתון זה.

טענה 2.1.37 התנאים הבאים שקולים:

1. A ו־B בלתי תלויים;
2. A ו־ B^c בלתי תלויים;
3. A^c ו־ B בלתי תלויים;
4. A^c ו־ B^c בלתי תלויים.

למאורע A, נסמן $A^0 = A$ ו־$A^1 = A^c$.

הגדרה 2.1.38 המאורעות $A_1, \dots, A_n$ הם **בלתי תלויים במשותף** אם לכל בחירת פרמטרים $\alpha_1, \dots, \alpha_n \in \{0, 1\}$ מתקיים $P(A_1^{\alpha_1} \cap \dots \cap A_n^{\alpha_n}) = P(A_1^{\alpha_1}) \cdots P(A_n^{\alpha_n})$.

לפי טענה 2.1.37, A, B 'בלתי תלויים במשותף' אם ורק אם הם בלתי תלויים.

טענה 2.1.39 המאורעות $A_1,\dots,A_n$ בלתי תלויים במשותף, אם ורק אם לכל $I \subseteq \{1,\dots,n\}$,

$$P(\bigcap_{i\in I} A_i) = \prod_{i\in I} P(A_i).$$

טענה 2.1.40 אם המאורעות $A_1,\dots,A_n$ בלתי תלויים במשותף, אז כל תת־קבוצה שלהם בלתי תלויה במשותף. **הדרכה.** מספיק להוכיח ש־$A_1,\dots,A_{n-1}$ בלתי תלויים; או הפעל את טענה 2.1.39.

דוגמא 2.1.41 נסמן ב־Ω את המרחב הכולל וקטורים $(x_1,x_2,x_3) \in \{0,1\}^3$ שסכומם מודולו 2 הוא 0; כך $|\Omega| = 4$. נסמן ב־A_i את המאורע $\{(x_1,x_2,x_3):\ x_i = 0\}$. אז כל A_i, A_j $(i \neq j)$ הם בלתי תלויים, אבל A_1,A_2,A_3 אינם בלתי תלויים במשותף.

תרגיל 2.1.42 על־פי ההגדרה, $A_1,\dots,A_n$ בלתי תלויים במשותף אם התנאי

$$P(A_1^{\alpha_1} \cap \dots \cap A_n^{\alpha_n}) = P(A_1^{\alpha_1}) \cdots P(A_n^{\alpha_n})$$

מתקיים עבור 2^n וקטורים שונים $(\alpha_1,\dots,\alpha_n)$. הראה שמספיק לבדוק $2^n - n - 1$ תנאים כדי להבטיח אי־תלות במשותף (האם זהו המספר הקטן ביותר של תנאים שיש לבדוק?) ואם ידוע ש־$A_1,\dots,A_n$ בלתי תלויים בזוגות?

דיון 2.1.43 תושבת אנגליה בשם סאלי קלארק הורשעה ב־1999 ברצח שני ילדיה, שמתו בזה אחר זה מוות בעריסה. ההרשעה היתה מבוססת על כך שהסיכוי למוות בעריסה הוא כ־1:8500; מכאן שהסיכוי לשני מקרי מוות בעריסה הוא $1{:}8500^2 \approx 1{:}73000000$; לכן ההסבר הסביר היחיד הוא שהגב' קלראק הרגה את הילדים. מה דעתכם? (ההרשעה הושארה על כנה בערעור הראשון שהתקיים בסוף 2000, ובוטלה בתחילת 2003. הגב' קלארק פיתחה סדרה של הפרעות פסיכיאטריות ומתה ב־2007, בגיל 42, מהרעלת אלכוהול.)

2.2 משתנים מקריים

2.2.1 משתנה יחיד

משתנה מקרי הוא פונקציה $X:\Omega \to \mathbb{R}$ (בהמשך נוסיף את התנאי שהפונקציה תהיה "מדידה"; במרחבי הסתברות בדידים זה לא רלוונטי). בנוסחת ההסתברות השלמה טיפלנו בהסתברות של מאורע בנוכחות חלוקה של המרחב לקבוצות זרות. משתנה מקרי משרה חלוקה כזו, משום שהקבוצות

$$X^{-1}(a) = \{\omega \in \Omega:\ X(\omega) = a\}$$

הן מאורעות זרים, והאיחוד שלהן על־פני כל הערכים האפשריים של a הוא המרחב כולו. אפשר לחשב את ההסתברות של כל מאורע כזה, שאפשר לסמן בכל הדרכים הבאות:

$$P(X=a)=P(X(\omega)=a)=P(\{\omega:\ X(\omega)=a\})=P(X^{-1}(a)).$$

לפי הגדרת ההסתברות של מאורע,

$$P(X=a)=\sum_{\omega:X(\omega)=a}P(\omega). \qquad (2.4)$$

דוגמא 2.2.1 **המשתנה המציין** של מאורע $A\subseteq\Omega$ מוגדר לפי $X_A(\omega)=1$ אם $\omega\in A$, ו־$X_A(\omega)=0$ אחרת. לפיכך, $P\left(X_a=1\right)=P\left(A\right)$.

דוגמא 2.2.2 כל קבוע c הוא משתנה מקרי, אם נחשוב עליו כפונקציה $\Omega\to\mathbb{R}$ המקבלת באופן זהותי את הערך c.

דוגמא 2.2.3 אם $a_1,\dots,a_n$ מספרים ו־$p_1,\dots,p_n$ מספרים חיוביים שסכומם 1, אז $P\left(X=a_i\right)=p_i$ מגדיר משתנה מקרי. זהו המשתנה המקרי הכללי ביותר על מרחב סופי. ההכללה למקרה הבן־מניה מיידית (קח סדרה a_i וטור חיובי p_i המסתכם ל־1).

הגדרה 2.2.4 יהי $X:\Omega\to\mathbb{R}$ משתנה מקרי על מרחב הסתברות בדיד (Ω,P). הפונקציה

$$a\mapsto P\left(X=a\right)$$

נקראת **פונקציית ההתפלגות של** X.

פונקציית ההתפלגות קובעת מהם הערכים ש־X מקבל, ובאיזו הסתברות מתקבל כל אחד מהם.

דוגמא 2.2.5 להתפלגות של משתנה מקרי המקבל ערכים שלמים בקטע $[1,n]$ בהסתברויות שוות קוראים **התפלגות אחידה**. כדי לציין שיש למשתנה התפלגות כזו, מסמנים $X\sim U[1,n]$ (יש להבין מההקשר שמדובר במשתנה בדיד, משום שבהמשך נשתמש באותו סימון בדיוק עבור משתנים רציפים).

טענה 2.2.6 תהי $f:a\mapsto P(X=a)$ פונקציית ההתפלגות של משתנה מקרי (המוגדר על כל $\mathbb{R}$). אז עוצמת הקבוצה $\{a:\ f(a)>0\}$ בת־מניה לכל היותר.

הוכחה. נסמן $T_n=\left\{a\in\mathbb{R}:\ f(a)>\frac{1}{n}\right\}$, אז $\{a:\ f(a)>0\}=\cup_{n=1}^{\infty}T_n$, אבל כל T_n סופית כי $\frac{1}{n}|T_n|=\sum_{a\in T_n}\frac{1}{n}\leq\sum_{a\in T_n}P(X=a)=P(T_n)\leq 1$. $\square$

הערה 2.2.7 (בניית משתנה מקרי מפונקציית ההתפלגות שלו) יהי X משתנה מקרי עם פונקציית התפלגות $f(a) = P(X = a)$. מפונקציית ההתפלגות אי אפשר לשחזר את המרחב, משום ששמות האברים ב־Ω הלכו לאיבוד; ובכל זאת, אפשר לבנות מרחב הסתברות חדש, ומשתנה מקרי חדש, ש־f היא (גם) פונקציית ההתפלגות שלו.

אכן, נגדיר $\tilde{\Omega} = \{a \in \mathbb{R} :\ f(a) > 0\} \subseteq X(\Omega)$ (מדוע זו הכלה, ולא שוויון?). לפי טענה 2.2.6, $\tilde{\Omega}$ היא קבוצה סופית או בת מניה. נגדיר $\tilde{P}(a) = f(a)$. זו פונקציה חיובית, ו־

$$\begin{aligned}\sum_{a\in\tilde{\Omega}} \tilde{P}(a) &= \sum_{a\in\mathbb{R}} f(a) = \sum_{a\in\mathbb{R}} P(X=a) = \sum_{a\in\mathbb{R}} P(\{\omega\in\Omega :\ X(\omega)=a\}) \\ &= \sum_{a\in\mathbb{R}} \sum_{\{\omega :\ X(\omega)=a\}} P(\omega) = \sum_{\omega\in\Omega} P(\omega) = 1;\end{aligned}$$

מכאן ש־$(\tilde{\Omega}, \tilde{P})$ הוא מרחב הסתברות (כל הסכומים שהם כביכול על־פני $a \in \mathbb{R}$ הם למעשה סכומים בני מניה). נגדיר $\tilde{X} : \tilde{\Omega} \to \mathbb{R}$ להיות פונקציית הזהות, $\tilde{X}(a) = a$. פונקציית ההתפלגות של $\tilde{X}$ היא $b \mapsto \tilde{P}(\tilde{X} = b) = \tilde{P}(\left\{a \in \tilde{\Omega} :\ a = b\right\}) = \tilde{P}(b) = f(b)$, כלומר f.

ל־$(\tilde{\Omega}, \tilde{P})$ ול־$\tilde{X}$ שבנינו לעיל יש יתרון על־פני (Ω, P) ו־X, בכך ש־$\tilde{\Omega}$ הוא מרחב של מספרים, וש־$\tilde{X}$ היא פונקציה פשוטה במיוחד. איננו מאמצים את השיטה הזו באופן כללי, משום שלעתים קרובות יש בתמונה כמה משתנים מקריים, ואנחנו זקוקים למרחב יציב שאפשר להפעיל עליו את כולם. חשיבותה של ההערה בכך שכאשר נדון בתכונות של משתנים מקריים, לא יהיה לנו צורך ללוות כל הגדרה בהגדרה מקבילה של מרחב ההסתברות; אלו, כביכול, מוגדרים מעצמם.

תרגיל 2.2.8 נניח ש־$\Omega = \{1, 2, 3, 4, 5, 6\}$. הגדר משתנה מקרי על המרחב Ω, כך ש־$P(X = 0) = P(X = 1) = P(X = 2) = \frac{1}{3}$. כמה משתנים כאלה יש?

דיון 2.2.9 הרהר במשך 20 שניות בהבדל שבין 'למשתנים המקריים X, Y יש אותה התפלגות' לבין 'המשתנים המקריים X, Y שווים'.

אם X משתנה מקרי, אז לכל פונקציה $f : \mathbb{R} \to \mathbb{R}$, גם $f(X)$ משתנה מקרי (שוב, בהמשך נדרוש שהפונקציה תהיה "מדידה"). לפי ההגדרה מתקיים

$$P(f(X) = b) = \sum_{a : f(a)=b} P(X = a).$$

הערה 2.2.10 אם $B \subseteq \Omega$ מאורע, אפשר להגדיר את המשתנה המותנה $X|B$, עם פונקציית ההתפלגות $P(X = a|B) = P(\{\omega :\ X(\omega) = a\}|B)$.

2.2.2 התפלגות משותפת

אם נתונים שני משתנים מקריים $X,Y:\Omega\to\mathbb{R}$, אפשר להגדיר מאורעות $\{X=a,Y=b\}$, ולקבל **פונקציית התפלגות משותפת** $(a,b)\mapsto P(X=a,Y=b)$.

המאורע $X=a,Y=b$ מסמן את החיתוך

$$\{\omega:\ X(\omega)=a,Y(\omega)=b\}=\{\omega:\ X(\omega)=a\}\cap\{\omega:\ Y(\omega)=b\}.$$

את פונקציית ההתפלגות המשותפת של שני משתנים אפשר לתאר במערך דו־ממדי.

הערה 2.2.11 מתקיים $P(X=a)=\sum_b P(X=a,Y=b)$, וכך גם $P(Y=b)=\sum_a P(X=a,Y=b)$.

לכל ערך b של Y, $X|Y=b$ הוא משתנה מקרי מותנה, וההתפלגות המותנית שלו היא $P(X=a|Y=b)=\frac{P(X=a,Y=b)}{P(Y=b)}$.

תרגיל 2.2.12 בכד יש כדור אדום וכדור כחול. בכל פעם מוציאים כדור באקראי, ומחזירים אותו עם כדור נוסף באותו צבע. הוכח שהתפלגות מספר הכדורים האדומים בכל שלב היא אחידה. (**רמז.** אינדוקציה).

אי־תלות

המשתנים X,Y **בלתי־תלויים** אם $P(X=a,Y=b)=P(X=a)P(Y=b)$ לכל a,b;

טענה 2.2.13 התנאים הבאים על משתנים מקריים X,Y הם שקולים:

1. X,Y בלתי תלויים;
2. ההתפלגות של $X|Y=b$ היא אותה התפלגות לכל b;
3. כל זוג של מאורעות $X=a$ ו־ $Y=b$ הם בלתי תלויים.

שימו לב: ההתפלגות של X וההתפלגות של Y, יחד, קובעות את ההתפלגות המשותפת בתנאי שידוע ששני המשתנים בלתי תלויים. אחרת טענה זו אינה נכונה.

הערה 2.2.14 את הערה 2.2.7 אפשר להכליל לשני משתנים על־ידי בניית מרחב התפלגות משותף, שבו המשתנים בלתי תלויים מתוך ההגדרה: אם Ω,Ω' מרחב הסתברות, גם $\Omega\times\Omega'$ הוא כזה, אם מגדירים $P(\omega,\omega')=P(\omega)P(\omega')$. אם X,Y המוגדרים על מרחב המכפלה מתפצלים דרך הרכיב הראשון והשני בהתאמה, אז הם בלתי תלויים.

טענה 2.2.15 אם X,Y בלתי תלויים, אז לכל $f:\mathbb{R}\to\mathbb{R}$ גם $f(X),Y$ בלתי תלויים.
מכאן גם ש־$f(X),g(Y)$ בלתי תלויים לכל שתי פונקציות $f,g:\mathbb{R}\to\mathbb{R}$.

הגדרה 2.2.16 *אם X,Y,Z משתנים מקריים, נאמר ש־Y* ***מפריד את*** *X,Z אם ידיעת Y הופכת את X,Z לבלתי תלויים, כלומר*

$$\forall y \forall x,z: \quad P\left(X=x,Z=z|Y=y\right)=P\left(X=x|Y=y\right)P\left(Z=z|Y=y\right).$$

תרגיל 2.2.17 תן דוגמא למשתנים מקריים X,Y,Z כך ש־

1. X,Z תלויים, ו־Y מפריד אותם.
2. X,Z בלתי תלויים, אבל Y אינו מפריד אותם (ואפילו יותר מזה: $X|Z=z$ ו־$Y|Z=z$ תלויים לכל ערך z של Z).

דוגמא 2.2.18 (n מ"מ שכל d מהם בלתי תלויים במשותף, וכל $d+1$ מהם תלויים.) נבחר שדה סופי $\mathbb{F}_q$ ו־n ערכים שונים $a_1,\dots,a_n\in\mathbb{F}_q$ (השדה צריך להיות גדול מספיק, כמובן). נבחר משתנים מקריים בלתי תלויים ובעלי התפלגות אחידה, $X_0,\dots,X_{d-1}$. נגדיר את הפולינום המקרי $P(t)=X_0+X_1t+\cdots+X_{d-1}t^{d-1}$, ובעזרתו את המשתנים המקריים $Y_i=P(a_i)$. לכל $i_1,\dots,i_d$ שונים, ההתפלגות המשותפת של $(Y_{i_1},\dots,Y_{i_d})$ היא אחידה (משום שלפי ההפיכות של מטריצת ואן־דר־מונדה, למערכת המשוואות $P(a_i)=b_i$ יש פתרון יחיד עבור ה־X_j, לכל בחירה של ה־b_i); אבל לכל $i_1,\dots,i_{d+1}$, כל d מבין $Y_{i_1},\dots,Y_{i_{d+1}}$ קובעים את האחרון.

סכום ומכפלה

כפי ש־$f(X)$ משתנה מקרי לכל פונקציה $f:\mathbb{R}\to\mathbb{R}$, גם $f(X,Y)$ משתנה מקרי לכל $f:\mathbb{R}^2\to\mathbb{R}$. בפרט הסכום $X+Y$ והמכפלה XY הם משתנים מקריים.

את זה אפשר להכליל לסכום של כמה משתנים מקריים; בהמשך נפגוש דוגמא חשובה: הממוצע $\bar{X}=\frac{1}{n}(X_1+\cdots+X_n)$.

תרגיל 2.2.19 אוסף המשתנים המקריים על מרחב Ω הוא מרחב וקטורי, שממדו, אם Ω סופי, $|\Omega|$.

תרגיל 2.2.20 מטילים שתי קוביות בלתי תלויות, X,Y. איך מתפלג הסכום $X+Y$?

2.2.3 תוחלת של משתנה מקרי בדיד

למשתנה מקרי אפשר (בדרך כלל) להתאים מספר מייצג, ה**תוחלת** של המשתנה, המשקלל את הערכים שהמשתנה מקבל על־פי ההסתברות שלהם. בהמשך נראה שתוחלת תופסת תפקיד מרכזי בכל חישוב הסתברותי, ואף נוכיח שאם מגרילים ערכים רבים של משתנה מקרי, ממוצע התוצאות הולך ומתקרב לתוחלת. את מושג התוחלת הכניס לשימוש המתמטיקאי והאסטרונום ההולנדי, כריסטיאן הויגנס (1629-1695).

הגדרה 2.2.21 ה**תוחלת** של משתנה מקרי בדיד X מוגדרת לפי

$$\mathbf{E}(X) = \sum_{\omega\in\Omega} P(\omega)X(\omega) = \sum_a P(X=a)a,$$

בתנאי שהטור מתכנס בהחלט.

כאן, ובהמשך, בסכום על a הכוונה לסכום על־פני כל הערכים שבהם ההסתברות $P(X=a)>0$; מספר הערכים האלה בן מניה לכל היותר (טענה 2.2.6), ולכן הסכום מוגדר היטב.

ההגדרות אכן שקולות זו לזו, לפי (2.4):

$$\begin{aligned}
\sum_a P(X=a)a &= \sum_a P(\{\omega:\ X(\omega)=a\})a \\
&= \sum_a \sum_{\omega:X(\omega)=a} P(\omega)a \\
&= \sum_{\omega\in\Omega} P(\omega)X(\omega).
\end{aligned}$$

דוגמא 2.2.22 אם מרחב הערכים של X הוא אינסופי, הטור לא בהכרח מתכנס. למשל, חישבו על המשתנה המקבל ערך n בהסתברות $\frac{6}{\pi^2 n^2}$.

דוגמא 2.2.23 אם $a_1,\dots,a_n$ מספרים ו־$P(X=a_i)=\frac{1}{n}$, אז $\mathbf{E}(X)=\frac{a_1+\cdots+a_n}{n}$. כלומר, התוחלת מכלילה את הממוצע. באופן כללי יותר אם $P(X=a_i)=p_i$ עבור $p_1,\dots,p_n$ חיוביים שסכומם 1, אז $\mathbf{E}(X)=p_1a_1+\cdots+p_na_n$.

תרגיל 2.2.24 אדם הקונה כרטיס הגרלה עשוי לזכות בפרסים שונים, שלכל אחד מהם הסתברות משלו. את גובה הזכיה אפשר לתאר כמשתנה מקרי X. הסבר מדוע **המחיר ההוגן** של הכרטיס הוא $\mu=\mathbf{E}(X)$ (אפילו אם תצטרך להגדיר באופן מתמטי מהי הגינות).

תרגיל 2.2.25 בסרט "התייר" (2010) אומר מפקד המשטרה למפקח: הונו של [הפושע] אלכסנדר מוערך ב־$\$744M$, אותם אפשר יהיה להחרים אם יתפס. מכיוון שהמבצע ללכידתו עלה עד כה $\$8M$, רציונלי מצדי להמשיך את המבצע אם יש סיכוי של לפחות 1% להצלחה.

מה דעתך על השימוש בתוחלת במקרה זה?

הערה 2.2.26 לכל פונקציה, $\mathbf{E}(f(X)) = \sum_a P(X=a)f(a)$.

הוכחה. התוחלת של המשתנה המקרי $f(X)$ שווה, לפי ההגדרה, ל־

$$\sum_{\omega\in\Omega} P(\omega)f(X(\omega)) = \sum_a \left(\sum_{\omega:X(\omega)=a} P(\omega)\right) f(a) = \sum_a P(X=a)f(x).$$

□

באותו אופן גם

$$\mathbf{E}(g(X,Y)) = \sum_{a,b} P(X=a, Y=b)g(a,b), \tag{2.5}$$

וכך לפונקציות בכל מספר משתנים.

טענה 2.2.27 1. לכל קבוע c, $\mathbf{E}(c) = c$.

2. $\mathbf{E}(\alpha X) = \alpha \mathbf{E}(X)$.

3. נניח ש־$X \geq 0$, כלומר X הוא משתנה מקרי חיובי. אז $\mathbf{E}(X) \geq 0$; ואם $\mathbf{E}(X) = 0$, אז $X = 0$ בהסתברות 1.

טענה 2.2.28 לכל שני משתנים מקריים, $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$.

הוכחה. לפי (2.5),

$$\begin{aligned}
\mathbf{E}(X+Y) &= \sum_{a,b} P(X=a,Y=b)(a+b) \\
&= \sum_{a,b} P(X=a,Y=b)a + \sum_{a,b} P(X=a,Y=b)b \\
&= \sum_a \left(\sum_b P(X=a,Y=b)\right) a + \sum_b \left(\sum_a P(X=a,Y=b)\right) b \\
&= \sum_a P(X=a)a + \sum_b P(Y=b)b \\
&= \mathbf{E}(X) + \mathbf{E}(Y).
\end{aligned}$$

□

האדיטיביות של התוחלת היא אחד העקרונות המועילים ביותר בתורת ההסתברות. היא מאפשר לחשב תוחלת של התפלגויות מסובכות ביותר, בלי להבין את ההתפלגות, על־ידי פירוק המשתנה לסכום של דברים שאפשר לחשב.

תרגיל 2.2.29 מסדרים n אנשים באקראי. מי שגבוה מכל מי שלפניו, מרים יד. מה תוחלת מספר האנשים שמרימים יד? (הדרכה: העזר במשתנים מציינים.)

תרגיל 2.2.30 בהמשך לבעיית המזכירה המבולבלת (תרגיל 2.1.11), שבה חישבנו את הסיכוי לכך שמספר נקודות השבת של תמורה מקרית יהיה 0, מה תוחלת מספר נקודות השבת?

תרגיל 2.2.31
1. בהגרלה עומדים לבחור כרטיס זוכה באקראי מכל אחד משני הכדים. כל משתתף רשאי להטיל את הכרטיס שלו לאיזה כד שירצה. לאיזה כד תבחר להטיל את הכרטיס שלך? (**פתרון**. לזה שמספר הכרטיסים בו הוא הקטן ביותר, כמובן.)
2. ואם בכל כד n כרטיסים, ולך שני כרטיסים, האם תעדיף להטיל את שניהם יחד, או לפזר את הסיכויים בין הכדים? (**פתרון**. עדיף לפזר; שני כרטיסים באותו כד מתחרים זה בזה.)
3. נניח שעליך להטיל את שני הכרטיסים בתחילת המשחק, והמארגנים ישלימו את מספר הכרטיסים בכל כד ל־n. האם תעדיף להטיל אותם יחד, או בנפרד? (**פתרון**. אם המטרה היא להגדיל את תוחלת מספר הזכיות, זה לא משנה. אבל אם המטרה היא להגדיל את הסיכוי לזכות בפרס אחד לפחות, עדיף להצמיד את הכרטיסים.)

טענה 2.2.32 אם $\Omega = A_1 \cup \cdots \cup A_n$ הוא פירוק של המרחב לקבוצות זרות, אז $\mathbf{E}(X) = \sum_i P(A_i)\mathbf{E}(X|A_i)$.

הוכחה. זו תוצאה של נוסחת ההסתברות השלמה (2.2). לפי ההגדרה

$$\begin{aligned}
\mathbf{E}(X) &= \sum_a P(X=a)a \\
&= \sum_a \sum_i P(X=a|A_i)P(A_i)a \\
&= \sum_i P(A_i) \sum_a aP(X=a|A_i) \\
&= \sum_i P(A_i)\mathbf{E}(X|A_i).
\end{aligned}$$

□

יהיו X,Y משתנים מקריים. הסימון $X|Y$ פירושו - המשתנה המקרי X, בהנתן הערך של Y, שאותו מסמנים לצורך העניין ב־Y במקום לתת לו שם נפרד. זהו, אם כך, משתנה מקרי, שההתפלגות שלו תלויה בערך של Y. משום כך, **התוחלת המותנית** $\mathbf{E}(X|Y)$ של המשתנה המקרי $X|Y$ היא בעצמה פונקציה של Y.

טענה 2.2.33 (חוק התוחלת החוזרת) $\mathbf{E}(X) = \mathbf{E}(\mathbf{E}(X|Y))$.

הוכחה. לפי טענה 2.2.32,

$$\mathbf{E}(X) = \sum_b P(Y=b)\mathbf{E}(X|Y=b) = \mathbf{E}(\mathbf{E}(X|Y)).$$

$\square$

הערה 2.2.34 גם הטענה $\mathbf{E}(\mathbf{E}(X)|Y) = \mathbf{E}(X)$ נכונה, אך היא טריוויאלית: התוחלת $\mathbf{E}(X)$ היא ערך מספרי, ולכן המשתנה המקרי המותנה $\mathbf{E}(X)|Y$ הוא קבוע, השווה תמיד ל־$\mathbf{E}(X)$; ממילא גם התוחלת שלו היא $\mathbf{E}(X)$.

מסקנה 2.2.35 אם X,Y בלתי תלויים אז

$$\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y).$$

הוכחה. $\mathbf{E}(XY) = \mathbf{E}(\mathbf{E}(XY|X)) = \mathbf{E}(X\mathbf{E}(Y|X)) = \mathbf{E}(X\mathbf{E}(Y)) = \mathbf{E}(X)\mathbf{E}(Y)$, כי ל־$Y|X$ ההתפלגות של Y לכל (ערך של) X. $\square$

תרגיל 2.2.36 ב־$3/2011$ הוגשה תביעה נגד מפעל הפיס בטענה שעל המארגנים לפרסם את העובדה שכרטיסי גירוד מסדרה מסויימת כבר זכו בפרסים, על־מנת שאנשים יידעו שאין עוד טעם לקנות אותם. נניח, לשם פשטות הניתוח, שמדפיסים n כרטיסים בסדרה, ואחד מהם זוכה. הכרטיסים נמכרים בזה אחר זה.

1. נניח שהמחיר ההוגן לכרטיס (בשיטה הקיימת, לפיה אין מפרסמים נתוני מכירות חלקיים) הוא 1; הראה שגובה הפרס הוא n. במלים אחרות, הפרס שווה לסכום מחירי הכרטיסים.
2. נניח שהכרטיסים נמכרים בזה אחר זה, ומפעל הפיס מודיע ברגע שהכרטיס הזוכה נמכר (כך שאנשים יחדלו מלקנות אותם). מה תהיה תוחלת ההכנסות ממכירת כרטיסים?

3. נניח שהתביעה מתקבלת, ומשרד הפיס מתחיל לפרסם נתונים שוטפים על מכירת הכרטיסים. אם נמכרו כבר i כרטיסים שלא זכו, אז המחיר ההוגן כבר אינו 1, אלא $\frac{n}{n-i}$ (הסבר מדוע). מה תהיה תוחלת ההכנסה ממכירת כרטיסים במקרה כזה? (הדרכה: סמן ב־X_i את ההכנסה מן הכרטיס ה־i: $\frac{n}{n-i}$ אם הפרס טרם הופיע בשלב ה־i, ו־0 אחרת. המשתנים X_i תלויים, אבל ההכנסה הכוללת היא $W = \sum_{i=1}^{n} X_i$, ו־$\mathbf{E}(W) = \sum \mathbf{E}(X_i)$ לפי אדיטיביות התוחלת.)

תרגיל 2.2.37 מגרש החניה באוניברסיטה הוא בצורת שורה ארוכה, עם סימונים ברווח קבוע. אורכה של מכונית הוא כפליים המרחק בין שני סימונים. הנהגים חונים על־פי הסימונים, אבל אינם מתחשבים באלו שיבואו אחריהם (למשל, אם נותרה שורה פנויה באורך 4, כדוגמת oooo, הנהג עשוי לחנות בכל המקומות xxoo, oxxo, ooxx - בסיכויים שווים; שורה פנויה באורך 1 תשאר ריקה, ואם נותרה שורה פנויה באורך 2 או 3, היא תספיק כמובן למכונית אחת בלבד). כתוב את נוסחת הרקורסיה עבור התוחלת של מספר המכוניות שיזכו למקום חניה. העזר במחשב כדי למצוא את התוחלת כאשר $n = 1000$. (הגרסה הרציפה של שאלה זו קשורה ל"קבוע החניה של Renyis", שערכו כ־0.74759.)

2.2.4 שונות

הגדרה 2.2.38 **השונות** של משתנה מקרי X היא

$$\mathbf{V}(X) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{E}(X^2) - \mathbf{E}(X)^2.$$

הערה 2.2.39

$\mathbf{V}(X+\beta) = \mathbf{V}(X)$.

$\mathbf{V}(\alpha X) = \alpha^2 \mathbf{V}(X)$.

טענה 2.2.40 לכל משתנה מקרי X,

1. $\mathbf{E}(X^2) \geq \mathbf{E}(X)^2$; לכן $\mathbf{V}(X) \geq 0$.
2. $\mathbf{V}(X) = 0$ אם ורק אם X קבוע בהסתברות 1.

(ההכללה של סעיף 1 ידועה בשם אי־שוויון הולדר: אם $f'' \geq 0$ אז $\mathbf{E}(f(X)) \geq f(\mathbf{E}(X))$.)

תרגיל 2.2.41 1. אדם יכול לקנות כרטיס הגרלה שמחירו 100 ש"ח, ובו זוכים, בסיכוי של אחד ל־100000, בסכום של 8000000 ש"ח (הכרטיס אינו הוגן: אם מפעל ההגרלות ימכור 100000 כרטיסים ואחד מהם יזכה, ישאר בידיו רווח של 2000000 ש"ח). הראה שתוחלת הרכוש של קונה הכרטיס יורדת, בעוד שהשונות עולה. לכן בקניית כרטיס הגרלה אנו קונים עליה של השונות (התרגשות, מתח) תמורת הפסד תוחלת.

2. אדם יכול לקנות פוליסת ביטוח במחיר 100 ש"ח: אם ביתו ישרף (אירוע שהסיכויים לו הם אחד ל־100000), תשלם לו חברת הביטוח 8000000 ש"ח (הפוליסה אינה הוגנת: אם חברת הביטוח תמכור 100000 ביטוחים ובית אחד ישרף, ישאר בידה רווח של 2000000 ש"ח). הראה שתוחלת הרכוש של קונה הכרטיס יורדת, בעוד שהשונות יורדת. לכן בקניית פוליסת ביטוח אנו קונים את ירידת השונות (בטחון, שקט נפשי) תמורת הפסד תוחלת.

3. נתח את התפלגות הרכוש של מי שקונה גם כרטיס הגרלה וגם פוליסת ביטוח. מה התוחלת שלה? מה השונות שלה?

טענה 2.2.42 נוסחת פירוק השונות: $\mathbf{V}(X) = \mathbf{V}(\mathbf{E}(X|Y)) + \mathbf{E}(\mathbf{V}(X|Y))$.

הוכחה. נחשב לפי חוק התוחלת החוזרת:

$$\mathbf{V}(\mathbf{E}(X|Y)) = \mathbf{E}(\mathbf{E}(X|Y)^2) - \mathbf{E}(\mathbf{E}(X|Y))^2 = \mathbf{E}(\mathbf{E}(X|Y)^2) - \mathbf{E}(X)^2,$$

$$\mathbf{E}(\mathbf{V}(X|Y)) = \mathbf{E}(\mathbf{E}(X^2|Y)) - \mathbf{E}(X|Y)^2) = \mathbf{E}(X^2) - \mathbf{E}(\mathbf{E}(X|Y)^2),$$

ולכן הסכום שווה ל־$\mathbf{V}(X)$. □

דוגמא 2.2.43 איך מתפרקת השונות כאשר X,Y בלתי תלויים? וכאשר X הוא פונקציה של Y? ואם $X = Y$?

תרגיל 2.2.44 נסמן $W = \mathbf{E}(X|Y)$ ו־ $T = \mathbf{E}(W|X)$. הוכח ש־ $\mathbf{V}(T) \leq \mathbf{V}(X)$.
פתרון. $\mathbf{V}(T) = \mathbf{V}(\mathbf{E}(W|X)) \leq \mathbf{V}(W) = \mathbf{V}(\mathbf{E}(X|Y)) \leq \mathbf{V}(X)$.

תרגיל 2.2.45 1. נניח ש־(X_0, X_1) מקבלים את הערכים $(0,0),(0,1),(1,0),(1,1)$ בסיכויים $\frac{1-\rho}{4}, \frac{1-\rho}{4}, \frac{1+\rho}{4}$, ו־$\frac{1+\rho}{4}$, בהתאמה, כאשר $0 < \rho < 1$ [כלומר $X_0, X_1 \sim b(\frac{1}{2})$ עם $\rho(X_0,X_1) = \rho$]. הראה ש־$\mathbf{E}(X_i|X_j) = \rho X_j + \frac{1-\rho}{2}$ (כאשר $\{i,j\} = \{0,1\}$). נגדיר את הסדרה $X_{n+1} = \mathbf{E}(X_n|X_{n-1})$. הוכח שלכל $n \geq 1$, $X_n = \frac{1-\rho^{n-1}}{2} + \rho^{n-1} X_{n \pmod 2}$.

2. ($\star$) יהיו $Y = X_0, X = X_1$ משתנים מקריים על מרחב Ω. נגדיר סדרה X_n כבסעיף הראשון. חקור את הסדרה X_n; הראה שסדרת השונויות מונוטונית יורדת. האם בהכרח $\mathbf{V}(X_n) \to 0$?

2.2.5 שונות משותפת ומקדם המתאם

הגדרה 2.2.46 **השונות המשותפת** של משתנים מקריים X,Y היא המספר

$$\mathrm{Cov}(X,Y)=\mathbf{E}(XY)-\mathbf{E}(X)\mathbf{E}(Y)=\mathbf{E}((X-\mathbf{E}(X))(Y-\mathbf{E}(Y))).$$

דוגמא 2.2.47 מקדם המתאם של משתנים מקריים מכליל את מקדם המתאם של מדגם, בדיוק כפי שהתוחלת של משתנה מקרי מכלילה את הממוצע של מדגם (דוגמא 2.2.23).

הגדרה 2.2.48 X,Y **בלתי מתואמים** אם $\mathrm{Cov}(X,Y)=0$.

טענה 2.2.49 Cov היא תבנית ביליניארית על מרחב המשתנים המקריים.

1. סימטריות: $\mathrm{Cov}(Y,X)=\mathrm{Cov}(X,Y)$.
2. אדיטיביות: $\mathrm{Cov}(X,Y+Z)=\mathrm{Cov}(X,Y)+\mathrm{Cov}(X,Z)$.
3. הומוגניות: $\mathrm{Cov}(X,\alpha Y)=\alpha\mathrm{Cov}(X,Y)$.

טענה 2.2.50 לכל משתנה מקרי X, $\mathrm{Cov}(X,X)=\mathbf{V}(X)\geq 0$, עם שוויון אם ורק אם X קבוע (בהסתברות 1). פירושו של דבר הוא שהשונות המשותפת היא "מכפלה פנימית על מרחב המנה של המשתנים המקריים מודולו סקלרים".

אם X,Y בלתי תלויים אז הם גם בלתי מתואמים (מסקנה 2.2.35).

תרגיל 2.2.51
1. יתכן ש־X,Y תלויים ובלתי מתואמים. **הדרכה**. קח X להיות המשתנה המקרי מקבל את הערכים $-1,0,1$ בהסתברויות $\frac{1}{4},\frac{1}{2},\frac{1}{4}$ בהתאמה. נסמן $Y=|X|$. הראה ש־$\mathrm{Cov}(X,Y)=0$.
2. יתכן ש־X,Y תלויים למרות ש־Y בלתי מתואם עם כל פונקציה של X. **הדרכה**. יהי X משתנה מקרי כלשהו, ו־Z המטבע ההוגן בעל ערכים ± 1, שאינו תלוי ב־X. קח $Y=XZ$. הראה ש־$\mathrm{Cov}(f(X),Y)=0$ לכל פונקציה של X.
3. יתכן ש־X,Y תלויים למרות שכל X^n בלתי מתואם עם כל Y^m. **הדרכה**. דוגמא נגדית, לא קלה, אפשר למצוא ב־The American Statistician, **63**(3), (2009), Letter to the editor, http://www.tandfonline.com/doi/abs/10.1198/tast.2009.09051.
4. אם כל פונקציה רציפה של X בלתי מתואמת עם כל פונקציה רציפה של Y, אז X,Y בלתי תלויים.

הגדרה 2.2.52 מקדם המתאם של משתנים מקריים X,Y (שאינם קבועים) מוגדר לפי $\rho(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathbf{V}(X)\mathbf{V}(Y)}}$.

טענה 2.2.53 לכל שני משתנים מקריים, $|\rho(X,Y)|\leq 1$.

הוכחה. נסמן $t=\frac{\mathrm{Cov}(X,Y)}{\mathbf{V}(Y)}$; אז

$$\begin{aligned}\mathbf{V}(X)\cdot(1-\rho(X,Y)^2) &= \frac{\mathbf{V}(X)\mathbf{V}(Y)-\mathrm{Cov}(X,Y)^2}{\mathbf{V}(Y)}\\ &= \mathrm{Cov}(X,X)-2t\mathrm{Cov}(X,Y)+t^2\mathrm{Cov}(Y,Y)\\ &= \mathrm{Cov}(X-tY,X-tY)\\ &= \mathbf{V}(X-tY)\geq 0.\end{aligned}$$

□

הערה 2.2.54 $\mathbf{V}(X+Y)=\mathbf{V}(X)+\mathbf{V}(Y)+2\mathrm{Cov}(X,Y)$. בפרט,

$$\mathbf{V}(X+Y)=\mathbf{V}(X)+\mathbf{V}(Y)$$

אם ורק אם X,Y בלתי מתואמים.

טענה 2.2.55 יהיו $X_1,\dots,X_n$ משתנים מקריים בלתי מתואמים, שלכולם אותה תוחלת $\mathbf{E}(X_i)=\mu$ ואותה שונות $\mathbf{V}(X_i)=\sigma^2$. נסמן $\bar{X}=\frac{X_1+\cdots+X_n}{n}$.
אז $\mathbf{E}(\bar{X})=\mu$ ו-$\mathbf{V}(\bar{X})=\frac{\sigma^2}{n}$.

דוגמא 2.2.56 נניח ש-$X_1,\dots,X_n$ משתנים מקריים בלתי מתואמים בעלי אותה שונות σ^2. נסמן, כרגיל, $\bar{X}=\frac{1}{n}(X_1+\cdots+X_n)$.

1. אם למשתנים אותה תוחלת, אז $\mathbf{E}(X_i-\bar{X})=0$.
2. $\mathrm{Cov}(\bar{X},X_i-\bar{X})=0$.
3. $\mathrm{Cov}(X_i-\bar{X},X_j-\bar{X})=-\frac{1}{n}\sigma^2$ לכל $i\neq j$.
4. $\mathbf{V}(X_i-\bar{X})=\frac{n-1}{n}\sigma^2$.

תרגיל 2.2.57 (הכללה של טענה 2.2.42) יהיו X,Y,Z משתנים מקריים על מרחב Ω.

$$\mathrm{Cov}(X,Y)=\mathrm{Cov}(\mathbf{E}(X|Z),\mathbf{E}(Y|Z))+\mathbf{E}(\mathrm{Cov}(X,Y|Z)).$$

בסימון $\mathrm{Cov}(X,Y|Z)$ הכוונה היא לשונות המשותפת של X,Y בהנתן Z, היינו $\mathrm{Cov}(X,Y|Z)=\mathbf{E}(XY|Z)-\mathbf{E}(X|Z)\mathbf{E}(Y|Z)$.

תרגיל 2.2.58 יהיו X, Y משתנים מקריים על מרחב Ω.

1. אם X, Y לא מתואמים אז לכל משתנה Z, $\mathbf{E}(X|Z), \mathbf{E}(Y|Z)$ לא מתואמים.
2. לכל X, Y, $Y - \mathbf{E}(Y|X)$ לא מתואם עם X (**רמז.** קח $Z = X$ בסעיף הקודם).
3. $\mathrm{Cov}(X, \mathbf{E}(Y|X)) = \mathrm{Cov}(X, Y)$.

תרגיל 2.2.59 נניח ש־$X = 0, 1$ עם $P(X = 1) = \alpha$; $Y = 0, 1$ עם $P(Y = 1) = \beta$, $\rho(X, Y) = \rho$.

1. חשב, במונחי α, β, ו־ρ את ההתפלגות המשותפת של X, Y;
2. חשב את $\mathbf{E}(X)$, $\mathbf{E}(XY)$, $\mathbf{V}(X)$, $\mathbf{E}(X|Y = 0)$, $\mathbf{E}(X|Y = 1)$, $\mathbf{E}(X|Y)$, ו־$\mathbf{V}(\mathbf{E}(X|Y))$;
3. בדוק ש־ $\mathbf{E}(\mathbf{E}(X|Y)) = \mathbf{E}(X)$;
4. הראה ש־$\mathbf{V}(\mathbf{E}(X|Y)) = \rho^2\alpha\alpha'$, והוכח ש־ $\mathrm{Cov}(X, Y)^2 \leq \frac{1}{4}\mathbf{V}(\mathbf{E}(X|Y))$.

נסמן $\theta(X; Y) = \frac{V(E(X|Y))}{V(X)}$. **לפי טענה 2.2.42, ערך זה נמצא תמיד בקטע** $[0, 1]$.

תרגיל 2.2.60 הראה שאם X, Y מקבלים רק שני ערכים אז $\theta(X; Y) = \rho(X, Y)^2$.

תרגיל 2.2.61 לכל α, β, $\theta(\alpha X + \beta; Y) = \theta(X; Y)$.
לכל $g : \mathbb{R} \to \mathbb{R}$ הפיכה, $\theta(X; g(Y)) = \theta(X; Y)$.

תרגיל 2.2.62 **הראה שאם** $\theta(X; Y) = 0$, **אז** $\mathrm{Cov}(X, Y) = 0$.

טענה 2.2.63 (גרסה חזקה של טענה 2.2.53) לכל שני משתנים מקריים X, Y, **מתקיים** $\rho(X, Y)^2 \leq \theta(X; Y)$.

הוכחה. קח $t = \frac{\mathrm{Cov}(X,Y)}{\mathbf{V}(Y)}$, **אז לפי תרגיל 2.2.58**,

$$
\begin{aligned}
0 &\leq \mathbf{V}(\mathbf{E}(X - tY|Y)) = \\
&= \mathbf{V}(\mathbf{E}(X|Y) - tY) = \\
&= \mathbf{V}(\mathbf{E}(X|Y)) - 2t\mathrm{Cov}(\mathbf{E}(X|Y), Y) + t^2\mathbf{V}(Y) = \\
&= \mathbf{V}(\mathbf{E}(X|Y)) - 2t\mathrm{Cov}(X, Y) + t^2\mathbf{V}(Y) = \\
&= \mathbf{V}(\mathbf{E}(X|Y)) - \frac{\mathrm{Cov}(X, Y)^2}{\mathbf{V}(Y)} = \frac{\theta(X; Y) - \rho(X, Y)^2}{\mathbf{V}(X)}.
\end{aligned}
$$

□

תרגיל 2.2.64 הראה שלכל $0 \leq r \leq \theta \leq 1$, למשתנים המקריים המוגדרים להלן יש $\theta(X;Y) = \theta$ ו־$\rho(X,Y)^2 = r$: X מקבל את הערכים $0,1$, ו־Y את הערכים $-1,0,1$;

$$P(X=0,Y=-1) = \frac{(\theta+r)^2}{4(\theta+r)^2},$$

$$P(X=0,Y=1) = \frac{(3\theta+r)(\theta-r)}{4(\theta+r)^2},$$

$$P(X=1,Y=0) = \frac{(3\theta+r)r}{r^2+3\theta r+\theta-r},$$

$P(X=1,Y=1) = P(X=1,Y=-1) = 0$, ו־$P(X=0,Y=0)$ משלים את ההסתברויות ל־1.

תרגיל 2.2.65 (*) מה הקשר בין $\theta(X;Y)$ ו־$\theta(Y;X)$?

2.3 התפלגויות בדידות

2.3.1 התפלגות אחידה

ראו דוגמא 2.2.5.

תרגיל 2.3.1 למשתנה $X \sim U[1,n]$ יש תוחלת $\mathbf{E}(X) = \frac{n+1}{2}$ ושונות $\mathbf{V}(X) = \frac{n^2-1}{12}$.

תרגיל 2.3.2 **פרדוקס המעטפות** הוא מעין־פרדוקס המדגים שיקולים הסתברותיים שגויים, ושניתן לפתרו, בהנחות מסויימות, בעזרת חישובי תוחלת זהירים.

אדם בוחר ערך X, ומושיט לנו שתי מעטפות: באחת X ובשניה $2X$. אנו בוחרים אחת מהן, בסיכויים שווים. האם כדאי לנו לפתוח אותה, או להחליפה במעטפה האחרת? נניח שהסכום שאנו רואים הוא Y; בהנתן Y, $X = Y$ או $X = \frac{1}{2}Y$. במקרה הראשון ההחלפה תעלה את הרווח ב־Y; אחרת נעבור למעטפה ובה $\frac{1}{2}Y$, והפסדנו $\frac{1}{2}Y$. הממוצע של שתי האפשרויות האלה הוא $\frac{5}{4}Y$, ולכן כדאי להחליף. אלא שנימוק זה נכון באותה מידה לכל Y: לאחר שנפתח את המעטפה, תמיד כדאי יהיה להחליף. אם כך, אין צורך לפתוח אותה. אבל אם אין פותחים את המעטפה, מה ההבדל בינה לבין האחרת? אפשר להפעיל את אותם שיקולים גם על השניה, ולהחליף בחזרה. אם כך, איזו מעטפה עדיפה?

הניתוח לעיל שגוי, משום שבהנתן Y, הסיכויים למאורע $Y = X$ (הגורר רווח אם מחליפים) אינו דווקא $1/2$: הערך תלוי בהתפלגות ממנה נבחר X. (אם הצפיפות של X היא f_X, אז $P(X=Y|Y) = \frac{f_X(Y)}{f_X(Y)+f_X(\frac{1}{2}Y)}$).

נסמן ב־Y' את תכולת המעטפה השניה. הראה שאם מניחים ש־$\mathbf{E}(X)$ סופית, אז $\mathbf{E}(Y'-Y) = 0$, ולכן תוחלת הרווח מהחלפה היא אפס.

תרגיל 2.3.3 לא קיימת התפלגות אחידה על המספרים הטבעיים. **הדרכה.** אם $P\left(X=n\right)$ אינו תלוי ב־n, אז $\sum P\left(X=n\right)$ הוא אפס או אינסוף, ולא 1.

תרגיל 2.3.4 קרא על **חוק בנפורד** והצג את הנושא לכתה בחמש הדקות הראשונות של השעור הבא.

2.3.2 התפלגות ברנולי

יהי $0<p<1$; נסמן $q=1-p$. ההתפלגות שבה $X=0,1$ עם $P(X=1)=p$ ו־$P(X=0)=q$ נקראת **התפלגות ברנולי**. מסמנים $X\sim b(p)$.

טענה 2.3.5 למשתנה $X\sim b(p)$ יש $\mathbf{E}(X)=p$, $\mathbf{V}(X)=pq$.

תרגיל 2.3.6 ה**פער** של משתנה ברנולי $X\sim b(p)$ הוא $\Delta=q-p$ (כאשר $q=1-p$). נראה שהפער כפלי. נניח ש־$X\sim b(p)$ ו־$X'\sim b(p')$ הם שני משתני ברנולי בלתי תלויים. נגדיר $Y=X\oplus X'$, כלומר הסכום מודולו 2. הראה ש־$\Delta(Y)=\Delta(X)\Delta(X')$.

יהיו X,Y משתנים מקריים בלתי תלויים. נאמר ש־X **עדיף על** Y אם מתקיים $P(X>Y)>\frac{1}{2}$. נראה שיחס העדיפות אינו טרנזיטיבי.

תרגיל 2.3.7 נסמן $0.618\approx\frac{\sqrt{5}-1}{2}=\phi$, ההפכי של יחס הזהב. נגדיר $X_1,X_2\sim b(\phi)$, שני משתני ברנולי בלתי תלויים. נגדיר $X=3X_1$, $Y=2$ ו־$Z=4-3X_2$. הראה ש־X עדיף על Y, שעדיף על Z, שעדיף על X.
(למעשה $P(X>Y)=P(Y>Z)=P(Z>X)=\phi$).

תרגיל 2.3.8 הראה שאם X,Y משתנים בלתי תלויים, אז

$$\min\left\{P(X>0),P(Y>0),P(X+Y<0)\right\}\leq\phi.$$

הדרכה. אחרת $P(X>0),P(Y>0),P(X+Y<0)>\phi$, אבל אז

$$\begin{aligned}1-\phi &> 1-P(X+Y<0) \;\geq\; P(X+Y>0)\\ &\geq P(X>0,Y>0) \;=\; P(X>0)\cdot P(Y>0) \;>\; \phi^2 \;=\; 1-\phi.\end{aligned}$$

הערה. הקשר לתרגיל 2.3.6 אינו מיידי, משום ששם $X-Y,Y-Z,Z-X$ תלויים למרות ש־X,Y,Z בלתי תלויים.

תרגיל 2.3.9 מצא התפלגות משותפת של X,Y,Z, כך ש־$X,Y,Z\sim U[1,3]$, ובכל זאת $P\left(X>Y\right)=P\left(Y>Z\right)=P\left(Z>X\right)=\frac{2}{3}$. **הדרכה.** בחר את מרחב האפשרויות כך ש־X,Y,Z תמיד שונים זה מזה.

2.3.3 התפלגות בינומית

לכסום $X = Y_1 + \cdots + Y_n$ של משתני ברנולי בלתי תלויים $Y_i \sim b(p)$ יש **התפלגות בינומית**. משתנה כזה סופר את ההצלחות בסדרה של n נסיונות ברנולי בלתי תלויים. מסמנים $X \sim \mathrm{Bin}(n,p)$. מן השוויון $X = \sum Y_i \sim \mathrm{Bin}(n,p)$ קל לחשב את התוחלת והשונות,

$$\mathbf{E}(X) = np, \qquad V(X) = npq.$$

טענה 2.3.10 אם משתנים $X_1 \sim \mathrm{Bin}(n_1,p)$ ו־$X_2 \sim \mathrm{Bin}(n_2,p)$ הם בלתי תלויים אז $X_1 + X_2 \sim \mathrm{Bin}(n_1+n_2,p)$.

הוכחה. באמצעות המבנה של משתנה בינומי כסכום של משתני ברנולי, או בחישוב ישיר. □

תרגיל 2.3.11 אם $X \sim \mathrm{Bin}(n,p)$ אז $n - X \sim \mathrm{Bin}(n,q)$.

תרגיל 2.3.12 בכד n כדורים. בוחרים $X \sim \mathrm{Bin}(n,p)$, וצובעים X מן הכדורים.

1. מוציאים כדור מהכד. מה הסיכוי לכך שהוא צבוע?
2. הוצאנו כדור צבוע, וכעת מוציאים עוד כדור, ללא החזרת הכדור הראשון. מהם הסיכויים לכך שהכדור השני צבוע?
3. ומהם הסיכויים לכך שהכדור השני צבוע, אם הוצאנו כדור צבוע ראשון והחזרנו אותו לכד?

רעיון 2.3.13 מדי שנה נבחנים סטודנטים למשפטים בבחינות הלשכה של עורכי הדין: מאה תלמידים מכל אחת מחמש אוניברסיטאות. הסיכוי של בוגר התואר לעבור את הבחינות הוא 90%. אם האוניברסיטה המובילה בדירוג השנה אינה זו שהובילה את הדירוג בשנה שעברה, על המוסף הכלכלי של העיתון לבשר מיד על "מהפך: אוניברסיטה A עקפה את אוניברסיטה B בדירוג בחינות הלשכה". הערך את הסיכויים לכותרת כזו.

תרגיל 2.3.14 שני שחקנים מחלקים ביניהם n מטבעות: מטילים את כולם, השחקן הראשון מקבל את אלו שנפלו על 'עץ' והאחר את אלו שנפלו על 'פלי'. לאחר מכן מטיל כל שחקן שוב את המטבעות שלו, ושומר רק את אלו שנפלו שנית על הצד שלו; את האחרים הוא מחזיר לקופה.

1. איך מתפלג הרווח של השחקן הראשון?

2. חשב את השונות המשותפת ומקדם המתאם בין הרווח של השחקן הראשון לרווח של השחקן השני.

פתרון. $D\sim \mathrm{Bin}(n,\frac{1}{2})$ הוא מספר המטבעות בידי השחקן הראשון; ולאחר שנקבע D מגרילים $X_0\sim \mathrm{Bin}(D,\frac{1}{2})$ ו־$Y_0\sim \mathrm{Bin}(n-D,\frac{1}{2})$ באופן בלתי תלוי. 1. הסיכוי שמטבע יגיע לידי השחקן הראשון הוא רבע, ובגלל האי־תלות הרווח מתפלג $\mathrm{Bin}(n,\frac{1}{4})$. 2. לפי תרגיל 2.2.57,

$$\begin{aligned}
\mathrm{Cov}(X_0,Y_0) &= \mathrm{Cov}(\mathbf{E}(X_0|D),\mathbf{E}(Y_0|D))+\mathbf{E}(\mathrm{Cov}(X_0,Y_0)|D)\\
&= \mathrm{Cov}(\frac{1}{2}D,\frac{1}{2}(n-D))+0\\
&= -\frac{1}{4}\mathbf{V}(D)=-\frac{1}{16}n
\end{aligned}$$

, ולכן $\rho(X_0,Y_0)=-\frac{1}{3}$.

תרגיל 2.3.15 שיכור צועד מנקודת האפס לאורך הציר הממשי: הוא עושה צעד ימינה או צעד שמאלה, בסיכויים שווים. (המיקום שלו בזמן n הוא $2X_0-n$ כאשר $X_0\sim \mathrm{Bin}(n,\frac{1}{2})$.) הוכח ש־$\mathbf{E}(X)=0$ ו־$\mathbf{V}(X)=n$.

תרגיל 2.3.16 (הילוך שיכור דו־ממדי) שיכור שיוצא מראשית הצירים מבצע בכל פעם צעד באחד מארבעת הכיוונים המקבילים לצירים, בהסתברויות שוות. נסמן את מקומו בצעד לאחר n צעדים ב־(X,Y).

1. הסבר את המודל:

$$D\sim \mathrm{Bin}(n,\frac{1}{2}),\quad X_0|D\sim \mathrm{Bin}(D,\frac{1}{2}),\quad Y_0|D\sim \mathrm{Bin}(n-D,\frac{1}{2}),$$

$X=2X_0-D$, $Y=2Y_0-(n-D)$.

2. הראה ש־$\mathbf{E}(X)=\mathbf{E}(Y)=0$, $\mathbf{V}(X)=\mathbf{V}(Y)=\frac{n}{2}$.
3. הראה שתוחלת ריבוע המרחק של השיכור מהמרכז היא n.
4. הראה ש־$\mathrm{Cov}(X,Y)=0$. האם X,Y תלויים?

תרגיל 2.3.17 למנעול האופניים שלי יש ארבע חוגות של 10 ספרות. בכל פעם שאני נועל אותו, אני מתעצל ומסובב חוגה אחת, שתיים, שלוש או ארבע (בהסתברויות שוות), כל אחת מהן ימינה או שמאלה (בהסתברויות שוות), צעד אחד, שניים או שלושה (בהסתברויות שוות). מישהו עוקב ומציץ במספר שלי אחרי הנעילה. כמה הצצות הוא צריך כדי לפענח בסיכוי טוב את הקוד הסודי?
(הדרך הטובה ביותר להתגונן מפני התקפה זו היא לבחור מספר קבוע שאליו עוברים כשהמנעול סגור.)

תרגיל 2.3.18 עורך דינה של קימברלי מקרת'י, שהורשעה ברצח ונידונה למוות בטקסס ב־2001, טוען שההרשעה התקבלה על בסיס גזעני. לדבריו, במדינה שבה רבע מהאוכלוסיה שחורה, לא יתכן שרק אחד מ־13 המושבעים יהיה שחור (מוסף "צדק" של מקור ראשון, גליון 809, 8.2.2013.)

1. מה הסיכוי להרכב כזה של חבר המושבעים, בהנחה שזה הורכב באקראי?
2. נניח שמושבע שחור היה מרשיע אותה בסיכוי 0.6, בעוד שמושבע לבן היה מרשיע אותה בסיכוי 0.9. מהם סיכויי ההרשעה ההוגנים, אם בוחרים את חבר המושבעים באקראי? **הדרכה.** נסמן ב־X את מספר השחורים בחבר המושבעים, אז $X \sim \operatorname{Bin}(13, 0.25)$, וסיכויי ההרשעה בהנתן X הם $0.6^X 0.9^{13-X}$.

תרגיל 2.3.19 ה**פער** של משתנה ברנולי הוגדר בתרגיל 2.3.6. מטילים מטבע $2m+1$ פעמים, ומסמנים ב־Y את הכרעת הרוב. מהו הפער של Y, כתלות בפער ϵ של המטבע, בהנחה ש־ϵ זערורי ביחס ל־m? (בקירוב מסדר ראשון, הפער הוא $2\sqrt{\frac{m}{\pi}}\epsilon \sim \frac{m+1}{2^{2m}}\binom{2m+1}{m}\epsilon$.)

תרגיל 2.3.20 חטיבה מורכבת משלושה גדודים; גדוד מורכב משלוש פלוגות; פלוגה מורכבת משלוש מחלקות; מחלקה מורכבת משלושה צוותים; וצוות מורכב משלושה חיילים. לכל חייל יש מקצוע (צנחן או שריונר). המקצוע של יחידה מוגדר כמקצוען של רוב היחידות המרכיבות אותה. נניח שכל חייל בוחר באקראי ובאופן בלתי תלוי את מקצועו, כאשר הסיכוי להיות צנחן הוא $\frac{1+\epsilon}{2}$, ו־ϵ קטן מאד. מהו הסיכוי שהחטיבה היא חטיבת צנחנים?

תרגיל 2.3.21 במדינה מוכרת נערכות בחירות אחת לארבע שנים (או פחות). החליטו שבמקום לספור את כל הקולות, יבחרו בכל קלפי k פתקים (למשל $k=10$), באקראי ובהתפלגות אחידה, עם החזרה; הפתקים האלה, משוקללים לפי מספר המצביעים בקלפי, יקבעו את מספר המצביעים בבחירות לכל מפלגה.

נסמן ב־n_i את מספר המצביעים בקלפי i, וב־a_i את מספר המצביעים למפלגת "כח לסוקרים" באותה קלפי. מספר הקולות שנזקפים לטובת המפלגה הוא הסכום $X = \frac{1}{k}\sum n_i X_i$, כאשר $X_i \sim \operatorname{Bin}(n_i, \frac{a_i}{n_i})$. הראה שהתוחלת של X שווה למספר הקולות שתמכו במפלגה, וחשב את סטיית התקן. מה צריך להיות k כדי שהתוצאה תהיה, בוודאות קרובה, אמינה?

תרגיל 2.3.22 אדם מטיל קוביה הוגנת סטנדרטית, שבה סכום הנקודות על כל שתי פאות מקבילות הוא 7. לפני כל הטלה, עליו לבחור (ולא לספר לאף אחד) האם הוא מאמץ הפעם את הפאה העליונה או התחתונה של הקוביה. אחר־כך הוא מטיל את הקוביה, מדווח באיזו פאה בחר, ומקבל את הסכום המוצג שם בשקלים (לדוגמא, אם הקוביה מציגה 5 והוא בחר בפאה התחתונה, יוכל לומר את האמת

ולקבל שני שקלים, או לשקר ולקבל חמישה). ההטלה היתה מוצלחת מבחינת המשתתף, אם מספר הנקודות על הפאה שהוא אומר שבחר גבוה מזה של הפאה הנגדית. חוזרים על הניסוי עשרים פעם. כמה הצלחות צפויות בממוצע? כמה הצלחות יגרמו לכם לחשוד שהאדם משקר? (בסדרת ניסויים שערך דן אריאלי, נמצא שמספר ההצלחות הממוצע הוא כ־14).

2.3.4 התפלגות פואסון

הגדרה 2.3.23 למשתנה המקרי המקבל ערכים $0,1,2,\ldots$ יש **התפלגות פואסון** עם פרמטר λ אם $P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$; מסמנים $X\sim\mathrm{Poi}(\lambda)$.

תרגיל 2.3.24 הוכח שזוהי אכן התפלגות, כלומר, ש־$\sum_{n=0}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}=1$. **הדרכה**. טור טיילור.

למה 2.3.25 לכל k, $\frac{n!}{(n-k)!n^k}\to 1$.

נניח שמגדירים משתנים מקריים $X_n\sim\mathrm{Bin}(n,\frac{\lambda}{n})$. (פירוש אפשרי: מספר האגוזים בטבלת שוקולד, כשבוחנים אותה ברזולוציה הולכת וגדלה). אז לפי למה 2.3.25,

$$\lim_{n\to\infty}P(X_n=k)=P(Z=k)$$

כאשר $Z\sim\mathrm{Poi}(\lambda)$.

תרגיל 2.3.26 מספר האנשים הנכנסים לבנק במשך שעה מתפלג פואסונית עם תוחלת λ. כל אחד מהם משאיר את כרטיס הביקור שלו. כמה זוגות אפשר להרכיב מן הכרטיסים? כמה k־יות סדורות?
(הראה שאם $X\sim\mathrm{Poi}(\lambda)$ אז $\mathbf{E}(\frac{X!}{(X-k)!})=\lambda^k$).

טענה 2.3.27 נניח ש־$X\sim\mathrm{Poi}(\lambda)$ ו־$Y\sim\mathrm{Poi}(\lambda')$ הם משתנים בלתי תלויים.

1. $X+Y\sim\mathrm{Poi}(\lambda+\lambda')$.
2. ההתפלגות של $X|X+Y=n$ היא בינומית $\mathrm{Bin}(n,\frac{\lambda}{\lambda+\lambda'})$.

טענה 2.3.28 פיצול של תהליך פואסון: נניח ש־$N\sim\mathrm{Poi}(\lambda)$, ובהנתן N, $X\sim\mathrm{Bin}(N,p)$. נסמן $Y=N-X$. אז X,Y בלתי תלויים ובעלי התפלגות פואסון, $X\sim\mathrm{Poi}(\lambda p)$, $Y\sim\mathrm{Poi}(\lambda q)$.

הוכחה.

$$\begin{aligned} P(X=k) &= \sum_{n=k}^{\infty} P(X=k \mid N=n)P(N=n) \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k q^{n-k} \frac{e^{-\lambda}\lambda^n}{n!} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{m=0}^{\infty} \frac{1}{m!}(\lambda q)^m = \frac{(\lambda p)^k}{k!} e^{-\lambda p}; \end{aligned}$$

כלומר, $X \sim \mathrm{Poi}(\lambda p)$**. מאותה סיבה גם** $Y \sim \mathrm{Poi}(\lambda q)$**, ואי־התלות נובעת מכך ש־**

$$\begin{aligned} P(X=k, Y=m) &= P(X=k, N=k+m) \\ &= P(X=k \mid N=k+m)P(N=k+m) \\ &= \binom{k+m}{k} p^k q^m \frac{e^{-\lambda}\lambda^{k+m}}{(k+m)!} \\ &= \frac{e^{-\lambda p}(\lambda p)^k}{k!} \frac{e^{-\lambda q}(\lambda q)^m}{m!} = P(X=k)P(Y=m). \end{aligned}$$

□

תרגיל 2.3.29 במצב המתואר בטענה 2.3.28, חשב את $\mathrm{Cov}(X,N)$. תאר את ההתפלגות של N בהנתן X.

תרגיל 2.3.30
1. נניח ש־$\lambda < \lambda'$. מצא דרך להגדיר שני משתנים מקריים X, X' באופן ש־$X \sim \mathrm{Poi}(\lambda)$, $X' \sim \mathrm{Poi}(\lambda')$, ותמיד $X \le X'$. **הדרכה.** העזר בטענה 2.3.27. טכניקה זו נקראת **צימוד** (coupling).
2. הסק שאם $X \sim \mathrm{Poi}(\lambda)$ ו־$X' \sim \mathrm{Poi}(\lambda')$ כאשר $\lambda < \lambda'$, אז לכל n מתקיים $P(X \ge n) < P(X' \ge n)$.
3. הסק שלכל n הפונקציה $\sum_{k=n}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!}$ עולה עם λ.

תרגיל 2.3.31 שכנעו חבר שאת הטענה "כל כרטיס רביעי זוכה" אפשר לפרש בשתי משמעויות שונות: "רבע מהכרטיסים זוכים" ו־"ארבעה כרטיסים (רצופים?) מבטיחים זכיה". מה הסיכוי לזכיה בהגרלה שבה רבע מהכרטיסים זוכים, אם קונים ארבעה כרטיסים? ומה קורה כשמחליפים בכל מקום את המספר ארבע בארבעים? בארבע־מאות?

תרגיל 2.3.32 מטילים חצים לעבר לוח שיש בו 1000000 משבצות. באחת מהן טמון מגנט שבגללו הסיכוי לפגוע בה הוא 0.01; לשאר המשבצות סיכויים שווים. זורקים 1000 חצים. חסום מלמטה את הסיכוי לכך שהמשבצת הממוגנטת היא זו שתקלוט את מספר החצים הגדול ביותר.

תרגיל 2.3.33 נניח שבכל סיבוב הגרלה נבחר מספר זוכה בין 1 ל־600, בסיכויים שווים.

1. בכמה אפשר לשפר את הסיכויים לבחור את המספר שיזכה בסיבוב הבא, אם נחכה לסיבוב ה־1800 ואז נבחר מספרים שטרם עלו בגורל?
2. האם יתכן שישארו מספרים שלא יעלו בגורל אחרי 2400 סיבובים? (כן; בסיכוי 99.98%).

תרגיל 2.3.34 בהמשך לתרגילים 2.1.11 ו־2.2.30, מהי השונות של מספר נקודות השבת (העזר במשתנים מציינים). מהי, בקירוב, התפלגות מספר נקודות השבת של תמורה מקרית?

תרגיל 2.3.35 מתווך דירות שרצה להראות עד כמה קשה מצב השוק, הדגים (ב־1.8.2011): "בחודש האחרון עשיתי רק עסקה אחת של שכירות ובחודש רגיל אנחנו עושים שתיים־שלוש עסקות".

לאיזו משפחה שייכת ההתפלגות של מספר העסקאות בחודש? מהו, להערכתך, הפרמטר? עד כמה מפתיעה העובדה שדווקא בחודש מסויים התבצעה רק עסקה אחת?

תרגיל 2.3.36 שעור החולים בסרטן דם נדיר הוא אחד ל־5000 (אינדקס סיכון 12). איך מתפלג מספר החולים בבניין שגרים בו 100 דיירים? מה הסיכוי שבבניין כזה יחלו שלושה אנשים?

ברחוב החרובים 17 במעלה־נחושתיים התגלו שלושה מקרים של המחלה האמורה. האם זו ראיה לכך שגורמים סביבתיים מיוחדים מזיקים לדיירי הבנין? (הנח שבישראל כ־$70,000$ בניינים שבכל אחד מהם גרים כ־100 דיירים.)

תרגיל 2.3.37 על מרחב גדול W מגדירים **פונקציה מקרית** $f\colon W \to W$; כלומר, כל ערך $f(w)$ הוא משתנה מקרי בעל התפלגות אחידה המקבל ערכים ב־W, ומשתנים אלו בלתי תלויים. איך מתפלג, בקירוב, מספר המקורות $|f^{-1}(w)|$? ואיך מתפלג $|f^{-1}f(w)|$?

תרגיל 2.3.38 בסקר שנערך בין מנתחים, הודו 20% מהם שניתחו במהלך הקריירה שלהם את הצד הלא נכון לפחות פעם אחת. הסיכוי לשגיאה כזו הוא קבוע בכל ניתוח. כמה פעמים בממוצע עשה כל אחד מהטועים את השגיאה המדוברת, אם:

1. נניח שכל מנתח עשה עד כה אותו מספר של ניתוחים.

2. נניח, בניתוח מדוקדק יותר, שכל מנתח עורך אותו מספר ניתוחים במהלך הקריירה שלו, אבל הנשאלים נסקרו בשלב אקראי (בעל התפלגות אחידה) בקריירה.

תרגיל 2.3.39 לפי משפט המספרים הראשוניים, ה"סיכוי" של מספר טבעי x להיות ראשוני הוא $\frac{1}{\log x}$ (ביתר דיוק: אם $\pi(x)$ מסמן את מספר הראשוניים עד x, אז $\pi(x) \sim \frac{x}{\log x}$). תן הצדקה היוריסטית להשערת גולדבך, שלפיה כל מספר זוגי אפשר להציג כסכום של שני ראשוניים.

הדרכה. הראה שמספר הדרכים להציג מספר טבעי x כסכום של שני ראשוניים עולה (בקירוב היוריסטי) על $\frac{x}{2\log^2 x}$. בהנחה שמספר ההצגות מתפלג פואסונית, הסיכוי לכך שאין הצגות הוא בקירוב $\exp(-\frac{x}{2\log^2 x})$. האינטגרל על הערך הזה, מ־10000 עד אינסוף, קטן מ־10^{-23}.

2.3.5 התפלגות גאומטרית

אומרים ש־$X \sim \mathrm{G}(p)$ אם X הוא מספר הנסיונות עד להצלחה ראשונה בסדרת ניסויי ברנולי שהסתברות ההצלחה שלהם p. אז $P(X=k) = q^{k-1}p$, $k \geq 1$, כאשר $q = 1-p$.

כדי לבדוק שזו אכן התפלגות, יש לחשב:

$$\sum_{k=1}^{\infty} q^{k-1}p = p(1+q+q^2+\cdots) = \frac{p}{1-q} = 1.$$

טענה 2.3.40 נניח ש־$X \sim \mathrm{G}(p)$. אז $\mathbf{E}(X) = \frac{1}{p}$ ו־$\mathbf{V}(X) = \frac{q}{p^2}$.

הוכחה. **נסמן $h(x) = \frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$, ונגזור: $h'(x) = \frac{1}{(1-x)^2} = \sum_{n=0}^{\infty} nx^{n-1}$. לכן**

$$\mathbf{E}(X) = \sum_{k=1}^{\infty} kq^{k-1}p = p\sum_{k=0}^{\infty} kq^{k-1} = \frac{p}{(1-q)^2} = \frac{1}{p};$$

נגזור שוב ונקבל $h''(x) = \frac{2}{(1-x)^3} = \sum_{n=0}^{\infty} n(n-1)x^{n-2}$. לכן

$$\mathbf{E}(X(X-1)) = \sum_{k=1}^{\infty} k(k-1)q^{k-1}p = pq\sum_{k=1}^{\infty} k(k-1)q^{k-2} = \frac{2pq}{(1-q)^3} = \frac{2q}{p^2}.$$

מכאן ש־

$$\begin{aligned}\mathbf{V}(X) &= \mathbf{E}(X^2)-\mathbf{E}(X)^2 = \mathbf{E}(X(X-1))+\mathbf{E}(X)-\mathbf{E}(X)^2\\ &= \frac{2q}{p^2}+\frac{1}{p}-\frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}.\end{aligned}$$

□

תרגיל 2.3.41 נניח ש־ $|x| < 1$. הוכח ש־ $\sum_{n=0}^{\infty}\binom{n+k}{k}x^n = (1-x)^{-(k+1)}$.

טענה 2.3.42 נניח ש־$X \sim \mathrm{G}(p)$. אז לכל n, ההתפלגות המותנית של $X-n$ בהנתן $X > n$ היא $\mathrm{G}(p)$. תכונה זו נקראת **חוסר הזכרון** של ההתפלגות הגאומטרית.

ציטוט 2.3.43 "כדי להבטיח , למשל, זכייה בפרס הראשון בלוטו, בהנחה שממלאים אותו בצירופים שונים פעמיים בחודש חמש טבלאות בכל פעם, יש למלא טופס במשך *112* אלף שנים!" (ד"ר ברי כינורי, *TheMarker Cafe, 25.7.2012*).

תרגיל 2.3.44 מה תוחלת מספר הנסיונות עד לפעם הראשונה שהופיעו בסדרה גם הצלחה וגם כשלון? **הדרכה.** התשובה היא $\frac{1}{pq}-1$.

תרגיל 2.3.45 מתבוננים בסדרת ניסויי ברנולי בעלי הסתברות קבועה, p. מן הניסוי השני ואילך, מכריזים בכל פעם האם התוצאה "שווה" או "שונה" לתוצאה הקודמת. מה הסיכוי לתוצאה "שונה" ברגע מסויים? מה הסיכוי לשתי הכרזות "שונה" רצופות? האם ההכרזה בזמן $n+1$ תלויה בהכרזה בזמן n? ומה לגבי ההכרזות בזמן $n+2$ ו־n?

תרגיל 2.3.46 תוחלת מספר ההגרלות עד לזכיה ראשונה בהגרלה היא n. הערך את הסיכוי לזכות לפחות פעם אחת ב־αn הגרלות (כאשר $\alpha \ll 1$).

תרגיל 2.3.47
1. מטילים מטבע הוגן שוב ושוב. מה אורכו הצפוי של הרצף (היינו, סדרת תוצאות זהות) שיתחיל בהטלה הבאה?
2. הערך: כמה רצפים יהיו בסדרה באורך 1000 הטלות?
3. מה יהיה אורכו של הרצף הארוך ביותר בסדרה של 1000 הטלות? בצע סימולציה ממוחשבת כדי לקבל את התפלגות הרצף הארוך ביותר, והשווה אותה לתוצאות המוערכות.

תרגיל 2.3.48 נניח ש־X,Y מתפלגים גאומטרית, $X \sim \mathrm{G}(p_1)$, $Y \sim \mathrm{G}(p_2)$. הראה שההתפלגות של X בהנתן $X=Y$ היא $\mathrm{G}(p_1+p_2-p_1p_2)$.

באופן כללי יותר, אם $X_i \sim \mathrm{G}(p_i)$ הם משתנים בלתי תלויים עבור $i=1,\ldots,n$, אז ההתפלגות של X_1 בהנתן $X_1=\cdots=X_n$ היא $\mathrm{G}(1-(1-p_1)\cdots(1-p_n))$.

תרגיל 2.3.49 לניסוי שלוש תוצאות אפשריות: הצלחה מלאה, הצלחה חלקית, וכשלון. הסיכוי להצלחה מלאה הוא p, והסיכוי להצלחה חלקית הוא δ. חוזרים על הניסוי עד להצלחה המלאה הראשונה. נסמן ב־X את מספר הנסיונות שנערכו, וב־Y את מספר הנסיונות (מתוכם) שלא נכשלו.

1. מה ההתפלגות של X? **הדרכה.** כרגיל, $X \sim G(p)$.
2. כתוב את ההתפלגות המשותפת של X, Y. **הדרכה.** לכל $1 \leq k \leq n$, $P\left(Y=k,\, X=n\right) = \binom{n-1}{k-1}(1-\delta-p)^{n-k}\delta^{k-1}p$.
3. מה ההתפלגות של X, בהנתן שאף ניסוי לא נכשל? **הדרכה.** המאורע "אף ניסוי לא נכשל" הוא המאורע $X=Y$. כפי שראינו $P\left(X=Y=n\right) = \delta^{n-1}p$, ולכן $P\left(X=Y\right) = \frac{p}{1-\delta}$. מכאן ש־$P\left(X=n|X=Y\right) = \frac{P(X=Y=n)}{P(X=Y)} = \frac{\delta^{n-1}p}{p/(1-\delta)} = \delta^{n-1}(1-\delta)$, כלומר, $X|X=Y \sim \mathrm{G}(1-\delta)$. בפרט $\mathbf{E}(X|X=Y) = \frac{1}{1-\delta}$.
4. מה ההתפלגות של Y? **הדרכה.** $Y \sim \mathrm{G}(\frac{p}{p+\delta})$.
5. מצא את ההתפלגות של Y בהנתן X. **הדרכה.** $(Y-1|X) \sim \mathrm{Bin}(X-1,\delta)$.
6. מצא את ההתפלגות של X בהנתן Y (ראה תת־סעיף 2.3.6). **הדרכה.** $(X-Y|Y) \sim \mathrm{NB}(Y, 1-\delta-p)$.
7. הסבר כיצד יתכן שההתפלגות של Y שונה מן ההתפלגות של X בהנתן $X=Y$.

תרגיל 2.3.50 שני שחקנים מתחרים בקליעה למטרה. לראשון סיכויי הצלחה α, ולשני סיכויי הצלחה β (נסיונות הקליעה בלתי תלויים במשותף). במשחק ראשון קובעים שהמנצח הוא מי שנזקק לפחות נסיונות קליעה עד לפגיעה הראשונה (אם שניהם נזקקו לאותו מספר נסיונות, מוכרז תיקו). במשחק השני שניהם מנסים לקלוע: אם אחד קלע וחברו החטיא, הוא מוכרז כמנצח, ואם שניהם קלעו מוכרז תיקו. אם שניהם החמיצו, מנסים שוב.

הראה שהסתברויות הנצחון שוות בין שני המשחקים.

טענה 2.3.51 *נניח ש־$X \sim \mathrm{G}(\frac{1}{2})$. מהי התוחלת של 2^X? ושל $(-1)^X$?*

בתאוריה כלכלית ותורת קבלת ההחלטות, מקובל לומר שאדם בוחר בין אלטרנטיבות לפי זו שמביאה לו תוחלת גדולה יותר.

תרגיל 2.3.52 מפעל הפיס מציע משחק חדש: תמורת כרטיס בשווי 100 שקל, תזכה ב־2^X שקלים, כאשר $X \sim \mathrm{G}(\frac{1}{2})$. האם כדאי להשתתף במשחק? [שאלה זו קשורה לפרדוקס של סנקט פטרבורג.]

תרגיל 2.3.53 בהימורי ה"בראקט" של כדורסל המכללות בארה"ב, מנחשים את זהות המנצח בכל אחד מבין 63 משחקים, וגובה הפרס תלוי במספר ההצלחות הרצופות (מן המשחק הראשון ואילך.) בזכות ידע מוקדם, הסיכוי לנחש כל משחק אינו חצי אלא 60%. עד היום נמכרו מליארד טפסי הימורים. העריך מה מספר ההצלחות הרצוף הגבוה ביותר שהושג בהם. **הדרכה.** התוצאה הטובה ביותר היא בעלת סיכוי של כאחד למליארד, שהם $0.6^{-40.7}$; נצפה, אם כך, ל־40-41 הצלחות.

תרגיל 2.3.54 בעיית אוסף הקופונים: חברה גדולה מפיצה מספר שווה של קופונים מכל אחד מ־n סוגים. כמה קופונים צריך אדם לאסוף, בתוחלת, על־מנת להרכיב אוסף שיש בו לפחות קופון אחד מכל סוג? נסו להעריך כמה עליו לאסוף כדי להחזיק לפחות שני קופונים מכל סוג.

תרגיל 2.3.55 ב"משחק החיים" המשימה היא להתקל באקראי במכונית ששלוש הספרות האמצעיות של מספר הרישוי שלה הן 000, אחר־כך במכונית שהספרות שלה הן 001, וכן הלאה על 999, לפי סדר. בכמה מספרי מכוניות יש להבחין כדי להשלים את המשחק, בתוחלת? מה שונות משך המשחק?

תרגיל 2.3.56 במשחק הקוביות Craps מטילים בכל פעם זוג קוביות, ומחשבים את סכומן. מן הערכים הקיצוניים $2, 3, 11, 12$ מתעלמים לחלוטין. המשחק מורכב מסדרת משחקונים, שבכל אחד מהם הכללים זהים: בפתיחת המשחקון מטילים את הקוביה עד שמתקבל הערך הראשון מהקבוצה $4, 5, 6, 8, 9, 10$. בשלב השני של המשחקון, מטילים את הקוביות עד שמתקבל אחד משני סכומים: אם מתקבל הערך שנקבע בפתיחה, "זוכים" בו, ומתחילים את המשחקון הבא. אם מתקבל 7, המשחק מסתיים.

1. נניח שבפתיחת המשחקון נקבע הערך 4. מה הסיכוי לכך שמשחקון זה יסתיים בזכיה?
2. מה הסיכוי לכך שמשחקון נתון יסתיים בזכיה?
3. משחקון מסויים הסתיים בזכיה. מהם הסיכויים לכך שהערך הזכוי הוא 4?
4. מה התפלגות מספר המשחקונים במשחק? כמה משחקונים מתקיימים במשחק, בתוחלת?
5. מה הסיכוי לכך שהשחקן יזכה בערך 4 במהלך משחק?
6. כתוב סימולציה ממוחשבת שתעריך את הסיכוי לכך שהשחקן יזכה במהלך משחק בכל ששת הערכים האפשריים. (התשובה היא כ־1 ל־6100.)

הטענה הבאה שימושית בחישוב תוחלת של זמני המתנה:

טענה 2.3.57 יהי X משתנה מקרי המקבל ערכים שלמים חיוביים. הוכח ש־$\mathbf{E}(X) = \sum_{n=0}^{\infty} P(X \geq n)$.

הוכחה.

$$\begin{aligned}
\sum_{n=0}^{\infty} P(X \geq n) &= \sum_{n=0}^{\infty}\sum_{k=n}^{\infty} P(X = k) \\
&= \sum_{k=0}^{\infty}\sum_{n=0}^{k} P(X = k) \\
&= \sum_{k=0}^{\infty} kP(X = k) \;=\; \mathbf{E}(X).
\end{aligned}$$

□

תרגיל 2.3.58
1. עבור מספרים חיוביים $x_1, x_2, \ldots$, הראה שהמכפלה האינסופית $\prod_{n=1}^{\infty}(1+x_i)$ מתכנסת אם ורק אם $\sum x_i$ מתכנס. **הדרכה**. $x \leq (1+x) \leq e^x$.
2. נניח ש־$y_1, y_2, \ldots$ חיוביים. הראה ש־$\prod_{i=1}^{\infty}(1-y_i) > 0$ אם ורק אם $\sum \frac{y_i}{1-y_i}$ מתכנס, אם ורק אם $\sum y_i$ מתכנס.

תרגיל 2.3.59 מבצעים סדרת ניסויים בלתי תלויים, כך שההסתברות להצלחה בניסוי ה־n, אם הוא מתרחש, היא p_n $(q_n = 1 - p_n)$. נסמן ב־T את מספר הניסויים עד להצלחה הראשונה (ו־$T = \infty$ אם ההצלחה לעולם אינה מגיעה).

1. הראה ש־$P\left(T \geq n\right) = q_1 \cdots q_n$, וש־$P\left(T = n\right) = q_1 \cdots q_{n-1}p_n$.
2. אם $p_n = p$ הוא קבוע, אז $T \sim \mathrm{G}(p)$.
3. הסיכוי לכשלון בכל סדרת הניסויים הוא $\prod_{n=1}^{\infty} q_n$. לכן הסיכוי להצלחה בשלב (סופי) כלשהו הוא 1 אם ורק אם $\sum p_n = \infty$. **הדרכה**. תרגיל 2.3.58.
4. בכד יש a כדורים לבנים וכדור אחד שחור. בכל שלב מוציאים כדור באקראי; אם הכדור לבן, זוהי הצלחה, והניסוי מסתיים. אחרת, מוסיפים לכד כדור שחור אחד, וממשיכים בסדרה.

   (א) הראה שלכל $a \geq 1$, ההצלחה (בשלב סופי כלשהו) מובטחת.

   (ב) אם $a = 1$ אז $\mathbf{E}(T) = \infty$; וכאשר $a > 1$, $\mathbf{E}(T) = \frac{1}{a-1}$. **הדרכה**. העזר בטענה 2.3.57.

2.3.6 התפלגות בינומית שלילית

יהי $r \geq 1$. נסמן ב־X את מספר ההצלחות בסדרת ניסויי ברנולי בלתי תלויים בעלי הסתברות הצלחה p, עד לכשלון ה־r־י. על־מנת שהכשלון ה־r־י יחול לאחר k הצלחות, עליו להתרחש בנסיון ה־$r+k$, כשמתוך $r+k-1$ הנסיונות הקודמים יש בדיוק $r-1$ כשלונות. לפיכך

$$P\left(X=k\right)=\binom{k+r-1}{k}q^rp^k,$$

לכל $k \geq 0$. את ההתפלגות מסמנים ב־$X \sim \mathrm{NB}(r,p)$. היא נקראת **בינומית שלילית** משום ש־

$$P\left(X=k\right)=\binom{-r}{k}(-p)^kq^r$$

כאשר לכל $x \in \mathbb{R}$ מסמנים $\binom{x}{k}=\frac{x(x-1)\cdots(x-k+1)}{k!}$.

דוגמא 2.3.60 אם $r=1$, אז $P(X=k)=qp^k$, כלומר $\mathrm{NB}(1,p)=\mathrm{G}(q)-1$.

תרגיל 2.3.61 יהי r מספר טבעי. הראה ש־$\sum_{n=0}^{\infty}\binom{n+r}{r}t^n=(1-t)^{-r-1}$. **הדרכה.** עיין בהתפלגות הבינומית השלילית; או הוכח באינדוקציה על r על־ידי גזירת הטור.

אפשר לכתוב את X כסכום של משתנים גאומטריים: נסמן ב־Y_i את מספר ההצלחות בין הכשלון ה־$(i-1)$ לכשלון ה־i. אז $Y_i+1 \sim \mathrm{G}(q)$, ו־$X=Y_1+\cdots+Y_r$. כך אפשר לחשב תוחלת ושונות:

$$\mathbf{E}(X)=\frac{rp}{q}, \qquad \mathbf{V}(X)=\frac{rp}{q^2}.$$

תרגיל 2.3.62 נניח ש־$X_1,X_2,\dots \sim \mathrm{G}(p)$ הם משתנים בלתי תלויים. נסמן $S_r=X_1+\cdots+X_r$. הראה שלכל $n \geq r$, $P(S_r=n)=\binom{n-1}{n-r}p^rq^{n-r}$.

תרגיל 2.3.63 נסמן ב־T את מספר הרצפים בסדרה של n משתני ברנולי $b(\frac{1}{2})$ (למשל, מספר הרצפים בסדרה 0011110 הוא 3; סופרים גם את הרצף שלא הסתיים). חשב את ההתפלגות של T. **פתרון.** אורכו של הרצף ה־i הוא $X_i \sim \mathrm{G}(\frac{1}{2})$. נסמן $S_t=X_1+\cdots+X_t$, אז $T<t$ אם ורק אם $S_{t-1} \geq n$. לכן $P(T<t)=P(S_{t-1}\geq n)=\sum_{i=n}^{\infty}\binom{i-1}{i-t+1}2^{-i}$. [ראו טענה 2.6.31 לבעיה דומה שבה החישובים מסתדרים טוב יותר.]

תרגיל 2.3.64 מספר העכברים באסם, ועוד אחד, מתפלג גאומטרית עם תוחלת $\frac{1}{p}$ (כך שהסיכוי שאין עכברים הוא p). הסיכוי של כל עכבר להתפס במלכודת הוא α (נסמן $\beta=1-\alpha$).

כלומר, אם S הוא מספר העכברים באסם ו־X מספר העכברים שנתפסו, אז $S+1 \sim \mathrm{G}(p)$ ו־$X|S \sim \mathrm{Bin}(S,\alpha)$.

1. מצא את ההתפלגות של מספר העכברים שנתפסו, ואת התוחלת. [תשובה. ההתפלגות גאומטרית: $X+1 \sim \mathrm{G}(\frac{p}{1-\beta q})$; לכן $\mathbf{E}(X) = \frac{q\alpha}{p}$.]
2. נניח שנתפסו $X = k$ עכברים. מה ההתפלגות של המשתנה $Y = S - X$ הסופר כמה עכברים לא נתפסו? [תשובה. ההתפלגות היא בינומית שלילית: $Y|X \sim \mathrm{NB}(X+1, q\beta)$.] והתוחלת? [תשובה. $\mathbf{E}(Y|X) = \frac{(X+1)q\beta}{1-q\beta}$.]
3. מה ההתפלגות של $Y = S - X$? [תשובה. גאומטרית: $Y+1 \sim \mathrm{G}(\frac{p}{\beta q+p})$.]
4. הסבר מדוע החקלאי מקווה לתפוס כמה שפחות עכברים, למרות שכל עכבר שנתפס הוא עכבר אחד פחות המכרסם את התבואה.
5. מצא את השונות המשותפת $\mathrm{Cov}(X,Y)$. [תשובה. $\mathrm{Cov}(X,Y) = \alpha\beta(q/p)^2$.]

תרגיל 2.3.65 (בעיית הגפרורים) לסטטיסטיקאי n גפרורים בכל כיס. כשעולה צורך, הוא שולף גפרור באקראי מאחד הכיסים. חשב את התפלגות מספר הגפרורים שנותרו לו, כשהוא מרוקן את אחד הכיסים בפעם הראשונה.
(הכלליות אינה מוסיפה לקושי, ולכן אפשר להניח שיש לסטטיסטיקאי m גפרורים בכיס הימני, ו־n בשמאלי; ושהסיכוי לפנות לכיס ימין הם p, לעומת $q = 1-p$ לגשת לכיס שמאל.)

1. נניח שהסטטיסטיקאי ממשיך לגשת אל הכיסים באקראי עד ששניהם מתרוקנים. נסמן ב־X_R את מספר הפניות לכיס ימין עד לפניה ה־n־ית לכיס שמאל, וב־X_L את מספר הפניות לכיס שמאל עד לפניה ה־m־ית לכיס ימין. הוכח שמתקיים בדיוק אחד משני התנאים $X_L < n$ או $X_R < m$.
2. הראה ש־$X_L \sim \mathrm{NB}(n;p)$ ו־$X_R \sim \mathrm{NB}(m;q)$.
3. נסמן ב־Y את מספר הגפרורים שנותרו בכיס השני, כשהראשון מתרוקן. הראה ש־$Y = m - X_R$ אם $X_L = n$ ו־$Y = n - X_L$ אם $X_R = m$.
4. הסק משני הסעיפים הקודמים ש־
$$P(Y=k) = \binom{n+m-k-1}{m-1} p^{n-k} q^m + \binom{n+m-k-1}{n-1} p^n q^{m-k},$$
לכל $1 \leq k \leq \max\{n,m\}$. בפרט, אם $n = m$ ו־$p = q$, אז $p_{n,n}(k) = \binom{2n-1-k}{n-1} 2^{k-2n}$.
5. נסמן ב־$p_{n,m}(k)$ את הסיכוי ש־$Y = k$ כשמתחילים מכיסים שיש בהם n, m גפרורים. הראה באמצעות נוסחת ההסתברות השלמה ש־$p_{n,m}(k) = pp_{n,m-1}(k) + qp_{n-1,m}(k)$, ובדוק שהנוסחה שהתקבלה לעיל מקיימת תנאי זה.

6. איך היתה משתנה ההתפלגות, אם במקום לדבר על מספר הגפרורים שיש לסטטיסטיקאי בזמן שהוא מרוקן את הכיס הראשון, היינו מעוניינים במספר הגפרורים שיש לו כשהוא מגלה כיס ריק בפעם הראשונה?

2.3.7 התפלגות היפרגאומטרית

תרגיל 2.3.66 נניח ש־$X \sim \mathrm{Bin}(n,p)$, $Y \sim \mathrm{Bin}(m,p)$, בלתי תלויים. חשב את ההתפלגות של X בהנתן $X+Y=N$.

אומרים ש־$X \sim \mathrm{Hyp}(n;A,B)$ אם X מספר האדומים בבחירת n כדורים משק שיש בו $N=A+B$ כדורים, מהם A אדומים ו־B אחרים. הדגימה כאן היא ללא החזרה.

תרגיל 2.3.67 הראה שאם הדגימה היתה עם החזרה, היינו מקבלים התפלגות $X \sim \mathrm{Bin}(n,\frac{A}{N})$.

הדגימה שלנו היא, כאמור, ללא החזרה. המאורע $X=a$ מתרחש כשבוחרים a מתוך A הכדורים האדומים ו־$n-a$ מתוך B הכדורים הכחולים, ולכן

$$P(X=a)=\frac{\binom{A}{a}\binom{B}{n-a}}{\binom{N}{n}}.$$

תרגיל 2.3.68 תן הסבר קומבינטורי לכך ש־$\binom{N}{n}=\sum_{a=0}^{A}\binom{A}{a}\binom{B}{n-a}$.

את התוחלת אפשר לחשב באמצעות משתנים מציינים $X_1,\dots,X_n$, המוגדרים כך ש־$X_i=1$ אם ורק אם הכדור ה־i הוא אדום. כך $\mathbf{E}(X_i)=\frac{A}{A+B}$ לכל i. המשתנים X_i תלויים זה בזה, ובכל־זאת ל־$X=X_1+\cdots+X_n$ יש תוחלת $\mathbf{E}(X)=\frac{nA}{N}$. מכיוון ש־$\mathbf{E}(X_iX_j)=\frac{A(A-1)}{N(N-1)}$, $\mathrm{Cov}(X_i,X_j)=\frac{-AB}{N^2(N-1)}$, ולכן $\mathbf{V}(X)=\frac{nAB(N-n)}{N^2(N-1)}$.

הערה 2.3.69 תהי Ω קבוצה סופית, ותהי $S\subseteq\Omega$. אם מגרילים קבוצה $T\subseteq\Omega$ בגודל n, אז $|S\cap T|\sim \mathrm{Hyp}(n;|S|,|\Omega|-|S|)$.

אין צורך להניח שקבוצה אחת קבועה במקומה: תהי Ω קבוצה סופית. אם מגרילים קבוצות $S,T\subseteq\Omega$ כך ש־$|S|=n$, $|T|=m$, אז $|S\cap T|\sim \mathrm{Hyp}(n;m,|\Omega|-m)$.

מן התאור האחרון עולה שההתפלגויות $\mathrm{Hyp}(n;A,B)$, $\mathrm{Hyp}(A;n,A+B-n)$. בדומה לזה, אם $X\sim\mathrm{Hyp}(n;A,B)$ אז מספר הכדורים הכחולים הוא $n-X$, שהתפלגותו $\mathrm{Hyp}(n;B,A)$.

אם N גדול מאד, ההבדל בין דגימה עם החזרה לדגימה בלי החזרה הולך ומטשטש. תופעה זו מודגמת כאשר משכפלים את הרכב הכד: $\mathrm{Hyp}(n;\theta A,\theta B)\to\mathrm{Bin}(n,\frac{A}{A+B})$ כש־$\theta\to\infty$.

תרגיל 2.3.70 בכד יש $N = A+B+C$ כדורים, בשלושה צבעים. נסמן ב־(X,Y,Z) את מספר הכדורים בכל צבע כשבוחרים n כדורים.

1. $X \sim \mathrm{Hyp}(n; A, B+C)$, $Y \sim \mathrm{Hyp}(n; B, A+C)$.
2. $X+Y \sim \mathrm{Hyp}(n; A+B, C)$.
3. חשב את $\mathrm{Cov}(X,Y)$.

רעיון 2.3.71 מספר הדגים באגם קטן אינו ידוע. דגים 1000 דגים, ומסמנים להם את הסנפיר. אחרי שבוע דגים שוב 1000 דגים. נמצאו 17 דגים מסומנים. הערך את מספר הדגים באגם?
(את השיטה הזו אפשר להפעיל גם קצת אחרת, אם מוסיפים לאגם 1000 דגים מסומנים.)

תרגיל 2.3.72 הפרופסור קטרוס שופך לים התיכון כוס של מים סגולים. המים מתערבבים היטב. הערך כמה מולקולות של מים כאלה יהיו בכוס מי ים לאחר זמן. (שטח הים התיכון כ־2.5 מליון קמ"ר; עמקו הממוצע כ־1500 מ'; נפח הכוס כ־200 סמ"ק; המשקל האטומי של מים הוא 18; מספר אבוגדרו מוערך ב־$6\cdot 10^{23}$).

תרגיל 2.3.73 בחודש פברואר 2011 פגע רעש אדמה בעיר קרייסטצ'רץ' שבניו־זילנד, שבה 390000 תושבים. ברעש נפגעו כ־100 אנשים. בעיר שהו בעת הרעש כ־200 ישראלים. לכמה נפגעים ישראלים היית מצפה?

תרגיל 2.3.74 במחלקות שונות במפעל יש $2,3,2,4,1,2$ עובדים העונים לקריטריונים לפרס מסויים, שיזכו בו ארבעה. בסופו של דבר זוכים בפרס $1,1,2,0,0,0$ מועמדים ממחלקות אלה, בהתאמה. חשב את הסיכויים לכך שכך יהיה במקרה.

תרגיל 2.3.75 לאחרונה דווח (www.haaretz.co.il/hasite/spages/1236742.html) ש"חוקרים ישראלים גילו חומר המונע פוסט־טראומה": "מחקר ראשוני שבוצע באחרונה בחדר המיון במרכז הרפואי שיבא בקרב 24 נפגעי תאונות דרכים שהראו סימני חרדה במיון, מצא כי בקרב 12 נבדקים שטופלו בהורמון קורטיזול, אחד בלבד פיתח פוסט־טראומה, בהשוואה לשלושה בקרב 12 הנבדקים שלא טופלו בהורמון וקיבלו טיפול דמה". מהו הסיכוי לתוצאה שהתקבלה בניסוי, אם החומר אינו משפיע כלל?

תרגיל 2.3.76 עורכים סדרה של n ניסויי ברנולי בלתי תלויים בעלי סיכוי הצלחה חצי. אחר־כך מבצעים סדרה דומה של ניסויים, באותו אורך. מה הסיכויים לכך שבשתי הסדרות יהיה אותו מספר הצלחות? **הדרכה.** נסמן את מספרי ההצלחות ב־X,Y. מכיוון ש־$X,Y \sim \mathrm{Bin}(n,p)$, $P\left(X=Y\right) = \sum_{k=0}^{n}\binom{n}{k}^2 2^{-2n} = 2^{-2n}\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k} = 2^{-2n}\binom{2n}{n} \approx \frac{1}{\sqrt{\pi n}}$.

פרדוקס יום ההולדת

"'פרדוקס יום ההולדת'" הוא כינוי לתופעה מפתיעה: כבר בקבוצה של 23 אנשים (שתאריכי הלידה שלהם מתפלגים באופן אחיד ובלתי תלוי על 365 הימים האפשריים), הסיכוי לכך ששניים חולקים יום־הולדת משותף עולה על 50%. עובדה זו נדמית כפרדוקס, משום ש־23 הוא מספר קטן ביחס ל־365. נציג כמה דרכים לנתח את התופעה.

טענה 2.3.77 יהיו $X_1, \dots, X_m \sim U[1, m]$ משתנים מקריים בדידים בלתי תלויים. הסיכוי לכך שכל ה־X_i־ים שונים זה מזה קטן מ־$e^{-\frac{m(m-1)}{2n}}$, ולכן יורד מתחת לחצי כאשר $m \approx \sqrt{2\log(2)\, n}$.

הוכחה. כידוע, $e^{-x} \geq 1 - x$ לכל x ממשי. הסיכוי שווה למכפלה

$$\begin{aligned} p &= 1 \cdot \left(1 - \tfrac{1}{n}\right) \cdot \left(1 - \tfrac{2}{n}\right) \dots \left(1 - \tfrac{m-1}{n}\right) \\ &< e^{-\frac{1}{n}} \cdot e^{-\frac{2}{n}} \cdot \dots \cdot e^{-\frac{m-1}{n}} = e^{-(\frac{1}{n} + \frac{2}{n} + \dots + \frac{m-1}{n})} = e^{-\frac{m(m-1)}{2n}}. \end{aligned}$$

□

עבור $n = 365$, הסיכוי שלא תהיה אף התנגשות יורד מתחת לחצי כאשר $m \geq 23$, ואכן החסם בטענה 2.3.77 הוא $m \approx 22.5$.

אם מחשבים במקום זה את **תוחלת מספר ההתנגשויות** מקבלים תוצאה מאותו סדר גודל, עם קבוע מעט שונה:

הערה 2.3.78 נסמן (עבור $i < j$) ב־Δ_{ij} את המשתנה המקרי המציין את המאורע $X_i = X_j$. אז $\Delta_{ij} \sim b(\frac{1}{n})$, ולכן התוחלת של מספר ההתנגשויות $X = \sum_{i<j} \Delta_{ij}$ היא $\frac{1}{n}\binom{m}{2}$. כדי שהתוחלת תגיע ל־1 נדרש $m \approx \sqrt{2n}$.

תרגיל 2.3.79 הראה שהמשתנים Δ_{ij} בלתי־תלויים בזוגות. מצא תנאי על האינדקסים לכך ששלושה משתנים מקבוצה זו יהיו תלויים. הצע פירוש בשפה של תורת הגרפים לתלות של קבוצה כלשהי של משתנים.

הערה 2.3.80 אפשר להגדיר $T = \min_j \{\exists i < j : X_i = X_j\}$ – **זמן ההמתנה** להתנגשות הראשונה. מתברר ש־$\mathbf{E}(T) \approx \sqrt{\frac{\pi n}{2}}$.

את זמן ההמתנה להתנגשות הראשונה, $O(\sqrt{n})$, יש להשוות לזמן עד למילוי כל התאים, שחושב בתרגיל 2.3.54 להיות $O(n\log(n))$.

תרגיל 2.3.81 הערך, מה מספר האנשים שממנו והלאה הסיכוי ששלושה יחלקו יום הולדת משותף עולה על 50%?

הערה 2.3.82 בלוטו הישראלי בוחרים 6 מספרים מתוך 37 (נתעלם מן 'המספר הנוסף'). באוקטובר 2010 התקבלו בדיוק אותם המספרים שהתקבלו שמונה הגרלות קודם לכן. חשב את הסיכוי לאירוע (מהו לדעתך ה'אירוע' הרלוונטי?).

באותו אירוע היו גם חזרות על מספרים; למשל, המספר 14 חזר 6 פעמים בשמונה הגרלות (זהו אירוע מסובך שקשה לחשב את ההסתברות שלו באופן ישיר. לפי תוצאות סימולציה, הסיכוי לזה בשמונה הגרלות שהוגדרו מראש הוא כ־1:800).

מתברר שהסיכוי לכך שיופיעו בהגרלה נתונה לפחות שני מספרים עוקבים הוא 61%; בכתבה ב־$Ynet$ מסביר אחד המהמרים הכבדים ש"הנטייה של מספרים עוקבים לחזור היא גדולה מאוד, פרט שאינו מסתדר סטטיסטית בשום אופן".

תרגיל 2.3.83 לכל משתתף בכנס מייצרים מספר אישי אקראי, בטווח $[1,d]$. אם משתתפים בכנס n אנשים, מה צריך להיות d כך שהסיכוי להתנגשות יהיה נמוך למדי? ואם משתמשים באותה שיטה (עם אותו d) ב־t כנסים שונים, כאשר החשש מהתנגשות הוא רק בין המשתתפים של אותו כנס?

תרגיל 2.3.84 מייצרים 3000 שמות אקראיים לקבצים במערכת יוניקס. כל שם מורכב משבע אותיות אנגליות הנבחרות באקראי (באופן בלתי תלוי ובהתפלגות אחידה) מכל $52 = 2 \cdot 26$ האפשרויות (מערכת ההפעלה הזו מבחינה בין אותיות גדולות וקטנות). אחר־כך מעתיקים את כל הקבצים למערכת חלונות, שבה אין הבחנה בין אותיות גדולות לקטנות. מה הסיכויים לכך שאחד הקבצים או יותר יפגע מהתנגשות?

2.3.8 גרפים מקריים

המבנה של עצמים קומבינטוריים מקריים (כגון גרפים, עצים, תמורות, פונקציות) הוא מודל פשוט להתנהגות של מבנים מסובכים (למשל, מערך החברויות בפייסבוק הוא גרף שבו המשתמשים הם קודקודים המחוברים בקשת אם הם חברים. הגרף עצמו מסובך מאד, וכדי לחקור אותו ממדלים אותו כגרף מקרי בעל תכונות מתאימות). את העצמים הקומבינטוריים חוקרים באמצעות מגוון של טכניקות הסתברותיות, מהן בסיסיות ומהן מתוחכמות יותר.

נציג כאן על קצה המזלג דוגמא מרכזית אחת. "מודל Erdös-Renyi **לגרף מקרי**" הוא גרף על n קודקודים, שכל אחת מ־$\binom{n}{2}$ הצלעות האפשרויות שלו קיימת בהסתברות p, באופן בלתי תלוי באחרות. בחקירת המבנה של גרף כזה מקובל לקבוע את p כפונקציה של n, ולתת ל־n לשאוף לאינסוף.

1. כאשר $p = \frac{\lambda}{n^2}$ (מכאן ולהבא λ קבוע), מספר הקשתות בגרף סופי (למרות ש־n שואף לאינסוף). הקשתות מעטות כל־כך עד שאינן נפגשות (בהסתברות השואפת ל־1 כאשר n שואף לאינסוף, וכך בכל טענה בסעיף זה).

2. כאשר p עולה ל־$p = \frac{\lambda}{n^{3/2}}$ נוצרים בגרף מסלולים באורך 2. כאשר $p = \frac{\lambda}{n^{4/3}}$ נוצרים עצים בני שלוש קשתות (מכל הסוגים בבת־אחת), וכן הלאה. כל עוד $np \to 0$, אין בגרף מעגלים.

3. אלו נוצרים רק כאשר $p = \frac{\lambda}{n}$, ואז יש מעגלים בכל אורך (חסום). כאשר $\lambda < 1$ כל מרכיבי הקשירות הם פשוטים (עצים עם קשת נוספת אחת לכל היותר). הגודל של מרכיבי הקשירות הוא לכל היותר כפולה של $\log(n)$. כאשר $\lambda = 1$ מתרחש "מעבר פאזה" מסובך (שלא נתאר כאן). מיד אחר־כך, כאשר $\lambda > 1$, יש בגרף מרכיב קשירות ענק, שגודלו כפולה לינארית של n, ושאר המרכיבים עדיין פשוטים. בכל השלב הזה יש בגרף קודקודים מבודדים רבים: $e^{-\lambda}$ מכל הקודקודים הם מבודדים.

4. העידן הבא מגיע כאשר $p = \frac{\lambda \log(n)}{n}$. כמעט כל מרכיבי הקשירות הקטנים נדבקו למרכיב הענק זה מכבר. יש רק מספר סופי של קודקודים מבודדים, והגרף נעשה קשיר כשהאחרון ביניהם מצטרף למרכיב הקשירות הענק.

2.4 מרחב התפלגות כללי

עד כאן, מרחב ההתפלגות שלנו היה מרחב סופי או בן־מניה. אנו רוצים לטפל גם במקרה הכללי, שבו המרחב הוא למשל $\Omega = [0,1]$ או $\Omega = \mathbb{R}$. לפני שנציג את ההגדרות הדרושות, עלינו להסביר מדוע ההגדרה הבדידה אינה מתאימה למקרה הכללי.

2.4.1 סיכום על קבוצה שאינה בת־מניה

באופן נאיבי, אנחנו עלולים לנסות להגדיר הסתברות על קבוצה שאינה בת מניה בעזרת ההגדרה המקורית: נדרוש שהפונקציה $P : \Omega \to \mathbb{R}$ תקיים $\sum P(\omega) = 1$. הבעיה היא שסכום על פני קבוצת נקודות שאינה בת מניה, אינו יכול להיות סופי: אחרת כל אחת מהקבוצות $\Omega_n = \{\omega : P(\omega) > 1/n\}$ מוכרחה להיות סופית, והרי $\cup \Omega_n = \{\omega : P(\omega) \neq 0\}$. לכן P תהיה אפס מחוץ לקבוצה בת מניה.

תרגיל 2.4.1 לאור הקשיים בהגדרת התפלגות על קבוצות גדולות, האם יתכן שיש מידת הסתברות על $[0,1] \cap \mathbb{Q}$ שאין בה נקודות מיוחסות (כלומר, $P(\alpha) = 0$ לכל נקודה α במרחב)?

2.4.2 אי־קיומן של מידות אינווריאנטיות

העובדה שלא ניתן לסכם על קבוצות גדולות מדי מונעת מאיתנו להגדיר את ההסתברות ישירות על נקודות. אבל בסופו של דבר איננו מעוניינים בנקודות, אלא במאורעות,

ונוכל להגדיר את ההסתברות ישירות עליהם. לשם כך נרצה לשמור את האקסיומות שבטענה 2.1.8.

מתברר שגם נסיון זה אינו עולה יפה. נסמן ב־S^1 את הנקודות על מעגל היחידה, שאיבריו הם המספרים המרוכבים $e^{2\pi i\theta} = \cos(\theta) + i\sin(\theta)$ עבור $\theta \in \mathbb{R}$. נניח ש־$P : \mathcal{P}(S^1) \to \mathbb{R}$ היא פונקציה המקיימת את שתי ההנחות האמורות. מכיוון שמטרתנו להכליל את מושג אורך הקשת, שאינו מושפע מסיבוב, נניח שגם P היא **אינווריאנטית לסיבוב**: לכל $e^{2\pi i\theta} \in S^1$, $P(e^{2\pi i\theta}A) = P(A)$. נגדיר יחס שקילות על S^1: $x \sim y$ אם יש $\theta \in \mathbb{Q}$ כך ש־$y = e^{2\pi i\theta}x$. תהי X קבוצה הכוללת נציג אחד בדיוק מכל מחלקת שקילות (קיומה של קבוצה כזו דורש את **אקסיומת הבחירה**). לפי ההנחה, הקבוצה X זרה לכל הסיבובים שלה $e^{2\pi i\theta}X$ בזוויות רציונליות $\theta \in \mathbb{Q} \cap [0,1]$, ואיחוד כל הקבוצות האלה הוא S^1. לכן, לכאורה,

$$1 = P(S^1) = P(\cup_\theta e^{2\pi i\theta}X) = \sum_{\theta\in\mathbb{Q}\cap[0,1]} P(e^{2\pi i\theta}X) = \sum_{\theta\in\mathbb{Q}\cap[0,1]} P(X),$$

אבל זהו סכום בן־מניה על ערך קבוע: אם $P(X) = 0$ הסכום הוא אפס, ואם $P(X) > 0$ הסכום אינסופי.

הראינו שלא ניתן להכליל את מידת האורך באופן כזה שתהיה מוגדרת על כל תת־הקבוצות של S^1. מתברר שאפשר לפתור את הבעיה אם רק נוותר על ההנחה ש־$P(X)$ מוגדרת לקבוצות כגון X שבנינו לעיל. כדי לפתח את התאוריה, נתמקד כמובן בקבוצות שעבורן P כן מוגדרת.

הערה 2.4.2 לאקסיומת הבחירה המופיעה בדוגמא זו יש מסקנות אבסורדיות עוד יותר: למשל, פרדוקס בנך־טרסקי, המאפשר לפרק את הכדור ל־10 חלקים ולהרכיב מחמשת הראשונים ומחמשת האחרונים כדורים שלמים. כתוצאה מכך, לא קיימת מידה אינווריאנטית לסיבובים המוגדרת על כל תת־הקבוצות של פני הכדור S^2.

2.4.3 סיגמא־אלגברות

הגדרה 2.4.3 משפחה $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ הכוללת את Ω וסגורה ללקיחת משלים ולאיחוד בן־מניה, נקראת σ־אלגברה על Ω.

טענה 2.4.4 כל σ־אלגברה סגורה לפי כללי דה־מורגן לחיתוך בן־מניה, ובפרט גם לאיחוד וחיתוך סופי (משפחה הסגורה למשלים ולאיחוד סופי נקראת **אלגברה** של קבוצות).

דוגמא 2.4.5 אוסף הקבוצות הסופיות והקו־סופיות ב־$\mathbb{R}$, כלומר, אלה שהן או סופיות או שמשלימותיהן סופיות, הוא אלגברה שאינה σ־אלגברה.

מאידך, **אלגברת בורל** שנגדיר בהמשך היא σ־אלגברה שאינה סגורה לאיחוד כלשהו של קבוצות. מכאן שהסגירות לאיחוד בן־מניה חלשה ממש מסגירות לאיחוד כלשהו, אבל חזקה ממש מהסגירות לאיחוד סופי.

תרגיל 2.4.6 תהי $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ σ־אלגברה על Ω, ותהי $Y \subseteq \Omega$ תת־קבוצה. קבע את הקשר בין הקבוצות $\{A \in \mathcal{F}:\ A \subseteq Y\}$ ו־$\{A \cap Y:\ A \in \mathcal{F}\}$. הוכח שאחת מהן היא σ־אלגברה על Y. הראה שהן שוות אם $Y \in \mathcal{F}$.

תרגיל 2.4.7 הוכח שחיתוך משפחה כלשהי של σ־אלגברות על Ω הוא σ־אלגברה. מכאן שלכל אוסף T קיימת σ־אלגברה מינימלית המכילה את T; זוהי ה־σ־אלגברה **הנוצרת על־ידי** T.

תרגיל 2.4.8 **הצפיפות הטבעית** על תת־קבוצות $A \subseteq \mathbb{N}$ מוגדרת לפי $\delta(A) = \lim_{n\to\infty} \frac{|A\cap\{1,\dots,n\}|}{n}$, אם הגבול קיים. הראה שאוסף הקבוצות שיש להן צפיפות טבעית סגור לאיחוד וללקיחת משלים, אבל אינו σ־אלגברה. הראה שלקבוצת המספרים שמספר הספרות שלהם זוגי אין צפיפות טבעית.

2.4.4 מרחבי הסתברות

תהי $\mathcal{F}$ σ-אלגברה על המרחב Ω. פונקציה $P\colon \mathcal{F} \to \mathbb{R}$ היא **אדיטיבית** אם לכל $A, B \in \mathcal{F}$, אם $A \cap B = \emptyset$ אז $P(A \cup B) = P(A) + P(B)$; ו־**$\sigma$־אדיטיבית** אם לכל $A_1, A_2, \dots \in \mathcal{F}$ זרות בזוגות מתקיים $P(\cup A_n) = \sum_n P(A_n)$.

הגדרה 2.4.9 (האקסיומטיקה של קולמוגורוב) שלשה סדורה $(\Omega, \mathcal{F}, P)$, שבה Ω קבוצה כלשהי, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ היא σ-אלגברה, ו־ $P\colon \mathcal{F} \to \mathbb{R}$ נקראת **מרחב הסתברות** אם מתקיים:

1. P חיובית (כלומר $P(A) \geq 0$ לכל $A \in \mathcal{F}$);
2. $P(\Omega) = 1$;
3. P σ־אדיטיבית.

כמובן, $\cup A_n \in \mathcal{F}$ מכיוון ש־$\mathcal{F}$ סגורה לאיחוד בן־מניה. חושבים על $\mathcal{F}$ בתור קבוצת המאורעות - דברים שיש משמעות להסתברות שלהם. הפונקציה P נקראת **פונקציית הסתברות**.

תרגיל 2.4.10 אם P פונקציית הסתברות, אז לכל $A \subset B$ השייכים ל־$\mathcal{F}$ מתקיים $P(B-A) = P(B) - P(A)$.

תרגיל 2.4.11 תהי $P\colon \mathcal{F} \to \mathbb{R}$ פונקציה חיובית ואדיטיבית, ונניח ש־$P(\Omega) = 1$. הראה שהתנאים הבאים שקולים (בכל המקרים מדובר בקבוצות השייכות ל־$\mathcal{F}$):

1. P היא σ־אדיטיבית.

2. אם $B_1 \subseteq B_2 \subseteq \cdots$ שרשרת עולה ב־$\mathcal{F}$ אז $P(\cup B_n) = \lim_{n\to\infty} P(B_n)$;

3. אם $C_1 \supseteq C_2 \supseteq \cdots$ שרשרת יורדת ב־$\mathcal{F}$ אז $P(\cap C_n) = \lim_{n\to\infty} P(C_n)$.

כדי להבין טוב יותר את ההבדל בין ההגדרה הנוכחית לקודמת, נעיר שבדוגמאות שנראה בעתיד (עם $\Omega = \mathbb{R}$, למשל), יתקיים $P(\{a\}) = 0$ לכל $a \in \Omega$, ולכן גם $P(A) = 0$ לכל A בת־מניה. מבחינת תורת ההסתברות (של מרחבים כאלה), הקבוצות בנות המניה אינן קיימות.

תרגיל 2.4.12 שקול את ההבדל בין 'בלתי אפשרי' ($A \cap \Omega = \emptyset$), 'לא רלוונטי' ($A \not\subseteq \Omega$), 'בעל הסתברות לא מוגדרת' ($A \subseteq \Omega$, $A \notin \mathcal{F}$) ו'בעל הסתברות אפס' ($P(A) = 0$).

צמצום של מרחב הסתברות $(\Omega, \mathcal{F}, P)$ לתת־קבוצה B (כאשר $P(B) > 0$) נעשה כך: $(B, \mathcal{F}|_B, P(\cdot|B))$, כאשר $\mathcal{F} = \{C \in \mathcal{F} \colon C \subseteq B\}$, ו־ $P(C|B) = \frac{P(C)}{P(B)}$.

תרגיל 2.4.13 הוכח ש־$(B, \mathcal{F}|_B, P(\cdot|B))$ הוא מרחב הסתברות (ראה תרגיל 2.4.6).

2.4.5 בעיית ברטרנד

חוקר ההסתברות הצרפתי ג'וזף ברטרנד (1822-1900) שאל את השאלה הבאה: מעבירים מיתר אקראי במעגל. מה הסיכוי לכך שהוא יהיה ארוך יותר מצלע המשולש שווה הצלעות החסום במעגל?

1. אם בוחרים את המיתר דרך נקודות קצה אקראיות, הסיכוי הוא $1/3$.

2. אם מקצים את המיתר במאונך לנקודה אקראית על רדיוס אקראי, הסיכוי הוא $1/2$.

3. אם בוחרים את נקודת האמצע של המיתר באקראי (שים לב שנקודת האמצע קובעת את המיתר), הוא יהיה ארוך מן הצלע אם ורק אם הנקודה בתוך מעגל ברדיוס חצי, ולזה יש הסתברות $1/4$.

ה'פרדוקס' הזה נועד להדגים דבר אחד: אי אפשר לדבר על 'אקראיות' סתם, בלי להגדיר בדיוק באיזו התפלגות מדובר. זהו אחד השעורים החשובים ביותר בקורס: חשובה לא רק האוכלוסיה שממנה בוחרים, אלא גם אופן הבחירה הקובע את ההתפלגות.

2.5 משתנים מקרים רציפים

2.5.1 σ-אלגברת בורל

הגדרה 2.5.1 ה־σ-אלגברה על $\mathbb{R}$ הנוצרת על־ידי כל הקטעים הפתוחים נקראת אלגברת בורל; מסמנים אותה ב־$\mathcal{B}$.

טענה 2.5.2 אלגברת בורל כוללת את כל הקטעים והקרניים. אכן, לכל $a < b$,

1. $(a,\infty) = \cup_{n=1}^{\infty}(a, a+n) \in \mathcal{B}$;
2. $(-\infty, b) = \cup_{n=1}^{\infty}(b-n, b) \in \mathcal{B}$;
3. $[a,\infty) = (-\infty, a)^{\mathrm{c}} \in \mathcal{B}$;
4. $(-\infty, b] = (b,\infty)^{\mathrm{c}} \in \mathcal{B}$;
5. $[a,b) = [a,\infty) \cap (-\infty, b) \in \mathcal{B}$;
6. $(a,b] = (a,\infty) \cap (-\infty, b] \in \mathcal{B}$;
7. $[a,b] = [a,\infty) \cap (-\infty, b] \in \mathcal{B}$;
8. $\{a\} = [a,\infty) \cap (-\infty, a] \in \mathcal{B}$.
9. $\mathcal{B}$ כוללת את כל הקבוצות בנות־המניה או שמשלימותיהן בנות מניה.

טענה 2.5.3 אלגברת בורל היא גם ה־σ-אלגברה הנוצרת על־ידי הקרניים $(-\infty, b]$.

הוכחה. $\cup_{n=1}^{\infty}(-\infty, b-1/n] = (-\infty, b)$, ו־$(a,b) = (-\infty, b) \cap (-\infty, a)^{\mathrm{c}}$. $\square$

הגדרה 2.5.4 יהי $(\Omega, \mathcal{F}, P)$ מרחב הסתברות. פונקציה $X : \Omega \to \mathbb{R}$ היא משתנה מקרי, אם $X^{-1}(B) \in \mathcal{F}$ לכל $B \in \mathcal{B}$. [פונקציה כזו נקראת **מדידה**.]
לפי טענה 2.5.3, די בכך ש־$\{\omega :\ X(\omega) \leq b\} = X^{-1}((-\infty, b]) \in \mathcal{F}$ לכל b, ולפי טענה 2.5.2 כל הסתברות $P(a < X < b)$ מוגדרת באופן דומה.

שימו לב שהשאלה אם X הוא משתנה מקרי אינה תלויה כלל בפונקציית ההסתברות P.

דוגמא 2.5.5 לכל פונקציית הסתברות P, פונקציית הזהות $\mathbb{R} \to \mathbb{R}$ היא משתנה מקרי על המרחב $(\mathbb{R}, \mathcal{B}, P)$.

לכל פונקציה רציפה (או בעלת מספר סופי של נקודות אי־רציפות) f, אם X משתנה מקרי אז גם $f(X)$ משתנה מקרי. בנושא זה תעמיקו יותר בקורס על תורת המידה.

2.5.2 פונקציית הצטברות

הגדרה 2.5.6 פונקציה $F:\mathbb{R}\to\mathbb{R}$ שהיא מונוטונית לא יורדת, רציפה מימין, ומקיימת

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

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

נקראת **פונקציית הצטברות** (לפעמים גם **פונקציית התפלגות מצטברת**).

אנו עוסקים בפונקציות הצטברות בגלל ההתאמה ביניהן לבין משתנים מקריים:

טענה 2.5.7 יהי X משתנה מקרי על $\mathbb{R}$. אז הפונקציה $F_X(b) = P(X \leq b)$ היא פונקציית הצטברות.

הוכחה. הרציפות מימין נובעת מזה ש־ $\bigcap_n(-\infty, b+1/n] = (-\infty, b]$. □

טענה 2.5.8 תהי F_X פונקציית הצטברות. אז $P(X \leq b) = F(b)$ מגדיר משתנה מקרי על $\mathbb{R}$.

מאיחוד הקטעים $(-\infty, b-1/n)$, נובע ש־

$$P(X = b) = F(b) - \lim_{x\to b^-} F(x). \tag{2.6}$$

מסקנה 2.5.9 F רציפה אם ורק אם $P(X = b) = 0$ לכל b.

הערה 2.5.10 1. לפונקציית הצטברות יש לכל היותר מספר בן מניה של נקודות אי־רציפות (זה נובע מנוסחה (2.6)).

2. קבוצת נקודות אי־הגזירות הוא ממידה אפס (משפט לבג).

3. מאידך יתכן שיהיו $\aleph$ נקודות אי־גזירות (פונקציית קנטור).

2.5.3 פונקציית צפיפות

הגדרה 2.5.11 פונקציה חיובית $f:\mathbb{R}\to\mathbb{R}$ כך ש־ $\int_{-\infty}^{\infty} f(x)dx = 1$ נקראת **פונקציית צפיפות**.

פונקציית צפיפות מאפשרת להגדיר פונקציה $F(b) = \int_{-\infty}^{b} f(t)dt$, שהיא פונקציית הצטברות גזירה בכל נקודה. זו, בתורה, מגדירה משתנה מקרי בעל סיכוי אפס לכל נקודה. להיפך, בנקודות שבהן F גזירה, מתקיים $f(b) = F'(b)$, וזה מגדיר פונקציית צפיפות.

פונקציית צפיפות מספקת תאור יעיל של המשתנה, ואפשר לחשב ממנה את כל התכונות שלו. הצפיפות מתארת את הסיכוי של המשתנה לקבל ערך בקטע קטן: סביר יותר ליפול לקטע הכולא ערכי צפיפות גדולים מאשר לכזה הכולא ערכים קטנים. בניסוי שבו "מרססים" נקודות על פי ההתפלגות, צפיפות הנקודות תהיה פרופורציונלית לפונקציית הצפיפות (זוהי מסקנה מהחוק החלש של המספרים הגדולים, משפט 2.10.1).

2.5.4 תוחלת ושונות

התוחלת של משתנה מקרי X עם צפיפות f_X מוגדרת לפי

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} f_X(t)tdt,$$

בתנאי שהאינטגרל הזה מתכנס. באופן כללי יותר, התוחלת של פונקציה $g(X)$ היא $\mathbf{E}(g(X)) = \int_{-\infty}^{\infty} f_X(t)g(t)dt$.

השונות מוגדרת כרגיל: $\mathbf{V}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2$.

טענה 2.5.12 יהי X משתנה מקרי המקבל ערכים חיוביים. אז

$$\mathbf{E}(X) = \int_{0}^{\infty} (1 - F_X(t))dt$$

כאשר $F_X(x) = P\left(X \leq x\right)$ היא פונקציית ההצטברות.

הוכחה. מבצעים החלפות משתנים:

$$\begin{aligned}
\mathbf{E}(X) &= \int_0^\infty t f_X(t)dt \\
&= \int_0^\infty \int_0^t f_X(t)dsdt \\
&= \int_0^\infty \int_s^\infty f_X(t)dtds \\
&= \int_0^\infty P(X > s)ds \\
&= \int_0^\infty (1 - F_X(s))ds.
\end{aligned}$$

□

תרגיל 2.5.13 הראה כיצד מכלילה טענה 2.5.12 (על הקשר בין התוחלת לפונקציית ההצטברות) את טענה 2.3.57 (על קשר דומה במקרה הבדיד).

תרגיל 2.5.14 חזור לבעיית ברטרנד (סעיף 2.4.5) והראה שתוחלת האורך של המיתר האקראי היא $\frac{4}{\pi}$, $\frac{\pi}{2}$ או $\frac{4}{3}$, בהתאמה לאופן ההגרלה שלו: אם קובעים נקודת קצה ומגרילים נקודת קצה שניה באופן אחיד על היקף המעגל, אם מרחקו מהראשית מתפלג אחיד, או אם נקודת המרכז שלו מתפלגת באופן אחיד במעגל.

תרגיל 2.5.15 (אפקט הספגטי - גרסה בדידה) מספר הילדים במשפחה, שנסמן ב־X, מתפלג לפי התפלגות מסויימת. שואלים ילד אקראי כמה ילדים במשפחה שלו. הסבר מדוע התשובה, שאותה נסמן ב־$\hat{X}$, היא בעלת התפלגות שונה משל X; הראה ש־$P\left(\hat{X} = k\right) = \frac{P(X=k)k}{\mathbf{E}(X)}$. הסק ש־$\mathbf{E}(\hat{X}) = \mathbf{E}(X) + \frac{\mathbf{V}(X)}{\mathbf{E}(X)}$.

כעת נניח ש־$Y \sim \mathrm{Bin}(X, \frac{1}{2})$ הוא מספר הבנים במשפחה. הראה שבתוחלת, מספר האחים של בן אקראי, לרבות הוא־עצמו, גדול ממספר הבנים במשפחה אקראית ב־$\frac{1}{2} + \frac{\mathbf{V}(X)}{2\mathbf{E}(X)}$, בעוד שמספר האחיות של בן אקראי גדול ממספר הבנות במשפחה אקראית ב־$\frac{1}{2} - \frac{\mathbf{V}(X)}{2\mathbf{E}(X)}$.

תרגיל 2.5.16 (אפקט הספגטי - גרסה רציפה) בקערה ענקית יש אטריות ספגטי שהאורך שלהן מתפלג לפי פונקציית צפיפות f, עם תוחלת μ. כשדוגמים אטריה, היא מתקבלת בהסתברות פרופורציונלית לאורך שלה: אם אטריה אחת ארוכה מהשניה פי k, אז הסיכוי ששיני המזלג יינעצו דווקא בה גדול פי k.

1. הוכח שלאורך של אטריה נדגמת יש צפיפות $g(t) = \frac{1}{\mu} f(t)t$. מצא את תוחלת האורך של אטריה נדגמת.

2. חשוב מה יקרה אם להתפלגות f יש תוחלת אינסופית.

3. הצע שיטה לאמוד את μ על־ידי דגימת אטריות (ראה סעיף 3.1.1).

2.5.5 התפלגות משותפת

פורמלית, כדי לטפל בזוגות של משתנים מקריים עלינו להכליל את אלגברת בורל אל המרחב $\mathbb{R}^2$. ההכללה אינה מסובכת, אבל לא ניגע בה כאן.

הגדרה 2.5.17 **פונקציית צפיפות משותפת** של שני משתנים מקריים ממשיים היא פונקציה $f:\mathbb{R}^2\to\mathbb{R}$, חיובית, המקיימת

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1.$$

תהי $f_{X,Y}$ פונקציית הצפיפות המשותפת של X,Y. הפונקציה $f_X(x)=\int f_{X,Y}(x,y)dy$ היא **הצפיפות השולית** של X, ובדומה $f_Y(y)=\int f_{X,Y}dx$ היא הצפיפות השולית של Y.

הערה 2.5.18 נניח ש־$f_{X,Y}$ היא פונקציית הצפיפות המשותפת של X,Y, ו־f_X הצפיפות השולית של X. לכל x,

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

היא **הצפיפות המותנית** של Y (בהינתן $X=x$), ובדומה עבור X. (שימו לב לדמיון של נוסחה זו לחוק בייס (2.3)).

גם להיפך: אם לכל x ידוע ש־$Y|X=x$ הוא בעל צפיפות $g_x(\cdot)$, אז הצפיפות המשותפת היא $f_{X,Y}(x,y)=f_X(x)g_x(y)$, ולכן הצפיפות של Y היא

$$f_Y(y) = \int f_X(x)g_x(y)dx.$$

נוסחה זו היא האנלוג הרציף של נוסחת ההסתברות השלמה (2.2).

דוגמא 2.5.19 שוברים מקל שאורכו יחידה אחת, בנקודה אקראית X (עם התפלגות אחידה). את החלק שאורכו X שוברים שוב בנקודה אקראית (עם התפלגות אחידה), כך שמתקבלים חלקים באורך $A, X-A$. מה ההתפלגות של X בהינתן A? והתוחלת?

פתרון. לפי ההנחה, הצפיפות של X היא $f_X(x)=1$ בקטע $0\le x\le 1$. בהינתן $X=x$, $f_{A|X=x}(a)=\frac{1}{x}$ בקטע $0\le a\le x$. לכן $f_{A,X}(a,x)=f_X(x)f_{A|X=x}(a)=\frac{1}{x}$ בתחום $0\le a\le x\le 1$. לכן פונקציית הצפיפות של A היא $f_A(a)=\int_a^1 f_{A,X}(a,x)dx=\int_a^1\frac{1}{x}dx=-\log(a)$. לכל a, $f_{X|A=a}(x)=\frac{f_{A,X}(a,x)}{f_A(a)}=\frac{1}{-\log(a)}\frac{1}{x}$. מכאן ש־$\mathbf{E}(X|A=a)=\int_a^1 f_{X|A=a}(x)xdx=\int_a^1\frac{1}{-\log(a)}dx=\frac{1-a}{-\log(a)}$.

תרגיל 2.5.20 שוברים מקל שאורכו 1 בנקודה אקראית (בעלת התפלגות אחידה). אחר־כך שוברים כל אחד מהחלקים בנקודה אקראית, וחוזרים על תהליך זה n פעמים, כך שמתקבלים 2^n חלקים. מהו מקדם המתאם בין שני חלקים מן הדור ה־n, שנפרדו זה מזה בדור ה־m? הראה ששני חלקים שנפרדו בדור הראשון או השני מתחרים זה בזה (המקדם שלילי), ואילו שני חלקים שהיו יחד בדור השני (כלומר הגיעו מאותו רבע) הם שותפים (המקדם חיובי).

אי־תלות

דוגמא 2.5.21 לכל שתי פונקציות צפיפות f_X, f_Y, הפונקציה $(x,y) \mapsto f_X(x)f_Y(y)$ היא פונקציית צפיפות משותפת.

הגדרה 2.5.22 X, Y בלתי תלויים אם $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

טענה 2.5.23 התנאים הבאים שקולים:

1. X, Y בלתי תלויים;
2. קיימות פונקציות g_1, g_2 כך ש־$f_{X,Y}(x,y) = g_1(x)g_2(y)$;
3. לכל x, x', $f_{Y|X=x} = f_{Y|X=x'}$;
4. לכל y, y', $f_{X|Y=y} = f_{X|Y=y'}$.

2.5.6 המקרה המעורב

אפשר לטפל בהתפלגות המשותפת של שני משתנים מקריים, X, T, גם כאשר אחד מהם רציף והשני בדיד.

נראה, לדוגמא, כיצד להפוך הסתברויות במקרה שבו לכל n, המשתנה המקרי המותנה $T|X=n$ הוא בעל פונקציית צפיפות, $f_{T|X=n}(t)$.

הערה 2.5.24 נניח ש־T משתנה מקרי רציף, ושלכל t נתונה ההתפלגות (הבדידה) של $X|T=t$. אז הצפיפות של T בהנתן $X=k$ היא $f_{T|X=k}(t) = \frac{P(X=k|T=t)}{P(X=k)}$.

אכן, $P\left(T \leq t | X=k\right) = \frac{P(X=k, T\leq t)}{P(X=k)} = \frac{P(X=k|T\leq t)P(T\leq t)}{P(X=k)} = \frac{\int_{-\infty}^{t} P(X=k|T=s)ds}{P(X=k)}$, והצפיפות מתקבלת מגזירה ביחס ל־t.

הערה 2.5.25 האינטגרל הבא שימושי מפעם לפעם; אפשר להוכיח אותו באינדוקציה על n.

$$\int_0^1 x^k(1-x)^{n-k}dx = \frac{k!(n-k)!}{(n+1)!}$$

תרגיל 2.5.26 (אמידה בשיטת בייס: משתנה בינומי בעל הסתברות לא ידועה) נניח שמאורע מתרחש בהסתברות קבועה אבל לא ידועה. יש הטוענים שכאשר ההסתברות אינה ידועה, הבחירה הטבעית היא להניח שיש לו התפלגות אחידה.

המודל המתאים הוא שהמשתנה $P \sim U[0,1]$ הוא פרמטר בעל התפלגות אחידה, וההתפלגות של X בהנתן P נתונה על־ידי $X \sim \mathrm{Bin}(n,P)$. כלומר, לכל p, $P\left(X=k|P=p\right)=\binom{n}{k}p^k(1-p)^{n-k}$.

1. הראה ש־X מתפלג באחידות על הערכים $0,\dots,n$. **רמז.** העזר באינטגרל $\int_0^1 \binom{n}{k}p^k(1-p)^{n-k}dp=\frac{1}{n+1}$.

2. כתוב את הצפיפות המותנית $f_{P|X=k}(p)$.

   פתרון. $f_{P|X=k}(p)=(n+1)\binom{n}{k}p^k(1-p)^{n-k}$.

3. הוכח ש־$\mathbf{E}(P|X=k)=\frac{k+1}{n+2}$ (עובדה זו קרויה **אומד בייסיאני** של P). **פתרון.** על־פי הערות 2.5.24 ו־2.5.25,

$$\begin{aligned}\mathbf{E}(P|X=k) &= \int_0^1 f_{P|X=k}(p)=\int_0^1 (n+1)\binom{n}{k}p^{k+1}(1-p)^{n-k}dp\\ &= \frac{(n+1)\binom{n}{k}}{\binom{n+1}{k+1}}\int_0^1\binom{n+1}{k+1}p^{k+1}(1-p)^{n-k}dp\\ &= \frac{(n+1)\binom{n}{k}}{(n+2)\binom{n+1}{k+1}}=\frac{k+1}{n+2}.\end{aligned}$$

4. שחקן משתתף בתוכנית הגרלות. מאחר שאף אחד מעשרת הכרטיסים שקנה לא זכה בפרס, הוא מעריך (על פי הסעיף הקודם) שהסיכוי לזכות הוא $\frac{1}{11}$. שותפו טוען שזה הסיכוי שלא לזכות בפרס הגדול דווקא (בהגרלה אפשר לזכות בחמישה פרסים שונים), ולכן הסיכוי לזכות *בפרס הגדול* הוא כ־$\frac{1}{11}$. לא יתכן ששניהם צודקים. מה דעתך?

תרגיל 2.5.27 במשחק טניס ההכרעה מתקבלת לפי רוב הנצחונות בחמש מערכות. בכל מערכה, הסיכוי של שחקן א' לנצח את שחקן ב' הוא P, באופן בלתי תלוי במערכות האחרות, כאשר $P \sim U[0,1]$. נניח ששחקן א' ניצח בשתי המערכות הראשונות. מה הסיכוי שלו לנצח במשחק כולו?

(לפי כתבה שהתפרסמה בניו־יורק טיימס, ראו *http://www.haaretz.co.il/sport/other/.premium-1.2330828*, "ב־94% מהמשחקים בשלושת הסיבובים הראשונים של הרולאן גארוס בשנה שעברה ניצח השחקן שהוביל $2-0$ במערכות". התשובה לתרגיל זה היא 95%.)

תרגיל 2.5.28 כמו בתרגיל 2.5.27, שני שחקנים משחקים $2n+1$ מערכות בלתי תלויות, והנצחון נקבע בשיטת הרוב. הסיכוי של השחקן הראשון לנצח בכל משחקון (באופן בלתי תלוי) הוא P, כאשר $P \sim U[0,1]$. מהם סיכויי השחקן לנצח במשחק (כלומר, לצבור רוב ב־$2n+1$ המערכות) אם הוא ניצח במערכה הראשונה?
פתרון. $\frac{3n+2}{4n+2}$.

וכעת דוגמא שבה טיפוסי המשתנים הפוכים:

תרגיל 2.5.29 (משתנה פואסוני בעל פרמטר מעריכי) נניח ש־$X \sim \mathrm{Exp}(\mu)$, ואילו $Y|X \sim \mathrm{Poi}(X)$. הוכח ש־$Y+1 \sim \mathrm{G}(\frac{1}{1+\mu})$. **הדרכה.** $P\left(Y=n\right) = \int_0^\infty \frac{e^{-x}x^n}{n!}\frac{1}{\mu}e^{-x/\mu}dx = \frac{1}{1+\mu}(\frac{\mu}{1+\mu})^n$.

2.5.7 סטטיסטיי הסדר

ראו תת־סעיף 2.6.1, שבו מוצגים סטטיסטיי הסדר של ההתפלגות האחידה.

2.5.8 טרנספורמציה של משתנה

המקרה החד־ממדי

אם $h:\mathbb{R}\rightarrow\mathbb{R}$ פונקציה הפיכה ו־$Y=h(X)$ אז הצפיפות של Y היא $f_Y(h(t)) = f_X(t)h'(t)^{-1}$.

תרגיל 2.5.30 הוכח את הטענה למקרה ש־h מונוטונית עולה. **הדרכה.** העזר בפונקציית ההצטברות.

דוגמא 2.5.31 $f_{aX}(t) = a^{-1}f_X(a^{-1}t)$.

דוגמא 2.5.32 $X \sim U[0,1]$. חשב את הצפיפות של $Y = -\log X$.

המקרה הדו־ממדי

תהי $h=(h_1,h_2)$ פונקציה הפיכה מתחום פתוח במישור לתחום פתוח אחר (תחום פתוח הוא קבוצה הכוללת, סביב כל נקודה, עיגול ברדיוס חיובי כלשהו; באופן טיפוסי זו קבוצה המוגדרת על־ידי אי־שוויונים מהצורה $\varphi(x,y)>0$). נניח שהנגזרות החלקיות קיימות, ונגדיר $J_h(t,s) = \det\begin{pmatrix} \frac{\partial h_1}{\partial x} & \frac{\partial h_1}{\partial y} \\ \frac{\partial h_2}{\partial x} & \frac{\partial h_2}{\partial y}\end{pmatrix}$, כשהנגזרות מחושבות בנקודה (x,y); זהו **היעקוביאן** של h.

כאשר מגדירים משתנים מקריים חדשים $(U,V)=h(X,Y)$, הצפיפות המשותפת היא $f_{U,V}(h(x,y))=f_{X,Y}(x,y)J_h(x,y)^{-1}$. הסיבה לכך היא כלל החלפת המשתנים, שאפשר לתאר באופן סכימטי כך:

$$\int_D f(h(u))du=\int_{h(D)} f(x)J^{-1}(x)dx.$$

דוגמא 2.5.33 נניח ש־X,Y משתנים מקריים בלתי תלויים בעלי צפיפויות f_X,f_Y. אפשר להגדיר $U=X+Y$ ו־$V=Y$; אז $J=1$, והצפיפות המשותפת של $X+Y,Y$ היא $f_{X+Y,Y}(u,v)=f_{X,Y}(u-v,v)=f_X(u-v)f_Y(v)$. הצפיפות השולית של $X+Y$ היא, לפי ההגדרה,

$$f_{X+Y}(u)=\int f_{X+Y,Y}(u,v)dv=\int f_X(u-v)f_Y(v)dv.$$

זוהי **הקונוולוציה** של f_X,f_Y. השווה לדוגמא 2.2.20.

דוגמא 2.5.34 $X,Y\sim U[0,1]$ בלתי תלויים; נגדיר $U=\cos(2\pi X)\sqrt{-2\log Y}$ ו־$V=\sin(2\pi X)\sqrt{-2\log Y}$. אז $J_h=\frac{-2\pi}{y}$ והרי $y=e^{-\frac{1}{2}(u^2+v^2)}$. לכן U,V בלתי תלויים; בהמשך נראה שהם מתפלגים התפלגות הנקראת "נורמלית".

2.6 התפלגויות רציפות חשובות

2.6.1 התפלגות אחידה

אומרים ש־X מתפלג התפלגות אחידה בקטע $[a,b]$ (וכותבים $X\sim U[a,b]$) אם $f_X(t)=\frac{1}{b-a}I_{[a,b]}(t)$. שימו לב שהסימון זהה לזה של ההתפלגות האחידה הבדידה, ויש לפענח לאיזו התפלגות מתכוונים מן ההקשר.

קל לחשב ש־$\mathbf{E}(X)=\frac{a+b}{2}$, ו־$\mathbf{V}(X)=\frac{(b-a)^2}{12}$.

טענה 2.6.1 אם X הוא משתנה מקרי בעל פונקציית הצטברות רציפה F, אז $F(X)$ מתפלג $U[0,1]$.

הוכחה. מכיוון ש־F מונוטונית עולה,

$$P(F(X)\leq a)=P(X\leq F^{-1}(a))=F(F^{-1}(a))=a;$$

כאן $F^{-1}(a)=\inf_t\{F(t)=a\}$. $\square$

טענה 2.6.2 תהי F פונקציית הצטברות רציפה. נניח ש־$U \sim U[0,1]$. אז $F^{-1}(U)$ בעל ההתפלגות F.

שילוב שתי הטענות מאפשר לעבור מכל התפלגות רציפה לכל התפלגות רציפה אחרת, תוצאה בעלת חשיבות מרכזית בהדמיה:

תרגיל 2.6.3 תהיינה F,G שתי פונקציות הצטברות רציפות. אם F היא פונקציית ההצטברות של משתנה מקרי X, אז G היא פונקציית ההצטברות של $G^{-1}F(X)$.

תרגיל 2.6.4 עיין במושגים "התפלגות אחידה על הרציונליים בקטע סגור", או "משתנה בעל התפלגות אחידה $U[0,1]$ המקבל רק ערכים רציונליים", לאור תרגיל 2.4.1.

סטטיסטיי הסדר

אם $X_1,\dots,X_n$ משתנים מקריים, **סטטיסטיי הסדר** הם אותם ערכים, לאחר סידור מחדש, כך ש־$X_{(1)} \le \cdots \le X_{(n)}$. למשל, $X_{(1)} = \min\{X_i\}$, $X_{(n)} = \max\{X_i\}$.

טענה 2.6.5 אם f_X היא פונקציית הצפיפות של המשתנים הבלתי תלויים $X_1,\dots,X_n$, אז פונקציית הצפיפות של סטטיסטי הסדר $X_{(k)}$ היא $n\binom{n-1}{k-1}f_X(t)F_X(t)^{k-1}(1-F_X(t))^{n-k}$.

הוכחה.

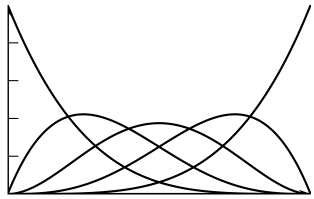
$$\begin{aligned}
P(X_{(k)} \le t) &= \sum_{A\subseteq\{1,\dots,n\},|A|\ge k} P(\forall i\in A: X_i\le t;\ \forall i\notin A: X_i>t) \\
&= \sum_{A\subseteq\{1,\dots,n\},|A|\ge k} F_X(t)^{|A|}(1-F_X(t))^{n-|A|} \\
&= \sum_{i=k}^{n}\binom{n}{i}F_X(t)^i(1-F_X(t))^{n-i}.
\end{aligned}$$

הנגזרת של פונקציה זו, לפי t, היא כנטען לעיל. □

תרגיל 2.6.6 1. נניח ש־$X_1,\dots,X_n \sim U[0,1]$. הראה שפונקצית הצפיפות של סטטיסטי הסדר $X_{(k)}$ היא $f_{X_{(k)}}(t) = n\binom{n-1}{k-1}t^{k-1}(1-t)^{n-k}$ (ראה איור 2.1).

2. הוכח ש־$\mathbf{E}(X_{(k)}) = \frac{k}{n+1}$. **הדרכה.** העזר באינטגרל שבתרגיל 2.5.25.

[הטענה $\mathbf{E}(X_{(n)}) = \frac{n}{n+1}$ היתה נושא עבודת הדוקטור של ניקולאוס ברנולי ה־I, ב־1709].



איור 2.1: צפיפות סטטיסטיי הסדר עבור $X_1,\dots,X_5 \sim U[0,1]$.

3. בדוק ש־$\mathbf{V}(X_{(k)}) = \frac{k(n+1-k)}{(n+1)^2(n+2)}$.

*4. מה מקדם המתאם בין $X_{(i)}, X_{(j)}$, כאשר $i<j$?

תרגיל 2.6.7 הכלל את טענה 2.6.5: נניח ש־$X_1,\dots,X_{n-1}$ בעלי התפלגות F, ו־X_n בעל התפלגות G, עם פונקציות צפיפות f,g בהתאמה. סטטיסטיי הסדר מוגדרים באותו אופן. הראה שהצפיפות של $X_{(k)}$ היא

$$f_{X_{(k)}}(t) = \binom{n-1}{k-1} f(t)F(t)^{k-1}(1-F(t))^{n-k}\left[\frac{g(t)}{f(t)} + (k-1)\frac{G(t)}{F(t)} + (n-k)\frac{1-G(t)}{1-F(t)}\right].$$

תרגיל 2.6.8 נניח ש־$X_1,\dots,X_{n-1} \sim U[0,1]$, ו־$X_n \sim U[0,a]$ כאשר $0<a<1$ קבוע. הראה שפונקציית הצפיפות של המקסימום $X_{(n)} = \max\{X_1,\dots,X_n\}$ היא $f_{X_{(n)}}(t) = \begin{cases} nt^{n-1}\frac{1}{a} & 0<t<a \\ (n-1)t^{n-2} & a<t<1 \end{cases}$. חשב את התוחלת $\mathbf{E}(X_{(n)}) = \frac{n^2-1+a^n}{n(n+1)}$.

הציון הסופי בתחרות בינלאומית לסטודנטים הוא הציון הטוב ביותר שהשיג אחד החברים בה, ועוד הציון הממוצע. בנבחרת שותפים $n-1$ חברים שווי יכולת (שהציון של כל אחד מהם מתפלג $U[0,1]$). יש מועמד נוסף, חלש יותר, שהתפלגות הציונים שלו היא $U[0,a]$, כאשר $a<1$. הראה שכדאי לצרף את המועמד הנוסף לנבחרת אם ורק אם $2a^n > (n+1)(1-2a)$. (בתחרות ה־IMO הציון הוא סכום שלוש התוצאות הטובות ביותר ועוד הממוצע).

תרגיל 2.6.9 בוחרים שתי נקודות על מקל (באופן אקראי ובלתי תלוי, בהתפלגות אחידה), ושוברים אותו בהן לשלושה חלקים. מה הסיכוי שאפשר להרכיב מן החלקים האלה משולש? (ואם שוברים את המקל ב־n נקודות, מה הסיכוי שאפשר להרכיב מן החלקים מצולע?)

תרגיל 2.6.10 1. בוחרים שתי נקודות על מקל (באופן בלתי תלוי ובהתפלגות אחידה). אחר־כך מסמנים על המקל $n-1$ נקודות במרחקים שווים זו מזו ומקצות המקל, ושוברים אותו ל־n חלקים לפי החלוקה שהתקבלה. מה הסיכוי ששתי הנקודות שייכות לאותו חלק?

2. בוחרים שתי נקודות על מקל. אחר־כך מסמנים על המקל $n-1$ נקודות, גם כן באקראי ובאופן בלתי תלוי. שוב שוברים את המקל לפי החלוקה שהתקבלה. מה הסיכוי ששתי הנקודות שייכות לאותו חלק?

תרגיל 2.6.11 יונתן יכול להגיע לעבודה בכל אחד משני אוטובוסים. זמן ההמתנה הממוצע לכל קו הוא 30 דקות. מה יהיה זמן ההמתנה הממוצע לקו הראשון מבין השניים? ואם זמן ההמתנה לקו אחד הוא a, ולשני b?

תרגיל 2.6.12 אלעד חוזר מהעבודה באחד משני אוטובוסים, שיש להם לוח זמנים קבוע: הם יוצאים לדרך פעם בשעה. הוא מגיע אל התחנה בזמן אקראי (בעל התפלגות אחידה), ועולה על האוטובוס הראשון שעוצר בתחנה. כיצד יתכן שהוא עולה על קו 15 רק 10% מהפעמים, ועל קו 52 בשאר הזמן?

תרגיל 2.6.13 נניח ש־$X_1,\dots,X_n$ הם משתנים מקריים בלתי תלויים בעלי פונקציית הצטברות F. לפי תרגיל 2.6.1, $F(X_1),\dots,F(X_n) \sim U[0,1]$. הסק מתרגיל 2.6.6 שהערך המקסימלי מבין ה־$F(X_i)$ הוא בקירוב $1-\frac{1}{n}$. פירושו של דבר הוא שבכל התפלגות, אם מגרילים n ערכים, סביר שאחד מהם יהיה קיצוני במידה כזו שהסיכויים נגדו הם בערך $n:1$.

תרגיל 2.6.14 משתנים מקריים רציפים $X_1,X_2,\dots$ הם בלתי תלויים ושווי התפלגות. מגדירים $N=n$ כאשר n הוא הקטן ביותר כך שעבורו $X_n < X_{n-1}$. מה התוחלת של N?

הדרכה. המאורע $\{N>n\}$ שווה למאורע $\{X_1<\dots<X_n\}$. לכן $\mathcal{P}(N>n)=\frac{1}{n!}$. לפי טענה 2.3.57, $\mathbf{E}(N)=\sum_{n=0}^{\infty}\frac{1}{n!}=e$.

תרגיל 2.6.15 בספורט שוויוני (מקצוע תחרותי חדש שתוכנן למנוע יתרון לא הוגן מבעלי הכשרון), כל שחקן מגריל משתנה מקרי בעל התפלגות $U[0,1]$. התוצאות מוכנסות מיד למחשב ונשמרות בסוד. המחשב מדווח בכל פעם שמתחרה קובע שיא עולם חדש, כלומר כאשר הוא משיג תוצאה טובה יותר מכל קודמיו. עד היום נרשמו m תוצאות. נסמן ב־N את מספר התוצאות שיירשמו עד לשיא העולם הבא ($N\geq 1$). הסבר מדוע $P\left(N\geq k\right)=\frac{m}{m+k-1}$. הוכח ש־$\mathbf{E}(N)=\infty$.

2.6.2 התפלגות מעריכית

למשתנה X יש **התפלגות מעריכית**, עם הפרמטר λ, אם פונקציית הצפיפות שלו היא $f_X(x) = \frac{1}{\lambda}e^{-x/\lambda}$. במקרה כזה מסמנים

$$X \sim \mathrm{Exp}(\lambda).$$

אפשר לחשב ש־$P(X > t) = e^{-t/\lambda}$. הזמן t שעבורו $P(X > t) = 0.5$ נקרא **זמן מחצית החיים**, ושווה ל־$\log(2)\lambda$.

התוחלת של $X \sim \mathrm{Exp}(\lambda)$ היא $\mathbf{E}(X) = \lambda$, והשונות $\mathbf{V}(X) = \lambda^2$.

טענה 2.6.16 תכונת חוסר הזכרון של ההתפלגות המעריכית: נניח ש־$X \sim \mathrm{Exp}(\lambda)$. אז $(X - a)|(X > a)$ מתפלג $\mathrm{Exp}(\lambda)$.

תרגיל 2.6.17 נניח שהזמן בין רעידות אדמה חזקות באגן הירדן שעל השבר הסורי־אפריקאי, מתפלג מעריכית, עם תוחלת 100 שנה. הרעידה האחרונה היתה לפני כ־80 שנה. מתי צפויה (בתוחלת) הרעידה הבאה? (אגב: במציאות, הזמן בין רעידות אדמה אינו לגמרי חסר זכרון, משום שמדובר בשחרור פתאומי של אנרגיה הנאצרת באופן מדורג).

תרגיל 2.6.18 נניח ש־$X \sim \mathrm{Exp}(1)$. מה ההתפלגות של e^{-X}? השווה לתרגיל 2.5.32.

תרגיל 2.6.19 קו אוטובוס יוצא מן התחנה מדי 30 דקות. לפיכך, זמן ההמתנה של נוסע המגיע באקראי לתחנה מתפלג אחיד והוא יחכה בממוצע 15 דקות.

לאחר זמן משנים את מתכונת הנסיעה, כך שבכל פעם שאוטובוס יוצא לדרכו, מפעילים שעון עצר מעריכי ומשחררים את האוטובוס הבא לפי הערך המתקבל (שהוא בעל התפלגות מעריכית עם תוחלת 30). מספר האוטובוסים הממוצע ביממה יהיה כפי שהיה בעבר. מה תהיה תוחלת זמן ההמתנה של הלקוחות לפי השיטה החדשה? כיצד זה יתכן? (הסבר את הקשר לתרגיל 2.5.16).

תרגיל 2.6.20 $X, Y \sim \mathrm{Exp}(\theta)$. חשב את $P(X < \alpha Y)$.

תרגיל 2.6.21 מדען צופה במטר מטאורים, שבו יש (בתוחלת) α הופעות בדקה. התצפית נפסקת כשנשבר העפרון, שמשך החיים שלו מעריכי עם תוחלת μ דקות. נסמן ב־X את משך זמן הצפיה וב־Y את מספר המטאורים שנצפו; כך $X \sim \mathrm{Exp}(\mu)$ ו־$Y|X \sim \mathrm{Poi}(\alpha X)$.

1. הראה שהתוחלת של Y היא $\mathbf{E}(Y) = \alpha\lambda$.
2. הראה ש־$Y + 1 \sim \mathrm{G}(\frac{\alpha\lambda}{1+\alpha\lambda})$. **הדרכה**. ראה תרגיל 2.5.26.

3. נניח ששני מדענים עובדים במקביל (וצופים כל אחד בחלקת שמיים אחרת).
הראה ש־$P\left(Y_1=0, Y_2=k|X_1>X_2\right)=\frac{1}{(\alpha\lambda+1)^2}\left(\frac{\alpha\lambda}{2(\alpha\lambda+1)}\right)^k$.
הסק ש־$P\left(Y_1=0|X_1>X_2\right)=\frac{2}{(\alpha\lambda+1)(\alpha\lambda+2)}$, בעוד ש־$P\left(Y_1=0\right)=\frac{1}{\alpha\lambda+1}$.
4. הראה ש־$P\left(Y_1=k, Y_2=0|X_1>X_2\right)=\frac{2-2^{-k}}{(\alpha\lambda+1)^2}\left(\frac{\alpha\lambda}{\alpha\lambda+1}\right)^k$.
הסק ש־$P\left(Y_2=0|X_1>X_2\right)=\frac{2}{\alpha\lambda+2}$.

סדרת משתנים מעריכיים

טענה 2.6.22 נניח שהמשתנים X_1, X_2 מתפלגים מעריכית, $X_1 \sim \mathrm{Exp}(\lambda_1)$ ו־$X_2 \sim \mathrm{Exp}(\lambda_2)$, והם בלתי תלויים. אז $\min(X_1, X_2) \sim \mathrm{Exp}(\lambda')$, כאשר $\frac{1}{\lambda'}=\frac{1}{\lambda_1}+\frac{1}{\lambda_2}$.

מסקנה 2.6.23 נניח ש־$X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$ בלתי תלויים. אז סטטיסטי הסדר $Y_1 = \min(X_1, \ldots, X_n) \sim \mathrm{Exp}(\lambda/n)$. בפרט, $\mathbf{E}(Y_1)=\frac{\lambda}{n}$.

נניח ש־$X_1, \ldots, X_n \sim \mathrm{Exp}(\lambda)$ בלתי תלויים. מתכונת חוסר הזכרון נובע שבהנתן Y_i, Y_{i+1} הוא המינימום של $n-i$ משתנים מעריכיים, ולכן $Y_{i+1}-Y_i \sim \mathrm{Exp}((n-i)\lambda)$. מכאן ש־$\mathbf{E}(Y_{i+1}-Y_i)=\frac{\lambda}{n-i}$.

מסקנה 2.6.24 $\mathbf{E}(Y_i)=\left(\frac{1}{n}+\cdots+\frac{1}{n-i+1}\right)\lambda$.

בפרט, $\mathbf{E}(\max(X_1, \ldots, X_n))=(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n})\lambda$.

תרגיל 2.6.25 בכד ענק יש כדורים רבים (מספרם סופי), שמהם p שחורים והשאר לבנים. מסננים את הכד באופן הבא: דוגמים כדור; אם הוא לבן מחזירים אותו לכד, ואם הוא שחור זורקים אותו החוצה. כמה פעמים יש לחזור על התהליך עד שפרופורציית הכדורים השחורים יורדת ל־p'?

התפלגות גמא

אומרים שמשתנה מקרי Y הוא בעל **התפלגות גמא** עם פרמטרים n, λ אם פונקציית הצפיפות שלו היא $f_Y(y)=\frac{y^{n-1}}{(n-1)!\lambda^n}e^{-y/\lambda}$. מסמנים $Y \sim \Gamma(n, \lambda)$.

טענה 2.6.26 פונקציית ההצטברות של משתנה $Y \sim \Gamma(n, \lambda)$ היא

$$F_Y(y)=P\left(Y \leq y\right)=\sum_{i=n}^{\infty}\frac{x^i}{i!\lambda^i}e^{-x/\lambda}.$$

תרגיל 2.6.27 ההתפלגות $\Gamma(1, \lambda)$ אינה אלא ההתפלגות המעריכית $\mathrm{Exp}(\lambda)$.

טענה 2.6.28 לכל n, m, אם $Y \sim \Gamma(n, \lambda)$ ו־$Y' \sim \Gamma(m, \lambda)$ הם משתנים מקריים בלתי תלויים, אז $Y + Y' \sim \Gamma(n + m, \lambda)$.

הוכחה. קונוולוציה. □

מסקנה 2.6.29 נניח ש־$X_1, X_2, \dots \sim \mathrm{Exp}(\lambda)$ הם משתנים בלתי תלויים. לכל n, הסכום $S_n = X_1 + \cdots + X_n$ מתפלג $\Gamma(n, \lambda)$.

מסקנה 2.6.30 אם $Y \sim \Gamma(n, \lambda)$ אז $\mathbf{E}(Y) = n\lambda$ ו־$\mathbf{V}(Y) = n\lambda^2$.

הקשר להתפלגות פואסון

נניח ש־$X_1, \dots \sim \mathrm{Exp}(\lambda)$, בלתי תלויים. לכל n, נסמן כמקודם $S_n = X_1 + \cdots + X_n$. עבור ערך קבוע t, נסמן $N_t = \max_n \{S_n < t\}$. אם התפלגות הזמן בין תקלות היא מעריכית, אז N_t סופר כמה תקלות יהיו בפרק זמן נתון.

טענה 2.6.31 $N_t \sim \mathrm{Poi}(t/\lambda)$.

הוכחה. את $P(N_t \geq n) = P(S_n \leq t)$ חישבנו בטענה 2.6.26. מכאן נובע ש־

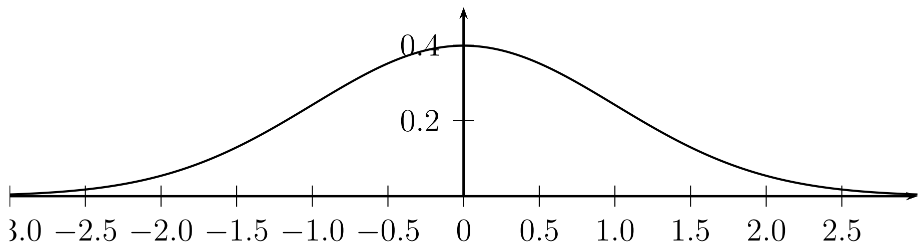
$$P(N_t = n) = P(N_t \geq n + 1) - P(N_t \geq n) = \frac{(t/\lambda)^n e^{-t/\lambda}}{n!}.$$

□

תרגיל 2.6.32 העזר בחוסר הזכרון של ההתפלגות המעריכית כדי לחלץ מטענה 2.6.31 הוכחה חדשה לתרגיל 2.3.27(1).

הכיוון ההפוך: אם מפזרים מספר פואסוני של נקודות בקטע נתון, המרחקים מתפלגים מעריכית.

תרגיל 2.6.33 נניח שמספר האירועים בקטע בכל אורך מתפלג פואסונית (עם פרמטר התלוי רק באורך הקטע). הוכח שהתלות באורך הקטע היא ליניארית.



איור 2.2: גרף ההתפלגות הנורמלית הסטנדרטית

2.6.3 התפלגות נורמלית

הגדרת התפלגות נורמלית סטנדרטית $N(0,1)$: $f_Z(t) = \frac{1}{c}e^{-t^2/2}$, כשאיננו יודעים בשלב זה את c.

ציטוט 2.6.34 "ואכן, מה לגאומטריה של המעגל ולסטטיסטיקה של בני אדם? ככל שידיעתי מגעת, איש לא הסביר זאת מעולם" (פרק ביוקנן, לשעבר עורך *Nature*, "האטום החברתי" עמ' 68).

נניח ש־ $T \sim U[0,2\pi]$ ו־ $f_R(t) = e^{-t^2/2}t$ עבור $t \geq 0$, בלתי תלויים. נגדיר $X = R\cos(T)$ ו־ $Y = R\sin(T)$ (היעקוביאן הוא R). מהחלפת המשתנים יוצא $X, Y \sim N(0,1)$, כשהם בלתי תלויים, ו־ $c = \sqrt{2\pi}$.

הגדרה 2.6.35 משתנה שהצפיפות שלו $f(t) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(t-\mu)^2}{2\sigma^2}}$ הוא משתנה נורמלי, בעל התפלגות $N(\mu,\sigma^2)$. כל אחת מן ההתפלגויות במשפחה דו־פרמטרית זו נקראת **התפלגות נורמלית**.

טענה 2.6.36 אם $Z \sim N(0,1)$ אז $X = \mu + \sigma Z$ הוא בעל ההתפלגות $N(\mu,\sigma^2)$.

טענה 2.6.37 נניח ש־$Z \sim N(0,1)$. הראה ש־$\mathbf{E}(Z) = 0$ ו־$\mathbf{V}(Z) = 1$.

תרגיל 2.6.38 נניח $Z \sim N(0,1)$. הראה ש־$\mathbf{E}(|Z|) = \sqrt{2/\pi}$.

לפי טענה 2.6.36, אם $X \sim N(\mu,\sigma^2)$, אז $\mathbf{E}(X) = \mu$ ו־$\mathbf{V}(X) = \sigma^2$. בפרט, התפלגות נורמלית מוגדרת על־פי התוחלת והשונות שלה.

תרגיל 2.6.39 X מתפלג נורמלית. ידוע ש־$P(X < 0.9) = 0.9$, $P(X < 0.7) = 0.7$. מצא את $P(X < 0.5)$ [תשובה: 40.8%].

תרגיל 2.6.40 ההישגים של פרחי־טיס בטיסת אימון הם בעלי התפלגות נורמלית סטנדרטית, ובלתי תלויים זה בזה. מדריך הטיסה משבח את החניך אם הישגיו בטיסה מסויימת היו a סטיות תקן מעל לממוצע, ונוזף בו אם היו a סטיות תקן מתחת לממוצע (כאן $a > 0$ קבוע). חניך טס שתי טיסות.

1. על־פי נתוני השאלה, מהם הסיכויים לכך שיחול שיפור בהישגים (הטיסה השניה טובה יותר מן הראשונה)?
2. הערך את הסיכוי לשיפור אם המדריך שיבח את החניך בפעם הראשונה; הערך את הסיכוי לשיפור אם המדריך נזף בו.
3. מה דעתך על המשוב הפנימי של מדריך הטיסה, האומר לעצמו 'בכל פעם שאני משבח את החניך, הישגיו יורדים. בכל פעם שאני נוזף, ההישגים משתפרים. מכאן שהשבחים מזיקים והנזיפות מועילות'?

טבלאות של התפלגות נורמלית

האינטגרל $\int e^{-x^2}$ אינו אלמנטרי, ולכן יש להעזר בטבלאות לחישוב הערכים שלו.

$$\begin{aligned} P(Z > z) &= \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &\leq \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-zt/2} dt \\ &\leq \sqrt{\frac{2}{\pi}} \frac{1}{z} e^{-z^2/2} \end{aligned}$$

קירוב שימושי לערכים גדולים: $P(Z > z) \sim \frac{1}{z+\frac{1}{2}} e^{-z^2/2}$.

תרגיל 2.6.41 המשך את תרגיל 2.6.13: הראה שאם $X_1, \dots, X_n \sim N(0,1)$ אז תוחלת הערך המקסימלי של X_i היא, בקירוב גס, $\mathbf{E}(\max X_i) \approx \sqrt{2 \log n}$. למשל, אם $n \approx e^{50}$, אז $\max X_i \approx 10$.

פונקציית ההצטברות של התפלגות *Gumbel* הסנדרטית היא $F_G(x) = e^{-e^{-x}}$. אם $X_1, \dots, X_n \sim N(0,1)$ בלתי תלויים, אז עבור הקבועים $b_n = \Phi^{-1}(1 - 1/n)$ ו־ $a_n = 1/(n\phi(b_n))$, התפלגות המקסימום $X_{(n)}$ נתונה על־ידי

$$P\left(X_{(n)} \leq a_n x + b_n\right) \to F_G(x),$$

כאשר $\Phi(\cdot)$ היא פונקציית ההצטברות של ההתפלגות הנורמלית, ו־$\phi(\cdot)$ פונקציית הצפיפות שלה.

תרגיל 2.6.42 בספטמבר *2014* הצביעו תושבי סקוטלנד כדי לקבוע האם יפרדו מעל הממלכה המאוחדת הבריטית. "בסקוטלנד חיים כ־7000 יהודים [מתוך *5000000* תושבים], כך שגם במשאל שעשוי להיות מאד צמוד, הם לא צפויים להכריע לשום כיוון" (אנשיל פפר, "הארץ", *18/9/2014*, http://www.haaretz.co.il/news/world/europe/.premium-1.2437985). נניח שלחצי מהתושבים זכות בחירה. נניח שהתושבים הלא־יהודים מצביעים באקראי ובהסתברות חצי. הראה שאם יצביעו כאיש אחד, הסיכוי של היהודים לקבוע את תוצאות ההצבעה גדול מ־99%.

תרגיל 2.6.43 חזור לציטוט 1.2.13. אם ציון מנת המשכל מתפלג נורמלית עם תוחלת 100, מהי סטיית התקן?

התפלגות נורמלית דו־ממדית

טענה 2.6.44 משפחת ההתפלגויות הנורמליות סגורה לחיבור: אם X,Y שני משתנים נורמליים בלתי תלויים, אז גם $X+Y$ מתפלג נורמלית. ביתר פירוט, אם $X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$ והם בלתי תלויים, אז $X+Y \sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

הוכחה. הפעל את ההתמרה $U=cX+sY$, $V=-sX+cY$ כאשר $s^2+c^2=1$. $\square$

תרגיל 2.6.45 נניח ש־$X \sim N(0,\sigma^2)$ ובהנתן X, $Y|X \sim N(\alpha\pi^2X,\pi^2)$, עבור קבועים מתאימים α,σ,π. נסמן $\theta = 1+\alpha^2\pi^2\sigma^2$, $\sigma'^2=\theta\pi^2$, $\pi'^2=\theta^{-1}\sigma^2$ (שים לב ש־$\pi\sigma=\pi'\sigma'$). העזר בתרגיל 2.5.18 כדי להראות ש־Y מתפלג נורמלית; הסק מנוסחת השונות החוזרת 2.2.42 (או מחישוב האינטגרל) ש־$Y \sim N(0,\sigma'^2)$. הפעל שוב את תרגיל 2.5.18 כדי להראות שבהנתן Y, $X|Y \sim N(\alpha\pi'^2Y,\pi'^2)$.

תרגיל 2.6.46 במוח יש 25 אזורים שאפשר למצוא בהם הבדל בין גברים לנשים; נסמן ב־$X_1,\dots,X_{25}$ את גדלי האזורים אצל נבדק מסויים. לאחר נירמול, גודלו של כל אזור מתפלג $N(0.2,1)$ אצל גברים ו־$N(-0.2,1)$ אצל נשים. הראה אצל גברים הסכום $X=\sum X_i$ חיובי בסיכוי של כ־99%; ואצל נשים, להיפך. הראה שהסיכוי שלכך שכל $X_i>0$ אצל גבר נתון הוא פחות מאחד למליון. העזר בנתונים (מפוברקים) אלה כדי לכתוב שתי פסקאות: אחת המוכיחה את קיומם של "מוח גברי" ו"מוח נשי", ואחת הלועגת לפתאים המאמינים בהבדלים כאלה, ומראה שהם חסרי ממשות ביולוגית. איזו פסקה נכונה?

(וראה: http://www.haaretz.co.il/news/science/.premium-1.2788500)

2.6.4 התפלגויות נוספות

הערה 2.6.47 שמה של התפלגות גמא מגיע מ**פונקציית גמא** $\Gamma(z)=\int_0^\infty t^{z-1}e^{-t}\,dt$, המקיימת את המשוואה הפונקציונלית $\Gamma(z+1)=z\cdot\Gamma(z)$. מכיוון ש־$\Gamma(1)=1$, מתקבל $\Gamma(n)=(n-1)!$ לכל n שלם. בנוסף לזה $\Gamma(\frac{1}{2})=\sqrt{\pi}$, ולכן עבור n איזוגי $\Gamma(n/2)=2^{-\frac{n-1}{2}}1\cdot3\cdot5\cdots(n-2)\sqrt{\pi}$.

1. נניח ש־$N(0,1) \sim Z_1,\dots,Z_n$ והם בלתי תלויים. לסכום הריבועים $W = Z_1^2 + \dots + Z_n^2$ יש התפלגות, הקרויה **התפלגות חי־בריבוע עם n דרגות חופש**, ומסומנת ב־χ_n^2. פונקציית הצפיפות היא $\frac{x^{n/2-1}e^{-x/2}}{\Gamma(n/2)2^{n/2}}$, כמו בהתפלגות $\Gamma(\frac{n}{2},2)$. התוחלת של $W \sim \chi_n^2$ היא $\mathbf{E}(W) = n$, שונות היא $\mathbf{V}(W) = 2n$.

הערה 2.6.48 פונקציית הצפיפות של התפלגות χ_1^2 היא $\frac{x^{-1/2}e^{-x/2}}{\sqrt{2\pi}}$, שאינה חסומה.

פונקציית הצפיפות של התפלגות χ_2^2 היא $\frac{e^{-x/2}}{2}$, הלוא היא ההתפלגות המעריכית עם תוחלת 2.

2. אם $Z \sim N(0,1)$ ו־$W \sim \chi_n^2$ בלתי תלויים, אז $T = \frac{Z}{\sqrt{W/n}}$ הוא בעל התפלגות הנקראת **התפלגות t של סטודנט**, עם n דרגות חופש ('סטודנט' היא הכינוי שאימץ לעצמו הסטטיסטיקאי ויליאם סילי גוסט, שפרסם את ההתפלגות ב־1908).

פונקציית הצפיפות של ההתפלגות הזו היא

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\,\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}.$$

כאשר n גדול (למשל $n = 30$), התפלגות זו קרובה להתפלגות הנורמלית. התוחלת היא $\mathbf{E}(T) = 0$ (כאשר $n > 1$; ב־ $n = 1$ אין תוחלת: האינטגרל אינו מתכנס) והשונות $\mathbf{V}(T) = \frac{n}{n-2}$.

3. להתפלגות היחס $X = \frac{U/n}{V/m}$ כאשר $U \sim \chi_n^2$ ו־$V \sim \chi_m^2$ קוראים **התפלגות** F, ומסמנים $X \sim F_{n,m}$. פונקציית הצפיפות היא

$$\frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}(\frac{n}{m})^{\frac{n}{2}}x^{\frac{n}{2}-1}(1+\frac{n}{m}\,x)^{-\frac{n+m}{2}}.$$

התוחלת היא $\mathbf{E}(X) = \mathbf{E}(\frac{1}{n}U)\mathbf{E}(m\frac{1}{V}) = \frac{m}{m-2}$. השונות היא $\mathbf{V}(X) = \frac{2m^2(m+n-2)}{n(m-2)^2(m-4)}$.

תרגיל 2.6.49 $T \sim t_n$ אם ורק אם $T^2 \sim F_{1,n}$.

תרגיל 2.6.50 $X \sim F_{n,m}$ אם ורק אם $\frac{1}{X} \sim F_{m,n}$.

תרגיל 2.6.51 בדוק שפונקציית הצפיפות של היחס $X = \frac{Z_1^2}{Z_2^2}$ כאשר $Z_1, Z_2 \sim N(0,1)$ היא $f_X(x) = \frac{1}{\pi\sqrt{x}(1+x)}$. הראה ש־$P\left(|Z_1| < a\,|Z_2|\right) = 2\arctan(\sqrt{a})$.

תרגיל 2.6.52 במדינה מסויימת נערכות בחירות מדי כמה שנים. כדי להבטיח הכרעה דמוקרטית באופן מלא, כל תושב בוחר באיזה משני המועמדים הוא תומך על־פי תוצאת הטלת מטבע (הוגן), המתבצעת זמן מה לפני הבחירות. בבוקר יום הבחירות מתרחש אירוע הגורם לכל תושב להפוך את החלטתו בהסתברות ϵ (ללא תלות בהחלטה הקודמת). העזר בתרגיל 2.6.51 כדי להראות שהסיכוי שתוצאת הבחירות תתהפך כתוצאה מן האירוע הוא $2\arctan(\sqrt{\frac{\epsilon}{1-\epsilon}}) \approx 2\epsilon$.

2.7 חסמים

2.7.1 אי־שוויון מרקוב

טענה 2.7.1 **יהי X משתנה מקרי חיובי בעל תוחלת μ. אז לכל a, $P(X \geq a\mu) \leq \frac{1}{a}$.** שוויון מתקבל (כאשר $a \geq 1$) רק עבור המשתנה שעבורו $P(X=0) = 1 - \frac{1}{a\mu}$ ו־$P(X=a\mu) = \frac{1}{a\mu}$.

תרגיל 2.7.2 יהי X משתנה מקרי חיובי. הראה ש־$P(X \geq 1) \leq \mathbf{E}(X)$.

תרגיל 2.7.3 תהי $X_1, X_2, \dots$ סדרת משתנים מקריים חיוביים עם תוחלת קבועה $\mathbf{E}(X_n) = \mu$. הראה ש־$P(\lim X_n = \infty) = 0$. **הדרכה.** נסמן $\epsilon = P(\lim X_n = \infty)$. בפרט לכל k קיים n כך ש־$\epsilon \leq P(X_n > k) \leq \frac{\mu}{k}$, ואם $\epsilon > 0$ אפשר לבחור $k > \mu/\epsilon$ ולקבל סתירה.

2.7.2 אי־שוויון צ'ביצ'ב

טענה 2.7.4 **לכל משתנה מקרי X (בעל תוחלת μ ושונות σ^2), ולכל קבוע k מתקיים**

$$P(|X-\mu| \geq k\sigma) \leq k^{-2}.$$

הוכחה. קח $Y = |X-\mu|$ באי־שוויון מרקוב. □

בפרט, אם $\mathbf{E}(X) = 0$ ו־$\mathbf{V}(X) = 1$, אז לכל k מתקיים $P(|X| \geq k) \leq k^{-2}$. להלן גרסה חד־צדדית של אי־השוויון הזה:

טענה 2.7.5 (אי־שוויון קנטלי) יהי X משתנה מקרי עם $\mathbf{E}(X) = 0$ ו־$\mathbf{V}(X) = 1$. לכל $k \geq 0$ מתקיים $P(X \geq k) \leq \frac{1}{k^2+1}$.

הוכחה. לכל $t \geq -k$ מתקיים לפי אי־שוויון מרקוב

$$\begin{aligned} P\left(X \geq k\right) &= P\left(\frac{X+t}{k+t} \geq 1\right) \\ &\leq P\left((\frac{X+t}{k+t})^2 \geq 1\right) \\ &\leq \mathbf{E}\left((\frac{X+t}{k+t})^2\right) = \frac{1+k^{-2}}{(k+t)^2} \end{aligned}$$

והתוצאה מתקבלת מבחירת $t = k^{-1}$, המביאה את אגף ימין למינימום. $\square$

תרגיל 2.7.6 לכל משתנה מקרי בעל תוחלת μ ושונות σ^2, ולכל $k > 0$, $P\left(X \geq \mu + k\right) \leq \frac{\sigma^2}{\sigma^2+k^2}$.

2.7.3 אי־שוויון צ'רנוף

יהי $X \sim \text{Bin}(m,p)$ משתנה בינומי. אי־שוויון צ'ביצ'ב נותן את החסם הריבועי

$$P\left(\left|\frac{X}{m} - p\right| \geq \epsilon\right) \leq \frac{pq}{m\epsilon^2}.$$

אי־שוויון צ'רנוב מספק שיפור מהותי של החסם הזה.
נסמן

$$D(x||y) = x\log\frac{x}{y} + (1-x)\log\frac{1-x}{1-y};$$

זוהי **סטיית** **Kullback-Leibler** בין ההתפלגויות $b(x)$ ו־$b(y)$.

הערה 2.7.7 לכל x, $D(x||x) = 0$. לכל $0 < x,y < 1$, $0 \geq D(x||y)$. קירוב טיילור סביב $\epsilon = 0$ נותן $D(p+\epsilon||p) \approx \frac{\epsilon^2}{2pq}$.

טענה 2.7.8 יהי $X \sim \text{Bin}(m,p)$ משתנה בינומי. לכל $\varepsilon > 0$,

$$P\left(\frac{X}{m} \geq p + \varepsilon\right) \leq e^{-D(p+\varepsilon||p)m}$$

ו־

$$P\left(\frac{X}{m} \leq p - \varepsilon\right) \leq e^{-D(p-\varepsilon||p)m}.$$

בקירוב, פירושו של דבר הוא ש־$P\left(\left|\frac{X}{m}-p\right|>\epsilon\right)$ חסום על־ידי $e^{-\epsilon^2 m/2pq}$; זו החלפה של החסם $1/k$ שקיבלנו מאי־שוויון צ'ביצ'ב בחסם $e^{-k/2}$.

הוכחה. אפשר להניח ש־$X = X_1+\cdots+X_m$ כאשר $X_1,\dots,X_m \sim b(p)$ והם בלתי תלויים. נסמן $p'=p+\epsilon$, $q'=1-p'$. לכל $\lambda>0$, $\frac{X}{m}\geq p'$ אם ורק אם $e^X \geq e^{mp'}$ אם ורק אם $\prod e^{X_i}\geq e^{mp'}$, אם ורק אם $\prod e^{\lambda X_i}\geq e^{\lambda mp'}$. המשתנה $\prod e^{\lambda X_i}$ חיובי, ולפי אי־שוויון מרקוב, $P\left(\prod e^{\lambda X_i}\geq e^{\lambda mp'}\right) \leq \frac{E[\prod e^{\lambda X_i}]}{e^{\lambda mp'}} = [\frac{E[e^{\lambda X_i}]}{e^{\lambda p'}}]^m = [\frac{pe^{\lambda}+q}{e^{\lambda p'}}]^m = [pe^{q'\lambda}+qe^{-p'\lambda}]^m$, משום ש־$P\left(X_i=1\right) = p$ ו־$P\left(X_i=0\right) = q$. נציב $\lambda = \log(\frac{qp'}{q'p})$ (זו נקודת המינימום של הפונקציה $pe^{q'\lambda}+qe^{-p'\lambda}$), ונקבל $P\left(\frac{X}{m}\geq p+\varepsilon\right) \leq \frac{p}{p'}(\frac{p'q}{pq'})^{mq'} = e^{-D(p+\varepsilon\|p)m}$.

אי־השוויון השני מתקבל מהחלפת המשתנים X_i ב־$X_i' = 1-X_i$ והפעלת אותו חסם. $\square$

2.8 פונקציה יוצרת מומנטים

2.8.1 מומנטים

יהי X משתנה מקרי. התוחלות $\mathbf{E}(X^n)$ נקראות **מומנטים** של X. כפי שכבר ראינו, התוחלת $\mu=\mathbf{E}(X)$ והשונות $\mathbf{E}((X-\mu)^2)$ הם בין התכונות החשובות ביותר של כל התפלגות. המומנטים $\mathbf{E}((X-\mu)^n)$ נקראים **מומנטים מרכזיים**.

זו הזדמנות להזכיר שהמומנטים לא בהכרח קיימים: המומנט ה־n־י קיים אם האינטגרל המגדיר את התוחלת המתאימה מתכנס.

טענה 2.8.1 אם למשתנה מקרי X בעל צפיפות f יש מומנט מסדר n, אז יש לו כל המומנטים מסדרים נמוכים יותר.

הוכחה. **לכל** k, $\int_{-1}^{1}\left|x^kf(x)\right| \leq \int_{-1}^{1}f(x)dx < 1$, **ולכן המומנט ה־**$k$ **קיים אם ורק אם** $\int_1^\infty x^kf(x)\,dx<\infty$ **ו־**$\int_{-\infty}^{-1}x^kf(x)\,dx<\infty$. **אבל כאשר** $k<n$ **אפשר להשוות בקרניים** $\left|\int_{-\infty}^{-1}x^kf(x)\,dx\right| \leq \left|\int_{-\infty}^{-1}x^nf(x)\,dx\right| < \infty$ **ובדומה** $\int_1^\infty x^kf(x)\,dx \leq \int_1^\infty x^nf(x)\,dx < \infty$. $\square$

תרגיל 2.8.2 נתון k שלם. תן דוגמא למשתנה מקרי חיובי שיש לו המומנטים מסדר k ומטה, ואין לו מומנט מסדר $k+1$. **הדרכה**. בחר משתנה מקרי בעל ערכים טבעיים, או משתנה רציף המקבל ערכים בקרן $[1,\infty)$.

דוגמא 2.8.3 הצפיפות $f_X(x)=\frac{1}{\pi(x^2+1)}$ מגדירה את **התפלגות קושי**. למרות הסימטריה, למשתנה בעל ההתפלגות הזו אין תוחלת. גם למשתנה המקרי $|X|$ אין תוחלת.

לעומת זאת יש משתנים שכל המומנטים שלהם מוגדרים, ובאלו נעסוק בהמשך הסעיף.

המומנט המרכזי השלישי, כשהוא מתוקנן לצורה $\gamma_1(X) = \frac{\mathbf{E}((X-\mu)^3)}{\sigma^3}$, נקרא **צידוד** **Skewness**. הוא מודד את חוסר הסימטריה של ההתפלגות: למשתנה בעל צידוד חיובי יש נטייה לקבל ערכים גבוהים ורחוקים מן הממוצע, יותר מאשר ערכים נמוכים הרחוקים ממנו, ובגרף של פונקציית הצפיפות ניכר שהזנב הימני ארוך ועבה יותר מן השמאלי. הצידוד אינו משתנה תחת הפעלת טרנספורמציה ליניארית.

המומנט המרכזי הרביעי, $\gamma_2(X) = \frac{\mathbf{E}((X-\mu)^4)}{\sigma^4} - 3$, נקרא **גבנוניות** **kurtosis**. גבנוניות גבוהה מתבטאת בחריגות גדולות ונדירות מן הממוצע, בעוד שגבנוניות נמוכה פירושה שהחריגות שכיחות אך קטנות יותר בעוצמתן. הגבנוניות מנורמלת כך שלהתפלגות הנורמלית יש גבנוניות 0.

2.8.2 פונקציה יוצרת מומנטים

אפשר להצמיד למשתנה מקרי פונקציות רציפות בדרכים שונות. גישה זו מאפשרת לרתום את התוצאות של האנליזה ההרמונית, לחקר משתנים מקריים. נתמקד כאן בדוגמא הנפוצה ביותר.

הפונקציה יוצרת המומנטים של משתנה מקרי X היא הפונקציה הממשית

$$M_X(t) = \mathbf{E}(e^{tX}).$$

(הפונקציה מוגדרת כאשר האינטגרל מתכנס, ולפי הנימוק של טענה 2.8.1 תחום ההתכנסות הוא קטע, אולי אינסופי ולאו דווקא סימטרי, סביב אפס).

משפט 2.8.4 ([1, 8.26]) אם הפונקציה יוצרת המומנטים של משתנה מקרי X קיימת (וסופית) בקטע פתוח סביב 0, אז כל המומנטים קיימים, ויש פיתוח טיילור $M_X(t) = \sum_{n=0}^{\infty} \frac{\mathbf{E}(X^n)}{n!} t^n$; בפרט $\mathbf{E}(X^n) = M_X^{(n)}(0)$.

מכאן, כמובן, שמה של הפונקציה הזו. למשל, $M_X(0) = 1$.

דוגמא 2.8.5 יהי $X \sim N(\mu, \sigma^2)$ משתנה בעל התפלגות נורמלית. נחשב את הפונקציה יוצרת המומנטים של X.

$$\begin{aligned}
M_X(t) &= E(e^{tX}) \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} e^{tx} dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{(x-(\mu+\sigma^2 t))^2 - (2\mu+\sigma^2 t)\sigma^2 t}{2\sigma^2}} dx \\
&= e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} dx = e^{\mu t + \sigma^2 t^2/2}.
\end{aligned}$$

בפרט, עבור המשתנה הנורמלי הסטנדרטי $Z \sim N(0,1)$, $M_Z(t) = e^{t^2/2}$. כלומר

$$\sum_{n=0}^{\infty} \frac{E(Z^n)t^n}{n!} = e^{t^2/2} = \sum_{m=0}^{\infty} \frac{t^{2m}}{2^m m!}$$

ולכן $\mathbf{E}(Z^{2m+1}) = 0$ לכל m, ואילו $E(Z^{2m}) = \frac{(2m)!}{2^m m!}$. למשל

$$\mathbf{E}(Z^2) = 1, \qquad \mathbf{E}(Z^4) = \frac{4!}{2^2 2!} = 3, \qquad \mathbf{E}(Z^6) = \frac{6!}{2^3 3!} = 15.$$

דוגמא 2.8.6 הפונקציה יוצרת המומנטים של

1. התפלגות מעריכית: $\frac{\theta}{\theta-t}$;
2. התפלגות פואסון: $e^{(e^t-1)\lambda}$.
3. התפלגות חי־בריבוע: $(1-2t)^{-k/2}$.

תרגיל 2.8.7 יהי X משתנה מקרי חיובי. אז $\lim_{t\to-\infty} M_X(t) = P(X=0)$.

טענה 2.8.8 אם X משתנה מקרי, אז לכל קבוע a מתקיים $M_{aX}(t) = M_X(at)$.

טענה 2.8.9 לפי טענה 2.2.15, אם X, Y בלתי תלויים, אז $M_{X+Y}(t) = M_X(t)M_Y(t)$.

תרגיל 2.8.10 אם $f \in \mathbb{R}\to\mathbb{R}$ פונקציה רציפה כך ש־$f(0) = 0$ ו־

$$f(x+y) + f(x-y) = 2f(x) + f(y) + f(-y),$$

אז קיימים קבועים a, b כך ש־$f(x) = ax^2 + bx$. **הדרכה**. נסמן $g(x) = f(x) + f(-x)$. הצבת $-x, -y$ במשוואה, וחיסור המשוואות, מביא ל־$g(x+y) = 2g(x) - g(x-y)$ עם $g(0) = 0$. באינדוקציה על n, נובע מכאן $g(nx) = ng(x)$ ולכן $g(x) = 2bx$ עבור קבוע c. מכיוון ש־$f(x) = x$ הוא פתרון למשוואה המקורית, גם ההפרש $\tilde{f}(x) = f(x) - bx$ הוא פתרון. אלא ש־$\tilde{f}(-x) = f(-x) + bx = f(x) - bx = \tilde{f}(x)$, ולכן

$$\tilde{f}(x+y) + \tilde{f}(x-y) = 2\tilde{f}(x) + 2\tilde{f}(y).$$

כעת נציב $x = ny$, ונקבל באינדוקציה $\tilde{f}(ny) = n^2\tilde{f}(y)$, ולכן גם $\tilde{f}(\frac{n}{m}y) = \frac{n^2}{m^2}\tilde{f}(y)$ ובסופו של דבר $\tilde{f}(x) = ax^2$ עבור קבוע a. לכן $f(x) = \tilde{f}(x) + bx = ax^2 + bx$.

משפט 2.8.11 תהי F התפלגות בעלת והתכונה הבאה: אם $X, Y \sim F$ בלתי תלויים, אז $X+Y, X-Y$ בלתי תלויים. נניח שהפונקציה יוצרת המומנטים של F רציפה. אז היא שווה לפונקציה יוצרת המומנטים של משתנה בעל התפלגות נורמלית.

הוכחה. לפי האי־תלות של $X+Y, X-Y$, לכל α, β מתקיים

$$\begin{aligned} M(\alpha)^2 M(\beta) M(-\beta) &= M_{X+Y}(\alpha) M_{X-Y}(\beta) \\ &= M_{\alpha(X+Y)+\beta(X-Y)}(1) \\ &= M_{(\alpha+\beta)X}(1) M_{(\alpha-\beta)Y}(1) \\ &= M(\alpha+\beta) M(\alpha-\beta). \end{aligned}$$

נסמן $L(t) = \log M_X(t)$, אז $L(0) = 0$ ומתקיים

$$2L(\alpha) + L(\beta) + L(-\beta) = L(\alpha+\beta) + L(\alpha-\beta). \tag{2.7}$$

לפי תרגיל 2.8.10, קיימים σ, μ כך ש־$L(t) = \mu t + \sigma^2 t^2/2$, ולכן $M(t) = e^{\mu t + \sigma^2 t^2/2}$, כפי שהיה להוכיח. $\square$

2.8.3 פונקציות יוצרות אחרות

הערה 2.8.12 וריאציות על נושא: מגדירים **פונקציה יוצרת** $\mathbf{E}(e^{izX})$ (כפונקציה מרוכבת); **ופונקציית מומנטים פקטוריאליים** המחשבת את התוחלות $\mathbf{E}(X(X-1)\cdots(X-m+1))$ (מתאימה בעיקר למשתנים בדידים המקבלים ערכים שלמים).

בסעיף זה נציג פונקציה יוצרת נוספת, שיש לה יתרונות בעיקר עבור משתנים בדידים.

הגדרה 2.8.13 יהי X משתנה מקרי. מגדירים $Q_X(t) = \mathbf{E}(t^X)$, כמובן כאשר הטור מתכנס.

בפונקציה הזו אין חידוש רב, משום ש־$Q_X(e^t) = M_X(t)$. אם X מקבל רק ערכים טבעיים, $0, 1, 2, \dots$, אז $Q_X(t) = \sum P(X=n)t^n$ הוא **טור חזקות פורמלי**, שהוא איבר של החוג $\mathbb{R}[[t]]$, וכך אנו פטורים משיקולי התכנסות.

תרגיל 2.8.14 יהי X משתנה מקרי.

1. $Q_X(1) = 1$;
2. $Q'_X(1) = \mathbf{E}(X)$;

3. לכל $k \in \mathbb{N}$, $Q_X^{(k)}(1) = \mathbf{E}[X(X-1)\cdots(X-k+1)]$.

תרגיל 2.8.15 הראה שהטור $Q_X(t)$ מתכנס לפחות בקטע $(-1,1]$, ותן דוגמא למשתנה שעבורו זה בדיוק תחום ההתכנסות.

תרגיל 2.8.16 חשב את $Q_X(t)$ עבור $X \sim \mathrm{Bin}(n,p)$, $X \sim \mathrm{G}(p)$. הראה שאם $X \sim \mathrm{Poi}(\lambda)$ אז $Q_X(t) = e^{(t-1)\lambda}$.

תרגיל 2.8.17 אם X, Y בלתי תלויים, אז $Q_{X+Y}(t) = Q_X(t)Q_Y(t)$.

תרגיל 2.8.18 לכל שני משתנים מקריים, $Q_Y(t) = E(Q_{Y|X}(t))$.

תרגיל 2.8.19 יהי Y משתנה מקרי המקבל ערכים טבעיים, ונניח שבהנתן Y, Z הוא סכום של Y משתנים מקריים בלתי תלויים $X_1, \ldots, X_Y$ בעלי אותה התפלגות. הראה ש־$Q_Z(t) = Q_Y(Q_X(t))$.

תרגיל 2.8.20 תהי F התפלגות על המספרים הטבעיים. נגדיר **עץ מקרי** (נקרא גם **תהליך גלטון־וטסון**) להיות עץ, שבו מספר הענפים היוצאים מכל ענף מתפלג לפי F. פורמלית, התהליך הוא סדרה $Y_0, Y_1, \ldots$, כך ש־$Y_0 = 1$ ולכל n, Y_{n+1} בהנתן Y_n הוא הסכום $\sum_{i=1}^{Y_n} X_i^{(n)}$ של Y_n משתנים מקריים בלתי תלויים המתפלגים כולם $X_i^{(n)} \sim F$.

הראה ש־$Q_{Y_n}(t) = Q_F^{(n)}(t)$, כאשר $Q^{(n)} = Q \circ \cdots \circ Q$, n פעמים.

נסמן $S = \sum_{n=0}^{\infty} Y_n$, מספר הקודקודים הכולל בעץ. הראה ש־$Q_S(t) = t \cdot Q_F(Q_S(t))$. (הדרכה: אפשר לפרק $S = 1 + S_1 + \cdots + S_{Y_1}$ כאשר S_i הם בעלי אותה התפלגות כמו S).

מצא את ההתפלגות של S בהנחה ש־$F = \mathrm{Poi}(\lambda)$. דון בהבדלים המהותיים בין המקרים $\lambda < 1$, $\lambda = 1$, $\lambda > 1$.

תרגיל 2.8.21 מכתב מסויים מצליח לשכנע כל אדם המקבל אותו לשלוח באקראי X עותקים של המכתב, כאשר X הוא משתנה מקרי בעל התפלגות קבועה. האם מכתב השרשרת ידעך או ישרוד? (הראה שהתשובה תלויה רק בתוחלת של X).

תרגיל 2.8.22 בשאלה 2.3.37, מה ההתפלגות של $\left|f^{-1}(f^k(w))\right|$, כאשר k מספר קטן יחסית?

תרגיל 2.8.23 נגדיר סדרת משתנים מקריים באופן הבא: $X_0 \sim \mathrm{Poi}(\lambda)$, ולכל $n > 1$, ההתפלגות של X_n בהנתן X_{n-1} היא $\mathrm{Poi}(X_n)$. חשב את $\mathbf{E}(X_n)$, את $\mathbf{V}(X_n)$ ואת $\mathrm{Cov}(X_n, X_m)$. הראה ש־ $P\left(X_n = 0\right) = e^{-\lambda a_n}$ כאשר $a_n = 1 - e^{-a_{n-1}}$. הוכח שהסדרה מתאפסת, בסופו של דבר, בהסתברות 1.

2.9 הלמה של בורל־קנטלי

נתבונן בסדרת מאורעות $A_1, A_2, \ldots$. מאורע הנגזר מהם נקרא **מאורע זנב** אם הוא אינו מושפע מאף קבוצה סופית של מאורעות. להלן שתי דוגמאות עיקריות. מסמנים

$$\limsup_n A_n = \bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right) = \{x:\ (\exists^{\infty} n)\, x \in A_n\};$$

זהו המאורע "מתרחשים אינסוף מבין המאורעות A_n". בדומה לזה, מגדירים

$$\liminf_n A_n = \bigcup_{n=1}^{\infty}\left(\bigcap_{m=n}^{\infty} A_m\right) = \{x:\ (\exists n_0)(n > n_0 \implies x \in A_n)\};$$

כלומר "המאורעות A_n מתרחשים ממקום מסויים ואילך". מכיוון ש־σ־אלגברות סגורות לאיחוד ולחיתוך, גם הגבול העליון והגבול התחתון הם מאורעות, וברור שהם אינם משתנים אם נשנה מספר סופי מבין המאורעות בסדרה.

חוק האפס־אחד של קולמוגורוב (שלא נוכיח כאן) קובע שההסתברות למאורע זנב היא תמיד אפס או אחד. כלומר, מאורע זנה הוא או "נכון" (מתרחשב בהסתברות אחד), או "לא נכון" (מתרחש בהסתברות אפס). הלמות של בורל־קנטלי עונות לשאלה מתי $\limsup A_n$ נכון, ומתי הוא לא נכון.

למה 2.9.1 (הלמה הראשונה של בורל־קנטלי) נניח ש־$\sum P(A_n) < \infty$. אז מתקיים $P(\limsup A_n) = 0$.

הוכחה. **נסמן ב־X_n את המשתנה המקרי המציין של A_n, וב־$N = \sum X_n$ את הסכום. כעת $\mathbf{E}(N) = \mathbf{E}(\sum X_n) = \sum \mathbf{E}(X_n) = \sum P\left(A_n\right)$, ולפי ההנחה זהו ערך סופי. לכן בהכרח $P\left(\limsup A_n\right) = P\left(N = \infty\right) = 0$.** $\square$

למה 2.9.2 (הלמה השניה של בורל־קנטלי) אם $\sum P(A_n) = \infty$ והמאורעות בלתי תלויים במשותף, אז $P(\limsup A_n) = 1$.

הוכחה. לכל n קבוע מתקיים ש־

$$\begin{aligned} P\left(\bigcup_{m=n}^{\infty} A_m\right) &= 1-P\left(\overline{\bigcup_{m=n}^{\infty} A_m}\right) \\ &= 1-P\left(\bigcap_{m=n}^{\infty} \overline{A_m}\right) \\ &= 1-\prod_{m=n}^{\infty} P\left(\overline{A_m}\right) \\ &= 1-\prod_{m=n}^{\infty}\left(1-P\left(A_m\right)\right)=1, \end{aligned}$$

משום שאם $\sum_{m=1}^{\infty} P\left(A_m\right)=\infty$ **אז** $\sum_{m=n}^{\infty} P\left(A_m\right)=\infty$ **לכל** n**, ומזה נובע שהמכפלה** $\prod_{m=n}^{\infty}(1-P\left(A_m\right))=0$.

אבל סדרת המאורעות $\bigcup_{m=n}^{\infty} A_m$ **היא סדרה יורדת, ולכן**

$$P\left(\limsup A_n\right)=P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right)=\lim_{n\to\infty} P\left(\bigcup_{m=n}^{\infty} A_m\right)=1.$$

□

לסיכום, לפחות כאשר המאורעות A_n בלתי תלויים במשותף, מתרחשים אינסוף מהם אם ורק אם סכום ההסתברויות $\sum P\left(A_n\right)$ הוא אינסוף.

2.10 חוקי המספרים הגדולים

אם $X_1, X_2, \ldots$ סדרה של משתנים מקריים, נסמן $\bar{X}_n=(X_1+\cdots+X_n)/n$.

2.10.1 החוק החלש של המספרים הגדולים

סדרת משתנים מקריים Y_n **מתכנסת בהסתברות** לקבוע μ (כותבים $Y_n \xrightarrow{P} \mu$), אם לכל $\varepsilon>0$,

$$\lim_{n\to\infty} P(|Y_n-\mu|<\varepsilon)=1.$$

במלים אחרות, לכל $\varepsilon>0$, הסיכוי לכך ש־$|Y_n-\mu|>\varepsilon$ שואף לאפס.

משפט 2.10.1 (החוק החלש של המספרים הגדולים) תהי $X_1, X_2, \dots$ סדרה של משתנים מקריים בלתי מתואמים, בעלי אותה תוחלת μ ושונות σ^2. אז $\bar{X}_n \xrightarrow{P} \mu$.

הוכחה. לפי אי־שוויון צ'ביצ'ב. $\square$

למעשה אין צורך להניח שונות קבועה: די בכך ש־ $\frac{1}{n^2}\sum \mathbf{V}(X_i) \to 0$.

ציטוט 2.10.2 "אם נטיל מטבע הוגן מליארד פעמים, יתכן גם רצף של עשרות אלפי 'עץ' - זהו החוק החלש" (ויקיפדיה, הערך "חוק המספרים הגדולים", עורך אלמוני).

החוק החלש אינו עוסק ברצפים של ערכים. אם מטילים מטבע מילארדי פעמים, צפוי שנתקל ברצף של 10־11 הטלות זהות (ראה תרגיל 2.3.47). מה שהחוק החלש יודע לומר במצב זה הוא שאם התמשכות סדרת ההטלות, הסיכוי שהערך הממוצע יפול בטווח $(0.4999, 0.5001)$ הולך וגדל, ושואף לאחד.

תרגיל 2.10.3 $X_1, \dots, X_{100}$ הם משתנים מקריים בלתי תלויים, בעלי תוחלת 0 ושונות 1. ידוע ש־$\frac{1}{50}(X_1 + \cdots + X_{50}) = 0.14$. מה התוחלת של $\frac{1}{50}(X_{51} + \cdots + X_{100})$?

הרהר בתפישה הנפוצה שלפיה המשך הסדרה אמור לתקן את ראשיתה, אם זו סוטה מן הממוצע.

2.10.2 החוק החזק של המספרים הגדולים

סדרת משתנים מקריים Y_n **מתכנסת כמעט תמיד** לקבוע μ (כותבים $Y_n \xrightarrow{a.s.} \mu$), אם $P(\lim Y_n = \mu) = 1$.

הערה 2.10.4 כאן צריך להוכיח שהדרישה $\lim Y_n = 0$ היא אכן מאורע (אחרת, הסתברות מניין). ואכן, המאורע הזה שווה ל־$\cap_d \cup_N \cap_{n \geq N} \{|Y_n| < 1/d\}$, ולכן שייך ל־$\sigma$-אלגברה שביחס אליה כל ה־$Y_n$ הם משתנים מקריים.

משפט 2.10.5 (החוק החזק של המספרים הגדולים) תהי $X_1, X_2, \dots$ סדרת משתנים מקריים בלתי תלויים בעלי תוחלת μ ושונות σ^2. אז $\bar{X}_n \xrightarrow{a.s.} \mu$.

את המשפט החשוב הזה לא נוכיח. גם כאן, אין צורך להניח שהשונות קבועה; די להניח שהיא חסומה.

כעת נוכל לטפל בשאלה האם סדרת משתנים מתכנסת לאפס:

טענה 2.10.6 התכנסות כמעט תמיד גוררת התכנסות בהסתברות: אם $Y_n \xrightarrow{a.s.} 0$ אז גם $Y_n \xrightarrow{P} 0$.

הוכחה. נקבע $\epsilon > 0$, ונסמן ב־ A_n את המאורע $|Y_n| < \epsilon$. אז

$$\begin{aligned} 1 &= P(\lim_{n\to\infty} Y_n = 0) \\ &= P(\forall\delta > 0\exists N\forall n \geq N : |Y_n| < \delta) \\ &= P(\cap_\delta \cup_N \cap_{n\geq N}\{|Y_n| < \delta\}) \\ &\leq P(\cup_N \cap_{n\geq N} A_n) \\ &= \lim_{N\to\infty} P(\cap_{n\geq N} A_n) \leq \lim_{N\to\infty} P(A_N); \end{aligned}$$

שים לב ש־$P(\cap A_n)$ אינו הגבול של $P(A_n)$, כי זו אינה סדרת מאורעות מונוטונית. $\square$

דוגמא נגדית לכיוון ההפוך:

טענה 2.10.7 נניח $Y_n \sim b(\frac{1}{n})$ בלתי תלויים. אז $Y_n \xrightarrow{P} 0$, אבל היא אינה מתכנסת כמעט תמיד: הסתברות ההתכנסות היא 0.

הוכחה. לכל $\epsilon > 0$, אפשר לחשב $P(Y_n < \epsilon) \geq P(Y_n = 0) = 1 - \frac{1}{n} \to 1$, ולכן $Y_n \xrightarrow{P} 0$. אבל מכיוון שזו סדרה בדידה, כדי שהיא תתכנס ל־0 נדרש שתהיה קבועה ממקום מסויים ואילך; לכל N, $P(\forall n \geq N : Y_n = 0) = \prod_{n\geq N} \frac{n-1}{n} = 0$, ולכן $P(\exists N : \forall n \geq N : Y_n = 0) = 0$. $\square$

תרגיל 2.10.8

$$\bigcap_{n=1}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right) = \bigcup_{k=1}^{\infty}\left(\bigcap_{n=k}^{\infty}\left(\bigcup_{m=n}^{\infty} A_m\right)\right).$$

2.10.3 משפט הגבול המרכזי

משפט 2.10.9 ([1, משפט 8.34]) אם $M_X(t)$ קיימת לכל t, אז הפונקציה קובעת את ההתפלגות. **הערה.** לא נוכיח בקורס.

סדרת משתנים מקריים Y_n **מתכנסת בהתפלגות** למשתנה מקרי Y (כותבים $Y_n \xrightarrow{D} Y$), אם לכל a,

$$\lim_{n\to\infty} P(Y_n \leq a) = P(Y \leq a).$$

טענה 2.10.10 אם $Y_n - Y \xrightarrow{P} 0$, אז $Y_n \xrightarrow{D} Y$.

טענה 2.10.11 אם $M_{Y_n}(t) \to M_Y(t)$ לכל t, אז $Y_n \xrightarrow{D} Y$.

משפט 2.10.12 (משפט הגבול המרכזי) תהי $X_1, X_2, \ldots$ סדרה של משתנים מקריים בלתי תלויים בעלי אותה התפלגות, שיש לה תוחלת μ ושונות σ^2. אז $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)$.

הוכחה. על־ידי טרנספורמציה ליניארית אפשר להניח $\mu = 0$, $\sigma = 1$. לכן $M_X(0) = 1$, $M'_X(0) = 0$, $M''_X(0) = 1$.

ההוכחה בהנחה ש־ $M_X(t)$ גזירה פעמיים ברציפות ב־$t = 0$. נחשב:

$$\begin{aligned}
\lim_{n\to\infty} M_{\sqrt{n}\cdot\bar{X}_n}(t) &= \lim_{n\to\infty} \mathbf{E}(e^{t\sqrt{n}\cdot\bar{X}_n}) \\
&= \lim_{n\to\infty} \mathbf{E}(e^{t(X_1+\cdots+X_n)/\sqrt{n}}) \\
&= \lim_{n\to\infty} \mathbf{E}(e^{tX/\sqrt{n}})^n \\
&= \lim_{n\to\infty} M(t/\sqrt{n})^n \\
&= \exp\left(\lim_{n\to\infty} \frac{\log M(t/\sqrt{n})}{n^{-1}}\right) \\
&\overset{L'hospital}{=} \exp(\lim_{n\to\infty} \frac{t}{2M(t/\sqrt{n})} \frac{\log M'(t/\sqrt{n})}{n^{-1/2}}) \\
&= \exp\left(\frac{t}{2}\lim_{n\to\infty} \frac{\log M'(t/\sqrt{n})}{n^{-1/2}}\right) \\
&\overset{L'hospital}{=} \exp\left(\frac{t}{2}\lim_{n\to\infty} tM''(t/\sqrt{(n)})\right) \\
&= \exp\left(\frac{t^2}{2}\right).
\end{aligned}$$

□

הערה 2.10.13 משפט *Berry-Esseen* (עם קבוע משופר לפי *Tyurin, 2010*) נותן חסם על גודל השגיאה במשפט הגבול המרכזי: לכל z,

$$\left| P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) - \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \right| \leq \frac{1}{2}\mathbf{E}\left(\left|\frac{X_1 - \mu}{\sigma}\right|^3\right) \frac{1}{\sqrt{n}}.$$

דוגמא 2.10.14 במקרה המיוחד $X_i \sim b(p)$, כמובן $\sum X_i \sim \mathrm{Bin}(n,p)$; במקרה זה משפט הגבול המרכזי קובע ש־ $X \sim \mathrm{Bin}(n,p)$ מתפלג בקירוב $N(np, npq)$. המשפט הזה נקרא 'משפט דה־מואבר–לפלס', או **הקירוב הנורמלי להתפלגות בינומית**.

תרגיל 2.10.15 נניח ש־X_i הם משתנים מקריים בלתי תלויים, בעלי סיכוי שווה לקבל את הערכים ± 1. הערך את גודלו של הסכום $X_1+\cdots+X_{1000000}$. מה הסיכויים לכך שהוא יפול בטווח $(-50,50)$?

תרגיל 2.10.16 את השאלה הבאה הציג סמואל פפיס (1633-1703) לניוטון. זורקים $6k$ קוביות, במטרה לקבל לפחות k פעמים 6. מתי הסיכוי גדול יותר: כאשר $k=1$, $k=2$ או $k=3$? חשב את הסיכוי כאשר $k\to\infty$.

תרגיל 2.10.17 דון בדוגמא $X_1,\dots,X_n \sim P(1)$, עם $\sum X_i \sim P(n)$, וקבל קירוב נורמלי להתפלגות פואסונית.

תרגיל 2.10.18 דון בדוגמא $X_1,\dots,X_n \sim P(\lambda/n)$, כאשר λ קבוע. כאן $\sum X_i \sim P(\lambda)$. האם נובע מזה שכל התפלגות פואסונית היא למעשה נורמלית?

תרגיל 2.10.19 מטילים קוביה מאתיים פעמים. הערך את הסיכויים לכך שסכום מאה ההטלות הראשונות יהיה שווה לסכום מאה ההטלות האחרונות. **הדרכה.** נסמן את שני הסכומים ב־S_1,S_2. לפי משפט הגבול המרכזי, $S_1,S_2 \sim N(350,\sigma^2)$ בקירוב עם $\sigma\approx 17.07$. מכיוון שהסכומים נופלים בדרך כלל בטווח של סטיית תקן או שתיים מן התוחלת, כלומר בטווח של כ־50 ערכים, הסיכוי הוא בסביבות 0.02. (השווה לתרגיל 2.3.76, שאפשר לפתור בדרך דומה.)

תרגיל 2.10.20 יהי λ קבוע. כדי לקבל משתנה פואסוני עם פרמטר λ, אפשר להגדיר $X_1,\dots,X_n \sim \mathrm{Poi}(\lambda/n)$, ולקחת $X=\sum X_i \sim \mathrm{Poi}(\lambda)$. הסבר מדוע לא נובע ממשפט הגבול המרכזי שהיחס $\frac{\mathrm{Poi}(\lambda)-\lambda}{\sqrt{\lambda}}$ מתפלג, בקירוב, $N(0,1)$.

2.11 שרשראות מרקוב

שרשראות מרקוב הם נושא חשוב ורחב היקף בתורת ההסתברות. בחרנו להציג את הנושא על קצה המזלג כבר בקורס זה, משום ששרשראות מרקוב פותחות, במאמץ לא גדול, פתח לפתרון של בעיות מעניינות וחשובות, הנמצאות מחוץ להישג ידן של השיטות שפגשנו עד כה.

הגדרה 2.11.1 **שרשרת מרקוב** היא סדרה של משתנים מקריים $X_0,X_1,\dots$, כך שלכל $n<m$, X_n מפריד את X_m מהווקטור $(X_0,\dots,X_{n-1})$ (ראה הגדרה 2.2.16).

כלומר, ההווה מנתק את הקשר בין העבר והעתיד. אנחנו נניח ש־

$$P\left(X_{n+1}=i|X_n=j\right)=A_{ij}$$

אינו תלוי ב־n (למרות שאפשר ללמוד גם שרשראות התלויות בזמן).

בהנתן מטריצה A_{ij} כך ש־$\sum_i A_{ij} = 1$ לכל j, אפשר להגדיר שרשרת מרקוב באינדוקציה - קובעים (או מגרילים) את X_0, ולכל n מגרילים את X_{n+1} בהתפלגות ש־X_n מכתיב.

דוגמא 2.11.2 אם מגרילים סדרת משתנים מקריים כך שהתפלגות X_{n+1} בהנתן $X_1, \dots, X_n$ תלויה רק ב־X_n, מתקבל שרשרת מרקוב.

נסמן ב־J את וקטור העמודה שכל רכיביו שווים ל־1. **וקטור התפלגות** על N מצבים הוא וקטור חיובי v באורך N, המקיים $J^{\mathrm{t}} v = 1$. לדוגמא, את ההתפלגות '$X = 2$ בוודאות' מקודד וקטור היחידה e_2. כל וקטור התפלגות הוא צירוף קמור של וקטורי היחידה.

טענה 2.11.3 תהי A מטריצת מרקוב; אז $J^{\mathrm{t}} A = J^{\mathrm{t}}$.

טענה 2.11.4 סכום הרכיבים בכל וקטור עצמי של ערך עצמי $1 \neq$ הוא אפס. אכן, מ־$Av = \lambda v$ נובע $J^{\mathrm{t}} v = J^{\mathrm{t}} A v = \lambda J^{\mathrm{t}} v$.

למה 2.11.5 אם $X_n \sim v$ אז $X_{n+1} \sim Av$.

זהו ניסוח מחדש של נוסחת ההסתברות השלמה.

באינדוקציה נובע מכאן ש־$X_{n+k} \sim A^k v$.

מסקנה 2.11.6 $P(X_{n+k} = i | X_n = j) = (A^k)_{ij}$.

בפרט, אם ההתפלגות של X_0 לפי וקטור u, אז ההתפלגות של X_n היא לפי הווקטור $A^n u$.

2.11.1 תאור גרפי

שרשרת מרקוב אפשר לתאר בגרף שקודקודיו הם המצבים, כשעל כל חץ כתובה ההסתברות למעבר; כלומר, על החץ מ־i ל־j תופיע ההסתברות $P(X_{n+1} = j | X_n = i)$. אם ההסתברות היא 0 אין צורך בחץ. אם כל ההסתברויות שוות, אפשר להשמיט את הערך. סכום כל ההסתברויות היוצאות מקודקוד צריך להיות 1, אלא אם מדובר בקודקוד מיוחד שתפקידו מובן מן ההקשר (ראו קודקודים סופגים להלן).

דוגמא 2.11.7 הגרף

$$\begin{tikzcd} 1 \arrow[loop left, "0.8"] \arrow[r, bend left, "0.2"] & 2 \arrow[l, bend left, "0.6"] \arrow[r, bend left, "0.4"] & 3 \arrow[l, bend left] \end{tikzcd}$$

מתאר שרשרת בת שלושה מצבים, שבו אפשר להשאר במצב 1 או לעבור ממנו למצב 2, ממצב 2 עוברים לאחד משני המצבים האחרים, וממצב 3 עוברים מיד למצב 2.

2.11.2 ההתפלגות הסטציונרית

התפלגות u היא יציבה אם $Au = u$. יהי v וקטור עצמי של ערך עצמי כלשהו; כלומר $Av = \lambda v$. אז $J^{t}v = J^{t}Av = \lambda J^{t}v$, ולכן יש שתי אפשרויות: סכום הרכיבים הוא אפס, או שהערך העצמי הוא 1.

לכל וקטור עצמי שסכומו אינו אפס, הערך העצמי הוא 1. אם A לכסינה, אז הווקטור $A^n u$ הולך ומתקרב להתפלגות הסטציונרית, וזאת לכל וקטור u.

דוגמא 2.11.8 בדוגמא 2.11.7, ההתפלגות הסטציונרית היא $(\frac{15}{22}, \frac{5}{22}, \frac{2}{22})$. פירושו של דבר שבטווח הארוך, השרשרת מבלה $15 : 5 : 2$ מהזמן במצבים $1, 2, 3$ בהתאמה.

דוגמא 2.11.9 נגדיר שרשרת מרקוב לפי המטריצה $A = \begin{pmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{pmatrix}$. חשב את $P(X_n = 1|X_{n+1} = 1)$, בהנחה ש־n גדול. היפוך הזמן דורש חישוב של $P(X_n = 1)$.

תרגיל 2.11.10
1. הראה שאם A מטריצת מרקוב 2×2 עם וקטור עצמי v של הערך העצמי 1, אז $A_{12}v_1 = A_{21}v_2$.
2. הראה שבשרשרת מרקוב עם שני מצבים, אם n גדול, אז
$$P(X_{n+1} = 1|X_n = 2) = P(X_n = 1|X_{n+1} = 2).$$
3. לכל k, אם n גדול מתקיים
$$P(X_{n+k} = 1|X_n = 2) = P(X_n = 1|X_{n+k} = 2).$$
4. הראה שתכונה זו אינה מתקיימת לתהליך 3×3, אפילו אם מניחים
$$P(X_1 = 2|X_0 = 3) = P(X_1 = 1|X_0 = 2) = 1.$$

תרגיל 2.11.11 שיכור מהלך על המספרים $0, \ldots, n-1$, כשבכל דקה הוא מבצע צעד ימינה או שמאלה (מודולו n). הראה שההתפלגות הסטציונרית היא ההתפלגות האחידה.

תרגיל 2.11.12 מהמר מתחיל את סדרת ההימורים שלו עם X_0 שקלים בכיסו. בכל משחקון, הוא מרוויח או מפסיד שקל אחד, בהסתברויות שוות. אם הוא מפסיד את כל כספו, הוא נזרק מן הקזינו. אם הוא מגיע להון של n שקלים, הוא שומר על רכושו ויוצא בראש מורם. הראה שהסיכוי לכך שהמהמר יצליח שווה ל-$\frac{X_0}{n}$.

דוגמא 2.11.13 הדוגמא הבאה קרויה "ספריית צטלין" ($Tsetlin$). בספריה מדף ארוך, שהספרן יושב בקצהו. בין הספרים יש פופולריים יותר ופחות: הסיכוי לבקש את ספר i הוא p_i, $p_1+\cdots+p_n=1$. מתוך עצלות, הספרן מחזיר את הספר תמיד לקצה המדף, כשהוא דוחק את שאר הספרים במורד הרשימה. עם הזמן, הספריה מתארגנת באופן כזה שהספרים המבוקשים קרובים יותר לתחילת המדף. זוהי שרשרת מרקוב על $n!$ הסידורים האפשריים, כאשר התמורה σ מתאימה לסידור הספרים $\sigma(1),\sigma(2),\ldots,\sigma(n)$.

לכל i, הסיכוי לעבור מסידור הספרים המתאים לתמורה $(\sigma(1),\ldots,\sigma(n))$ אל הסידור המתאים לתמורה $(\sigma(i),\sigma(1),\ldots,\sigma(i-1),\sigma(i+1),\ldots,\sigma(n))$, הוא $p_{\sigma(i)}$.

הוכח שההתפלגות הסטציונרית של הספרים על המדף היא זו המתקבלת מבחירת הספרים לתחילת המדף ואילך, באקראי וללא החזרה.

2.11.3 מאורעות שאינם תלויי זמן

ראינו בסעיף הקודם שחזקות של מטריצת המעבר מאפשרות לחשב היכן יבקר התהליך בעוד מספר קבוע של צעדים, ואפילו היכן הוא יבקר בעתיד הרחוק. אחת התכונות היפות של שרשראות מרקוב היא שאפשר לחשב בעזרתם את ההסתברות למאורעות שונים שאינם תלויים בזמן. למשל, עבור תהליך נתון:

1. מה ההסתברות שהתהליך יבקר בסופו של דבר במצב α?
2. מה ההסתברות לכך שהביקור הראשון במצב α יתרחש לפני הביקור הראשון במצב β?
3. מהם הסיכויים לכך שהביקור הראשון במצב α יתרחש בזמן זוגי?

לפני שנציג פתרון לכל הבעיות האלה, נדון בבעיה בראשונה. אם התהליך קשיר (כלומר, אפשר להגיע בהסתברות חיובית מכל נקודה לכל נקודה אחרת), הוא מבקר בסופו של דבר בכל מצב (ולכן מבקר בכל מצב אינסוף פעמים). דוגמא פשוטה לתהליך לא קשיר מתקבלת כאשר יש בו מצב סופג:

הגדרה 2.11.14 מצב t של שרשרת מרקוב נקרא **מצב סופג** אם אי אפשר לעזוב אותו, כלומר $P(X_{n+1}=t|X_n=t)=1$.

אם יש לתהליך קשיר מצב סופג יחיד, הוא יגיע אליו בסופו של דבר בוודאות. שאלה 1 נעשית מעניינת אם יש לתהליך כמה מצבים סופגים, משום שאז לא ברור באיזה מהם

יתקע התהליך בסופו של דבר. במקרים רבים אפשר להתאים את התהליך לבעיה. למשל, אם רוצים לדעת מה הסיכויים להגיע למצב α לפני מצב β, אפשר להפוך את שניהם למצבים סופגים, משום שממילא ברגע שהתהליך הגיע לאחד משניהם השאלה הוכרעה.

התשובה לשאלות שהוצגו בתחילת הסעיף, ואחרות, תלויה במצב ההתחלה. על־מנת לפתור שאלה כזו למצב התחלה נתון, יש לפתור אותה בבת־אחת לכל מצבי ההתחלה, על־ידי הגדרת משתנים מתאימים ובניה של מערכת משוואות לינארית על המשתנים.

נדגים זאת בפתרון הבעיה הראשונה: האם התהליך יגיע בסופו של דבר למצב α? נסמן

$$p_i = P\left(\exists n : X_n = \alpha | X_0 = i\right).$$

אז $p_\alpha = 1$. אם $\beta \neq \alpha$ הוא מצב סופג אז כמובן $p_\beta = 0$. באופן כללי, לכל $j \neq \alpha$,

$$\begin{aligned} p_j &= \sum_i P\left(\exists n : X_n = \alpha | X_1 = i, X_0 = j\right) P\left(X_1 = i | X_0 = j\right) \\ &= \sum_i P\left(\exists n : X_n = \alpha | X_1 = i\right) A_{ij} \\ &= \sum_i A_{ij} p_i = (A^{\mathrm{t}} p)_j. \end{aligned}$$

זוהי כמובן מערכת משוואות לינארית על המשתנים p_i, שאפשר אפילו להציג באופן מטריציאלי.

תרגיל 2.11.15 הראה שאם מסדרים את המצבים כך ש־α הוא האחרון וכותבים $A = \begin{pmatrix} B & v \\ w^{\mathrm{t}} & \alpha \end{pmatrix}$, אז מן המשוואה $J^{\mathrm{t}} A = J^{\mathrm{t}}$ יוצא $w = (I - B^{\mathrm{t}}) J_0$, כאשר J_0 מסמן את וקטור J ללא הרכיב האחרון, ולכן $(I - B^{\mathrm{t}})(\vec{p}_0 - J_0) = 0$. אם α מצב סופג, המשוואה שקולה ל־ $(I - A)(\vec{p} - J) = 0$.

דוגמא 2.11.16 שתי קבוצות שקולות מתחרות במשיכת חבל. יש ארבעה מצבים: נצחון לראשונה, יתרון לראשונה, יתרון לשניה, נצחון לשניה. הסיכוי לעבור מיתרון לנצחון הוא 0.2. נניח שיש יתרון לקבוצה הראשונה.

$$\begin{tikzcd} V_1 & L_1 \arrow[l, "0.2"'] \arrow[r, bend left, "0.8"] & L_2 \arrow[l, bend left, "0.8"] \arrow[r, "0.2"] & V_2 \end{tikzcd}$$

מה הסיכוי שהיא תנצח?

תרגיל 2.11.17 חייל משוטט באקראי על לוח שחמט בן 6×6 משבצות: בכל צעד הוא בוחר בין כל הכיוונים האפשריים בהסתברויות שוות. המשחק נעצר כשהחייל

נוגע באחת הדפנות, ואז נקבע המנצח: הצד "צפון ומזרח" או הצד "מערב ודרום". חשב את הסיכוי לנצחונו של השחקן הראשון אם החייל עומד ברביעיית המשבצות שבמרכז, קרוב יותר לפינה שלו. (תרגיל זה מדגים כמה חשוב לזהות סימטריות במצב: לאורך האלכסון שבין השחקנים ההסתברויות הן $1/2$, והלוח כולו סימטרי לשיקוף באלכסון האחר; במקום 16 נעלמים, יש כאן רק ארבעה.)

כדי לפתור את בעיה 3 (ובעיות נוספות בהמשך), נתבונן במשתנה המקרי

$$T = \min\{n: \ X_n = \alpha\},$$

ונסמן ב־A את המאורע "T זוגי". כעת נסמן

$$q_i = P(A \mid X_0 = i).$$

שוב אפשר לקבל מערכת משוואות ליניארית מנוסחת ההסתברות השלמה: $q_\alpha = 1$, ולכל $j \neq \alpha$,

$$\begin{aligned} q_j &= \sum_i P(A|X_1 = i, X_0 = j) P(X_1 = i | X_0 = j) \\ &= \sum_i P(A|X_1 = i) A_{ij} \\ &= \sum_i A_{ij}(1 - q_i), \end{aligned}$$

משום שהסיכוי לבקר ב־α לראשונה בזמן זוגי, אם $X_1 = i$, הוא הסיכוי לבקר שם לראשונה בזמן אי־זוגי בהנתן $X_0 = i$.

תרגיל 2.11.18 הילוך שיכור על ציר המספרים השלמים הוא תהליך מרקוב עם אינסוף מצבים, שבו עוברים ממצב m למצבים $m+1$ ו־$m-1$ בסיכוי $\frac{1}{2}$. הראה ששיכור המתחיל את ההילוך בנקודה m כלשהי יגיע בהסתברות 1 לנקודה 0. (אם p_m היא ההסתברות להגיע לראשית מהנקודה m, אז $p_{m+1} = \frac{1}{2}(p_{m+1} + p_{m-1})$ ומכאן ש־$p_m = 1 - m(1 - p_1)$.)

השפעת התוצאה על התהליך

הידיעה שמאורע מסויים התרחש, משנה את הסתברויות המעבר עצמן. לדוגמא, אם שמענו שקבוצת כדורגל מסויימת ניצחה במשחק ואנו צופים בו בשידור חוזר, נדע שהסיכוי למהלך מוצלח מצד הקבוצה שניצחה בסופו של דבר, הוא גדול מן הסיכוי שהיינו מקצים לאותה הצלחה אלמלא הידיעה המוקדמת.

חישוב ההסתברויות החדשות אינו מסובך. יהי A מאורע שאינו תלוי בזמן, ונניח ש־$p_i = P\left(A|X_0 = i\right)$. הסתברויות המעבר בתהליך המותנה הן

$$\begin{aligned} P\left(X_1 = j|A, X_0 = i\right) &= \frac{P\left(A, X_1 = j|X_0 = i\right)}{P\left(A|X_0 = i\right)} \\ &= \frac{P\left(A|X_1 = j, X_0 = i\right)P\left(X_1 = j|X_0 = i\right)}{P\left(A|X_0 = i\right)} \\ &= \frac{p_j}{p_i}A_{ji}, \end{aligned}$$

אלא אם i או j הם בעלי משמעות מיוחדת מבחינת המאורע A.

תרגיל 2.11.19 בנתוני שאלה 2.11.16, קבוצה 1 ביתרון, וידוע שהיא נצחה בסופו של דבר. מהם הסיכויים לכך שהיתרון יישמט מידה בסיבוב הראשון?

2.11.4 תוחלת זמן ההגעה

כדי לחשב את התוחלת $\mathbf{E}(T)$, כאשר T הוא המשתנה שהוגדר לעיל עבור מצב קבוע α, נבנה מערכת משוואות על התוחלות המותנות

$$e_i = \mathbf{E}(T|X_0 = i):$$

כמובן $e_\alpha = 0$, ולכל $j \neq \alpha$,

$$\begin{aligned} e_j &= \sum_i \mathbf{E}(T|X_1 = i, X_0 = j)P\left(X_1 = i|X_0 = j\right) \\ &= \sum_i A_{ij}\mathbf{E}(T|X_1 = i) \\ &= \sum_i A_{ij}\mathbf{E}(T + 1|X_0 = i) \\ &= 1 + \sum_i A_{ij}e_i. \end{aligned}$$

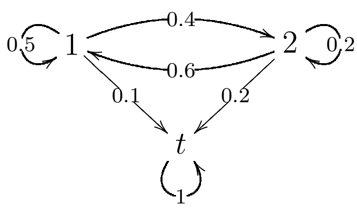
דוגמא 2.11.20 איך מתפלג המשתנה המקרי T עבור מטריצת המעבר $A = \begin{pmatrix} q & 0 \\ p & 1 \end{pmatrix}$? הראה ש־$e_1 = 1/p$.

תרגיל 2.11.21 בהילוך השיכור של תרגיל 2.11.18, הראה ש**תוחלת** זמן ההגעה מהנקודה 1 לנקודה־0 היא אינסופית (למרות שהסיכוי להגיע הוא 1). (אם e_n היא תוחלת זמן ההגעה לאפס מהנקודה m, אז $e_m = 1+\frac{1}{2}(e_{m+1}+e_{m-1})$ ומכאן ש־ $e_m = a(e_1-(m-1)) \geq 0$. הסק ש־$e_1 = \infty$.)

לאחר חישוב התוחלות אפשר לחשב גם את שונות זמן ההגעה, על־ידי חישוב התוחלת של T^2: נגדיר $m_i = \mathbf{E}(T^2|X_0 = i)$, כך ש־$m_\alpha = 0$, ולכל $j \neq \alpha$

$$\begin{aligned} m_j &= \sum_i \mathbf{E}(T^2|X_1 = i, X_0 = j)P\left(X_1 = j|X_0 = j\right) \\ &= \sum_i A_{ij}\mathbf{E}(T^2|X_1 = i) \\ &= \sum_i A_{ij}\mathbf{E}((T+1)^2|X_0 = i) \\ &= \sum_i A_{ij}(m_i + 2e_i + 1) \\ &= \sum_i (A_{ij}m_i) + 2e_j - 1. \end{aligned}$$

דוגמא 2.11.22 נמיה משוטטת בין שני שיחים הסמוכים למלכודת t, לפי הגרף הבא:



מה תוחלת הזמן עד שתיפול למלכודת? מהי השונות? התשובה תלויה כמובן בנקודת ההתחלה.

דוגמא 2.11.23 ארנב ושועל מטיילים בין ארבעה קודקודי ריבוע: בכל דקה, הארנב מקפץ אל הקודקוד שמימין או משמאל למקומו הנוכחי, והשועל מקפץ אף הוא באקראי ובסיכויים שווים, אך בהסתברות חצי הוא נשאר במקומו. אם שניהם נמצאים באותו קודקוד, השועל טורף את הארנב. מה תוחלת חייו של הארנב, בהתחשב במיקום היחסי של שניהם בתחילת הסיפור? (העזר בסימטריה כדי להציג את הבעיה כשרשרת מרקוב בת 4 מצבים).

כעת נניח שעל אותו ריבוע משוטט שועל נוסף, לפי אותם כללים. השועלים יתפסו את הארנב רק אם שלושת בעלי החיים נמצאים באותו קודקוד באותו זמן. מה תוחלת חייו של הארנב הזה?

תרגיל 2.11.24 ב**מגדלי האנוי** מעבירים בהדרגה מגדל של טבעות במתקן עם שלושה מוטות, כאשר טבעת גדולה לעולם אינה מונחת מעל טבעת קטנה ממנה. כמה צעדים נדרשים כדי להעביר מגדל בן שתי טבעות מן המוט השמאלי לימני? ומה קורה אם מעבירים את הטבעות באקראי (בכל פעם מבצעים אחד מהמהלכים החוקיים, בהסתברויות שוות)?

2.11.5 הרחבת זכרון

שרשרת מרקוב היא, מטבעה, חסרת זכרון. כדי לטפל בבעית שבהן נדרש זכרון, יש להרחיב את השרשרת על־ידי תוספת מצבים.

תרגיל 2.11.25 חשב את ההסתברות שבסדרת ניסויי ברנולי, עם הסתברות הצלחה p, יתרחשו שתי הצלחות רצופות לפני שמתרחשים שלושה כשלונות רצופים.

תרגיל 2.11.26 מטילים מטבע הוגן שוב ושוב. כמה זמן יש להמתין להופעת הרצף 0100, הנקרא משמאל לימין? וכמה זמן עד להופעת הרצף 0010?

הטענה הבאה מסבירה מדוע התוחלות בתרגיל 2.11.26 שוות זו לזו.

טענה 2.11.27 *מטילים קוביה כללית (מספר פאות סופי כלשהו, הסתברויות כלשהן). תוחלת זמן ההמתנה לרצף $v = v_1 \cdots v_t$ שווה לזו של זמן ההמתנה לרצף ההפוך $v' = v_t \cdots v_1$.*

הוכחה. **נסמן ב־T את תוחלת זמן ההמתנה. לפי טענה 2.3.57, $\mathbf{E}(T) = \sum P\left(T \geq n\right)$, וההסתברויות ל־$T \geq n$ שוות עבור v ועבור v' מטעמי סימטריה.** $\square$

תרגיל 2.11.28 בנתוני טענה 2.11.27, הראה שהסיכוי לכך שרצף v יופיע לפני רצף w שווה לסיכוי לכך שהרצף המנוגד v' יופיע לפני הרצף המנוגד w'. **הדרכה.** ההסתברות שווה לסכום ההסתברויות לכך ש־v מופיע לראשונה בזמן n ואילו w אינו מופיע כלל.

תרגיל 2.11.29 נחשב (ללא תהליכי מרקוב) את תוחלת זמן ההמתנה עד להופעה של רצף w, שאורכו n, בסדרה ארוכה של הטלות מטבע הוגן.

על פי [Solov'ev, *A combinatorial identity...*, Theory Prob Appl **11**, 276–282, (1966)]. סקירת מקורות בנושא זה: [Anant P. Godbole, Stavros G. Papastavridis, "Runs and Patterns in Probability: Selected Papers", p. 214].

1. נאמר שהופעה של w היא "נקיה" אם היא אינה חופפת את ההופעה הנקיה הקודמת (בפרט, ההופעה הראשונה היא נקיה; ברצף 0001100000 יש שתי הופעות נקיות של 000, ושתיים לא נקיות). נסמן ב־C את מספר ההופעות

הנקיות של w לאורך הסדרה, וב־D את מספר ההופעות הלא נקיות. מספר ההופעות הלא נקיות של w החופפות הופעה נקיה מסויימת הוא $\alpha = \sum_{n\neq k\in K} 2^{k-n}$, כאשר K היא קבוצת אורכי החפיפות של w עם עצמו (למשל אם $w = 000$ אז $\alpha = 2^{-1}+2^{-2}$). הראה ש־$\mathbf{E}(D) = \alpha\mathbf{E}(C)$.

2. הראה ש־$\mathbf{E}(C+D) = 2^{-n}\ell$, כאשר ℓ הוא אורך הסדרה.

3. הראה למספר המטבעות שיש להטיל מסופה של הופעה נקיה אחת עד לסופה של ההופעה הנקיה הבאה יש אותה התפלגות. נסמן את המשתנה המקרי הזה ב־X. הראה ש־$\mathbf{E}(C)\mathbf{E}(X) = \mathbf{E}(CX) \approx \ell$ (כלומר, היחס שואף ל־1).

4. הסק ש־$\mathbf{E}(X) = 2^n(1+\alpha) = \sum_{k\in K} 2^k$. למשל, תוחלת זמן ההמתנה ל־010101 היא $2^6+2^4+2^2$.

תרגיל 2.11.30 נתבונן בסדרת הטלות של מטבע הוגן. נסמן ב־$\alpha_{u|v}$ את תוחלת מספר הטלות עד להופעת הרצף u, אם הרצף האחרון שהוטל הוא v. בפרט נסמן $\alpha_u = \alpha_{u|\varepsilon}$ כאשר ε מסמן את הרצף הריק. הראה כי לכל שלושה רצפים u, v, w:

1. אם u סיפא של v אז $\alpha_{u|v} = 0$; אחרת $\alpha_{u|v} = 1 + \frac{1}{2}(\alpha_{u|v0} + \alpha_{u|v1})$.

2. אם v רישא של u, אז $\alpha_u = \alpha_v + \alpha_{u|v}$.

3. $\alpha_{u|wv} \leq \alpha_{u|v}$ ובפרט $\alpha_{u|v} \leq \alpha_u$.

דוגמא 2.11.31 כדי להדגים את השפעת גודל הזכרון על ההתפלגות, להלן חמישה טקסטים אקראיים שיוצרו על בסיס ההתפלגות של חמישה חומשי התורה (לפי נוסח המסורה, עם נקודה בסופי הפסוקים). ההתפלגות הראשונה נבנתה על־ידי הגרלת אותיות לפי התפלגות אחידה (הוספנו כמה רווחים כדי לשבור את השורות). בשניה הוגרלו האותיות לפי השכיחות היחסית שלהן. בשלישית כל אות מן השניה ואילך הוגרלה לפי השכיחות שלה כעוקבת אחרי האות הקודמת (כאשר הרווח והנקודה נחשבים כאותיות לכל דבר). בדוגמא הרביעית כל אות, מן השלישית ואילך, הוגרלה לפי השכיחות שלה כעוקבת אחרי שתי הקודמות לה, ובדוגמא האחרונה השתמשנו בהתפלגות הרביעיות, כך שההתפלגות של כל אות תלויה בשלוש הקודמות לה.

1. התפלגות אחידה: תבזגרםףזו באפקבססתההחדכהטדזזדףוףגהפאחטםעזופט חכפקסתכגצדההזה.למע פולרססקספהר דסגצהסתב.פנבנפחצפפ בדדם ם צביתשףגפהחטצאאזנצצאבםלסוסנטףםסס.ן תדי.קו ר.פנו.י .גוכוגסגדחינסטגחתונעצטטס.רע.ש טף.וגמחכידעדתזדפסאסצעבחתילעשבסמבחס.בף לפטקבתאהיזה פצ. כבדאעדידשיטרז.ווכויוסבימחסל...

2. **התפלגות האותיות:** ת רז לעתאהיתעשחד ה נהספחא השאאל שתרהילהן המוויבכירע ו אבנאבא ע.האהש ה ייהשססאהה יבתלהיסוילבהאביאוב כ ה ד וש ל צוד תייהלוונ ו .בו חאבויתה ראצפר תררדא ענבהסיאלכולשים שודם הדנ אטמשלונאדקוע וביסמב אאונחהאויה ההדש נולעיב אאאבלמעמיצמ ר ה.האהיה ושהשטנ דנ...

3. **התפלגות הזוגות:** אשישח א ו. א תהואכלבדת הס עוטואמבכהיוהועואתחבריא לדבכתן אכתעפה לך אתיוים שפישתך ים שם לנחת. משרח ל א בנוההעת מניהים והות. ל אמא ב תיהסמכל הקר א ארץ מקה ע י ועשפעשר היה בן עים ו ה ל. ח או ות אלי אשבחמחשרב. עלדהי בים השת הנעמר ים ע אשרבנתב איהאם ול אלתם ו י...

4. **התפלגות השלשות:** הקד וא תקב לא קדש לאמרועל כי יריב וישמך וצבאנו את הגידך. ובני. ועלה לותה ליעשה לך שבעלה חיוס אתו עשה בין. ונהוה ארצחו אשריתנו ושמן מצרים לו לאמר אישבעון רע טה אשה המו. כי יה השבענו ליו הואתולדו. ואת אדמושם. וית המחציר הואת בלך יב כבליל השית או ומרו אל יד ת...

5. **התפלגות הרביעיות:** ל יחרם אשר עד ותכלת הבאים עשית קרבנו רעהו עדת ארשה ועמוד ותכחשו את פרע מאת הגבריתם יקדש ומגרר ומיו את פרעה טרם ימים ונד נזר זניהם אשר לא תשמע בני ארון מכשפה לא תהיה הנתכס בחג הסרד בן חמשים. וישבים שה יפרש ובין המשקדיו תחשך ולא אליך גזלה על הנגב ויאמר אבן נ...

איך יראה טקסט אקראי המבוסס על ההתפלגות של עשיריות?

פרק 3

מבוא לסטטיסטיקה

3.1 אמידה

במקרים רבים אנו יודעים שההתפלגות של משתנה מקרי שייכת למשפחה מסויימת (למשל, התפלגות מעריכית או התפלגות נורמלית), אבל איננו יודעים את הפרמטרים המתאימים. הנחה בסיסית זו נקראת **מודל**. למשל, מספר המטאורים הנצפים מדי דקה בלילה מסויים מתפלג פואסונית. המודל קובע שהמספר $X \sim \mathrm{Poi}(\lambda)$, כאשר λ אינו ידוע. בניית המודל (סטטיסטי או אחר) היא הצעד הראשון בשיטה המדעית: בסופו של דבר $X_1, \dots, X_n$ יקבלו ערכים מספריים, ואז הם יחדלו מלהיות משתנים מקריים. הטבע אינו חושף את סודותיו בנקל, ולכן איננו יכולים לדעת כיצד התקבלו המספרים האלה; המודל מציע משפחה של אפשרויות לתהליך כזה, ומשאיר בידי הסטטיסטיקאי את הצורך להכריע מי מהן היא הסבירה ביותר.

תורת האמידה מעריכה פרמטרים של ההתפלגות על־פי ערכים הנדגמים ממנה. כפי שראינו, למשל, בהתפלגות הנורמלית, הפרמטרים עשויים להיות מומנטים של ההתפלגות, כמו התוחלת והשונות, אבל לגיטימי לנסות לאמוד גם דברים כמו p/q בהתפלגות בינומית, $e^{-\lambda}$ בהתפלגות פואסון, וכן הלאה.

בתורת האמידה, ובהסקה סטטיסטית בכלל, ה**אוכלוסיה** היא כלל הערכים הרלוונטיים (למשל, אורך החיים של כל הנורות שמייצר המפעל), וה**מדגם** הוא אוסף (בדרך כלל קטן) של ערכים שנאספו מן האוכלוסיה.

הניתוח המתמטי של המדגם, כמייצג של האוכלוסיה, מבוסס על ההנחה שהדגימה אינה מוטה (כלומר, לכל פרט יש סיכוי שווה להופיע במדגם) - אחרת יש לבצע התאמות שונות ומשונות. זהו אינו קורס בסטטיסטיקה מעשית, ולכן לא נעסוק בהרחבה בהטיות דגימה אפשרויות. מכיוון שפטור בלא כלום אי אפשר, נסתפק ברמזים: לא כל הפרטים באוכלוסיה זמינים לצרכי דגימה (חסרי בית; תושבי חו"ל שיגיעו ליום הבחירות); תהליך הדגימה יוצר הטיות (בדגימה לפי מספרי טלפון יש לבעלי שני קווים סיכוי מוגבר

להופיע; בדגימת אנשים בתחנת רכבת או בסניף דואר מגיעים לנוסעי רכבות ושולחי דואר); אנשים נוטים לשקר בסקרים (למשל בנושא שכר, הרגלים אישיים, אמונות, 'האם עברת הטרדה מינית') או להבין את השאלות אחרת מן הפרשן, כשעוסקים בהעדפות ודעות התשובה אינה יציבה (האם אתה אוהב כרוב? לפעמים.); ועוד ועוד.

3.1.1 אמידה נקודתית

אומד נקודתי עבור פרמטר מסויים הוא פונקציה של המדגם שאינה תלויה בפרמטר (כל פונקציה כזו נקראת **סטטיסטי**) שערכה אמור להיות קרוב לפרמטר המבוקש. לדוגמא, אם המדגם כולל $X_1,\dots,X_n \sim N(\mu,\sigma^2)$, כאשר μ,σ^2 אינם ידועים, אז X_3 ו־$\frac{X_1+X_2}{2}$ הם אומדים אפשריים עבור μ. ערך של האומד נקרא **אומדן**.

כדי להבין איך פועל אומד נקודתי, יש לקרוא את התהליך בסדר הנכון. הפרמטר θ הוא קבוע, אבל אינו ידוע לנו (אחרת לא היה צורך לאמוד אותו). עם זאת, הפרמטר קובע את ההתפלגות F_θ (היינו, זוהי משפחה של התפלגויות, ולכל ערך מתאימה התפלגות אחרת), ונתוני המדגם נאספים מתוך ההתפלגות F_θ. לאומד, המורכב מנתוני המדגם, יש התפלגות, שגם היא תלויה ב־θ. בפרט, התוחלת של האומד תלויה ב־θ (אבל לא בנתוני המדגם - הרי זו תוחלת).

הגדרה 3.1.1 נניח ש־$X_1,\dots,X_n \sim F_\theta$, כאשר θ הוא פרמטר הקובע את ההתפלגות. סטטיסטי $T = t(X_1,\dots,X_n)$ הוא **אומד חסר הטיה** של הפרמטר θ אם לכל ערך של θ, התוחלת של T בהנחה ש־$X_1,\dots,X_n \sim F_\theta$ שווה ל־θ.
(באופן כללי יותר, סטטיסטי שהתוחלת שלו היא תמיד $\tau(\theta)$ הוא אומד חסר הטיה ל־$\tau(\theta)$.)

דוגמא 3.1.2 הזמן הנדרש לתוכנית מחשב מסויימת להתחיל לרוץ מתפלג מעריכית עם תוחלת $30 + 12\theta$ שניות, כאשר θ הוא מספר הווירוסים המתרוצצים בזכרון. יהיה נאיבי מצידנו לצפות שזמן הריצה $X \sim \mathrm{Exp}(\theta)$ יהיה בדיוק $30 + 12\theta$, ובכל זאת, $T = \frac{1}{12}(X-30)$ הוא אומד חסר הטיה למספר הווירוסים, משום שהתוחלת שלו היא θ.
תרגיל. מצא אומד חסר הטיה המקבל ערכים שלמים.

תרגיל 3.1.3 בכל התפלגות, הממוצע $\frac{1}{n}(X_1+\dots+X_n)$ הוא אומד חסר הטיה של התוחלת.

תרגיל 3.1.4 במה עדיף הממוצע $\frac{1}{n}(X_1+\dots+X_n)$ על־פני האומד X_2, אם ממילא שניהם חסרי הטיה?

תרגיל 3.1.5 כשרוצים לאמוד את מספר התומכים במפלגת 'כרוב וחסה', מניחים שיש באוכלוסיה שעור מסויים, p, של תומכים. כשבוחרים את המדגם האקראי, הסיכוי של כל משתתף במדגם להשתייך לקבוצת התומכים הוא p, ולכן תוצאות

המדגם מתפלגות $X_1,\dots,X_n \sim b(p)$. בדוק שהממוצע הוא אומד חסר הטיה ל־p. מה שונותו? איך היא תלויה ב־n (האם כדאי לקחת מדגם גדול פי 100?) איך תלויה השונות ב־p, ומה משמעות התוצאה הזו בבואנו לאמוד את מספר התומכים במפלגה שעל גבול אחוז החסימה?

תרגיל 3.1.6 מה דעתך על פרסום (אפרת וייס, 3/11/2002, ***Ynet***) שכותרתו "מחקר: 141,710 נשים מוכות בישראל"?

תרגיל 3.1.7 חשב את התוחלת של שונות המדגם $s^2 = \frac{1}{n}\sum(X_i - \bar{X})^2$, והסק שזהו אינו אומד חסר הטיה לשונות האוכלוסיה σ^2.

דוגמא 3.1.8 נניח ש־$X_1,\dots,X_n \sim N(\mu,\sigma^2)$.

1. הממוצע הוא אומד חסר הטיה לתוחלת μ.
2. $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ הוא אומד חסר הטיה של השונות σ^2.

למרות ש־$\mathbf{E}(S^2) = \sigma^2$, אין סיבה לצפות ש־$\mathbf{E}(S) = \sigma$; אכן, לא קיים אומד חסר הטיה של סטיית התקן σ בהתפלגות כללית.

תרגיל 3.1.9 רוצים לאמוד את האחוז p של הילדים הפוחדים מחושך. חוששים שהילדים לא יענו על שאלה כזו בכנות, ולכן מבקשים מהם לפעול כדלקמן: כל ילד יטיל מטבע; אם יצא 'עץ' הוא יאמר את התשובה הנכונה, ואם יצא 'פלי' הוא יטיל מטבע שני, ויתן תשובה אקראית (כן או לא) על־פי התוצאה שקיבל שם. הילדים לא יחששו לשתף פעולה, משום שתשובה חיובית יכולה להתקבל גם כתוצאה מהטלת מטבע.

1. אם ילד עונה שהוא פוחד מחושך, מה הסיכוי שזה אכן כך?
2. כתוב את הסטטיסטי המתקבל באופן כזה, ובנה ממנו אומד חסר הטיה של הפרמטר המבוקש. (מהי שונות האומד? השווה אותה לשונות של האומד שהיה מתקבל אם אפשר היה לסמוך על התשובות כלשונן.)

רעיון 3.1.10 פיקוד העורף מבצע תרגיל לבדיקת צופרי האזעקה. אם יודיעו מראש שכל אזרח שאינו שומע את האזעקה מתבקש להתקשר ולהתריע על כך, קווי הטלפון בפיקוד יקרסו. נתח את האפשרות שרק שהאזרחים שמספר הזהות שלהם מסתיים ב־345 יתבקשו להודיע במקרה הצורך. מה דעתך על האלטרנטיבה, שלפיה רק אלו שמספר הזהות שלהם מתחיל ב־345 יתקשרו?

תרגיל 3.1.11 בבחירות לנשיאות בארצות הברית זוכה המועמד שצבר את מספר האלקטורים הגדול ביותר. מדינה i ($i = 1, \dots, k$, כאשר למשל $k = 50$) מגיעים n_i אלקטורים, מתוך $N = \sum n_i$ אלקטורים בסך־הכל. הסיכוי של מועמד A לנצח במדינה i (על־פי הסקרים המקומיים) הוא p_i בדיוק. נסמן $p^0 = p$ ו־$p^1 = (1-p)$. הראה שהסיכוי של A לנצח בסופו של דבר הוא $P = \sum_{(\epsilon_1,\dots,\epsilon_k)\in\{0,1\}^k : \sum \epsilon_i n_i > \frac{1}{2}N} (p_1^{\epsilon_1} \cdots p_k^{\epsilon_k})$.

קבוצת הווקטורים $(\epsilon_1, \dots, \epsilon_k) \in \{0,1\}^k$ המקיימים את התנאי $\sum \epsilon_i n_i > \frac{1}{2}N$ היא מסובכת למדי, וחישוב ישיר של P עשוי להיות בלתי אפשרי. הצע דרך מהירה לאמוד את הפרמטר P. **הדרכה.** בכל אחת מ־1000 החזרות על הניסוי, מטילים k מטבעות.

השוואת אומדים

המודל קובע ש־$X_1, \dots, X_n \sim F_\theta$. נניח ש־$T = t(X_1, \dots, X_n)$ הוא אומד חסר הטיה של θ. כדי להעריך את איכות האומד, ולבחור בין כמה אומדים אפשריים, מתבוננים בשונות $V_\theta(T)$. גם כאן חשוב להדגיש שהשונות תלויה ב־θ, ולכן יתכן שאומד מסויים יהיה בעל שונות נמוכה משל אומד אחר עבור ערכים מסויימים של הפרמטר, אבל בעל שונות גבוהה יותר במקומות אחרים.

הגדרה 3.1.12 אם $V_\theta(T_1) \leq V_\theta(T_2)$ לכל ערך של θ, אומרים ש־T_1 **עדיף** על T_2.

מן ההגדרה מובן שבהנתן שני אומדים, יתכן שאף אחד מהם אינו עדיף על משנהו.

דוגמא 3.1.13 הראה שהאומד $\bar{X}$ עדיף על $\frac{X_1+X_n}{2}$ בתור אומדים לתוחלת בהתפלגות $N(\mu, \sigma^2)$.

תרגיל 3.1.14 כל אחד מן המשתנים המקריים הבלתי תלויים $X_1, \dots, X_n$ הוא אומד חסר הטיה לגודל מסויים, עם שונות $\mathbf{E}(X_i) = \sigma_i^2$. צירוף ליניארי $Y = \alpha_1 X_1 + \cdots + \alpha_n X_n$ הוא אומד חסר הטיה אם $\alpha_1 + \cdots + \alpha_n = 1$. מצא את האומד בעל השונות הקטנה ביותר מצורה זו. הראה שעבורו $\mathbf{V}(Y) = (\sum \sigma_i^{-2})^{-1}$, וערך זה קטן מכל אחת מן השונויות σ_i^2.

תרגיל 3.1.15 נניח ש־$X_1, \dots, X_n \sim U[0, \theta]$.

העזר בתרגיל 2.6.6 כדי להראות שכל אחד מסטטיסטיי הסדר המתוקנים $\frac{n+1}{k} X_{(k)}$ מהווה אומד חסר הטיה לפרמטר θ, בעל שונות $\frac{n+1-k}{(n+2)k}\theta^2$. השווה את האומד $\frac{n+1}{n}\max\{X_i\} = \frac{n+1}{n}X_{(n)}$ לכפליים הממוצע $\frac{2}{n}(X_1 + \cdots + X_n)$. איזה אומד עדיף?

הגדרה 3.1.16 אומד חסר הטיה שהוא עדיף על כל אומד חסר הטיה אחר, נקרא **אומד חסר הטיה בעל שונות מינימלית במידה שווה** (ובקיצור, על־פי ראשי התיבות של השם האנגלי, $UMVUE$).

ההפתעה הגדולה היא שעבור מודלים רבים, אומד כזה אכן קיים; מקוצר היריעה לא נוכל להרחיב בכיוון זה.

אומד נראות מקסימלית

נניח שהצפיפות של המשתנים X_i היא $f(x;\theta)$, כאשר θ פרמטר לא ידוע הקובע את ההתפלגות. **הנראות** של המדגם $X_1,\ldots,X_n$ היא המכפלה $L(X_1,\ldots,X_n;\theta) = \prod f(X_i;\theta)$.

אומד הנראות המקסימלית של θ הוא הפונקציה $T = t(X_1,\ldots,X_n)$ המביאה את $L(X_1,\ldots,X_n;t(X_1,\ldots,X_n))$ למקסימום.

באופן מעשי, כדי למצוא את אומד הנראות המקסימלית, אפשר לגזור לפי θ את $\log L(X_1,\ldots,X_n;\theta)$, ולהשוות לאפס.

תרגיל 3.1.17 1. מצא את אומד הנראות המקסימלית עבור μ כאשר המדגם מגיע מהתפלגות נורמלית $X_1,\ldots,X_n \sim N(\mu,1)$.

2. מצא את אומד הנראות המקסימלית עבור θ אם $X_1,\ldots,X_n \sim N(\theta,\theta^2)$.

3. מצא את אומד הנראות המקסימלית עבור θ אם $X_1,\ldots,X_n \sim U(0,\theta)$.

3.1.2 רווחי סמך

באמידה נקודתית מציעים סטטיסטי (היינו, פונקציה של המדגם), האמור "לקלוע" אל הערך של הפרמטר. באמידת רווח המטרה שונה:

בעיה 3.1.18 נאסף מדגם בלתי תלוי $X_1,\ldots,X_n \sim F_\theta$, מהתפלגות השייכת למשפחה מוכרת של התפלגויות, התלויה בפרמטר θ שערכו אינו ידוע. יש ליצור מנתוני המדגם רווח (T_1,T_2), כאשר T_1,T_2 הם סטטיסטים, כך שהסיכוי

$$P_\theta(T_1 < \tau(\theta) < T_2)$$

שווה למספר $1-\alpha$ שנקבע מראש.

הקטע (T_1,T_2) נקרא **רווח סמך** בעל **רמת מובהקות** α עבור $\tau(\theta)$. גודל טיפוסי ל־α הוא 0.05 או 0.01.

אנו מסמנים $P_\theta(\cdot)$ כדי להדגיש שההסתברות למאורע תלויה בפרמטר: זוהי ההסתברות במקרה ש־$X_1,\ldots,X_n \sim F_\theta$.

לאחר ביצוע הליך הדגימה בפועל, המשתנים המקריים $X_1,\ldots,X_n$ מקבלים ערכים מספריים, וכך הופכים גם קצות הקטע T_1,T_2 למספרים, נאמר t_1,t_2.

זו מכשלה נפוצה לומר שבמקרה זה, "הסיכוי לכך שהפרמטר θ נמצא בין t_1 ל־t_2 הוא $1-\alpha$". ניסוח זה שגוי בתכלית, משום שלפרמטר אין התפלגות - הוא מספר (וגם אם הוא נקבע על־פי התפלגות כלשהי, התפלגות זו אינה נלקחת בחשבון בחישוב הרווח). אם כך, הסיכוי לכך שהפרמטר יהיה בין שני מספרים הוא או אפס או אחד

(גם אם איננו יודעים איזו אפשרות היא הנכונה). א־פריורי, הסיכוי לכך שהפרמטר יהיה שייך לקטע הוא בדיוק $1-\alpha$; אבל לאחר מעשה, גורל הניסוי כבר נגזר, לשבט (בסיכוי α) או לחסד, והוא איננו מאורע הסתברותי.

השיטה הפשוטה ביותר לבניית רווח סמך היא **שיטת הכמות הצירית**, שלפיה אנו מוצאים פונקציה של המדגם ושל הפרמטר, שהתפלגותה אינה תלויה בפרמטר.

3.1.3 רווחי סמך לפרמטרים של ההתפלגות הנורמלית

דוגמא 3.1.19 נניח ש־$X_1,\dots,X_n \sim N(\mu,\sigma^2)$. אז ההתפלגות של $Q=\frac{X_1-\mu}{\sigma}$ אינה תלויה בפרמטרים, ולכן היא כמות צירית.

נחוצים לנו סימונים שיהפכו את פונקציית ההצטברות של ההתפלגויות הרלוונטיות:

הגדרה 3.1.20 לכל $0<\gamma<1$, z_γ הוא המספר (היחיד) כך שמשתנה מקרי $Z\sim N(0,1)$ מקיים $P(Z<z_\gamma)=\gamma$.

בדומה לזה, $t_{m,\gamma}$ מוגדר כך שמשתנה מקרי $Z\sim t_m$ מקיים $P(Z<t_{m,\gamma})=\gamma$, וכך גם $P(W<\chi^2_{m,\gamma})=\gamma$ כאשר $W\sim\chi^2_m$.

את הערכים z_α אפשר לקרוא בטבלה מתאימה. הערכים השכיחים, שכדאי לזכור בעל־פה, הם $z_{0.95}=1.645$, $z_{0.975}=1.96$, $z_{0.99}=2.33$. בגלל הסימטריות של ההתפלגות הנורמלית, $z_{1-\alpha}=-z_\alpha$. אם $\alpha<0.5$, $z_\alpha<0<z_{1-\alpha}$.

רווח סמך לתוחלת

נניח ש־$X_1,\dots,X_n\sim N(\mu,\sigma^2)$ משתנים נורמליים בלתי תלויים. אנו מבקשים רווח סמך עבור התוחלת μ, שאינה ידועה. נסמן $\bar{X}_n=\frac{X_1+\cdots+X_n}{n}$.

אם השונות σ^2 ידועה, אז $Q=\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ היא כמות צירית, בעלת התפלגות $N(0,1)$.

לפי הגדרה 3.1.20 של $z_{1-\frac{\alpha}{2}}$, $P(\left|\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right|>z_{1-\frac{\alpha}{2}})=\alpha$, ומכאן נובע ש־

$$P(\left|\bar{X}_n-\mu\right|>z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})=\alpha.$$

מסקנה 3.1.21 ש־$X_1,\dots,X_n\sim N(\mu,\sigma^2)$ משתנים נורמליים בלתי תלויים בעלי שונות ידועה σ, אז $(\bar{X}_n-z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\bar{X}_n+z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$ הוא רווח סמך בעל רמת מובהקות α לפרמטר μ.

(זוהי רק דוגמא אחת למשפחה שלמה של רווחי סמך עבור μ, שכולם בעלי אותה רמת מובהקות: $(\bar{X}_n+z_\gamma\frac{\sigma}{\sqrt{n}},\bar{X}_n+z_{1-\alpha+\gamma}\frac{\sigma}{\sqrt{n}})$, כאשר $0\leq\gamma\leq\alpha$ הוא מספר כלשהו. המקרה הקיצוני $\gamma=0$ נותן את רווח־הסמך החד־צדדי $(-\infty,\bar{X}_n+z_{1-\alpha}\frac{\sigma}{\sqrt{n}})$, והמקרה $\gamma=\alpha$ נותן $(\bar{X}_n+z_\alpha\frac{\sigma}{\sqrt{n}},\infty)$.)

תרגיל 3.1.22 חשוב איך משפיע שינוי בכל אחד מהפרמטרים α, σ, n על רוחבו של הרווח לתוחלת. נסח את המסקנה במלים (ככל שרמת המובהקות קטנה יותר, ...; ככל שהמדגם גדול יותר, ...)

תרגיל 3.1.23 את המשקל הממוצע של אדם בוגר בישראל אפשר להעריך באמצעות מדגם בגודל 200 ('אפשר להעריך' פירושו שאפשר לקבל רווח סמך ברוחב מסויים וברמת מובהקות מסויימת). מה צריך להיות גודל המדגם אם רוצים להעריך את המשקל הממוצע של אדם מבוגר בארצות־הברית, שאוכלוסיה מרובים פי 50?

כאשר השונות אינה ידועה, איננו יכולים להשתמש ברווח הסמך של מסקנה 3.1.21: יש לבחור בכמות צירית שאינה מערבת את השונות σ. עלינו להחליף את σ^2 באומד שלו; למשל, האומד

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

של טענה 3.1.8.

משפט 3.1.24 **נניח ש־$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ משתנים נורמליים בלתי תלויים. אז:**

1. $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.
2. $\frac{n-1}{\sigma^2} S^2 = \sum (\frac{X_i - \bar{X}_n}{\sigma})^2 \sim \chi^2_{n-1}$.
3. $\bar{X}_n$ ו־S הם בלתי תלויים.
4. $Q = \frac{\bar{X}_n - \mu}{S/\sqrt{n}}$ הוא משתנה מקרי בעל התפלגות t_{n-1}.

הוכחת החלק השלישי והרביעי של המשפט חורגת מגבולות הקורס שלנו: היא עושה שימוש בהתפלגות הנורמלית הרב־ממדית, שהיא התפלגות משותפת של כמה משתנים שכל אחד מהם מתפלג נורמלית.

כמקודם, המספר $t_{n-1,1-\frac{\alpha}{2}}$ מקיים $P(|Q| > t_{n-1,1-\frac{\alpha}{2}}) = \alpha$.

מסקנה 3.1.25 **נניח ש־$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ משתנים נורמליים בלתי תלויים בעלי שונות σ שאינה ידועה. אז $(\bar{X}_n - t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \bar{X}_n + t_{n-1,1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}})$ הוא רווח סמך בעל רמת מובהקות α לפרמטר μ.**

תרגיל 3.1.26 אם הבחירה בידינו, נעדיף רווח סמך (בעל אותה רמת מובהקות) לפרמטר מסויים, ככל שהוא קצר יותר. אורכו של רווח הסמך של מסקנה 3.1.25 אינו קבוע מראש, ולכן ההשוואה אינה מובנת מאליה. בכל זאת, איזה רווח סמך עדיף לדעתך, זה של מסקנה 3.1.21, או זה של מסקנה 3.1.25? האם הרווח העדיף יהיה בהכרח קצר יותר? (נסה להעריך את הסיכויים שלא כך יהיה, בהנחה ש־$\alpha = 0.05$).

רווח סמך להפרש תוחלות

נניח ש־$X_1,\dots,X_n \sim N(\mu_1,\sigma_1^2)$ ו־$Y_1,\dots,Y_m \sim N(\mu_2,\sigma_2^2)$, כולם בלתי תלויים. לפעמים אין צורך באמידה של כל אחת מהתוחלות בנפרד, ומעוניינים רק בהפרש $\mu_1-\mu_2$.

דוגמא 3.1.27 נניח ש־$X_1,\dots,X_n$ הם ציונים של תלמידים שלמדו סטטיסטיקה בשיטה ישנה, ואילו $Y_1,\dots,Y_m$ הם ציוני תלמידים שלמדו לפי שיטה חדשה. לציונים האבסולוטיים יש משמעות נמוכה, משום שהם תלויים ברמת הקושי של הבחינה. ההפרש הוא בעל משמעות רבה יותר, משום שהוא מלמד האם המעבר לשיטה החדשה מועיל, מזיק, או אינו משפיע על התוצאות.

1. **שונויות ידועות:** $\bar{X}_n-\bar{Y}_m \sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m})$, ולכן $\bar{X}_n-\bar{Y}_m \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}$ הם קצוות של רווח סמך ברמת מובהקות α ל־$\mu_1-\mu_2$.

2. **שונויות לא ידועות אבל שוות:** נניח ש־$\sigma_1=\sigma_2=\sigma$. אז $\bar{X}_n-\bar{Y}_m \sim N(\mu_1-\mu_2,\left(\frac{1}{n}+\frac{1}{m}\right)\sigma^2)$, אבל σ אינו ידוע; למרבה המזל, אם נסכם נקבל $\frac{1}{\sigma^2}((n-1)S_X^2+(m-1)S_Y^2)\sim\chi^2_{n+m-2}$, ולכן

$$\bar{X}_n-\bar{Y}_m \pm t_{n+m-2,1-\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n}+\frac{1}{m}\right)\frac{(n-1)S_X^2+(m-1)S_Y^2}{n+m-2}}$$

הם קצוות של רווח סמך כבסעיף הקודם.

3. **שונויות לא ידועות:** אם השונויות אינן ידועות ואין סיבה להניח שהן שוות, אז לא ידוע רווח סמך מדוייק, ומשתמשים במקום זה בקירוב Satterthwaite (1964):

$$Q=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

מתפלג בקירוב בהתפלגות t_ν, כאשר $\nu=\frac{(S_X^2/n+S_Y^2/m)^2}{\frac{(S_X^2/n)^2}{n-1}+\frac{(S_Y^2/m)^2}{m-1}}$.

יש מקרה נוסף שבו רוצים לתת רווח סמך להפרש תוחלות ־ אם **המדגמים מזווגים**, כלומר, $X_1,\dots,X_n\sim N(\mu,\pi^2)$ ו־Y_i, בהנתן X_i, מתפלג נורמלית $N(X_i+\delta,\sigma^2)$. את ההנחה האחרונה אפשר לנסח גם כך: $Y_i-X_i\sim N(\delta,\sigma^2)$.

מסקנה 3.1.28 אם המדגמים מזווגים ורוצים לאמוד את הפרש התוחלות, אז מפעילים רווח סמך לתוחלת בהתפלגות נורמלית על ההפרשים Y_i-X_i.

רווח סמך לשונות

אם רוצים לבנות רווח סמך לתוחלת והשונות אינה ידועה, יש לאמוד אותה. לפעמים ממקדים את תשומת הלב בשונות עצמה, ולשם כך מתבקשת יצירת רווח סמך גם עבורה.

נפעל שוב לפי שיטת הכמות הצירית. לפי משפט 2.3.1.24, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$. מכאן שלכל α,

$$P\left(\chi^2_{n-1,\frac{\alpha}{2}} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{n-1,1-\frac{\alpha}{2}}\right) = 1-\alpha,$$

ולכן

מסקנה 3.1.29

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)S^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)$$

הוא רווח סמך לשונות, ברמת מובהקות α.

הסימטריות כאן היא בכך שהסיכוי שהקטע יפול מימינו של σ^2 שווה לסיכוי שהקטע יפול לשמאלו.

ומה אם רוצים להשוות שתי שונויות? נניח ש־$X_1,\dots,X_n \sim N(\mu_1,\sigma_1^2)$, $Y_1,\dots,Y_m \sim N(\mu_2,\sigma_2^2)$, כולם בלתי תלויים. לפי משפט 3.1.24, $\frac{(n-1)S_X^2}{\sigma_1^2} \sim \chi^2_{n-1}$ ובאותו אופן $\frac{(m-1)S_Y^2}{\sigma_2^2} \sim \chi^2_{m-1}$, ולכן היחס $\left(\frac{S_X^2}{S_Y^2}\right) / \left(\frac{\sigma_1^2}{\sigma_2^2}\right)$ מתפלג $F_{n-1,m-1}$.

מסקנה 3.1.30

$$\left(\frac{S_X^2/S_Y^2}{F_{n-1,m-1,1-\frac{\alpha}{2}}}, \frac{S_X^2/S_Y^2}{F_{n-1,m-1,\frac{\alpha}{2}}}\right)$$

הוא רווח סמך ליחס השונויות σ_1^2/σ_2^2 ברמת מובהקות α.

רווח סמך לפרופורציה

נניח ש־$X \sim \mathrm{Bin}(n,p)$, כאשר הפרופורציה p אינה ידועה (אפשר לחשוב על המשתנים $X_1,\dots,X_n \sim b(p)$ בתור נתוני המדגם הבלתי תלוי, אלא שבפועל אנו משתמשים בנתוני המדגם רק דרך הסכום שלהם X).

לפי משפט הגבול המרכזי, $\bar{X} = \frac{1}{n}X$ מתפלג בקירוב נורמלית, $N(p,\frac{pq}{n})$ כאשר $q = 1-p$. הפרמטר p, אם כך, הוא התוחלת של ההתפלגות, אלא שאי אפשר להפעיל כאן את השיטות של תת־סעיף 3.1.3 משום שהשונות תלויה בתוחלת. במקום זה מקובל להשתמש בשני קירובים, לפי טווח הערכים של p.

טענה 3.1.31 יהי $X \sim \mathrm{Bin}(n,p)$. אז לכל α, $(\bar{X} - \frac{z_{1-\frac{\alpha}{2}}}{2\sqrt{n}}, \bar{X} + \frac{z_{1-\frac{\alpha}{2}}}{2\sqrt{n}})$ הוא רווח סמך עבור p, ברמת מובהקות α או פחות.

התוצאה נובעת מן החסם הטריוויאלי $pq \leq \frac{1}{4}$: אילו היינו יודעים ש־$\bar{X} \sim N(p, \frac{1}{4n})$, רמת המובהקות היתה בדיוק α. מכיוון שהשונות בפועל קטנה יותר, רמת המובהקות טובה יותר.

תרגיל 3.1.32 מה צריך להיות גודל המדגם, אם רוצים לקבל רווח סמך ברוחב של מנדט אחד לכנסת לאחוז התומכים במפלגה מסויימת שרמת המובהקות שלו 0.01?

אם ידוע מראש ש־p קיצוני (נאמר $p < 0.1$ או $p > 0.9$), החסם $pq \leq \frac{1}{4}$ מפריז ברוחב רווח הסמך; אפשר להניח ש־$\bar{X}$ אומד את p במידה מתקבלת על הדעת של דיוק לצורך השונות, ולהשתמש בזה כדי לקבל רווח סמך לפרמטר עצמו:

טענה 3.1.33 כאשר p קיצוני, $(\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\bar{X}(1-\bar{X})}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\bar{X}(1-\bar{X})}{\sqrt{n}})$ הוא רווח סמך עבור p, ברמת מובהקות מקורבת של α.

תרגיל 3.1.34 חזור על תרגיל 3.1.32 אם רוצים לאמוד את שיעור המצביעים של מפלגה המתנדנדת סביב אחוז החסימה 2%.

3.2 בדיקת השערות

בדיקת השערות היא פרוצדורה המשתמשת בנתוני מדגם שנאסף מהתפלגות, כדי להכריע בהשערה לגבי הפרמטרים של ההתפלגות. למשל, אפשר לבנות פרוצדורה שתכריע האם משקלו הממוצע של קלח כרוב שנמכר ברשת מסויימת גדול מ־2.2 קילוגרם, על־פי מדגם של קלחי כרוב שנמכרו באותה רשת.

הבדיקה מכריעה בין שתי השערות לגבי ההתפלגות: **השערת האפס** H_0 שהיא לרוב ההשערה השמרנית אותה מנסים לדחות, והשערה אלטרנטיבית H_1. לרוב, H_1 היא התופעה שאותה רוצים להוכיח. דחיית H_0 נחשבת להצלחה, משום שהיא מוכיחה את H_1, ברמת מובהקות שנקבעה מראש. כמו בדגימה, גם בבדיקת השערות יש שגיאות שצריך להזהר מהן. אחת הדוגמאות המפורסמות היא 'הטיית המגירה', המתארת את הנטיה של חוקרים (ועורכים) להשאיר במגירה תוצאות שליליות, ולפרסם רק תוצאות חיוביות.

בפרק זה נניח שהשערת האפס **נקודתית**, כלומר קובעת את ההתפלגות באופן חד־משמעי (היא השערה מהצורה $\theta = 5$, ולא $\theta \geq 5$ או $\theta \in (5,7)$. גם השערות מורכבות אפשר לבדוק, אבל לא נטפל בהכללה זו כאן.)

דוגמא 3.2.1 שעור האנשים בעלי הפרעת אישיות חרדתית באוכלוסיה הוא 3%. חוקר משער ששיעורם בקרב בני מזל טלה שונה מן השיעור בקרב בני המזלות האחרים. הפרמטר הלא־ידוע הוא ההסתברות p לכך שבן מזל טלה יפתח הפרעת אישיות חרדתית. השערת האפס תהיה $H_0 : p = 0.03$. ההשערה האלטרנטיבית היא $H_1 : p \neq 0.03$.

3.2.1 השערות ושגיאות

לבדיקת השערות יש שתי תוצאות אפשריות: **דחיה** של השערת האפס, ואי־דחיה שלה. לאפשרות השניה קוראים לפעמים "קבלה" של H_0, אבל חשוב להבין שאין בכוחה של בדיקת השערות להוכיח טענה סטטיסטית.
גם במציאות יש שתי אפשרויות: או שהשערות האפס נכונה, או שההשערה האלטרנטיבית היא הנכונה. בהתאם לכך, יש ארבע אפשרויות:

1. H_0 נכונה, והליך בדיקת ההשערות דוחה אותה. אזעקת שווא. זוהי טעות, הנקראת **טעות מסוג ראשון**.
2. H_0 היא הנכונה, והליך בדיקת ההשערות לא דחה אותה. בהתחשב בנסיבות, זו כמובן התוצאה הרצויה של ההליך.
3. H_1 נכונה, והליך בדיקת ההשערות דוחה את H_0, ובכך מוכיח את H_1. הניסוי הצליח.
4. H_1 נכונה, והליך בדיקת ההשערות לא דוחה את H_0, ובכך מאשש בטעות את H_0. גם זו טעות, הנקראת **טעות מסוג שני**.

דוגמא 3.2.2 חוקר רוצה להוכיח שתוספת ויטמין K לתזונת פגים מעלה את המשקל שלהם מעבר לתזונה הרגילה. ידוע שהוויטמין הזה אינו מזיק. נסמן את העליה במשקל מעבר לצפוי ב־X. השערת האפס (שאותה רוצים לדחות) קובעת שהתוחלת של X היא אפס, כלומר $H_0 : \mu = 0$. כנגדה, ההשערה האלטרנטיבית היא $H_1 : \mu > 0$. נפרש את ארבע האפשרויות, לפי הסדר שבו הן מופיעות לעיל.

1. הוויטמין אינו מועיל, והחוקר 'הוכיח' שהוא כן מועיל. נעשה כאן חוכא ואטלולא מן השיטה המדעית. טעות מסוג ראשון היא טעות חמורה.
2. הוויטמין אינו מועיל, והחוקר נכשל בנסיונו להוכיח את ההיפך. החוקר בזבז את זמנו ואת כספן של קרנות המחקר, אבל הגיע בסופו של דבר לתוצאה הנכונה.
3. הוויטמין מועיל, והחוקר מצליח לדחות את השערת האפס ולהוכיח שזה אכן כך. כולם מרוצים: המחקר מתפרסם בספרות המקצועית, ותזונת הפגים משתפרת.

4. הוויטמין מועיל, אלא שהניסוי לא הצליח להוכיח זאת. זו טעות מסוג שני - לא נעים (בזבוז זמן וכסף), אבל לא נורא (בשנה הבאה יבצע מישהו אחר מחקר על קבוצת תינוקות גדולה פי ארבעה, ואולי יצליח להוכיח את האפקט).

דוגמא 3.2.3 הסיווג של טעות מסוג ראשון כחמורה יותר מן הטעות מסוג שני אינו תורה מסיני. לדוגמא, אם מנסים לחזות פריצת מלחמה באמצעים סטטיסטיים, עדיף לדחות בטעות את השערת האפס (ולגייס את המילואים לחינם), מאשר לקבל בטעות את השערת האפס (ולספוג מתקפת פתע).

דוגמא 3.2.4 לקראת תחילת שנת הלימודים האקדמית, סטודנט אינו בטוח האם הקורס שנרשם אליו מתקיים בסמסטר הראשון או בסמסטר השני. הוא מתכוון להכריע על־ידי סקר בין סטודנטים אחרים (שרבים מהם מבולבלים לפחות כמוהו). דון במשמעות של טעות משני הסוגים במקרה זה.

3.2.2 הליך הבדיקה

נציג תאור מפורט (אם כי פשטני לפרקים) של הצעדים הדרושים לבדיקת השערות.

- רקע תאורטי:
  - זיהוי והגדרת התופעה הנמדדת; זהו המשתנה המקרי X.
  - קביעת המודל, על־פי שיקולים תאורטיים, היורסיטיים וניסויים; המודל הוא משפחת ההתפלגויות $\{F_\theta\}$, שהמשתנה X מתפלג לפי אחת מהן.
  - מנסחים את השערת האפס, $H_0 : \theta = \theta_0$, כאשר θ_0 הוא הערך של הפרמטר שאותו מנסים **לשלול**.
  - מנסחים את ההשערה האלטרנטיבית לפי הידע התאורטי; בדרך כלל $H_1 : \theta \neq \theta_0$, אבל יתכן גם $H_1 : \theta > \theta_0$, $H_1 : \theta = \theta_1$, וכדומה.
  - קובעים את רמת המובהקות α. פירוש הדבר הוא ש־α תהיה ההסתברות לטעות מסוג ראשון.
- רקע סטטיסטי:
  - על פי משפחת ההתפלגויות והפרמטר שבו מדובר, קובעים סטטיסטי (למשל, אומד של הפרמטר).
  - על־פי ההשערות ורמת המובהקות, קובעים **אזור דחיה**, שהוא תחום הערכים שאם הסטטיסטי יפול לתוכו נכריז על דחיית השערת האפס.
  - קובעים את גודל המדגם n, בהתחשב באילוצים תקציביים או בחישוב המבוסס על הסיכוי הרצוי לטעות מסוג שני (שאותו מסמנים ב־β).

- ביצוע הניסוי:
  - אוספים נתוני מדגם $X_1, \ldots, X_n$.
  - מחשבים את הסטטיסטי S.
  - בודקים האם הסטטיסטי נופל לאזור הדחיה.
- פרשנות התוצאות:
  - אם הסטטיסטי נפל לאזור הדחיה, דוחים את השערת האפס. זוהי 'הוכחה' שהשערת האפס אינה נכונה. אם ההשערה נכונה, הסיכוי לדחיה הוא α.
  - אם הסטטיסטי נפל מחוץ לאזור הדחיה, ההשערה אינה נדחית.

הערה 3.2.5 אזור הדחיה מאופיין על־ידי התכונה הבאה: אם H_0 נכונה, אז הסיכוי של סטטיסטי המבחן S ליפול לאזור הדחיה שווה ל־α.

הערה 3.2.6 ככל שאזור הדחיה קטן יותר, נעשה קשה יותר לדחות את השערת האפס, וכך קטן הסיכוי לטעות מסוג ראשון, וגדל הסיכוי לטעות מסוג שני - ולהיפך.

הערך המקובל של α הוא 5%. אם כך, רמת המובהקות המקובלת היא 5%, ויש סיכוי של 5% לטעות מסוג ראשון. כ־5% מן התוצאות החיוביות המוכרזות בספרות המדעית כמובהקות, הוכרזו ככאלה בטעות; ראו http://xkcd.com/882/ לאיור הכשל הזה (אכן רופאים אינם מאמצים שיטות טיפול חדשות על־פי תוצאות של מחקר יחיד, אלא מחכים שהתוצאה תחזור בניסויים אחרים).

ציטוט 3.2.7 "נעשו מחקרים ובהם בדקו יותר ממאה פרמטרים שמאבחנים באמצעות הרורשך, 95% מהם התגלו כלא תקפים מדעית, ורק 5% תקפים" (אביבה לורי, "מרד הפסיכולוגים", "הארץ", *13.12.2007*).

תרגיל 3.2.8 נניח שבודקים סדרה של השערות על מכונה לייצור מספרים אקראיים; בכל המקרים השערת האפס נכונה, ומשערים, כרגיל, שהיא שגויה. איזה אחוז מההשערות "ימצאו "תקפות מבחינה מדעית"?

תרגיל 3.2.9 מפעם לפעם מציגים נתונים השוואתיים על שביעות הרצון של סטודנטים ממוסדות הלימוד שלהם. (1) כיצד היית מציג את התוצאות של סקר שבו התקבלו הערכים הממוצעים א=8.41, ב=8.37, ג=7.91, ד=7.88, ה=7.74? **פתרון**. הייתי חוזר אל הסטטיסטיקאי ודורש לקבל גם את מספר הנבדקים וסטיות התקן. (2) ואם נשאלו 500 תלמידים (100 מכל מוסד) וסטיית התקן בכל המקרים היא 1.5? **פתרון**. יש לחשב עבור כל שני מוסדות האם ההפרש ביניהם מובהק, ולחלק אותם לקבוצות בהתאם.

3.2.3 בדיקת השערות בעזרת רווחי סמך

נסכם: רווח סמך הוא אזור שהפרמטר נמצא בתוכו בסיכוי $1-\alpha$. אזור דחיה הוא אזור ש(אם השערת האפס נכונה) הסטטיסטי נופל מחוץ לו בסיכוי $1-\alpha$. לפי משפט הסיכום הבא, בדיקת השערות ובניית רווחי סמך הם צדדים שונים של אותו מטבע:

משפט 3.2.10 נניח ש־$X_1,\dots,X_n \sim F_\theta$ הם נתוני מדגם בלתי תלויים, כאשר הפרמטר θ אינו ידוע. נניח ש־(T_1,T_2) הוא רווח סמך ברמת מובהקות α. אז $\{\theta_0 \not\in (T_1,T_2)\}$ הוא אזור דחיה אפשרי להשערה $H_0:\theta=\theta_0$.

במלים פשוטות, אפשר לדחות את ההשערה $H_0:\theta=\theta_0$ אם θ_0 אינו נופל בתוך רווח סמך של θ, והסיכוי לטעות מסוג ראשון היא המשלים α של רמת המובהקות של רווח הסמך.

המשפט מאפשר לתרגם כל רווח סמך שבנינו בסעיף הקודם להליך לבדיקת השערה לגבי פרמטר מתאים: התוחלת, הפרש תוחלות, השונות, וכן הלאה.

דוגמא 3.2.11 (בדיקת השערות על התוחלת בהתפלגות נורמלית, כאשר השונות σ^2 ידועה) כשבוחנים את $H_0:\mu=\mu_0$ כנגד ההשערה האלטרנטיבית $H_1:\mu\neq\mu_0$, אזור הדחיה הוא $\left|\bar{X}-\mu_0\right| > z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$. לעומת זאת כשההשערה חד־צדדית, למשל $H_1:\mu>\mu_0$, אזור הדחיה הוא $\bar{X}-\mu_0 > z_{1-\alpha}\frac{\sigma}{\sqrt{n}}$.

נשווה בין אזורי הדחיה של ההשערה הדו־צדדית להשערות חד־צדדיות כאשר $\alpha=0.05$. הערכים בטבלה מתייחסים להכרעה (H_1 לדחייה, H_0 לאי־דחייה), על־פי הערך של הסטטיסטי $Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$, בהתאם להשערה האלטרנטיבית.

		-1.96		-1.645		0		1.645		1.96	
$H_1:\mu\neq\mu_0$	H_1		H_0			H_0			H_0		H_1
$H_1:\mu<\mu_0$	H_1		H_1			H_0			H_0		H_0
$H_1:\mu>\mu_0$	H_0		H_0			H_0			H_1		H_1

דוגמא 3.2.12 (בדיקת השערות על התוחלת בהתפלגות נורמלית, כאשר השונות לא ידועה) פועלים בדיוק כמו בדוגמא 3.2.11 פרט לשני הבדלים: את הערך $\frac{\sigma^2}{n}$, שאינו ידוע, מחליפים באומד חסר ההטיה $\frac{s^2}{n-1}$; ואת ערכי הטבלה $z_{1-\alpha}$ וכדומה מחליפים ב־$t_{n-1,1-\alpha}$.

נסיים בהסבר הרמז שניתן למעלה על הקשר בין גודל המדגם לסיכוי לטעות מסוג שני. נסמן ב־ $\Phi(z)=P(Z<z)$ את פונקציית ההצטברות של המשתנה הנורמלי הסטנדרטי $Z\sim N(0,1)$. כלומר, $\Phi(z_\gamma)=\gamma$.

טענה 3.2.13 עבור התפלגות נורמלית $N(\mu,\sigma^2)$, עם שונות ידועה, משערים את השערת האפס $H_0:\mu=\mu_0$, וכנגדה את ההשערה האלטרנטיבית $H_1:\mu=\mu_1$. נניח ש־$\mu_0<\mu_1$. עבור אזור הדחיה $\{\bar{X}_n>t\}$[1], הסיכוי לטעות מסוג ראשון הוא $\alpha=\Phi(-\frac{t-\mu_0}{\sigma/\sqrt{n}})$, והסיכוי לטעות מסוג שני הוא $\beta=\Phi(-\frac{\mu_1-t}{\sigma/\sqrt{n}})$.

מכאן מתקבלת נוסחת ה־*trade-off* הקושרת בין α ל־β (אם n קבוע):

$$z_\beta+z_\alpha=-\frac{\mu_1-\mu_0}{\sigma/\sqrt{n}};$$

ככל שהסיכוי לטעות מסוג ראשון יורד, הסיכוי לטעות מסוג שני עולה, ולהיפך.

יישום נוסף: כדי להגיע לסיכויי שגיאה α,β, כאשר μ_0,μ_1,σ קבועים, יש לבחור $n\geq\left(\frac{\sigma}{\mu_1-\mu_0}\right)^2(z_\alpha+z_\beta)^2$.

דוגמא 3.2.14 נניח, בדוגמא הקודמת, ש־$\sigma/\sqrt{n}=0.2$, $\mu_0=0$ ו־$\mu_1=1$. מצא את β אם $\alpha=0.05$.

התרגיל הבא מסביר, דרך דוגמא, מדוע איננו מספקים נוסחה דומה לזו של טענה 3.2.13 למקרה של ההשערה האלטרנטיבית $H_1:\mu\neq\mu_0$.

תרגיל 3.2.15 תנובת החלב הממוצעת של פרה היא כ־14 ליטר חלב ביום, עם סטיית תקן 3. מה הסיכוי לכך שממוצע תנובת החלב של מאה פרות יהיה, בדיוק של שלוש ספרות אחרי הנקודה, 14.000? האם תוצאה כזו מוכיחה שתוחלת תנובת החלב היא 14 ולא 14.00002? מה צריך להיות גודל המדגם שיפריך את הטענה $\mu=14$, אם התוחלת האמיתית היא 14.00002?

[1] ככל שדוגמים ערך גדול יותר של $\bar{X}$, יותר מתקבל על הדעת שהתוחלת היא μ_1; לכן לוקחים את אזור הדחיה להיות קרן ימנית.

ביבליוגרפיה

[1] מרצבך־שמרון.