פתח דבר

המוח הוא מכונה מופלאה, המורכבת מיחידות פיזיקליות – ביולוגיות, והמייצרת התנהגות, תחושה ותפיסה. הבנה עמוקה של הקשר בין ההיבט הפיזיקלי של המוח לבין התוצר ההתנהגותי שלו, תובעת פיתוח מצע תאורטי מוצק. אכן, חסרה לנו "תאוריה של המוח". זוהי משימתו העיקרית של תחום המחקר הקרוי "חישוביות עצבית".

הספר הנוכחי מהווה נדבך חשוב לביסוס גישות תאורטיות לתאור ולניתוח פעולת מערכת העצבים. זהו אחד מפירותיו המוצלחים של המרכז ל"חישוביות עצבית" ושל מסלול הדוקטורט ב"מדעי המוח: חישוב ועיבוד מידע", באוניברסיטה העברית.

הקורס "*מבוא לעיבוד מידע ולמידה*" הניתן במסגרת זו, אשר נבנה בשום – שכל על ידי פרופ' תלי תישבי, מהווה בסיס לספר הנוכחי. תלמידי הדוקטור של תלי, לידרור טרוינסקי וגל צ'צ'יק, טרחו להביא את היצירה לידי גמר. תודת המרכז, על תלמידיו ומרציו, שלוחה לעושים במלאכה.

אני מקווה שהספר הזה, המגרה את המחשבה והנעים למראה, יעודד נוספים מאיתנו לבסס את קורפוס החישוביות העצבית בכתובים. ובכך, בלימוד ובכתוב, נתרום תרומה מכרעת לעיצוב דור חדש של תלמידים וחוקרים השותפים לבניית "תאוריה של המוח".

עידן שגב
ראש המרכז ל"חישוביות עצבית"

הקדמה

תחום החישוביות העצבית פועל להבנת התהליכים החישוביים שרקמות עצביות מבצעות כדי לעבד קלט מהסביבה, לקבל החלטות בתנאי אי ודאות, ולהוציא אותן אל הפועל. כדי להצליח בביצוע מטלות מורכבות אלו נדרשת מערכת העצבים ללמוד תכונות של העולם ולייצג אותן באופן יעיל. האופי הסטוכסטי של הקלטים המתקבלים מהעולם, ובמקביל של הפעילות העצבית עצמה, הביאו לכך שמודלים הסתברותיים של העולם הפגינו הצלחה הן במערכות אינטיליגנטיות מלאכותיות והן במחקר מערכות עצביות בבעלי חיים. מסיבות אלו, ידע בהסתברות וסטטיסטיקה, בלמידה חישובית, ובתורת המידע, מהווים היום בסיס הכרחי למחקר של מערכות עצביות, טבעיות ומלאכותיות.

החוברת הנוכחית נבנתה בהתבסס על סדרת הרצאות שניתנה על ידי פרופ' נפתלי תשבי במסגרת הקורס "*מבוא לעיבוד מידע ולמידה*". קורס זה נבנה על ידי פרופ' תשבי כאחד מקורסי הליבה של מסלול הדוקטורט ל- "מדעי המוח: חישוב ועיבוד מידע" במרכז לחישוביות עצבית באוניברסיטה העברית. סדרה של תקצירי ההרצאות נכתבה לפני מספר שנים על ידי לידרור טרוינסקי בהתבסס על הרצאות אלו, והחוברת הנוכחית, שנכתבה על ידי גל צ'צ'יק, מבוססת במדה רבה על תקצירים אלו. החשיבות והצורך באיסוף החומר לכדי קובץ אחיד ואמין המכסה את החומר באופן מסודר הביאו להוצאת החומר במבנה הנוכחי, תהליך שנמשך לסירוגין במהלך שלושת השנים האחרונות.

כבכל פרוייקט מורכב, אנשים רבים עזרו ביצירתו. החוברת נערכה ונכתבה על ידי גל צ'ציק על בסיס הרצאותיו של פרופ תשבי וסדרת תקצירים שנכתבו על ידי לידרור טרוינסקי. יורם זינגר ודנה רון תרמו לגרסה המוקדמת של החוברת. איתי גת תרם את פרק התהליכים המרקוביים החבויים, אמיר גלוברזון עזר בכתיבת הפרק על בחירת מודל, פרופ' עידן שגב תמך ועודד, מיכאל פינק עזר רבות ויצר את רוב האיורים בספר, עמית מילר הגיה את הגרסה הסופית וטיפל בהוצאה לאור, תמר צ'צ'יק הדפיסה את כתב היד, מיכל בן שחר סייעה בהגהת הגרסאות הראשונות, ואורן שריקי תרם מנסיונו בהוצאת חוברת קודמת של המסלול. תודות לעליזה שדמי, רותי סוצ'י ואלינור אטרקצ'י שסייעו בהיבטים שונים, ותודות לתלמידי הקורס "*מבוא לעיבוד מידע ולמידה*", שעזרו בהגהת הגרסאות השונות של כתב היד, וסייעו לסלק טעויות תוכניות וצורניות: דוד טלבי, יובל טסה, יעל ניב, אדם ספירו ואבישלום שליט. תודה מיוחדת למיכאל בראוטבר ויהושע רוזנברג. כל אלו סייעו בהפחתת השגיאות שנפלו, אך אלו שבוודאי נותרו, הן כמובן כולן באחריותנו.

גל צ'צ'יק
לידרור טרוינסקי
נפתלי תשבי

פרק א'

רקע בהסתברות

1.1 מבוא

פרק זה מיועד לשמש תזכורת למספר מושגים בסיסיים בהסתברות, אך הוא מניח ידע פורמלי מוקדם שנרכש בקורס ראשון בהסתברות. בפרט, לא נדון כאן בתכונותיה הבסיסיות של פונקצית ההסתברות, במונחים בסיסיים מתורת הקבוצות או בקומבינטוריקה.

1.1.1 ההסתברות ופירושיה

מושג ההסתברות הוא מושג שכיח בחיי היום-יום, וקל להיתקל במשפטים כמו "יש סיכוי גבוה לגשם מחר". למרות זאת, למושג ההסתברות אין פירוש יחיד שהצליח להפוך למקובל על כל הסטטיסטיקאים והפילוסופים. לאורך השנים הוצעו מספר פירושים מתחרים להסתברות:

הסתברות כשכיחות היחסית של מאורעות

לעתים ניתן לפרש את ההסתברות להתרחשות תוצאה מסויימת של תהליך בתור השכיחות היחסית של התוצאה לעומת תוצאות אחרות אם נחזור על התהליך מספר רב של פעמים בתנאים דומים. לדוגמא, ההסתברות לקבל "1" בהטלת קוביית שש-בש היא 1/6 היות ואם נזרוק קוביה הוגנת פעמים רבות, אז בשישית מהמקרים נקבל "1". הגדרה זו מעוררת מספר בעיות קשות: ראשית, לא מוגדר היטב "מספר רב של פעמים". שנית, המושג "בתנאים דומים" הוא בעייתי במיוחד. ברור כי התנאים הפיסיקליים אינם יכולים להיות זהים (כי אז הקוביה תיפול תמיד על אותו הערך), ולכן חייב להיות מרכיב מקרי בתהליך. מצד שני הגדרה כזו לא יכולה להתאים לתהליכים שאינם יכולים לחזור מספר רב של פעמים, כמו למשל הניסיון להעריך את הסיכוי שאדם יחלה באבעבועות רוח.

הפירוש הקלאסי להסתברות

הגישה הקלאסית מבוססת על הרעיון של תוצאות שיש להן אותה סבירות ומקנה ערך הסתברותי לתוצאה מסויימת לפי מספר הדרכים הזרות ושוות-ההסתברות לקבלת אותה התוצאה. למשל, כשמטילים קוביה, יש סבירות שווה לכל אחת מהתוצאות, ולכן עלינו לייחס לכל תוצאה את ההסתברות שישית. לעומת-זאת, יש שלוש דרכים זרות לקבלת תוצאה זוגית, ולכן ההסתברות לקבלת תוצאה זוגית היא 1/2 = 3/6. באופן כללי אם ישנן k דרכים שונות בהן אפשר להשיג תוצאה מסוימת, מתוך n דרכים אפשריות בכלל, אז ההסתברות לתוצאה תהיה k/n. גם פירוש זה מעלה בעיות עקרוניות כאשר מנסים לבנות תורה פורמלית של הסתברות, במיוחד כאשר אין אפשרות להגדיר את הדרכים הזרות ושוות ההסתברות לקבלת התוצאה, כמו למשל הניסיון להעריך את הסיכוי ל"ראש" במטבע מוטה.

הסתברות כמדד להערכה סובייקטיבית

על פי גישה זו, ההסתברות שאדם מייחס למאורע מייצגת את הערכתו לגבי הסיכוי שמאורע כזה יתרחש. הסתברות כזו מסתמכת על ניסיונו האישי והסובייקטיבי של האדם. לדוגמא, ישנם אנשים המאמינים כי ההסתברות ל "1" בהטלת קוביה היא ששית. תחת הנחות מסוימות, ניתן אכן לבנות תורה פורמלית של הסתברות מתוך גישה זו, אבל ההסתמכות על תפישה סובייקטיבית מקשה לבנות בסיס אובייקטיבי להערכת הסתברויות.

הגישה האקסיומטית

למרות המחלוקת בנוגע למשמעות שיש לייחס להסתברות, הרי שברגע שהוסכם על ייחוס הסתברויות לתוצאות מסוימות בניסוי, ניתן לבנות תורה מתמטית קונסיסטנטית של הסתברות.
על פי הגישה האקסיומטית, ההסתברות מוגדרת כמידה על פני מרחב מדגם, הנדרשת לקיים מספר אילוצים תוך שימוש בטרמינולוגיה של תורת הקבוצות. גישה זו מאפשרת להתייחס להסתברות כאל הרחבה טבעית של הלוגיקה.

מטרת החלקים הבאים בפרק זה היא לתאר מקצת התוצאות בתורת ההסתברות.

1.2 הסתברות מותנית ונוסחת בייס

1.2.1 הסתברות מותנית

נניח שבוצע ניסוי שמרחב תוצאותיו האפשריות הוא Ω. נטפל כעת בדרך שבה ההסתברות למאורע A משתנה כאשר אנחנו יודעים שמאורע אחר B קרה. הסתברות חדשה זאת של המאורע A נקראת "**ההסתברות המותנית של המאורע A בהינתן המאורע B**" ומסומנת $P(A|B)$. כאשר אנחנו יודעים כי המאורע B קרה, אזי אנחנו יודעים שתוצאת הניסוי היא אחת מהתוצאות הנכללות ב- B. מכאן שלצורך הערכת הסתברות המאורע A, אנו צריכים לקחת בחשבון רק תוצאות

מתוך B שבהן מתקיים המאורע A. מתבקש אם כן להגדיר את ההסתברות המותנית כחלק היחסי של תוצאות B המקיים את המאורע A.
ובאופן פורמלי,

$$\Pr(A\,|\,B)=\frac{\Pr(A\cap B)}{\Pr(B)} \tag{1.1}$$

נשים לב כי עבור מאורעות בלתי תלויים $\Pr(A\cap B)=\Pr(A)\Pr(B)$ ומהצבה בנוסחה לעיל נקבל $\Pr(A|B) = \Pr(A)$.

דוגמא: הטלת זוג קוביות

נניח כי הוטלו שתי קוביות, וידוע לנו כי הסכום S של שתי ההטלות הוא אי-זוגי. נחשב את הסיכוי כי סכום ההטלות קטן מ- 8. אם נסמן ב- A את המאורע "S<8" וב- B את המאורע "Sהוא אי זוגי" אזי המאורע $A\cap B$ הוא המאורע "S שווה ל- 3, 5 או 7". מתוך ההתפלגות של סכום שתי קוביות נוכל לרשום

$$\Pr(A\cap B)=\frac{2}{36}+\frac{4}{36}+\frac{6}{36}=\frac{12}{36}=\frac{1}{3}$$
$$\Pr(B)=\frac{2}{36}+\frac{4}{36}+\frac{6}{36}+\frac{4}{36}+\frac{2}{36}=\frac{18}{36}=\frac{1}{2}$$

ולכן

$$\Pr(A\,|\,B)=\frac{\Pr(A\cap B)}{\Pr(B)}=\frac{1/3}{1/2}=\frac{2}{3}$$

1.2.2 נוסחת בייס

מתוך הגדרת ההסתברות המותנית (1.1) נקבל כי מתקיים

$$\Pr(A\cap B)=\Pr(A\,|\,B)\Pr(B)$$
$$\Pr(A\cap B)=\Pr(B\,|\,A)\Pr(A) \quad ,$$

ומכאן נקבל נוסחה הנקראת **נוסחת בייס**

$$\Pr(A\,|\,B)=\frac{\Pr(B\,|\,A)\Pr(A)}{\Pr(B)}. \tag{1.2}$$

1.2.3 הסתברויות אפריוריות ואפוסטריוריות

בניתוח המקרה של הסתברויות מותנות, טיפלנו במאורע A עם הסתברות ידועה Pr(A) וראינו כיצד הסתברותו משתנה כאשר מתגלות לנו עובדות נוספות כגון העובדה שהמאורע B מתקיים. מצב זה שכיח במיוחד כאשר אנחנו מבצעים ניסויים כדי ללמוד על התפלגות של משתנה מקרי לא ידוע. במקרה כזה יש לנו ידע מוקדם או "ניחוש" על ההסתברות Pr(A) (שנקראת ההסתברות האפריורית למאורע A),

ואנחנו משתמשים בתצפיות מניסויים על מנת להסיק את ההסתברות ה"אמיתית" של A. ההסתברות המותנית בתוצאות הניסויים שביצענו נקראת ההסתברות האפוסטריורית.

דוגמא: שימוש בתצפית חדשה לעדכון ידע אפריורי

ידוע כי שיעור החולים בסרפדת גמדית באוכלוסיה הוא אחד לאלף. משרד הבריאות פיתח בדיקה חדשה לאבחון הסרפדת בעלת אמינות של 90 אחוז, וזאת במובן הבא: אם אדם חולה בסרפדת אזי יש סיכוי של 10 אחוז שהבדיקה תטען כי הוא בריא, ואם אדם בריא אזי יש סיכוי של 10 אחוז שהבדיקה תטען כי הוא חולה. באחד הימים, תוך שוטטות בקניון החביב עליך, שוכנעת על ידי סוכן מכירות ממולח לבצע את הבדיקה, כשקיבלת את התוצאות בדואר, הופתעת לגלות שאובחנת כחולה סרפדת. מהו הסיכוי כי אכן נדבקת בסרפדת? אנשים רבים יניחו כי ההסתברות קרובה ל- 90 אחוז. אולם התחשבות בהסתברות האפריורית ובנוסחת בייס יפחית את הבהלה בלבך: נסמן ב- A את המאורע כי אדם חולה בסרפדת, וב- B את המאורע כי בדיקת משרד הבריאות טוענת כי חלית בסרפדת. נשתמש בנוסחת ההסתברות השלמה $P(B)=P(B|A)P(A)+P(B|A^c)P(A^c)$ ונקבל

$$
\begin{aligned}
\Pr(A\mid B) &= \frac{\Pr(B\mid A)P(A)}{\Pr(B\mid A)\Pr(A)+\Pr(B\mid A^c)\Pr(A^c)} \\
&= \frac{(0.9)(0.001)}{(0.9)(0.001)+(0.1)(0.999)} \\
&= 0.009
\end{aligned}
$$

מכאן, שאמנם ההסתברות שהינך חולה גדלה פי 9, אך ההסתברות האפוסטריורית למחלה היא עדיין קטנה במידה מרגיעה.

1.3 משתנים מקריים ופונקציות התפלגות

משתנה מקרי (מ"מ - random variable) הוא משתנה שיכול לקבל מספר ערכים, וערכו נקבע על פי תוצאה של תהליך או ניסוי שתוצאתו אינה ידועה מראש. לדוגמא, המספר שיופיע על קוביית שש-בש לאחר שנטיל אותה, אינו ידוע לפני ההטלה. הקבוצה הכוללת את כל הערכים האפשריים של המשתנה המקרי נקראת **מרחב המדגם** וסימונו Ω. **מאורע** הוא תת קבוצה של מרחב המדגם, דהיינו קבוצת תוצאות אפשריות בניסוי. מאורע יכול לכלול תוצאה אחת בלבד או מספר תוצאות. לדוגמה המאורע "תוצאה זוגית בהטלת הקוביה" מכיל את המאורעות הפשוטים {"2","4","6"}.

1.3.1 משתנים מקריים בדידים ורציפים

משתנה מקרי בדיד, הוא משתנה מקרי שיכול לקבל מספר סופי או לכל היותר בן מניה של ערכים. לדוגמה: תוצאת הטלה קוביה (6 ערכים אפשריים), או מספר פוטנציאלי הפעולה שאירעו בתא עצב בפרק זמן נתון (לכל היותר מספר בן מניה של ערכים).

פונקצית ההסתברות של מ"מ (probability function – p.f.) קובעת לכל ערך אפשרי x של המשתנה המקרי הבדיד, את הסיכוי p(x) לקבלת ערך זה. היא מקיימת את הדרישות כי

א. $\sum_{\{x\in\Omega\}} p(x)=1$

ב. $0\le p(x)\le 1 \qquad \forall x\in\Omega$

פונקצית ההתפלגות המצטברת של מ"מ (cumulative distribution function – c.d.f) היא הפונקציה $F(k)=P(x<k)$. פונקציה היא פונקציה מונוטונית המקיימת

$$\lim_{x\to-\infty}F(x)=0\ ,\ \lim_{x\to\infty}F(x)=1\ .$$

פונקצית ההתפלגות המצטברת של משתנה מקרי בדיד תהיה "פונקצית מדרגות" עם נקודות אי רציפות בכל מקום בו המ"מ מקבל את הערך בהסתברות חיובית.

משתנה מקרי רציף הוא משתנה שמקבל ערכים בתחום רציף. לדוגמה, הזמן שעובר בין התרחשות שני פוטנציאלי פעולה יכול לקבל ערכים ממשים חיוביים כלשהם. עבור משתנה מקרי רציף הסיכוי לקבל ערך מסוים (למשל לקבל "בדיוק" 3) הוא אפס, אך פונקצית ההתפלגות המצטברת שלו מוגדרת היטב והיא פונקציה רציפה בתחום הערכים שהמשתנה המקרי מקבל. לאור זאת מגדירים את פונקצית צפיפות ההסתברות, המודדת את הסיכוי היחסי לקבל כל ערך של המ"מ. פונקצית הצפיפות (probability density function- p.d.f.), מוגדרת כנגזרת של פונקצית ההתפלגות המצטברת, דהיינו

$$f(x)\equiv\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x} \tag{1.3}$$

והיא מודדת את הסיכוי היחסי לקבל ערכים בסביבה של הערך x. פונקצית הצפיפות מקיימת כי $\int_{-\infty}^{\infty} f(x)dx=1$ אבל f(x) עשויה לקבל ערכים גדולים מאחד (שכן רק כאשר כופלים את הצפיפות באינטרוול הרלבנטי מקבלים הסתברות).

משתנה מקרי מעורב

קיימים משתנים מקריים שאינם בדידים או רציפים. לדוגמא, מ"מ המקבל בסיכוי חצי את הערך אפס, ובסיכוי חצי ערכים המתפלגים אחיד בקטע [1,2] אינו בדיד ואינו רציף. פונקצית ההתפלגות המצטברת מאפשרת טיפול נוח גם במשתנים כאלו.

1.3.2 התפלגות משותפת והתפלגויות שוליות

במקרים רבים אנחנו מעונינים בתכונות של מספר משתנים בעת ובעונה אחת, ואז נתעניין בהתפלגות המשותפת של מספר משתנים מקריים. במקרה של שני משתנים בדידים, התפלגות זו הנה פונקציה דו ממדית $f(x,y):R^2 \rightarrow [0,1]$, המתאימה הסתברות לכל זוג ערכים (x,y)

$$f(x,y) = \Pr(X = x \cap Y = y)$$

והיא מקיימת $\sum_{\{x\}} \sum_{\{y\}} f(x,y) = 1$ כאשר סוכמים על כל הערכים האפשריים לזוגות (x,y).

התפלגות שולית

לעתים ידועה לנו ההתפלגות משותפת של שני משתנים, ואנו מעונינים למצוא מתוכה את ההתפלגות של אחד המשתנים. התפלגות זו נקראת ההתפלגות השולית, והיא מקיימת

$$f(x) = \sum_y f(x,y) \qquad \text{or} \qquad f(x) = \int_y f(x,y)dy$$

אי תלות

שני משתנים מקריים יקראו בלתי תלויים אם מתקיים $f(x,y) = f(x)f(y)$ לכל זוג (x,y).

1.3.3 סכום משתנים מקריים

נתונים לנו שני משתנים מקריים X ו- Y שלהם פונקציות התפלגות f_X ו- f_Y בהתאמה ונגדיר משתנה מקרי חדש Z שיהיה סכומם: Z=X+Y. מהי פונקצית ההתפלגות של הסכום? הסיכוי ש- Z יקבל ערך מסוים z, הוא סכום כל קומבינציות של ערכי X ו- Y המסתכמים ל- z. נניח למשל כי המשתנים רציפים ונרשום באופן פורמלי

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx \qquad (1.4)$$

האינטגרל באגף ימין במשוואה נקרא ה**קונבולוציה** של הפונקציות f_X ו- f_Y.

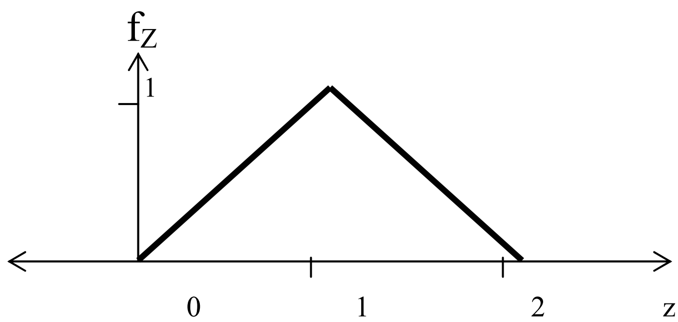
דוגמא: סכום של משתנים מקריים מפולגים אחיד

יהיו X ו- Yמפולגים אחיד בקטע $[0,1]$. נחשב את התפלגות הסכום שלהם

$$f_X(x)=\begin{cases}1 & for \quad 0\le x\le 1\\ 0 & otherwise\end{cases}$$

$$f_Y(z-x)=\begin{cases}1 & for \quad 0\le z-x\le 1\\ 0 & otherwise\end{cases}$$

$$f_Z(z)=\begin{cases}0 & z\le 0\\ \int_0^z dx=z & 0\le z\le 1\\ \int_{z-1}^1 dx=2-z & 1\le z\le 2\\ 0 & 2\le z\end{cases}$$



1.4 מומנטים

1.4.1 התוחלת

הגדרה:

ה**תוחלת** (mathematical expectation) של משתנה מקרי X בדיד מוגדרת כ-

$$E(X)=\sum_{\{x\}} x\cdot p(x) \qquad (1.5)$$

התוחלת היא הממוצע של הערכים ש- X מקבל. עבור משתנה מקרי רציף התוחלת מוגדרת באופן טבעי כ-

$$E(X)=\int_{-\infty}^{\infty} x\cdot f(x)dx \qquad (1.6)$$

דוגמא: תוחלת של משתנה מקרי בדיד

יהי X משתנה המתאר תוצאות הטלת קוביה, דהיינו מקבל הערכים 1 עד 6 בהסתברות ששית כל אחד. אזי התוחלת שלו היא $E(X) = (1+..+6)/6=4.5$. ובמקרה זה התוחלת אינה אחד מהערכים אותם המ"מ יכול לקבל.

דוגמא: משתנה מקרי חסר תוחלת

במקרה הרציף נאמר שלמשתנה מקרי יש תוחלת רק אם האינטגרל מתכנס. נביט למשל במ"מ בעל ההתפלגות הבאה (הנקראת התפלגות קושי Cauchy)

$$f(x)=\frac{1}{\pi(1+x^2)} \qquad for \quad -\infty\le x\le\infty$$

ניתן לוודא כי האינטגרל של פונקציה זו מתכנס ושווה לאחד, אבל האינטגרל של התוחלת אינו מתכנס, כלומר $\int_{-\infty}^{\infty}|x|f(x)dx=\infty$ (לא נוכיח זאת כאן, אבל נזכיר כי ידוע כי אינטגרל על הפונקציה x^{-1} אינו מתכנס בעוד שאינטגרל על הפונקציה x^{-2} אכן מתכנס). הסיבה לכך שלהתפלגות זו אין תוחלת היא כדלקמן: נשים לב כי הזנבות של ההתפלגות מתקרבים לאפס עבור ערכי x גדולים מאוד או קטנים מאוד, ודבר זה מבטיח כי האינטגרל על ההתפלגות יתכנס. אולם, הזנבות אינם שואפים לאפס חזק מספיק, כך כאשר מכפילים אותם בערך המתאים של x האינטגרל מתבדר.

תכונות התוחלת

התוחלת היא אופרטור ליניארי ולכל זוג מ"מ X ו- Y בעלי תוחלת ולכל זוג קבועים a ו- b מתקיים:

$$E(aX+bY)=aE(X)+bE(Y) \qquad (1.7)$$

1.4.2 השונות

הגדרה

יהי X מ"מ עם תוחלת μ, אז ה**שונות** (Variance) שלו מוגדרת כ-

$$Var(X)=E\left[(X-\mu)^2\right] \qquad (1.8)$$

סטיית התקן של מ"מ מוגדרת כשורש הריבועי של השונות, ונהוג לסמן $\sigma^2=Var(X)$.

תכונות השונות

א. היות והשונות היא תוחלת של מ"מ אי-שלילי $(X-\mu)^2$ אזי השונות היא אי שלילית.

ב. השונות שווה לאפס אם ורק אם קיים המ"מ X הוא קבוע (דהיינו קיים c כך ש- $\Pr(X=c)=1$).

ג. לכל שני קבועים a ו- b מתקיים $Var(aX+b)=a^2Var(X)$.

ד. לכל מ"מ X השונות מקיימת $Var(X)=E(X^2)-E^2(X)$.

ה. אם $X_1,\dots,X_n$ הם מ"מ **בלתי תלויים** אזי השונות של הסכום היא סכום השונויות $Var\left(\sum X_i\right)=\sum Var(X_i)$.

1.4.3 השונות המשותפת

יהיו X ו- Y שני מ"מ בעלי תוחלת ושונות שאינם בהכרח בלתי תלויים. נביט בשונות של סכומם:

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

האיבר האחרון נקרא **השונות המשותפת** (covariance) ומסומן $Cov(X,Y)$.
קיבלנו אם כן כי השונות המשותפת מוגדרת על ידי

$$Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)] \tag{1.9}$$

מקדם המתאם

מקדם המתאם של שני משתנים X ו- Y מוגדר באופן הבא

$$\rho=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\,. \tag{1.10}$$

טענה: $-1\le\rho\le 1$

הוכחה: יהיו X ו- Y מ"מ בעלי תוחלת אפס ושונות אחת: E(X)=E(Y)=0, Var(X)=Var(Y)=1 , ונביט בשונות של הסכום ושל ההפרש

$$\begin{aligned}
&0\le Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)\\
&\Rightarrow -2\le 2Cov(X,Y)\\
&0\le Var(X-Y)=Var(X)+Var(Y)-2Cov(X,Y)\\
&\Rightarrow 2Cov(X,Y)\le 2
\end{aligned}$$

קיבלנו כי עבור מ"מ כאלו מתקיים $-1\le Cov(X,Y)\le 1$.
יהיו X ו- Y משתנים כלשהם, נביט במשתנים המתוקננים $X^*=(X-\mu_X)/\sigma_X$ ו- $Y^*=(Y-\mu_Y)/\sigma_Y$ שלהם יש תוחלת אפס ושונות אחת.

כעת נרשום

$$\begin{aligned}
Cov(X^*,Y^*)&=\frac{Cov(X-\mu_X,Y-\mu_Y)}{\sigma_X\sigma_Y}\\
&=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}=\rho
\end{aligned}$$

וקיבלנו כי מקדם המתאם אכן מקיים הנדרש.

משתנים בלתי מתואמים ומשתנים בלתי תלויים

שני משתנים שהשונות המשותפת שלהם היא אפס, נקראים בלתי מתואמים, והם מקיימים כי תוחלת המכפלה שווה למכפלת התוחלות. נשים לב כי זוהי דרישה חלשה יותר מהדרישה לאי תלות. למעשה, אי תלות גוררת אי-תאימות: אם X ו- Y מ"מ ב"ת, נוכל לרשום

$$
\begin{aligned}
E(XY) &= \sum_{\{x,y\}} xy\Pr(x,y) \\
&= \sum_{\{x\}}\sum_{\{y\}} x\Pr(x)y\Pr(y) \\
&= \sum_{\{x\}} x\Pr(x)\sum_{\{y\}} y\Pr(y) \\
&= E(X)E(Y)
\end{aligned}
$$

והוכחנו כי X ו-Y בלתי תלויים $\Leftarrow$ X ו- Y בלתי מתואמים.

הכיוון ההפוך אינו בהכרח נכון, היות והדרישה לקיום שוויון של התוחלות חלשה בהרבה מהדרישות הנחוצות כדי ששני משתנים יהיו בלתי תלויים. זאת היות ולצורך אי תלות נדרוש כי **לכל זוג ערכים** (x,y) יתקיים Pr(x,y)=Pr(x)Pr(y). על מנת להדגיש את ההבדל, שימו לב שאם X מקבל n ערכים אפשריים ו- Y מקבל m ערכים אפשריים, אזי הדרישה לאי תלות נותנת לנו סדר גודל של $m\cdot n$ אילוצים, בעוד השוויון בתוחלות הוא אילוץ בודד בלבד.

1.4.4 מומנטים מסדר כללי

לכל משתנה מקרי X ולכל שלם חיובי k התוחלת $m_k = E\left(X^k\right)$ נקראת **המומנט ה- k** של X (או **המומנט מסדר k**). ראינו כבר כי המומנט הראשון של מ"מ הוא התוחלת שלו, וכן ראינו כי השונות של מ"מ נקבעת על ידי שני המומנטים הראשונים $Var\left(X\right) = E\left(X^2\right) - E^2\left(X\right)$.

נאמר שהמומנט ה- k של מ"מ קיים אם התוחלת $E(X^k)<\infty$. קל לראות כי אם X חסום ובפרט אם הוא מקבל מספר סופי של ערכים, אזי כל המומנטים קיימים. משפט שלא נוכיח כאן קובע כי אם קיים המומנט ה- k אזי קיימים כל המומנטים הקטנים מ- k.
כעת יהי X מ"מ בעל תוחלת μ, אזי מגדירים את המומנט המרכזי ה- k על ידי התוחלת $E[(X-\mu)^k]$. על פי ההגדרה, השונות היא המומנט המרכזי השני.

החשיבות של אפיון ההתפלגות על ידי מומנטים נעוצה בעובדה שבתנאים מסוימים ניתן לעבור מהייצוג של התפלגות שבו עסקנו עד כה – ייצוג על ידי סדרת הערכים $P(x_1),P(x_2),\ldots$ לייצוג על ידי סדרת המומנטים $m_1,m_2,\ldots$ (וזאת באנלוגיה לטרנספורם פורייה למשל). קיימות תכונות רבות של התפלגויות שאותן קל יותר להוכיח במרחב המומנטים מאשר בייצוג בו השתמשנו עד כה.

1.5 התפלגויות מיוחדות

1.5.1 התפלגות ברנולי

משתנה מקרי המתפלג ברנולי מקבל שני ערכים אפשריים, האחד בהסתברות p והשני בהסתברות (1-p):

$$(1.11)\quad f(x)=\begin{cases} p & for \quad x=1 \\ 1-p & for \quad x=0 \end{cases}$$

משתנה כזה מתאר למשל את ההתפלגות של תוצאת "ראש" בהטלת מטבע. באופן כללי יותר, הוא מתאר ההתפלגות של מאורע מהסוג "האם תוצאה קרתה".

1.5.2 התפלגות בינומית

נניח שאנחנו מבצעים סדרה של n ניסויים בלתי תלויים כאשר כל ניסוי יכול להצליח בסיכוי p או להיכשל בסיכוי (1-p). כפי שהוגדר לעיל, כל אחד מניסויים אלו מתפלג ברנולי. כעת יהי X משתנה מקרי הסופר את מספר הניסויים המוצלחים מתוך n הניסויים. נאמר ש- X מתפלג בינומית ונסמן $X \sim B(n,p)$, כאשר ההסתברות לקבל k הצלחות היא

$$\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \quad \forall k=0,1,\ldots,n$$

ולכן פונקצית ההתפלגות של X היא

$$(1.12)\quad f(x)=\begin{cases} \binom{n}{x}p^x(1-p)^{n-x} & for \quad x=0,1,\ldots,n \\ 0 & otherwise \end{cases}$$

תכונות

התוחלת של מ"מ בינומי היא n·p, והשונות שלו היא n·p·(1-p).

1.5.3 התפלגות פואסון

משתנה מקרי המתפלג פואסונית עם פרמטר קצב λ (נסמן (X~Poisson(λ) מקבל ערכים אי שליליים בהסתברות:

$$(1.13)\quad \Pr(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \quad \forall k=0,1,\ldots$$

דהיינו פונקצית ההתפלגות שלו היא

$$(1.14)\quad f(x)=\begin{cases} \frac{\lambda^x}{x!}e^{-\lambda} & for \quad x=0,1,\ldots \\ 0 & otherwise \end{cases}$$

המשתנה המקרי הפואסוני שימושי במיוחד בטיפול בתהליכים מקריים חסרי זכרון, עליהם נרחיב בפרק 5.

קירוב פואסוני להתפלגות בינומית

נראה כעת כי כאשר הערך של n גדול, ואילו ההסתברות להצלחה p קרובה לאפס, ההתפלגות הבינומית עם פרמטרים n ו- p ניתנת לקירוב על ידי ההתפלגות הפואסונית.
נניח ש- X מ"מ המתפלג בינומית X~B(n,p) ונרשום את ההתפלגות

$$\Pr(X=k)=\frac{n(n-1)(n-2)...(n-k+1)}{k!}p^k(1-p)^{n-k}$$

נסמן $\lambda=n\cdot p$ ונוכל לרשום

$$\Pr(X=k)=$$
$$=\frac{1}{k!}\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdots\frac{n-k+1}{n}\right)n^k p^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$
$$=\frac{\lambda^k}{k!}\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdots\frac{n-k+1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}$$

כעת נביט בגבול בו $n\to\infty$ ו- $p\to 0$ כך שמתקיים $n\cdot p=\lambda$. ונטפל בנפרד באיברים השונים במכפלה לעיל:

א. היות ו- k קבוע אזי

$$\lim_{n\to\infty}\left(\frac{n}{n}\cdot\frac{n-1}{n}\cdot\frac{n-2}{n}\cdots\frac{n-k+1}{n}\right)=1$$

ב. היות ו- λ ו- k קבועים אז

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{k}=1$$

ג. מחשבון אינפיטיסימלי ידוע כי

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}=e^{-\lambda}$$

וקיבלנו

$$\Pr(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}\qquad\forall x=0,1,...$$

התפלגות פואסון מתאימה אם כן לתיאור מספר האירועים בתהליכים בהם יש מספר רב של משתנים מקריים בלתי תלויים או שמקורם בתהליך רציף. דוגמא מקובלת היא התפלגות מספר אירועי הפירוק בחומר רדיואקטיבי. בחומר כזה, ישנו מספר עצום של אטומים וכל אחד מהם מתפרק בהסתברות p קטנה מאוד, אך $n\cdot p$ הוא מספר בסדר גודל סביר.

דוגמא: שרשרת פוטנציאלי פעולה כתהליך פואסוני

נניח את המודל הפשטני (והשגוי) הבא לתיאור שרשרת פוטנציאלי פעולה בתא עצב. נניח כי פוטנציאלי פעולה הם ארועים ממוקדים שמשכם אפס זמן (למשל

נתייחס רק לרגע השיא של פוטנציאל הפעולה), ונניח כי התא יכול לירות בכל רגע באופן בלתי תלוי ביריותיו ברגעי זמן אחרים, אך הוא יורה בקצב ממוצע של λ פ"פ לשניה. היות והיריות הן בלתי תלויות, אזי בפרק זמן בן T שניות יהיו λT יריות בממוצע (דהיינו קצב הירי הוא λT פ"פ ל- T שניות) ואנו יכולים לחלק את הזמן לפרקים קטנים כך שבכל פרק זמן יהיה לכל היותר אירוע פ"פ בודד.
במקרה זה מספר הפ"פ מתפלג פואסונית עם פרמטר הקצב λ.

טענה: התוחלת של מ"מ פואסוני עם פרמטר λ היא λ

$$E(X)=\sum_{k=0}^{\infty}k\Pr(X=k)=\sum_{k=1}^{\infty}k\Pr(X=k)$$
$$=\sum_{k=1}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}$$

נחליף משתנים $y = k-1$ ונקבל

$$E(X)=\lambda\sum_{y=0}^{\infty}\frac{\lambda^y}{y!}e^{-\lambda}=\lambda\sum_{y=0}^{\infty}\Pr(X=y)=\lambda\cdot 1=\lambda$$

טענה: השונות של מ"מ פואסוני עם פרמטר λ היא λ

נחשב ראשית את התוחלת הבאה, באופן דומה לחישוב התוחלת בטענה הקודמת

$$E[X(X-1)]=\sum_{k=0}^{\infty}k(k-1)\Pr(X=k)$$
$$=\sum_{k=2}^{\infty}k\Pr(X=k)=\lambda^2\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda}$$

גם כאן נחליף משתנים $y = k-2$ ונקבל

$$E[X(X-1)]=\lambda^2\sum_{y=0}^{\infty}\frac{\lambda^y}{y!}e^{-\lambda}=\lambda^2$$

כעת היות ו- $E[X(X-1)]=E[X^2]-E[X]$, אזי קיבלנו כי $E[X^2] = \lambda^2+\lambda$. ולכן השונות מקיימת

$$Var(X)=E[X^2]-E^2[X]=\lambda^2+\lambda-\lambda^2=\lambda\ .$$

1.5.4 התפלגות אקספוננציאלית

משתנה X מתפלג אקספוננציאלית עם פרמטר θ אם הוא מקבל ערכים ממשים חיוביים עם צפיפות הסתברות

$$(1.16)\quad f(x)=\begin{cases}\theta e^{-\theta} & for \quad x\ge 0\\ 0 & otherwise\end{cases}$$

ההתפלגות האקספוננציאלית היא הגבול הרציף של ההתפלגות הגיאומטרית.
למ"מ בעל התפלגות אקספוננציאלית יש תוחלת $1/\theta$ ושונות $1/\theta^2$.

1.5.5 התפלגות גאמא

פונקצית גאמא

לכל מספר α נגדיר את פונקצית גאמא (Gamma Function) $\Gamma(\alpha)$ באופן הבא:

$$(1.17)\quad \Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx\ .$$

ניתן להראות כי הערך של אינטגרל זה הוא סופי לכל α חיובי. נטפל כעת במספר תכונות של ההתפלגות.

טענה: לכל $\alpha>1$, $\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$

הוכחה: על ידי אינטגרציה בחלקים. אם נסמן $u=x^{(\alpha-1)}$ ו- $dv=e^{-x}$ אז נקבל $du=(\alpha-1)x^{(\alpha-2)}$ ו- $v=-e^{-x}$ ולכן

$$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}dx=[-x^{(\alpha-1)}e^{-x}]_0^\infty+(\alpha-1)\int_0^\infty x^{\alpha-2}e^{-x}dx$$
$$=(\alpha-1)\Gamma(\alpha-1)$$

מסקנה: $\Gamma(n)=(n-1)!$ לכל שלם $n>1$ וכן $\Gamma(1)=1$

הוכחה: עבור $n=1$ המסקנה נובעת מתוך האיטגרל על ההתפלגות האקספוננציאלית. עבור $n>1$ המסקנה נובעת באינדוקציה.

התפלגות גאמא

מ"מ X הוא בעל התפלגות גאמא עם פרמטרים α ו- β, אם פונקצית ההתפלגות שלו היא

$$(1.18)\quad f(x)=\begin{cases}\dfrac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & for \quad x\ge 0\\ 0 & for \quad x<0\end{cases}$$

האינטגרל של התפלגות זו הוא 1, היות ומהגדרת פונקצית גאמא נובע כי

$$(1.19) \quad \int_0^\infty x^{\alpha-1}e^{-\beta x}dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$$

את המומנטים של X בעל התפלגות גאמא קל לחשב מתוך שתי המשוואות האחרונות באופן הבא

$$E(X^k) = \int_0^\infty x^k f(x)dx = \frac{\beta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{\alpha+k-1}e^{-\beta x}dx$$
$$= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha+k)}{\beta^{\alpha+k}} = \frac{\Gamma(\alpha+k)}{\beta^k\Gamma(\alpha)}$$

בפרט, נקבל כי

$$(1.20) \quad E(X) = \frac{\alpha}{\beta}$$

$$(1.21) \quad Var(X) = \frac{\alpha(\alpha+1)}{\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2}$$

מקרים מיוחדים

ההתפלגות האקספוננציאלית אותה תיארנו בסעיף קודם, היא מקרה פרטי של התפלגות גאמא עם $\alpha = 1$.

התפלגות גאמא עם $\alpha = \frac{1}{2}, \beta = \frac{1}{2}$ נקראת גם התפלגות חי בריבוע (Chi-square) - χ^2. זוהי התפלגות שימושית ביותר בסטטיסטיקה, ונפגש בה שוב בהמשך הקורס.

1.5.6 ההתפלגות הנורמלית

ההתפלגות הנורמלית שנגדיר מיד, היא ללא ספק ההתפלגות ששימושיה הם הנפוצים ביותר בסטטיסטיקה ולכך מספר סיבות. ראשית, להתפלגות הנורמלית יש תכונות מתמטיות נוחות: במקרה שמדגם נלקח מתוך התפלגות נורמלית, אזי ניתן לגזור במפורש ובאופן אנליטי תכונות שונות של פונקציות של המדגם בהן נטפל בפרקים הבאים. שנית, התפלגויות רבות בטבע מתנהגות בקירוב כהתפלגויות נורמליות. שלישית, ישנם טיעונים המציעים כי בהינתן מידע מוגבל על פונקצית ההתפלגות, הרי שהתפלגות נורמלית תהיה קירוב "טוב". ולבסוף, הסיבה החזקה מכל היא משפט הגבול המרכזי שאותו נתאר בהמשך. משפט זה קובע כי אם אנו צופים במדגם גדול הנלקח מהתפלגות כלשהי (!), אזי לפונקציות רבות וחשובות של המדגם יש התפלגות נורמלית. למשל אם מדגם נלקח מהתפלגות כלשהי בעלת שונות סופית, הרי שהממוצע שלו מתפלג בקירוב נורמלית.

הגדרה

משתנה מקרי רציף X הוא בעל התפלגות נורמלית (חד ממדית) עם פרמטר תוחלת μ ופרמטר סטיית תקן σ אם פונקצית הצפיפות שלו היא

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (1.22)$$

ונסמן $X \sim N\left(\mu, \sigma^2\right)$. כאשר $\mu = 0$ ו- $\sigma = 1$ אז ההתפלגות נקראת נורמלית סטנדרטית. נהוג לסמן את ההתפלגות של מ"מ נורמלי סטנדרטי ב- $\phi(x)$ ואת פונקצית ההתפלגות המצטברת שלו ב- $\Phi(x)$.

ההתפלגות הנורמלית מהווה קירוב להתפלגות הבינומית עבור n גדול בסביבת μ. דהיינו פונקצית ההתפלגות הנורמלית מעריכה היטב את ההסתברות לקבל ערכים של X הקרובים יחסית ל- μ, אך לא עבור ערכים בזנבות ההתפלגות.

סכום משתנים נורמליים

יהיו X ו- Y מ"מ המתפלגים נורמלית $X \sim N(\mu_X, \sigma^2_X)$, $Y \sim N(\mu_Y, \sigma^2_Y)$, דהיינו

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-\mu_X)^2}{2{\sigma_X}^2}\right)$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left(-\frac{(y-\mu_Y)^2}{2{\sigma_Y}^2}\right)$$

נרשום את התפלגות הסכום Z=X+Y

$$\begin{aligned}
f_Z(z) &= \frac{1}{\sqrt{2\pi}\sigma_X}\frac{1}{\sqrt{2\pi}\sigma_Y} \\
&\quad \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu_X)^2}{2{\sigma_X}^2}\right) \exp\left(-\frac{([z-x]-\mu_Y)^2}{2{\sigma_Y}^2}\right) dx \\
&= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\frac{{\sigma_Y}^2(x-\mu_X)^2 + {\sigma_X}^2([z-x]-\mu_Y)^2}{2{\sigma_X}^2{\sigma_Y}^2}\right) dx \\
&= \frac{1}{2\pi\sigma_X\sigma_Y} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2({\sigma_Y}^2+{\sigma_X}^2) - 2x({\sigma_Y}^2\mu_X + {\sigma_X}^2(z-\mu_Y))}{2{\sigma_X}^2{\sigma_Y}^2}\right) \\
&\quad \exp\left(-\frac{{\sigma_Y}^2{\mu_X}^2 + {\sigma_X}^2(\mu_Y - z)^2}{2{\sigma_X}^2{\sigma_Y}^2}\right) dx
\end{aligned}$$

קיבלנו בתוך האקספוננט תבנית ריבועית ב- x, כלומר ביטוי מהצורה $\exp(ax^2+bx+c)$, עם המקדמים

$$a=\frac{\sigma_X^2+\sigma_Y^2}{2\sigma_X^2\sigma_Y^2}$$

$$b=\frac{{\sigma_Y}^2\mu_X+{\sigma_X}^2(z-\mu_Y)}{2\sigma_X^2\sigma_Y^2}$$

$$c=\frac{{\sigma_Y}^2{\mu_X}^2+{\sigma_X}^2(\mu_Y-z)^2}{2\sigma_X^2\sigma_Y^2}$$

האינטגרל של אקספוננט של תבנית ריבועית מקיים

$$\int_{-\infty}^{\infty}\exp\left[-(ax^2+bx+c)\right]dx=\sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a}-c\right)$$

(ניתן להוכיח זאת על ידי "השלמה לריבוע" ושימוש באינטגרל של התפלגות נורמלית). ולכן אנחנו מקבלים כי ההתפלגות של z אף היא התפלגות מהצורה הנורמלית (היות והמקדם c מכיל איבר ריבועי ב- z והמקדם b מכיל איבר לינארי ב- z). כדי לקבל את הצורה המדויקת של ההתפלגות, נציב את המקדמים a,b,c ונקבל כי ההתפלגות של Z היא

$$f_Z(z)=\frac{1}{\sqrt{2\pi}({\sigma^2}_X+{\sigma^2}_Y)^{1/2}}\exp\left(-\frac{(z-(\mu_X+\mu_Y))^2}{2({\sigma^2}_X+{\sigma^2}_Y)^{1/2}}\right)$$

כלומר סכום של שני מ"מ נורמליים בלתי תלויים מתפלג נורמלית עם תוחלת שהיא סכום התוחלות ושונות שהיא סכום השונויות ($Z\sim N(\mu_X+\mu_Y,\ \sigma^2_X+\sigma^2_Y)$.
ניתן להראות באינדוקציה כי סכום של n מ"מ נורמלים בלתי תלויים מתפלג נורמלית עם תוחלת שהיא סכום התוחלות ושונות שהיא סכום השונויות.

משפט הגבול המרכזי

יהיו $X_1,\ldots,X_n$ משתנים מקריים בלתי תלויים ושווי התפלגות היוצרים מדגם בגודל n מהתפלגות נתונה בעלת תוחלת μ ושונות σ^2, ונסמן את הממוצע שלהם על ידי $\overline{X_n}=\frac{1}{n}\sum_{i=1}^{n}X_i$ אזי לכל קבוע k מתקיים

$$\lim_{n\to\infty}\Pr\left(\frac{\overline{X_n}-\mu}{\sigma/\sqrt{n}}\le k\right)=\Phi(k)\qquad(1.23)$$

כאשר $\Phi(x)$ היא פונקצית ההתפלגות המצטברת של מ"מ נורמלי סטנדרטי.

הפירוש של משוואה (1.16) הוא כדלקמן: אם לוקחים מדגם גדול מהתפלגות כלשהי בעלת תוחלת μ וסטית תקן σ (ולא משנה אם זוהי התפלגות בדידה או רציפה), אזי ההתפלגות של המ"מ $\sqrt{n}\left(\overline{X_n}-\mu\right)/\sigma$ היא בקירוב נורמלית סטנדרטית. לא נוכיח את המשפט כאן, אך ניתן למצוא שלד של ההוכחה המתואר באופן קריא ביותר ב- DeGroot. נעיר כי הקירוב הוא טוב בעיקר סביב התוחלת והוא פחות טוב בזנבות של ההתפלגות.

1.5.7 התפלגות נורמלית רב ממדית

משתנים מקריים רב ממדים

עד כה טיפלנו בהתפלגויות של משתנה מקרי סקלרי, דהיינו חד ממדי. נעבור כעת לתאר התפלגויות רב ממדיות. מ"מ רב ממדי הוא פשוט וקטור של מ"מ בעלי התפלגות משותפת. למעשה טיפלנו כבר בשני משתנים מקריים X ו- Y בעלי התפלגות משותפת f(x,y). המקרה הרב ממדי הוא הרחבה טבעית למקרה של n משתנים מקריים $\left(X_1,\ldots,X_n\right)$ שלהם התפלגות משותפת $f\left(x_1,\ldots,x_n\right)$. ערך של משתנה מקרי רב ממדי שקול לוקטור של ערכי המשתנים המקריים הבודדים.

התוחלת של מ"מ רב ממדי מוגדרת כוקטור התוחלות של המ"מ הבודדים

$$E(X^{(n)})=E(X_1,\ldots,X_n)=(E(X_1),\ldots,E(X_n))\,. \tag{1.24}$$

כעת יהיו $X=\left(X_1,\ldots,X_n\right)$ ו- $Y=\left(Y_1,\ldots,Y_m\right)$ מ"מ רב מימדים ממימד n ו- m בהתאמה, אזי **מטריצת השונות המשותפת** שלהם, תהיה מטריצה ממימד n·m כך שהתא ה-ij שלה מכיל את השונות המשותפת Cov(Xi,Yj)

$$V=\begin{pmatrix} Cov(X_1,Y_1) & Cov(X_1,Y_2) & \cdots & Cov(X_1,Y_m) \\ \vdots & Cov(X_2,Y_2) & \ddots & \vdots \\ Cov(X_n,Y_1) & \cdots & \cdots & Cov(X_n,Y_m) \end{pmatrix} \tag{1.25}$$

מטריצת השונות המשותפת למ"מ רב ממדי X תהיה V(X,X), והיא מקיימת כי היא חיובית בהחלט וסימטרית.

נדגים נושאים אלו על ידי טיפול בהתפלגות נורמלית רב ממדית.

הגדרה

נאמר שלמ"מ n ממדי X יש התפלגות נורמלית רב ממדית אם צפיפות ההסתברות של הערך $\mathbf{x}=(x_1,\ldots,x_n)$ נקבעת על ידי פונקצית צפיפות ההתפלגות הבאה

$$(1.26)\qquad f(\mathbf{x})=\frac{1}{(2\pi)^{n/2}\sqrt{\det(V)}}\exp\left[\frac{-1}{2}(\mathbf{x}-\mu)V^{-1}(\mathbf{x}-\mu)^T\right]$$

כאשר V היא מטריצת חיובית לחלוטין ו- μ וקטור n ממדי כלשהו. הוקטור $\mathbf{x}$ הוא וקטור שורה ממימד n ואילו המטריצה V^{-1} היא מטריצה ריבועית ממימד n ולכן הביטוי בתוך האקספוננט הוא סקלר.

תכונות

א. התוחלת של מ"מ נורמלי רב ממדי בעל פונקצית הצפיפות זו קיימת והיא μ .
ב. מטריצת השונות המשותפת של מ"מ נורמלי רב ממדי בעל פונקצית הצפיפות זו קיימת והיא V.
ג. כל אחד מהמשתנים X_i מתפלג התפלגות נורמלית חד ממדית עם תוחלת μ ושונות $V(X_i,X_i)$.

1.5.8 התפלגות לוג נורמלית

נתאר תחילה את המקרה הכללי של חישוב התפלגות של פונקציה של משתנה מקרי. יהי X משתנה מיקרי המתפלג לפי פונקצית התפלגות $f(x)$, ויהי Y משתנה מקרי המקבל ערכים $y = g(x)$, כאשר $g(x)$ היא פונקציה רציפה וגזירה. אזי פונקצית צפיפות ההתפלגות של y

$$f_y(y)=\frac{f_x(x\,|\,g(x)=y)}{|\,g'\,|}=\frac{f_x(g^{-1}(y))}{|\,dy/dx\,|}$$

כעת, יהי X משתנה מקרי המתפלג נורמלית עם ממוצע μ_x וסטית-תקן σ_x

נגדיר $y=g(x)\equiv e^x$, ואז $x=\log(y)$ והנגזרת של y לפי x מקיימת $g'(x)=e^x$ ולכן

$$f(y)=\frac{1}{y\sigma_x\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\ln y-\mu_x}{\sigma_x}\right)^2\right]$$

להתפלגות זו שימושים חשובים בתהליכים פיסקאלים ופיננסים שונים בהם תוצאת התהליך הינה מכפלה של מספרים אקראיים.

1.6 זנבות של התפלגויות

חוק המספרים הגדולים קובע כי הממוצע של דגימות בלתי תלויות ושוות התפלגות מתקרב לתוחלת ההתפלגות ככל שמספר הדגימות גדל, $\frac{1}{n}\sum_{i=1}^{n}X_i\to E(X)$

ומשפט הגבול המרכזי קובע כי בסביבת התוחלת, יש לממוצע התפלגות שהיא בקירוב טוב נורמלית. אבל, עד כמה באמת הממוצע מרוכז סביב התוחלת ?
בפרק זה נתאר מספר חסמים על המשקל של זנבות ההתפלגות, דהיינו על ההסתברות שמשתנה מקרי יקבל ערך הרחוק מהתוחלת שלו.

1.6.1 חסם מרקוב - Markov Bound

יהי X משתנה מקרי אי שלילי ו- k מספר חיובי. חסם מרקוב הוא

$$\Pr\left(X \geq k\right) \leq \frac{E(X)}{k} \tag{1.27}$$

חסם מרקוב אינו תלוי בשום ידע על ההתפלגות של X, למעט העובדה כי הוא חיובי. כדי להוכיח, נרשום את הגדרת התוחלת (המקרה הרציף דומה)

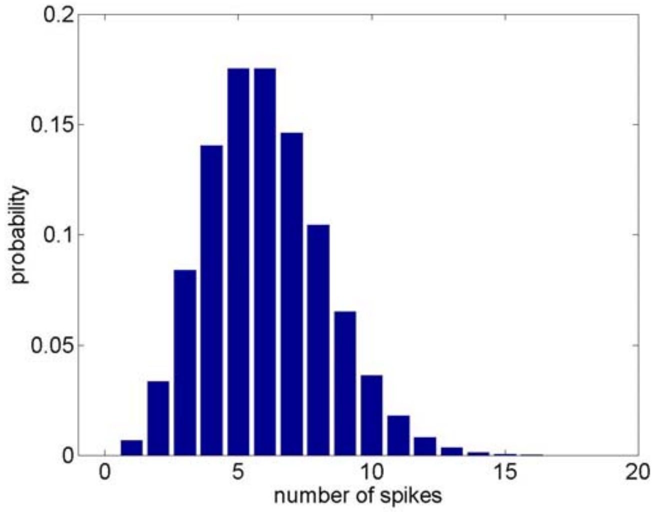
$$\begin{aligned} E\left(X\right) &= \sum_{x<k} x\Pr(X=x) \quad + \sum_{x\geq k} x\Pr(X=x) \\ &\geq \sum_{x\geq k} x\Pr(X=x) \\ &\geq \sum_{x\geq k} k\Pr(X=x) \\ &= k\sum_{x\geq k} \Pr(X=x) \\ &= k\Pr(X\geq k) \end{aligned}$$

וקיבלנו כנדרש

$$\frac{E\left(X\right)}{k} \geq \Pr(X \geq k)$$

דוגמא

לרוב, חסם מרקוב איננו להיט היסטרי. נסתכל למשל על משתנה מקרי פואסוני, המתאר את מספר פוטנציאלי הפעולה בתא היורה בקצב של 5 הרץ. על פי חסם מרקוב הסיכוי לקבל יותר מ- 10 יריות בשניה קטן מ- 5/10 דהינו חצי, בעוד שההסתברות האמיתית היא 0.0318. הסיכוי לקבל יותר מ- 20 יריות קטן מרבע, בעוד ההסתברות האמיתית $3.45*10^{-7}$ רחוקה ממנה בששה סדרי גודל.



התפלגות מספר המאורעות של משתנה מקרי פואסוני עם פרמטר קצב 5

1.6.2 חסם צ'בישב - Chebyshev Bound

חסם צ'בישב מספק לנו אינפורמציה טובה יותר על פיזור המשתנה המקרי מאשר חסם מרקוב, ומצליח לעשות זאת על ידי שימוש ביותר אינפורמציה על ההתפלגות.
חסם צ'בישב קובע כי

$$\Pr\left(\left|X - E\left(X\right)\right| \geq \delta\right) \leq \frac{Var\left(X\right)}{\delta^2}\,. \tag{1.28}$$

הוכחה

נגדיר את המשתנה המקרי החדש $Y=(X-E(X))^2$ (זהו משתנה מקרי חיובי) ונשים לב כי

$$\Pr\left(Y \geq \delta^2\right) = \Pr\left(\left(X - E(X)\right)^2 \geq \delta^2\right) = \Pr\left(\left|X - E(X)\right| \geq \delta\right)$$

כעת נשתמש בחסם מרקוב

$$\Pr\left(\left|X - E(X)\right| \geq \delta\right) = \Pr\left(Y \geq \delta^2\right) \leq \frac{E\left(Y\right)}{\delta^2} = \frac{Var\left(X\right)}{\delta^2}$$

כאשר השתמשנו בכך שהשונות של X היא התוחלת של Y.

דוגמא

טפל שוב במשתנה הפואסוני מהדוגמה הקודמת שלו פרמטר קצב 5. על פי חסם צ'בישב הסיכוי לקבל יותר מ- 10 יריות בשניה קטן מ-

$$\Pr(X-5\geq 5)\leq \frac{Var(X)}{25}=\frac{5}{25}=0.2$$

והסיכוי לקבל יותר מ- 20 יריות בשניה קטן מ-

$$\Pr(X-5\geq 15)\leq \frac{Var(X)}{225}=\frac{5}{225}=0.0222$$

זהו שיפור ביחס לחסם מרקוב אך עדיין החסם מאוד לא הדוק.

1.6.3 חסם צ'רנוף - Chernoff Bound

נטפל כעת בסדרה של משתני ברנולי $X_1,\ldots,X_n$, שלכל אחד מהם הסתברות הצלחה של p_i (וכמובן הסתברות $(1-p_i)$ לכשלון), ונסמן ב- X את סכום המשתנים. נזכיר כי כאשר כל הסתברויות ההצלחה זהות, אז X מתפלג בינומית עם פרמטרים $B(n,p)$, וכאן יש לנו הרחבה פשוטה של המשתנה המקרי הבינומי. מתוך הגדרת התוחלת ברור כי התוחלת של X (שאותה נסמן ב- μ) שווה ל-

$$\mu \equiv E(X) = \sum_{i=1}^{n} p_i$$

וננסה כעת לחסום את משקלי הזנבות עבור משתנה מסוג זה.

חסם צ'רנוף עבור הזנב התחתון הוא

$$\Pr\left(X < (1-\delta)\mu\right) < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} . \tag{1.29}$$

ניתן לפשט חסם זה לחסם הדוק מעט פחות

$$\Pr\left(X < (1-\delta)\mu\right) < \exp\left(-\mu\delta^2/2\right) \tag{1.30}$$

הוכחה

ראשית נכתוב את אי השוויון כאי שוויון בין אקספוננטים, לאחר הכפלה בגורם $t>0$. ברגע זה עדיין לא ברור מדוע אנחנו מכפילים ב- t, אבל תוכלו לפחות לוודא כי השוויון הבא נכון עבור t חיובי.

$$\Pr\left[X < (1-\delta)\mu\right] = \Pr\left[\exp(-tX) < \exp(-t(1-\delta)\mu)\right]$$

כעת נשתמש באי שוויון מרקוב עבור אגף ימין לעיל

$$\Pr\left[X < (1-\delta)\mu\right] < \frac{E\left(\exp(-tX)\right)}{\exp(-t(1-\delta)\mu)}$$

כעת, נשים לב כי האקספוננט $\exp(-tX)$ מורכב ממכפלות של משתנים מקריים מהצורה $\exp(-tX_i)$, ומשתנים מקריים אלו הם בלתי תלויים היות ו- X_i הם בלתי תלויים. עובדה זאת היא לב ליבו של חסם צ'רנוף. התוחלת של מ"מ בלתי תלויים היא מכפלת התוחלות ולכן

$$\Pr\left[X < (1-\delta)\mu\right] < \frac{\prod_{i=1}^{n} E\left(\exp(-tX_i)\right)}{\exp(-t(1-\delta)\mu)}$$

כעת התוחלת של כל רכיב היא

$$E\left(\exp(-tX_i)\right) = p_i \exp(-t) + (1-p_i)\exp(0) = 1 - p_i(1-e^{-t})$$

וכדי לפשט נשתמש באי השוויון $1-a<\exp(-a)$ עם $a=p_i(1-\exp(-1))$ ונקבל

$$E\left(\exp(-tX_i)\right) < \exp\left(p_i(e^{-t}-1)\right)$$

ולכן

$$\prod_{i=1}^{n} E\left(\exp(-tX_i)\right) < \prod_{i=1}^{n} \exp\left(p_i(e^{-t}-1)\right)$$
$$= \exp\left(\sum_{i=1}^{n} p_i(e^{-t}-1)\right)$$
$$= \exp\left(\mu(e^{-t}-1)\right)$$

הצבה בחזרה בחסם הכולל נותנת

$$\Pr\left[X<(1-\delta)\mu\right] < \frac{\exp\left(\mu(e^{-t}-1)\right)}{\exp(-t(1-\delta)\mu)}$$
$$= \exp\left(\mu(e^{-t}+t-t\delta-1\right)$$

וכעת הגיע הזמן לבחור את t כך שנקבל חסם הדוק ככל האפשר: נגזור את $(e^{-1}+t-t\delta-1)$ ביחס ל- t ונשווה לאפס ונקבל כי ה- t האופטימלי הוא $t=\ln\left(1/(1-\delta)\right)$.
שימוש בהצבה זאת נותן

$$\Pr\left[X<(1-\delta)\mu\right] < \exp\left(\mu\left[(1-\delta)+(1-\delta)\ln\left(\frac{1}{(1-\delta)}\right)-1\right]\right)$$

ופישוט אלגברי יתן לנו כנדרש

$$\Pr\left(X<(1-\delta)\mu\right) < \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}$$

כדי לקבל חסם אלגנטי יותר, נרצה להפטר מהאיבר המגושם במכנה. נביט על לוג המכנה $(1-\delta)\ln(1-\delta)$ ונשתמש בטור טיילור של הלוג

$$\ln(1-\delta) = -\delta-\delta^2/2-\delta^3/3-\dots$$

נכפול בחזרה ב- $(1-\delta)$ ונקבל

$$(1-\delta)\ln(1-\delta) = -\delta+\delta^2/2+(\text{all positive values}) > -\delta+\delta^2/2$$

נפעיל בחזרה את האקספוננט ונקבל

$$(1-\delta)^{(1-\delta)} > \exp\left(-\delta+\delta^2/2\right)$$

וכעת נציב באי השוויון שמצאנו קודם ונקבל כנדרש

$$\Pr\left(X<(1-\delta)\mu\right)<\left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}<\left(\frac{\exp\left(-\delta\right)}{\exp\left(-\delta+\delta^2/2\right)}\right)^{\mu}$$
$$=\exp\left(-\mu\delta^2/2\right)$$

חסם על הזנב העליון

באופן דומה להוכחה לעיל ניתן להראות כי

$$(1.31)\quad \Pr\left(X>(1+\delta)\mu\right)<\left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}.$$

גם את החסם הזה ניתן לפשט לחסם הדוק מעט פחות

$$(1.32)\quad \Pr\left(X>(1+\delta)\mu\right)<\exp\left(-\mu\delta^2/4\right)$$

1.6.4 השוואה

נשווה בין החסמים שפיתחנו לעיל עבור סדרת הטלות באורך n של מטבע הוגנת. נרצה להעריך מה ההסתברות ש- 75 אחוז מהמטבעות נפלו על "עץ".

חסם מרקוב

$$\Pr\left(X\ge 3n/4\right)\le\frac{E(X)}{3n/4}=\frac{n/2}{3n/4}=\frac{1}{6}$$

חסם צ'בישב

$$\Pr\left(X\ge 3n/4\right)\le\frac{Var(X)}{\left(n-3n/4\right)^2}=\frac{n/4}{(n/4)^2}=\frac{4}{n}$$

חסם צ'רנוף

$$\Pr\left(X\ge 3n/4\right)\le\exp\left(-E(X)\frac{\left(1/4\right)^2}{2}\right)=\exp\left(-\frac{n}{64}\right)$$

וקיבלנו כי בעוד אי שוויון מרקוב נותן חסם שאיננו תלוי ב- n ואי שוויון צ'בישב נותן חסם התלוי ב- 1/n הרי שאי שוויון צ'רנוף נותן חסם היורד אקספוננציאלית עם n.

תרגילים

1. בתוך שק ישנם 1000 מטבעות, שמתוכם 999 "רגילים" (משמע - ההסתברות לקבל "פלי" היא 0.5) ומטבע אחד מזויף (שתמיד נותן "פלי"). נניח שהוצאתם מטבע אחד, הטלתם אותו 10 פעמים ובכולן קיבלתם "פלי" - מהי ההסתברות שהמטבע מזויף?

2. בסין מותר, בדרך כלל להביא לעולם רק ילד אחד. מאחר שבחברה הסינית המסורתית קיימת עדיפות לבנים על-פני בנות, משפחות שנולדה להן רק בת עשויות לנקוט בצעדים קיצוניים. על מנת להקל על האוכלוסיה, הוחלט במחוז בייסצ'ואן שבסין לאפשר לכל משפחה להוליד ילדים עד שיולד הבן הראשון. כלומר המשפחות במחוז זה הן מהטיפוס:
   בן.
   בת, בן.
   בת, בת, בן.
   וכו'...
   האם במחוז הנ"ל יש יותר בנים או בנות?

3. יהי X_1 משתנה מקרי המתפלג פואסונית $Poisson(\lambda_1)$ ו- X_2 מתפלג פואסונית $Poisson(\lambda_2)$. מהי ההתפלגות של $Y = X_1 + X_2$? הוכח.

4. יהי X משתנה מקרי המתפלג אקספוננציאלית
   $$f(x) = \exp(-x) \quad \text{for all } x > 0$$
   חשב את ההתפלגות של המשתנה המקרי $Y=X^{1/2}$.

5. יהי X משתנה מקרי שלו פונקצית ההתפלגות הבאה
   $$f(x) = \frac{k}{\left(1+x^3\right)} \quad \text{for all } x > 0$$
   א. מצא ערך לקבוע k כך שהפונקציה f תהיה פונקצית התפלגות.
   ב. חשב את השונות של X.

6. הוכח את אי שוויון שוורץ
   $$E(UV)^2 \le E(U^2)E(V^2)$$

7. יהיו X ו- Y משתנים מקריים שהתוחלות והשוניות שלהם קיימות, הוכח כי מתקיים
   $$Var(Y) = E_x\left[Var\left(Y \mid X\right)\right] + Var_x\left[E\left(Y \mid X\right)\right]$$

8. נתונות 2 מעטפות, הנראות זהות מבחוץ. באחת x ש"ח ובשנייה $2x$ ש"ח. פתחנו אחת מהן ומצאנו בה 100 ש"ח. יש לנו כעת אפשרות לבחור אם להחליף או לא להחליף מעטפות.
   א. נסחו את הבעיה כבעיית הכרעה בייסיאנית.
   ב. האם כדי להחליף מעטפות?

9. נתון משתנה מקרי דו-נורמלי $\underline{X}$ שלו ההתפלגות, $N_{\underline{X}}\left(\underline{m},\Sigma\right)$,

$$\underline{X}=\begin{pmatrix}x_1\\x_2\end{pmatrix},\underline{m}\begin{pmatrix}m_1\\m_2\end{pmatrix},\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$$

הראה כי ההתפלגות השולית אף היא נורמלית, עם

$$p\left(x_1\right)=N_{x_1}\left(m_1,\sigma_1^2\right)$$

גם ההתפלגות המותנית היא נורמלית, עם

$$p\left(x_1\mid x_2\right)=N_{x_1}\left(m_1+\rho\sigma_1\left(x_1-m_2\right)/\sigma_2;\sigma_1^2\left(1-\rho^2\right)\right)$$

10. נתונים לנו שני מספרים ממשיים שונים אך לא ידועים A ו-B. מראים לנו אחד מהם (בהסתברות שווה). האם יש דרך להכריע אם הוא הגדול או הקטן מביניהם בהסתברות שגיאה קטנה מחצי? הוכח. מה קורה אם A ו-B שלמים וערכם חסום?

תרגיל מחשב

יהי X_n משתנה מקרי הסופר את מספר הפעמים שמטבע הוגן נפל על "ראש" מתוך n הטלות מטבע. חשב את הערכים של $\Pr(X_{10}>6)$ $\Pr(X_{100}>60)$ ו- $\Pr(X_{1000}>600)$ והשווה לערכים המתקבלים מתוך חסמי מרקוב, צ'בישב וצ'רנוף.