תוכן העניינים:

מבוא - סטטיסטיקה תיאורית – בסיס כללי

הגדרה: סטטיסטיקה היא חקר שיטות לעיבוד נתונים.

עיבוד נתונים יכול להיעשות בשתי צורות:

תיאור הנתונים: כאשר אנו מסתכלים על רשימה ארוכה של נתונים גולמיים, קשה להניב מהסתכלות זו תכונות עבור רשימת, על כן, נצטרך לצמצם את הרשימה הארוכה לטבלה, גרף, מדדים המשקפים את מגמת הנתונים וכו'.

המטרה העיקרית בעיבודים סטטיסטים היא הסקה מהמדגם, הנתונים הגולמיים שנתקבלו, עבור כל האוכלוסייה כולה. מסקנות אלו יכולות להיות אישוש של השערות מסוימות או בניית מודלים מתקדמים. למשל, בעולם הביוסטטיסטיקה (סטטיסטיקה עבור מחקרים רפואיים) ישנם מחקרים בהם נרצה לדעת האם תרופה חדשה עוזרת או לא, יש להתייחס להשערה זו באופן רציני ביותר כיוון שאנו מדברים על חיי אדם. כדי לבצע ניסוי שכזה נצטרך לדעת מראש מהו גודל המדגם שיביא לתוצאות מספקות, נזכור שניסוי כזה מתבצע על אנשים ולכן נרצה כמות מועטת ככל שניתן אך כמות אנשים זאת – כמות האנשים **במדגם** - חייבת להיות מספקת על מנת שנוכל להסיק מסקנות כוללניות עבור כל האנשים – כל **האוכלוסיה**.

מהו מדגם? מדגם הינו אוסף של תצפיות (אנשים, בעלי חיים, חפצים וכו') שנאסף מאוכלוסיה מסוימת.

באיסוף המדגם מתוך האוכלוסיה צריכים להתקיים כמה תנאים:

1. המדגם צריך להיות מקרי – לכל פריט באוכלוסיה יש הסתברות שווה להיבחר למדגם. הבדל בין מקרי לקבוע - כאשר אנו זורקים קובייה אנו יודעים שהתנוחה בה הקובייה נופלת היא מקרית, למשל, אם הקובייה נפלה על הערך 2 אזי הערך 2 הינו קבוע אך העובדה שבגללה הקובייה נפלה דווקא על 2 הינה מקרית.
2. מדגם מייצג - מדגם חייב לייצג את כל תתי הקבוצות שבתוך האוכלוסייה. לדוגמא, אם נרצה לדעת את הדעה הפוליטית במדינה נצטרך לבדוק בכל סוגי החברה: דתיים חילונים, חרדים, ערבים, יהודים, אזורי מגורים שונים וכו' .
3. גודל המדגם – נשאף לגודל מדגם גדול ככל האפשר על מנת להפחית את הסיכוי לטעויות (כמו בדוגמא הראשונה שנתנו – בעיית מחקר התרופות)

משתני המחקר וסוגיהם:

ראשית, נפריד בין שני מושגים:

תצפיות: הנחקרים

משתנים: התכונה הנחקרת

למשל עבור מחקר בן אנו חוקרים את ציוניי הסטודנטים אזי, התצפיות הם הסטודנטים והמשתנים הם הציונים

ישנם סוגים שונים של משתנים:

1. משתנה נומינלי – משתנה קטגוריאלי. במשתנים אלו אנו משתמשים במספרים על מנת לייצג מצב מסוים, כך שלערך אין משמעות. למשל, משתנה בשם "מצב משפחתי" נוכל לקבוע כי המספר 1 מייצג רווק\ה המספר 2 מייצג נשוי\אה המספר 3 מייצג גרוש\ה והמספר 4 מייצג אלמן\ה. כמובן שאין משמעות למספרים, אין זה אומר שרווק קיבל ערך נמוך כי הוא פחות טוב...

2. משתנה אורדינאלי- ניתן לבטא במילים ובמספרים. המספרים מבטאים סדר (מהמילה order = סדר) אך צריך להיזהר, משתנה זה מבטא סדר אך אין הוא מספק את המרחק בין הקטגוריות. למשל, ציונים: טוב, טוב מאוד, מצוין.

3. משתנה אינטרוולי – למשתנה אינטרוולי יש חשיבות לסדר המספרים וגם למרחק ביניהם. למשל, אם מסתכלים על אינטרוולים של זמנים, אזי הדקה ה- 14 אינה פי 2 זמן מהדקה ה-7.

4. משתנה יחס – משתנה כמותי שבו יש משמעות למרווחים בין המשתנים וליחס ביניהם. למשל, גיל, משקל, גובה ועוד.

5. משתנה דמי (דמה) – ניתן לבצע מניפולציות על משתנים קיימים ולהמירם למשתנה שנוח לעבוד איתו. משתנה דמי הינו משתנה דיכוטומי (משתנה דיכוטומי, או ניתן לומר בינארי, מקבל שני ערכים בלבד, למשל, בדיקת תוצאות של טיפול באדם חולה, האם הטיפול הצליח או לא. אם הטיפול הצליח תינן למשתנה ערך השווה, למשל, ל-1 ואם לא הצליח אזי ניתן את הערך -0). למשל, אם יש לנו משתנה מצב משפחתי המוגדר כארבע קטגוריות: רווק, נשוי, גרוש ואלמן. אזי ניתן להפכו לדיכוטומי על ידי יצירת ארבעה משתנים חדשים (רווק, נשוי, גרוש ואלמן), כל אדם שציין בשאלון שהוא רווק אזי יקבל את הערך 1 במשתנה רווק ועבור שאר המשתנים יקבל את הערך 0, באותו אופן אדם שציין בשאלון שהוא גרוש, אזי יקבל את הערך 1 במשתנה גרוש ובשאר המשתנים יקבל את הערך 0 וכך הלאה.

אנו נמיין את המשתנים לשתי קטגוריות כוללות:

1. משתנה איכותי – משתנה שערכיו הם פריטים לא מספריים (או שעבר קידוד, אך עדיין הוא איכותי כיוון שאין משמעות לערך שקבל).

2. משתנה כמותי – נפצל מקרים:

2.1. משתנה כמותי בדיד – משתנה זה יכול לקבל מספר סופי של ערכים.

2.2. משתנה כמותי רציף – משתנה זה יכול לקבל אינסוף ערכים – משתנה יחס.

הצגת הנתונים

כמו שכבר אמרנו, קשה לקבל תמונה מסכמת של נתונים המתקבלים כרשימה ארוכה של מספרים. נרצה לחשוב על שיטה שבה נוכל לקבץ את הנתונים בצורה ממצה.

דוגמא 1 : נניח כי ישנם בכיתה 35 תלמידים. התקבלו ציוניהם במבחן בסטטיסטיקה 1 :
79,72,68,79,72,86,79,68,72,79,87,63,57,72,63,68,63,79,57,100,79,72,86,79,63,57,86,63
.79,72,90,50,79,79,57

כפי שאנו רואים קשה לדעת מהי מגמת ציוני הכיתה. לכן,ננסה לארגן את הנתונים בצורה טובה יותר, למשל נוכל לאגד את כל הנתונים לטבלת שכיחות באופן הבא : בעמודה השמאלית, נסדר בסדר יורד\עולה את הערכים של הציונים. בעמודה הימנית, נרשום כמה ציונים כאלו קיימים.

טבלה 1.

$f(x)$	x
1	100
1	90
1	87
3	86
10	79
6	72
3	68
5	63
4	57
1	50
35	סך הכל

נגדיר כמה מושגים :

x - הינו משתנה המקבל, לצורך העניין בדוגמא זו, ערכים שונים של ציונים.

$f(x)$ - שכיחות הופעת המשתנה, אצלנו כמה פעמים מופיע ציון מסוים, למשל הציון 86 מופיע 3 פעמים (תוכלו לבדוק!). אנו מסמנים את השכיחות באות f מן המילה : frequency.

החיסרון בהצגת כל הנתונים : כאשר יש נתונים רבים יהיה מאוד קשה להציג אותם בצורה זו כיוון שנקבל טבלה מאוד ארוכה כך שנחזור לבעיה המקורית. הפתרון למצב זה הינו קיבוץ הנתונים למחלקות – קטגוריות.

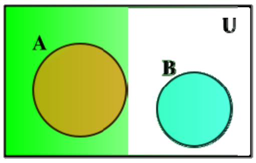
טבלה 2:

x	$f(x)$
93-100	1
85-92	5
77-84	10
69-76	6
61-68	8
53-60	4
45-52	1
סך הכל	35

על מנת לבנות טבלת מחלקות חייבים להתקיים מספר תנאים:

1. המחלקות חייבות להיות זרות.

תחילה ניתן הגדרה לקבוצות זרות (הגדרה זו לקוחה מעולם של "תורת הקבוצות"). קבוצות זרות: במתמטיקה קבוצות הן זרות נקראות כך אם אין להן איבר משותף למשל הקבוצות A ו- B הינן קבוצות זרות $B=\{4,5,6\}$, $A=\{1,2,3\}$.



נחזור לטבלה שלנו, ונשים לב כי אין איברים משותפים לשתי מחלקות (לא חשוב על אילו מחלקות נתמקד, בכל מקרה אין איברים משותפים).

2. החלוקה צריכה להיות ממצה – כלומר לא קיים ערך של x שאינו מופיע במחלקה מסוימת.
3. רוחב המחלקות נקבע ע"י הנוסחא הבאה: $x_{\max}-x_{\min}=$ רוחב מחלקה. נוסחא זו מניבה את האורך האופטימאלי של המחלקה.

כעת, נרצה לדעת כיצד בונים טבלה הדומה לטבלה 2. אך ראשית נגדיר כמה מושגים:

א. **גבול עליון**: הערך הגבוה בכל מחלקה

ב. **גבול תחתון**: הערך הנמוך בכל מחלקה

ג. **גבולות מדומים**: כאשר הגבול העליון של מחלקה אינו מתלכד עם הגבול התחתון של המחלקה שמעליה, מצב שכזה יוצר רווח בין המחלקות. במקרה שלנו – טבלה 2, יש גבולות מדומים כיוון שלמשל אין מחלקה עבור הערכים הנעים בין המספרים 92 עד 93.

ד. **גבולות אמיתיים**:כאשר אין רווח בין המחלקות. נוכל בקלות לבצע טרנספורמציה מטבלה בעלת גבולות מדומים לטבלה בעלת גבולות אמיתיים, נסביר זאת בהמשך – סעיף 4.

נקודת אמצע – נקודת אמצע של מחלקה מסוימת היא הממוצע של הגבול העליון והתחתון, בין אם מדובר בגבולות אמיתיים או מדומים.

טבלה 4:

$f(x)$	נקודת אמצע	x-גבולות מדומים	x - גבולות אמיתיים
1	96.5	93-100	92.5-100.5
5	88.5	85-92	84.5-92.5
10	80.5	77-84	76.5-84.5
6	72.5	69-76	68.5-76.5
8	64.5	61-68	60.5-68.5
4	56.5	53-60	52.5-60.5
1	48.5	45-52	44.5-52.5
35	סך הכל	סך הכל	סך הכל

שכיחות מצטברת: על מנת להציג את השכיחות המצטברת נוסיף עמודה בטבלה אשר בה נצבור את ערכיהם של השכיחויות מלמטה למעלה, מצב זה יביא לכך שעבור הערכים הגדולים ביותר של x נקבל את גודל המדגם כולו, במקרה שלנו 35. נסמן את השכיחות המצטברת ב- $F(x)$.

טבלה 5:

$F(x)$	$f(x)$	x
35=34+1	1	92.5-100.5
34=29+5	5	84.5-92.5
29=19+10	10	76.5-84.5
19=13+6	6	68.5-76.5
13=5+8	8	60.5-68.5
5=1+4	4	52.5-60.5
1	1	44.5-52.5
	35	סך הכל

שכיחות יחסית: נרצה לדעת מה היא השכיחות בכל מחלקה ביחס לכל גודל המדגם, נוכל לקבל זאת על ידי חילוק השכיחות בכל מחלקה בגודל המדגם כולו.

טבלה 6:

$\frac{f(x)}{n}$	$f(x)$	x
$1/35$	1	92.5-100.5
$5/35$	5	84.5-92.5
$10/35$	10	76.5-84.5
$6/35$	6	68.5-76.5
$8/35$	8	60.5-68.5
$4/35$	4	52.5-60.5
$1/35$	1	44.5-52.5
	35	סך הכל

שכיחות יחסית מצטברת: נרצה לדעת מה היא השכיחות המצטברת בכל מחלקה ביחס לכל גודל המדגם, נוכל לקבל זאת על ידי חילוק השכיחות המצטברת בכל מחלקה בגודל המדגם כולו. נשים לב כי פרופורציית השכיחות המצטברת עבור הערכים הגבוהים ביותר של x שווה ל-1, במקרה שלנו, עבור הערכים 92.5-100.5 אכן נקבל שכיחות יחסית מצטברת ששווה ל-1.

טבלה 7:

$\frac{F(x)}{n}$	$F(x)$	x
$35/35 = 1$	35	92.5-100.5
$34/35$	34	84.5-92.5
$29/35$	29	76.5-84.5
$19/35$	19	68.5-76.5
$13/35$	13	60.5-68.5
$5/35$	5	52.5-60.5
$1/35$	1	44.5-52.5
	35	סך הכל

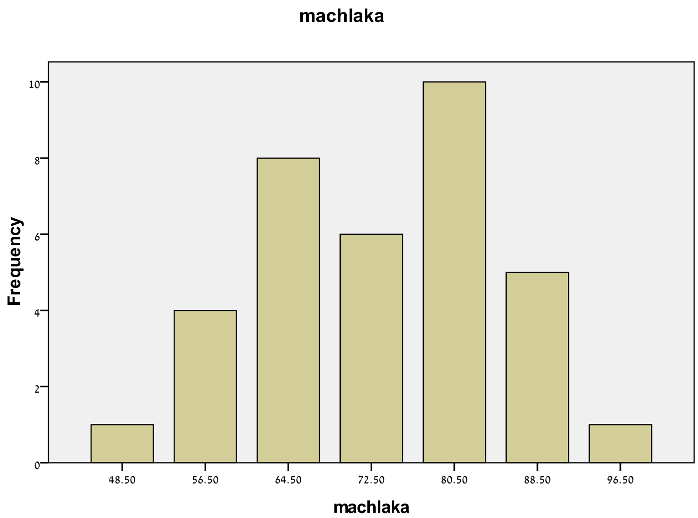
ניתן להציג את פרופורציות השכיחויות והשכיחויות המצטברות באחוזים.

גרפים: דיאגרמת מקלות, היסטוגרמה ודיאגרמת עוגה.

לפעמים מאוד נוח לנציג את הממצאים כגרף מסכם.

דיאגרמת מקלות: דיאגרמה זו מיועדת למשתנים כמותיים בדידים או אורינאלים, כאשר הציר האופקי מיצג את ציר ה- x והציר האנכי מייצג את השכיחות - $f(x)$.

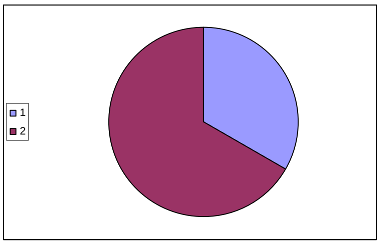
במקרה שלנו נוכל לבנות דיאגרמת מקלות ע״י שימוש בנקודת האמצע שהגדרנו לעיל בטבלה 4.



עוגה:הצגה זו מתאימה למשתנה איכותי-נומינלי. למשתנה הנחקר אין משמעות לסדר או למרחק. למשל, מין, מצב משפחתי, מקצוע וכו׳. למשל, בטבלת השכיחות הבאה:

f(x)	x
254	1
985	2

כאשר המשתנה מין **1**= **נקבה** ו- **2** =**זכר**. ניתן לראות כי במחקר זה יש יותר גברים.



היסטוגרמה:ההסטוגרמה מיועדת למשתנים רציפים, את הסטוגרמה נבנה מתוך טבלת השכיחויות בעלת גבולות אמיתיים, על מנת שלא יהיו רווחים בין המלבנים.

ניתן לחבר בין ראשי המקלות בדיאגרמת המקלות או את מרכזי המלבנים בהיסטוגרמה ולקבל קו שבור הנקרא פוליגון או מצולע. נשים לב כי כאשר רוחב המחלקות קבוע אזי המצולעים יהיו שווים, אך אם רוחב המחלקות לא יהיה הומוגני אז נקבל פוליגונים שונים (נראה זאת בדוגמא הבאה). חשוב להדגיש כי כאשר מדובר בהיסטוגרמה אשר רוחב המחלקות, בטבלת השכיחויות, שונה ממחלקה למחלקה, שטח המלבנים הוא זה שמייצג את השכיחות, ולכן המצולעים שנקבל – דיאגרמת המקלות וההסטוגרמה- יהיו שונים.

פונקצית צפיפות: כשמה כן היא, מתארת כמה צפוף עבור המשתנה בכל נקודה במרחב המדגם. אם נחפוץ למצוא את ההסתברות שמשתנה מקרי יימצא בטווח מסוים אזי נוכל לחשב את האינטגרל בשטח זה, משמע את השטח מעל קטע זה, על כן ברור כי המשתנה נוטה יותר לקבל ערכים שבהם הצפיפות גבוהה.

צפיפות: צפיפות היא השכיחות ליחידה אחת של המשתנה הנחקר. נסמן את הצפיפות באות d - density ונחשבה כך : $\frac{f(x)}{L}=d$ כאשר L הינו אורך המחלקה. במילים: צפיפות תחושב על-ידי חישוב היחס בין השכיחות לאורך המחלקה - צפיפות = שכיחות\ אורך מחלקה. דוגמא :

x	$f(x)$	L=אורך המחלקה	צפיפות = $\frac{f(x)}{L}=d$
0.5-3.5	32	3	$32/3=10.67$
3.5-6.5	146	3	$146/3=48.67$
6.5-16.5	190	10	$190/10=19$
16.5-36.5	32	20	$32/20=1.6$
סך הכל	400		

נוכל לראות כי הנתונים בטבלת השכיחות הבסיסית יוצרים התרשמות מוטעית, כתוצאה מכך שרוחב המחלקות אינו אחיד, למשל, עבור קבוצת הגיל 6.5-16.5 השכיחות הינה 190 שזאת השכיחות שלמראית עין הינה הערך הגדול ביותר, אך אין זה בולט כאשר חישבנו את הצפיפות או כאשר התבוננו בהסטוגרמה, אז קיבלנו תמונה אמיתית. לעומתה קבוצת הגילאים 3.5-6.5 היא מהווה את המלבן הגבוה ביותר בציור וערך הצפיפות שלה הכי גדול- 48.67. דבר המורה על העובדה כי קיים ריכוז גבוה של גילאים אלו באזור המדובר.

נזכיר כי אם היינו מחלקים את המחלקות באורכים שווים אזי ההסטוגרמה הייתה נראית בדיוק כמו דיאגרמת המקלות. כאשר אורך המחלקות שונה נקבל גרפים שונים, מצב זה מתרחש בעקבות העובדה שבהסטוגרמה ייצוג השכיחות בא לידי ביטוי גם האורך וגם בשטח.

מצולע שכיחויות:

מצולע שכיחות נוצר על ידי חיבור של כל נקודות האמצע של ההסטוגרמה. מעבר זה – מהסטוגרמה למצולע- נותן תמונה רציפה יותר של התופעה הנחקרת ובנוסף מראה את המגמתיות

דוגמא 3 : להלן נתונים על מספר ימי היעדרות של סטודנטים באוניברסיטה במשך שבוע, ב-40 שבועות : 14 16 15 14 11 18 21 17 19 18 11
21 19 19 15 17 16 16 13 17 14 15 15 17
19 18 16 14 18 13 16 20 19 16 21 12 13
17 21 17

א. הציגו את הנתונים בטבלת שכיחויות ומצאו נקודת אמצע, שכיחות יחסית ושכיחות יחסית מצטברת (חלקו למחלקות של 2 וקבעו שכל מחלקה כוללת את הגבול התחתון שלה).

ב. ציירו דיאגרמת מקלות מתאימה.

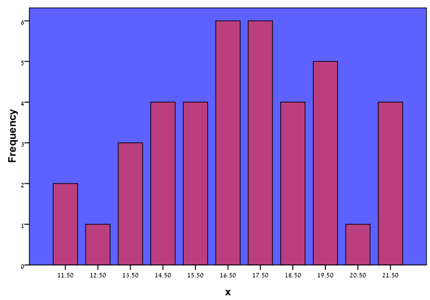
ג. ציירו הסטוגרמה מתאימה.

ד. ציירו מצולע מתאים

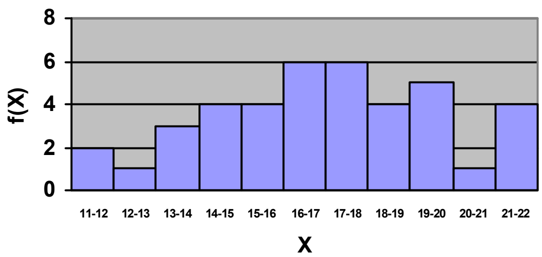
פתרון: א. טבלת השכיחויות:

x	$f(x)$	$F(x)$	נקודת אמצע	שכיחות יחסית	שכיחות מצטברת
11-12	2	2	11.5	1/20	1/20
12-13	1	3	12.5	1/40	3/40
13-14	3	6	13.5	3/40	3/20
14-15	4	10	14.5	1/10	1/4
15-16	4	14	15.5	1/10	7/20
16-17	6	20	16.5	3/20	1/2
17-18	6	26	17.5	3/20	13/20
18-19	4	30	18.5	1/10	3/4
19-20	5	35	19.5	1/8	7/8
20-21	1	36	20.5	1/40	9/10
21-22	4	40	21.5	1/10	1

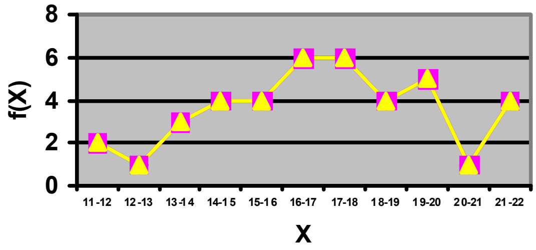
ב. דיאגרמת מקלות לפי נקודת האמצע:



הסטוגרמת שכיחויות



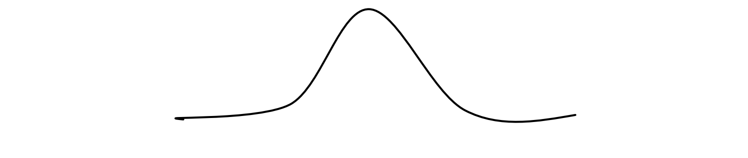
מצולע שכיחויות



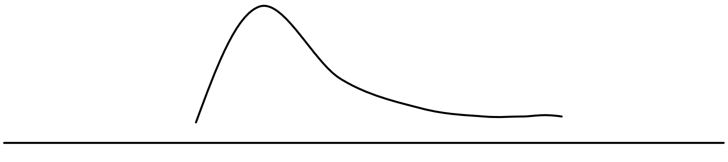
על מנת להבין את מושג הצפיפות ניתן דוגמא: נניח שנתון לנו משתנה המקבל כמות ערכים גדולה, למשל, נאספו הערכים של גובה כל הילדים בארץ ישראל במשך 10 שנים. את ערכים אלו תין להכניס לטבלת שכיחות בכמה אופנים, למשל, עם מחלקות רחבות כך שמספר המחלקות יהיה מועט, מחלקות שוות אורך, מחלקות שונות באורכן, או מחלקות צרות – ברור כי ככל שמספר המחלקות יהיה רב כך אורכן של המחלקות יהיה קטן. ישנו תהליך תיאוריטי די ברור שמתרחש: ככל שנקטין את רוחב המחלקה כך ההסטוגרמה ומצולע השכיחות יהיו יותר אינפורמטיביים, מפורטים יותר, משמעות הדבר כי נקבל דיוק רב עבור מצולע השכיחות. לשון אחר, ככל שכמות המחלקות ישאף לאינסוף כך נקבל מצולע שכיחות שקורב לפונקציה חלקה, הגובה מעל נקודה מסוימת נקרא צפיפות הערך המתאים לנקודה. חשוב להדגיש כי צפיפות אינה שכיחות. השכיחות מיוצגת על-ידי שטח המלבן ולכן תלויה גם ברוחבו. צפיפות היא שכיחות ליחידת x.

ישנן סוגים שונים של עקומות:

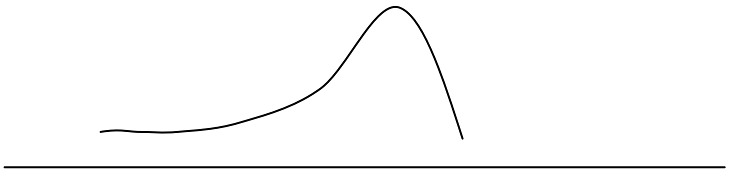
1. עקומה סימטרית פעמונית (חד שיאית)- רוב התצפיות מתרכזות במרכז ערכי x ומעט מהן בקצוות. פיזור התצפיות סביב המרכז הינו סימטרי.



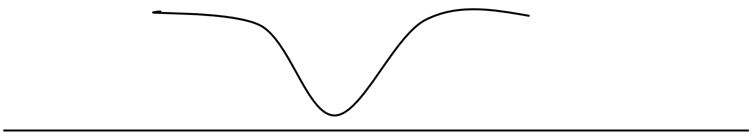
2. עקומה אסימטרית חיובית – רוב התצפיות מרוכזות בחלק ה"שמן" של העקומה והתצפיות החריגות מתקבלות בערכים הגבוהים של x



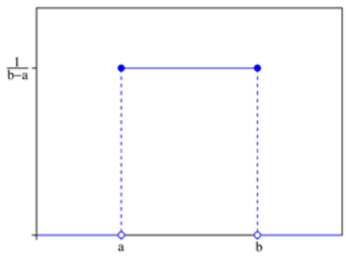
3. עקומה אסימטרית שלילית - רוב התצפיות מרוכזות בחלק ה"שמן" של העקומה והתצפיות החריגות מתקבלות בערכים הנמוכים של x



4. התפלגות סימטרית דו שיאית – נראית בצורת U הריכוז הגבוה נמצא הצדדים ובמרכז ישנו ריכוז נמוך של תצפיות



5. התפלגות אחידה – ריכוז שווה של תצפיות לכל אורך המשתנה x



6. התפלגות אסימטרית רב שיאית – ישנם כמה ערכי x בהם יש ריכוז גבוה של תצפיות (לא קיימת קונסיסטנטיות, עקביות, לאורך המשתנה x).

מדדים למיקום מרכזי

צעד ראשוני ובסיסי בהצגת נתונים ותמצותם הינו להציג ערך שישאף למקד את התכונה הנחקרת. אחד המדדים הידועים הינו ממוצע, למשל, כאשר הצטרכנו לספר על תעודת הבגרות שלנו, בדר"כ ציינו את הממוצע של המקצועות ולא ציינו ציון ציון של כל מקצוע ומקצוע.

כעת, נציג 4 מדדי מרכז – כל אחד ממדדים אלו ממוקם באזור מסוים בהתפלגות.

❖ שכיח (Mode): מסומן ב- MO. הערך ששכיחותו היא הגבוהה ביותר בהתפלגות. כאשר מדובר במשתנה כמותי רציף, אזי נמצא את השכיח על פי ערך הצפיפות הגבוה ביותר, ז"א ערך המקסימום בהתפלגות.

❖ אמצע הטווח (Midrange): מסומן ב- MR. ממוצע של שני הערכים הקיצוניים בהתפלגות. נוסחא: $MR=\frac{x_{\max}+x_{\min}}{2}$ (מדד זה פחות בשימוש).

❖ חציון (Median): מסומן ב- MD. הערך שמחצית מהמקרים קטנים או שווים לו ומחצית מהמקרים גדולים או שווים לו. ערך זה מחלק את ההתפלגות לשני חלקים שווים בשטחם (נזכור כי השטח מתאר את כל המקרים האפשריים).

* לא ניתן לחשב חציון למשתנה נומינלי.

בחישוב החציון נצטרך לעשות הפרדה בין סדרת ערכים לטבלת שכיחויות.

1. כאשר מדובר בסדרת ערכים אזי נבצע את הפעולות הבאות:

א. נסדר את הסדרה לפי סדר עולה.

ב. נמצא את מיקומו של החציון: אזי נבדיל בין המקרים

- אם מדובר במספר איברים זוגי אזי נמצא את הממוצע בין הערך שנמצא במקום ה- $\frac{n}{2}$ לערך שנמצא במקום $\frac{n+2}{2}$.
- אם מדובר במספר איברים אי-זוגי אזי נוסיף לערך זה 1 ואז נחלק ב-2. $\frac{n+1}{2}$

2. כאשר מדובר במשתנה אורדינאלי איכותי: נחשב תחילה את מיקומו של החציון: נעשה זאת ע"י חילוק של גודל המדגם (מספר התצפיות) ב-2 : $\frac{n}{2}$ (במקרה זה תמיד נחשב את $\frac{n}{2}$ בין אם מספר האיברים זוגי או אי זוגי). ונחפש בעמודת השכיחות המצטברת איכן נמצא המקרה.

3. כאשר מדובר במשתנה כמותי בדיד ננהג בדיוק כמו במשתנה האורדינאלי.

4. כאשר מדובר במשתנה כמותי רציף אזי נשתמש בנוסחא:

$$MD = \frac{\frac{n}{2} - F(x_{m-1})}{f(x_m)}\left(L_1 - L_0\right) + L_0$$

כאשר:

n - גודל המדגם

x_m - המחלקה בה נמצא החציון

L_1 - הגבול העליון האמיתי של מחלקה זו

L_0 - הגבול התחתון האמיתי של מחלקה זו

x_{m-1} - המחלקה הקודמת ל- x_m

$F(x_{m-1})$ - השכיחות המצטברת עד למחלקה x_{m-1}

$f(x_m)$ - שכיחות המחלקה x_m

❖ ממוצע (Mean): מסומן ב- $\overline{X}$. סוכמים את המשתנה ומחלקים במספר התצפיות

הלא הוא גודל המדגם. $\overline{X}=\frac{\sum_{i=1}^{n}X_i}{n}$. ניתן להציג את הממוצע גם כך

$\overline{X}=\frac{\sum_{i=1}^{n}f(x_i)\cdot x_i}{n}$ נשתמש בנוסחא זו כאשר השכיחות נתונה (כאשר ישנן מחלקות אזי נחשב את נקודת האמצע תחילה וכך נוכל להציב בנוסחא).

דוגמאות:

בטבלת שכיחות הבאה, נתונים הציונים של הסטודנטים בקורס סטטיסטיקה א'. מצא את אמצע הטווח. נתונים f_1 עד f_3 דוגמאות שונות למצבי שכיחויות שונים. מצא את השכיח בכל אחד מן המקרים:

i	x - ציון	f_1	f_2	f_3
1	6	20	**40**	**40**
2	7	20	20	**40**
3	8	20	2	10
4	9	20	8	5
5	10	20	30	5

אמצע הטווח= $\frac{6+10}{2}=8$

דוגמא 6: נתונים גלאי 400 אנשים. מצא את השכיח: מאחר ואורך המחלקות אינו זהה, נשתמש בצפיפות.

x	$f(x)$	L=אורך המחלקה	צפיפות = $\frac{f(x)}{L}=d$
0.5-3.5	32	3	$32/3=10.67$
3.5-6.5	146	3	$\boxed{146/3=48.67}$
6.5-16.5	190	10	$190/10=19$
16.5-36.5	32	20	$32/20=1.6$
סך הכל	400		

נוכל לראות כי ערך הצפיפות הגבוה ביותר הוא 48.67 אזי נסיק כי השכיח נמצא בטווח 3.5-6.5. מאחר ולא ניתן לקבוע שכיח כטווח, ניקח את נקודת האמצע: $\frac{3.5+6.5}{2}=5$

והיא מייצגת את השכיח.

דוגמא : נתונים גילאים של 7 ילדים :5, 8, 6, 2, 7, 6, 10. מצא את החציון :ראשית נסדר לפי סדר עולה : 2, 5, 6, 6, 7, 8, 10. כעת, נשתמש בנוסחא: $\frac{n+1}{2}\xrightarrow{n=7}\frac{7+1}{2}=4$ (השתמשנו בנוסחא זו כיוון שמדובר במספר אי זוגי של איברים - 7) לכן נסיק כי מיקומו של החציון במקום הרביעי 2, 5, 6, **6**, 7, 8, 10.

דוגמא : נתונים גילאים של 8 ילדים :5, 8, 6, 12, 2, 7, 6, 10. מצא את החציון :ראשית נסדר לפי סדר עולה : 2, 5, 6, 6, 7, 8, 10, 12. כעת, נשים לב כי מדובר במספר זוגי של איברים ולכן נמצא את הממוצע בין הערך שנמצא במקום ה- $\frac{n}{2}\xrightarrow{n=8}\frac{8}{2}=4$ לערך שנמצא במקום $\frac{n+2}{2}\xrightarrow{n=8}\frac{8+2}{2}=5$ לכן נסיק כי החציון הינו ממוצע של האיברים הממוקמים במקומות הרביעי והחמישי 2, 5, 6, **6**, **7**, 8, 10, 12 ולכן החציון הינו 6.5

דוגמא 9 : בטבלה הבאה ישנם נתונים עבור 24 מוצרים פגומים בגדלים שונים, מצא את החציון : ברור כי אנו מדברים על משתנה אורדינאלי מילולי

i	x - גודל המוצר	$f(x)$	$F(x)$
1	קטן	11	11
2	**בינוני**	**7**	**18**
3	גדול	5	23
4	ענק	1	24

נחשב את מיקומו של חציון : $\frac{n}{2}\xrightarrow{n=24}\frac{24}{2}=12$. נוכל לראות לפי העמודה כי האיבר ה12 "נופל" במחלקה מספר 2 המייצגת גודל מוצר בינוני.

דוגמא : בטבלה הבאה נתונים 45 תצפיות. מצא את החציון : תחילה נוכל לראות כי מדובר במשתנה כמותי רציף ולכן נחפש את החציון לפי השלבים :

i	x	$f(x)$	$F(x)$
1	4.5-3.5	2	2
2	3.5-4.5	5	7
3	4.5-5.5	14	21
4	**5.5-6.5**	**14**	**35**
5	6.5-7.5	4	39
6	7.5-8.5	3	42
7	8.5-9.5	3	45

נשים לב כי אם ישנם 45 תצפיות אזי החציון צריך להיות הערך שמעליו 22.5 תצפיות וכך גם מתחתיו. $MD = \frac{\frac{n}{2} - F(x_{m-1})}{f(x_m)}(L_1 - L_0) + L_0$

כאשר:

$n = 45$ - גודל המדגם

$x_m = x_4$ - המחלקה בה נמצא החציון

$L_1 = 6.5$ - הגבול העליון האמיתי של מחלקה זו

$L_0 = 5.5$ - הגבול התחתון האמיתי של מחלקה זו

$x_{m-1} \xrightarrow{m=4} x_3$ - המחלקה הקודמת ל- $x_m = x_4$

$F(x_{m-1}) \xrightarrow{x_{m-1}=x_3} F(x_3) = 21$ - השכיחות המצטברת עד למחלקה x_{m-1}

$f(x_m) \xrightarrow{x_m=x_4} f(x_4) = 14$ - שכיחות המחלקה x_m

נשאר רק להציב: $MD = \frac{22.5-21}{14}(6.5-5.5) + 5.5 = 5.6$.

דוגמא נוספת:

להלן נתונים על פתיחת תיקים במס הכנסה בישראל לפי גיל בשנת 2013: חשב חציון, ממוצע ושכיח לנתונים בטבלה

i	x	f(x)
1	20-25	2633
2	25-27	5487
3	27-34	1566
4	34-50	5684
5	50-52	4522

תחילה נחשב את עמודת השכיחות המצטברת

i	x	$f(x)$	$F(x)$
1	20-25	2633	2633
2	25-27	5487	8120
3	27-34	1566	9686
4	**34-50**	**5684**	**15370**
5	50-52	4522	19892

חציון: על מנת שנחשב את מדד החציון, נרצה לדעת באיזו מחלקה "נופל" החציון: $\frac{n}{2}\xrightarrow{n=19892}\frac{19892}{2}=9946$. על פי עמודת השכיחות המצטברת נוכל לראות כי החציון נמצא במחלקה הרביעית (המודגשת). נציב בנוסחא המתייחסת למשתנה כמותי רציף: $MD=\frac{\frac{n}{2}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)+L_0$ כאשר במקרה שלנו:

$n=19892$ - גודל המדגם

$x_m=x_4$ - המחלקה בה נמצא החציון

$L_1=50$ - הגבול העליון האמיתי של מחלקה זו

$L_0=34$ - הגבול התחתון האמיתי של מחלקה זו

$x_{m-1}\xrightarrow{m=4}x_3$ - המחלקה הקודמת ל- $x_m=x_4$

$F(x_{m-1})\xrightarrow{x_{m-1}=x_3}F(x_3)=9686$ -השכיחות המצטברת עד למחלקה x_{m-1}

$f(x_m)\xrightarrow{x_m=x_4}f(x_4)=5684$ - שכיחות המחלקה x_m

נחשב ונקבל: $MD=\frac{9946-9686}{5684}\left(50-34\right)+34=35$

ממוצע: על מנת לחשב את הממוצע נחשב את נקודת האמצע עבור כל מחלקה:

i	x	נקודת אמצע	$f(x)$	$F(x)$
1	20-25	22.5	2633	2633
2	25-27	26	5487	8120
3	27-34	30.5	1566	9686
4	34-50	42	5684	15370
5	50-52	51	4522	19892

נציב בנוסחא $\overline{X}=\frac{\sum_{i=1}^{n} f\left(x_i\right)\cdot x_i}{n}$ ונקבל:

$$\overline{x}=\frac{22.5\cdot 2633+26\cdot 5487+30.5\cdot 1566+42\cdot 5684+51\cdot 4522}{19892}\approx 36$$

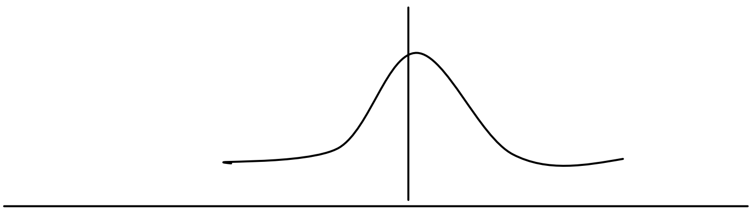
שכיח: על מנת למצוא את השכיח נוסיף עמודת צפיפות

i	x	נקודת אמצע	$f(x)$	L=אורך המחלקה	צפיפות = $\frac{f(x)}{L} = d$
1	20-25	22.5	2633	5	$2633/5 = 526.6$
2	25-27	26	5487	2	$5487/2 = 2743.5$
3	27-34	30.5	1566	7	$1566/7 = 223.7$
4	34-50	42	5684	16	$5684/16 = 355.25$
5	50-52	51	4522	2	$4522/2 = 2261$

הצפיפות הגדולה ביותר מתקבלת במחלקה השנייה ולכן השכיח הינו נקודת האמצע של מחלקה זו: 26.

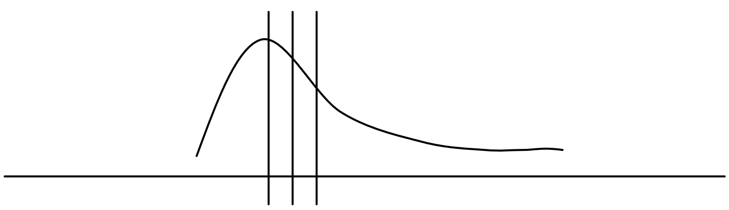
התייחסות לצורות ההתפלגויות השונות:

1. עקומה סימטרית פעמונית (חד שיאית)- שכיח = חציון = ממוצע



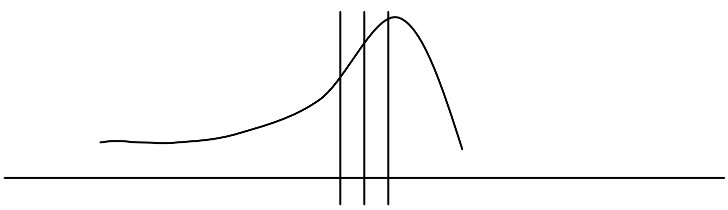
מ+ח+ש

2. עקומה אסימטרית חיובית – ממוצע> חציון> שכיח



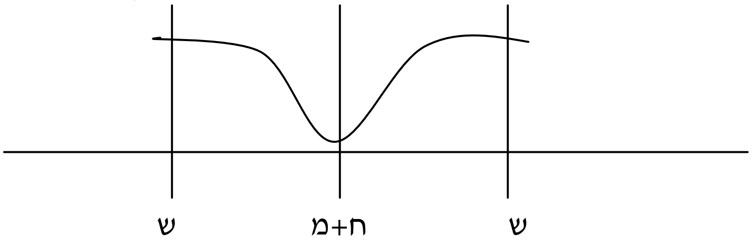
מ ח ש

3. עקומה אסימטרית שלילית – שכיח> חציון>ממוצע

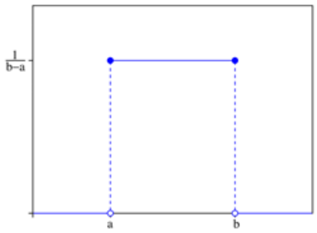


מ ח ש

4. התפלגות סימטרית דו שיאית – שני שכיחים ובמרכז החציון והממוצע מתלכדים



5. התפלגות אחידה – ממוצע, חציון ושכיח שווים לערך הממוצע של הקצוות



6. התפלגות אסימטרית רב שיאית – ישנם כמה ערכי x בהם יש ריכוז גבוה של תצפיות (לא קיימת קונסיסטנטיות, עקביות, לאורך המשתנה x).

חשוב להדגיש כי לא תמיד נוכל לחשב את כל המדדים, למשל אם מדובר בצבע עיניים אין טעם לחשב ממוצע או אמצע טווח, גם כאשר ישנו ערך מספרי, למשל, דירוג של סטודנטים המסיימים את המבחן : ראשון, שני, שלישי וכו'.

הממוצע הינו המדד המועדף כיוון שנוח לטיפול מתמטי וכך גם פונקצית הסיכון שלו - סכום ריבועי הסטיות (נזכור כי פונקצית ערך מוחלט לא גזירה בנקודה – 0 ולכן בעייתית).

משפט : ממוצע הסטיות של כל התצפיות מהממוצע שווה לאפס - $\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)}{n}=0$

הוכחה :

$$\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)}{n}=\frac{\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}\bar{x}}{n}$$

$$\frac{\sum_{i=1}^{n}x_i}{n}-\frac{\sum_{i=1}^{n}\bar{x}}{n}\xrightarrow{\frac{\sum_{i=1}^{n}x_i}{n}=\bar{x}}\bar{x}-\frac{\sum_{i=1}^{n}\bar{x}}{n}$$

$$\bar{x}-\frac{\left(n-1+1\right)\bar{x}}{n}=\bar{x}-\frac{n\bar{x}}{n}=\bar{x}-\bar{x}=0$$

דוגמא 11: עבור כל אחד מהסעיפים הבאים, קבע, האם ההתפלגות סימטרית, א-סימטרית חיובית או א-סימטרית שלילית.

א. ממוצע=85 חציון=84.7 שכיח=85.1 התפלגות סימטרית

ב. ממוצע=64 חציון=73 שכיח= 85 התפלגות א-סימטרית שלילית

ג. ממוצע=85 חציון= 73 שכיח=64 התפלגות א-סימטרית חיובית

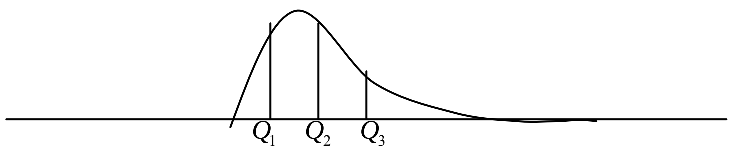
ד. ממוצע= 10 חציון=10 שכיח=10 התפלגות סימטרית

מדדי פיזור:

1. הטווח הכולל: הטווח מצביע על פיזור של ההתפלגות נסמנו באות R. $R = x_{\max} - x_{\min}$.

לדוגמא, אם ידוע כי טווח של ציונים נע בין 60-100 זה מעיד על פיזר קטן יותר מטווח ציונים של 40-100 (מדד זה פחות בשימוש)

2. הטווח הבינרבעוני: טווח שבו נמצאים 50% הערכיים האמצעיים. מדד זה הינו ההפרש בין הערך של הרבעון העליון לרבעון התחתון



כיצד נמצא אותו? תחילה נמצא את הרבעון העליון (אותו ערך ש- $\frac{3}{4}$ מהערכים קטנים ממנו). ואז נחשב את הרבעון התחתון (אותו ערך ש- $\frac{1}{4}$ מהערכים קטנים ממנו). כמו שכבר כתבנו הטווח הבינרבעוני הוא ההפרש בין הערך של הרבעון העליון לרבעון התחתון.

חישוב הרבעון העליון והתחתון דומה לחישוב החציון אך הפעם נציב $\frac{3n}{4}$, $\frac{n}{4}$ עבור רבעון עליון ותחתון בהתאמה.

למשל, רבעון עליון:

$$Q_3 = \frac{\frac{3n}{4} - F(x_{m-1})}{f(x_m)}\left(L_1 - L_0\right) + L_0$$

כאשר:

n - גודל המדגם

x_m - המחלקה בה נמצא הרבעון העליון

L_1 - הגבול העליון האמיתי של מחלקה זו

L_0 - הגבול התחתון האמיתי של מחלקה זו

x_{m-1} - המחלקה הקודמת ל- x_m

$F(x_{m-1})$ - השכיחות המצטברת עד למחלקה x_{m-1}

$f(x_m)$ - שכיחות המחלקה x_m

רבעון תחתון:

$$Q_1 = \frac{\frac{n}{4} - F(x_{m-1})}{f(x_m)}\left(L_1 - L_0\right) + L_0$$

כאשר:

n - גודל המדגם

x_m - המחלקה בה נמצא הרבעון התחתון

L_1 - הגבול העליון האמיתי של מחלקה זו

L_0 - הגבול התחתון האמיתי של מחלקה זו

x_{m-1} - המחלקה הקודמת ל- x_m

$F(x_{m-1})$ - השכיחות המצטברת עד למחלקה x_{m-1}

$f(x_m)$ - שכיחות המחלקה x_m

דוגמא 12: נתונה סדרת המספרים בת 24 איברים:

5,5,6,6,6,6,6,6,6,6,7,7,7,7,7,7,7,7,7,7,8,8,8,8,8 מצא את הרבעון העליון, החציון והרבעון התחתון:

פתרון: **רבעון תחתון**: $\frac{n}{4} = \frac{24}{4} = 6$ זאת אומרת שמיקומו של הרבעון התחתון הינו שישי

5,5,6,6,6,**6**,6,6,6,6,7,7,7,7,7,7,7,7,7,7,8,8,8,8,8 ולכן הרבעון התחתון הינו 6.

חציון: $\frac{n}{2} = \frac{24}{2} = 12$, $\frac{n+2}{2} = \frac{24+2}{2} = 13$. ז"א שהחציון הינו הממוצע בין האיברים שבמקום ה-12 וה-13: 5,5,6,6,6,6,6,6,6,6,7,**7,7**,7,7,7,7,7,7,7,8,8,8,8,8 ולכן החציון הינו 7.

רבעון עליון: $\frac{3n}{4} = \frac{3 \cdot 24}{4} = 18$ ז"א שמיקומו של הרבעון העליון במקום ה-18

5,5,6,6,6,6,6,6,6,6,7,7,7,7,7,7,7,**7**,7,7,8,8,8,8,8.

דוגמא 13 :נתונה טבלת השכיחות הבאה :

$F(x)$	$f(x)$	נקודת אמצע	x	i
2	2	200	100-300	1
10	8	400	300-500	2
19	9	600	500-700	3
21	2	800	700-900	4
22	1	1000	900-1100	5
25	3	1200	1100-1300	6
30	5	1400	1300-1500	7

מצא חציון, רבעון עליון, רבעון תחתון, עשירון עליון, עשירון תחתון, ממוצע.

פתרון : רבעון עליון : מיקום הרבעון העליון : $\frac{3\cdot 30}{4}=22.5$ ולכן :

$$Q_3=\frac{\frac{3n}{4}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)+L_0$$

כאשר :

n - גודל המדגם =30

x_m - המחלקה בה נמצא הרבעון העליון = x_6

L_1 - הגבול העליון האמיתי של מחלקה זו = 1300

L_0 - הגבול התחתון האמיתי של מחלקה זו = 1100

x_{m-1} - המחלקה הקודמת ל- x_m . $x_{m-1}=x_5$

$F(x_{m-1})$ - השכיחות המצטברת עד למחלקה x_{m-1} $F(x_{m-1})=22$

$f(x_m)$ - שכיחות המחלקה x_m $f(x_m)=3$

$$Q_3=\frac{22.5-22}{3}\left(1300-1100\right)+1100=1133.33$$

רבעון תחתון : מיקומו $\frac{30}{4}=7.5$. כעת, נוכל למצוא אותו :

$$Q_1=\frac{\frac{n}{4}-F(x_{m-1})}{f(x_m)}\left(L_1-L_0\right)+L_0$$

כאשר:

n - גודל המדגם = 30

x_m - המחלקה בה נמצא הרבעון התחתון = x_2

L_1 - הגבול העליון האמיתי של מחלקה זו = 500

L_0 - הגבול התחתון האמיתי של מחלקה זו =300

x_{m-1} - המחלקה הקודמת ל- x_m . $x_{m-1} = x_1$

$F(x_{m-1})$ - השכיחות המצטברת עד למחלקה x_{m-1} . $F(x_{m-1}) = 2$

$f(x_m)$ - שכיחות המחלקה x_m . $f(x_m) = 8$

$$Q_1 = \frac{7.5-2}{8}\left(500-300\right)+300 = 437.5$$

- כאשר מתקיים $Q_2 - Q_1 = Q_3 - Q_2$ אזי מדובר בהתפלגות סמטרית
- כאשר מתקיים $Q_3 - Q_2 > Q_2 - Q_1$ אזי מדובר בהתפלגות א-סמטרית ימנית
- כאשר מתקיים $Q_3 - Q_2 < Q_2 - Q_1$ אזי מדובר בהתפלגות א-סמטרית שמאלית

שונות וסטיית תקן:

המדד הנפוץ ביותר המורה על פיזור המשתנים הינו סטיית התקן או השונות.

סטיית תקן: שורש של סכום הסטיות מהממוצע בריבוע חלקי גודל המדגם.

$$s = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}$$

ניתן להציג את הנוסחא גם באופן הבא: $s = \sqrt{\frac{\sum_{i=1}^{n} f\left(x_i\right)\cdot x_i^2}{n} - \bar{X}^2}$

שונות: ערך סטיית התקן בריבוע. $s^2 = \frac{\sum_{i=1}^{n} f\left(x_i\right)\cdot x_i^2}{n} - \bar{X}^2$

דוגמאות:

דוגמא 14: נתונות התצפיות הבאות: 2,4,6,8,10 מצא את הטווח הכולל, את סטיית התקן ואת השונות:

פתרון אמצע הטווח: 10-2=8

על מנת למצוא את סטיית התקן נמצא תחילה את הממוצע $\overline{X}=\frac{2+4+6+8+10}{5}=6$ כעת, נמצא את סטיית התקן, להדגמה נראה על-ידי שתי הנוסחאות:

$$s=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n}}=\sqrt{\frac{(2-6)^2+(4-6)^2+(6-6)^2+(8-6)^2+(10-6)^2}{5}}=\sqrt{8}=2.82$$

$$s=\sqrt{\frac{\sum_{i=1}^{n}f(x_i)\cdot x_i^2}{n}-\overline{X}^2}=\sqrt{\frac{2^2+4^2+6^2+8^2+10^2}{n}-6^2}=\sqrt{8}=2.82$$

השונות שווה לסטיית התקן בריבוע ולכן ערך השונות שווה ל-8.

דוגמא 15: נתונה הטבלה הבאה, מצא את סטיית התקן.

i	x	נקודת אמצע	$f(x)$	x^2	$f(x_i)\cdot x_i^2$
1	1-3	2	2	4	$2\cdot 4$
2	3-5	4	2	16	$2\cdot 16$
3	5-7	6	3	36	$3\cdot 36$
4	7-9	8	4	64	$4\cdot 64$
5	9-11	10	1	100	$1\cdot 100$
$\text{sum}=\sum_{i=1}^{n}f(x_i)\cdot x_i^2$					504

כעת נמצא את הממוצע: $\bar{x}=\frac{2\cdot 2+2\cdot 4+6\cdot 3+8\cdot 4+10\cdot 1}{12}=6$

נשאר רק להציב בנוסחא את כל המרכיבים:

$$s=\sqrt{\frac{\sum_{i=1}^{n}f(x_i)\cdot x_i^2}{n}-\overline{X}^2}\xrightarrow{\sum_{i=1}^{n}f(x_i)\cdot x_i^2=504\ ,\ n=12\ ,\ \overline{X}=6}\sqrt{\frac{504}{12}-6^2}=2.44$$

כמה כללים חשובים:

1. כאשר כל הערכים זהים בערכם אזי הממוצע גם הוא זהה לערכים אלו וסטית התקן שווה ל-0.
2. כידוע סטיית תקן מבטאת את המרחק \ פיזור של התצפיות מהממוצע ולכן נוכל להסיק כי בתחום הסגור $\left[\bar{X}-s_x \ , \ \bar{X}+s_x\right]$ ימצאו רוב התצפיות.
3. סכום כל הסטיות מהממוצע שווה ל-0 (כבר הוכחנו)
4. כאשר מוסיפים תצפיות בקצוות ההתפלגות, ז״א מחוץ לתחום הסגור $\left[\bar{X}-s_x \ , \ \bar{X}+s_x\right]$ אזי הפיזור=סטיית התקן גדל\ה.
5. כאשר מוסיפים תצפיות בתוך ההתפלגות, ז״א בתוך התחום הסגור $\left[\bar{X}-s_x \ , \ \bar{X}+s_x\right]$ אזי הפיזור=סטיית התקן קטן\ה.
6. מעצם חישובה, סטיית התקן תמיד חיובית (גדולה או שווה ל-0).

דוגמא 18: נתון כי ממוצע של 100 סטודנטים הינו 90 וסטיית התקן שווה ל-5. איכן מתרכזים רוב ציוני הסטודנטים?

פתרון: נמצא את התחום $\left[\bar{X}-s_x \ , \ \bar{X}+s_x\right]$:

$\left[\bar{X}-s_x \ , \ \bar{X}+s_x\right] \xrightarrow{\bar{X}=90 \ , \ s_x=5} \left[90-5 \ , \ 90+5\right]=\left[85 \ , 95\right]$. ז״א שרוב הסטודנטים קיבלו ציונים הנעים בין 85 ל- 95.

אלגברה של מאורעות

הקדמה

הסתברות קלאסית: ניסוי עם כמה תוצאות אפשריות.

מרחב מדגם: כל האפשריות האפשריות עבור אירוע מסוים. את מרחב המדגם נסמן באות Ω.

הגדרה מתמטית: מרחב מדגם בדיד הוא זוג סדור (Ω, P) שבו Ω היא קבוצה סופית או בת מניה. $P:\Omega \to \mathbb{R}$ פונקציה חיובית המקיימת $\sum_{x\in\Omega} P(x) = 1$.

* ההגדרה אינה מתאימה למרחבים שאינם בני מניה.

כל תת קבוצה $A \subseteq \Omega$ נקראת מאורע המיצגת אירוע שיכול להתרחש.

דוגמא: זורקים מטבע שלוש פעמים. מהו מרחב המדגם?

פתרון: כידוע כל פעם שאנו זורקים קובייה ישנן שתי אפשרויות – עץ או פלי בהסתברות זהה, לכן, עבור שלוש זריקות ישנן הרבה אפשרויות. אפשרויות אלו נקראות מרחב במדגם. בכל סוגריים נכתוב אפשרות אחרת המתקבלת מתוך זריקת הקובייה שלוש פעמים, כאשר האות "ע" מסמלת קבלת עץ והאות "פ" מסמלת קבלת פלי. למשל, כאשר רושמים (פעפ) מתכוונים שבזריקה הראשונה התקבל פלי, בזריקה השנייה התקבל עץ ובזריקה השלישית התקבל פלי.

Ω = (עעע) ,(ععפ) ,(עפע) ,(פעע) ,(פפפ) ,(פפע) ,(פעפ) ,(עפפ)

מספר הצירופים כמספר הסוגריים ז"א – 8.

לכל זריקת מטבע ישנן 2 אפשרויות – עץ או פלי ואת המטבע זורקים שלוש פעמים לכן מרחב המדגם הינו $2^3 = 8$

דוגמא 2: נגדיר את המאורעות הבאים: A, B, C, ו- D באופן הבא:

המאורע A - כאשר קיבלנו שני עצים רצופים.

המאורע B - כאשר קיבלנו פלי בהטלה השנייה.

המאורע C - כאשר קיבלנו פלי בהטלה השלישית.

המאורע D - כאשר קיבלנו שלושה עצים רצופים.

כאשר נרשום P( ) אנו מתכוונים להציג את השאלה: מה ההסתברות שהמאורע שרשום בתוך הסוגריים התרחש?, למשל כאשר נכתוב P(A) נתכוון לשאלה: מה ההסתברות שהמאורע A התרחש?

הגדרה פורמאלית: נניח כי ישנם N תרחישים אפשריים, שווי הסתברות, ומהם A תרחישים שאנו מגדירים כהצלחה, אזי הסיכוי שמשהו מוצלח יתרחש הינו $\frac{A}{N}$

* הסתברות קלאסית של מאורע A מסומנת $p(A)$ ומוגדרת להיות $p(A) = \frac{\#A}{\#\Omega}$

חישוב מפורט:

A: {(עעע) (פעע) (עעפ)} $P(A) = \frac{3}{8}$

B: {(עפפ) (פפע) (פפפ) (עפע)} $P(B) = \frac{4}{8}$

C: {(עפפ) (פעפ) (פפפ) (עעפ)} $P(C) = \frac{4}{8}$

D: {(עעע)} $P(D) = \frac{1}{8}$

חיתוך בין מאורעות:

הגדרה: החיתוך של שתי קבוצות A ו – B היא הקבוצה אשר איבריה הם כל האיברים הנמצאים גם ב – A וגם ב – B, כלומר האיברים המשותפים ל – A ול – B, או ניתן לומר המקרים המשותפים בין שני המאורעות. קבוצת החיתוך תסומן $A \cap B$.

למשל: ההסתברות לחיתוך של מאורעות A ו- D שווה: $P(A \cap D) = \frac{1}{8}$ כיוון שאנו מחפשים את המקרים המשותפים שהתרחשו ב- A **<u>וגם</u>** ב- D, במקרה שלנו יש רק אירוע **אחד** שמשותף מתוך סך שמונה האפשרויות והוא "שלושה עצים רצופים" - {(עעע)} $= A \cap D$ ניתן לראות זאת על – ידי הסתכלות של המאורעות ומציאת המשותף ביניהם.

- כאשר קבוצת החיתוך של A ו – B היא קבוצה ריקה נאמר כי ל – A ול – B אין איבר משותף, $A \cap B = \{\varnothing\}$, והן מוגדרות כקבוצות זרות. בנוסף,ההסתברות למצב שכזה הינה אפס $P(A \cap B) = 0$. במקרה זה של קבוצות זרות, נוכל לומר כי $P(A \cup B) = P(A) + P(B)$

1. ההסתברות לחיתוך המאורעות A וגם B: כיוון שאין משותף למאורעות A ו- B אזי נסמן את החיתוך כקבוצה ריקה $A \cap B = \{\varnothing\}$ ומכאן נובע כי הסתברות להתרחשות זו הינה -0 ונסמן זאת כך: $P(A \cap B) = \frac{0}{8} = 0 \quad \rightarrow \quad A \cap B = \{\varnothing\}$

2. ההסתברות לחיתוך המאורעות C וגם B: מהסתכלות על המאורעות B ו- C נוכל לראות כי ישנם שני אברים משותפים {(פפפ) (עפפ)} $= B \cap C$ ולכן ההסתברות לחיתוך זה הינה שתי אפשרויות מתוך 8, ונסמן: $P(B \cap C) = \frac{2}{8}$.

{(פפפ) (עפפ)} $= B \cap C$

3. ההסתברות לחיתוך המאורעות A **וגם** D: $P(A \cap D) = \frac{1}{8}$ {(עעע)} $= A \cap D$

4. ההסתברות של חיתוך המאורעות B **וגם**– D הינה קבוצה ריקה ולכן ההסתברות שווה לאפס: $P(B \cap D) = 0 \quad B \cap D = \{\varnothing\} \rightarrow$.

5. ההסתברות של חיתוך המאורעות C וגם– D הינה קבוצה ריקה ולכן ההסתברות שווה לאפס: $D \cap C = \{\varnothing\} \quad \to P(D \cap C) = 0$

איחוד בין מאורעות:

הגדרה: כל המקרים שנכללים במאורעות הנתונים .

למשל: קבוצת האיחוד של המאורעות A ו – B הינה כל האיברים שנמצאים גם בקבוצה A וגם בקבוצה B (את האיברים שחופפים נכתוב פעם אחת ולא פעמיים!!!)

בדוגמא שלנו :

A: {(עעע) (פעע) (עעפ)}

B: {(עפפ) (פפע) (פפפ) (עפע)}

לכן קבוצת האיחוד הינה: $A \cup B =$ {(עעע)(עפפ) (פפע) (פפפ) (פעע) (עפע) (עעפ)}

ההסתברות להתרחשות "A איחוד B" הינה $P(A \cup B) = \frac{7}{8}$ כיוון שישנן 7 אפשרויות בקבוצת האיחוד וכל זאת מתוך 8 אפשרויות.

דוגמא 1: מצא/י את ההסתברות של $A \cup D$.

פתרון: ידוע כי

A: {(עעע) (פעע) (עעפ)}

D: {(עעע)}

ולכן קבוצת האיחוד הינה: $A \cup D =$ {(עעע) (פעע) (עעפ)} נשים לב כי את המאורע (עעע) רשמנו רק **פעם אחת** כיוון שזהו האיבר המשותף – נוכל לראות שאיבר זה הוא בדיוק החיתוך של מאורעות אלו - $A \cap D =$ {(עעע)} .

ניתן למצוא את ההסתברות לאיחוד על-ידי נוסחה:

נוסחת האיחוד:

$$P\left(A \cup B\right) = P(A) + P(B) - P(A \cap B)$$

נתבונן שוב בדוגמא הראשונה שהצגנו, בדוגמא זו הגענו למסקנה, ע"י הסתכלות על המאורעות עצמם כי $P(A \cup B) = \frac{7}{8}$, כעת, ננסה להגיע לאותה המסקנה על-ידי הנוסחא:

כאמור :

A: {(עעע) (פעע) (עעפ)} $\quad P(A) = \frac{3}{8}$

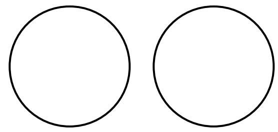
B: {(עפפ) (פפע) (פפפ) (עפע)} $\quad P(B) = \frac{4}{8}$

מאורעות זרים לכן החיתוך ריק: $A \cap B = \{\varnothing\} \quad \rightarrow \quad P(A \cap B) = \frac{0}{8} = 0$

כעת, נוכל להציב את כל הגורמים בנוסחה ולקבל:

$$P\left(A \cup B\right) = P(A) + P(B) - P(A \cap B) = \frac{7}{8}$$

$$\frac{3}{8} \quad + \quad \frac{1}{8} \quad - \quad 0$$

A  B

דוגמא 1 – לפי הנוסחה: באותו אופן נוכל לחשב את ההסתברות של $A \cup D$.

$A \cup D = \{$(עעע) (פעע) (עעפ)$\} \qquad \rightarrow \qquad P(A \cup D) = \frac{3}{8}$

לפי הנוסחה נוכל לראות שנקבל ערך זהה:

$$P\left(A \cup D\right) = P(A) + P(D) - P(A \cap D) = \frac{3}{8}$$

$$\frac{3}{8} \quad + \quad \frac{1}{8} \quad - \quad \frac{1}{8} \quad = \frac{3}{8}$$

ניתן להרחיב את הנוסחא עבור שלוש קבוצות:

$$P\left(A \cup B \cup D\right) = P(A) + P(B) + P(D) - P(A \cap D) - P(A \cap B) - P(B \cap D) + P(A \cap B \cap D)$$

כמובן שניתן להרחיב אף יותר עבור n קבוצות, בדר״כ נעשה זאת ע״י הגירסה ההסתברותית של נוסחת ההכלה וההדחה.

מאורע משלים:

הגדרה: קבוצת המשלים של מאורע A מסומנת ב - A^C. משמעות הסימון של הקבוצה A^C הינה "A משלים" או באנגלית "**A complete**". A^C - כל האיברים ב - Ω שלא נמצאים ב – A.

- ❖ **מעצם הגדרת המשלים ברור כי לכל A סכום ההסתברויות של A ו - A^C שווה ל- 1**

למשל בדוגמא שלנו: המשלים של המאורע A הינו A^C וכאמור המאורע A מוגדר להיות קבלת שני עצים רצופים, לכן המאורע A^C יהיה זריקת מטבע שלוש פעמים **ולא** לקבל שני עצים רצופים.

❖ נגדיר את A^C : {(פעפ) (עפפ) (פפע) (פפפ) (עפע)} $= A^C$, ההסתברות להתרחשות A^C הינה $P(A^C)=\frac{5}{8}$. נוכל לראות כי סכום ההסתברויות שווה ל – 1, כזכור $P(A)=\frac{3}{8}$.

❖ נגדיר את המשלים של B : {(פעפ) (פעע) (עעפ) (עעע)} $= B^C$ ולכן הסתברותו הינה: $P(B^C)=\frac{4}{8}$

❖ נוכל למצוא משלים של האיחוד A ו- B, נוכל לראות כי בקבוצת המשלים ישנו איבר אחד {(פעפ)} $= (A\cup B)^C$ ולכן ההסתברות הינה $P(A\cup B)^C=\frac{1}{8}$.

❖ נמצא את החיתוך של A^C ו - B^C, נוכל להתבונן במאורעות A^C ו - B^C ולמצוא את האיברים המשותפים: {(פעפ)} $= A^C\cap B^C$, כיוון שמדובר באיבר אחד ההסתברות שווה לשמינית : $P(A^C\cap B^C)=\frac{1}{8}$

חלק א' – שאלות בסיסיות

דוגמא: במשפחה של 4 ילדים מה ההסתברות ל (א) בת בכורה (ב) בת בכורה ובן זקונים (ג) לפחות בת אחת (ד) בת אחת בדיוק. ידוע כי הסתברות להולדת בן ובת שווים.

פתרון: במשפחה 4 ילדים, תחילה נוכל לכתוב את כל האפשרויות הקיימות עבור נתון זה

$\Omega =$ { (בת, בת, בת, בת) , (בן, בן, בן, בן)

(בן, בת, בת, בת), (בת, בן, בת, בת), (בת, בת, בן, בת), (בת, בת, בת, בן) ,

(בן, בן, בת, בת), (בן, בת, בן, בת), (בן, בת, בת, בן), (בת, בן, בת, בן), (בת, בן, בן, בת) ,

(בת, בת, בן, בן) ,

(בן, בן, בן, בת) , (בת, בן, בן, בן) , (בן, בן, בת, בן) , (בן, בת, בן, בן) }

ידענו מראש כי ישנן 16 אפשרויות, כמו שהגדרנו בהתחלה בכל לידה יש שתי אפשרויות או בן או בת – וכיוון שישנם ארבעה ילדים אז היו 4 לידות ולכן מרחב המדגם הינו $2^4=16$

כעת נוכל לענות על השאלות:

א. ההסתברות לבת בכורה: קיימים 8 מקרים בהם הבת ראשונה והם:

(בת, בת, בת, בת), (בת, בן, בת, בת), (בת, בת, בן, בת), (בת, בת, בת, בן), (בת, בן, בת, בן), (בת, בן, בן, בת) , (בת, בת, בן, בן) , (בת, בן, בן, בן) ולכן : $\frac{8}{16}=0.5$

ב. ההסתברות לבת בכורה ובן זקונים: קיימים 4 מקרים כאלו והם:

(בת, בת, בת, בן), (בת, בן, בת, בן), (בת, בת, בן, בן), (בת, בן, בן, בן) ולכן : $\frac{4}{16}=\frac{1}{4}$

ג. ההסתברות ל-לפחות בת אחת : כל המקרים חוץ ממקרה אחד - (בן, בן, בן, בן) ולכן יש 15 אפשרויות נחשב את ההסתברות $\frac{15}{16}=0.9375$

ד. ההסתברות לבת אחת בדיוק : יש 4 אפשרויות והם : (בן, בן, בן, בת) , (בת, בן, בן, בן) , (בן, בן, בת, בן) , (בן, בת, בן, בן) ולכן $\frac{4}{16}=\frac{1}{4}$

דוגמא : על מנת לקבל אישור על מעבר מפקולטה מסוימת לפקולטה למדעי המחשב צריך הסטודנט לעבור בהצלחה שתי בחינות. ההסתברות לעבור בהצלחה את הבחינה ראשונה היא 0.6 וההסתברות לעבור בהצלחה את הבחינה השנייה היא 0.7. ההסתברות של סטודנט מסוים לעבור בהצלחה לפחות אחת מן הבחינות היא 0.95. מה הסיכוי שסטודנט זה לא יקבל אישור על מעבר למחלקה למדעי המחשב?

פתרון: לפי הנתונים בשאלה כדי לקבל אישור מעבר צריך סטודנט מסוים לעבור בהצלחה 2 בחינות. המספרים המסומנים עם **קו תחתון הם הנתונים**.

לנתון 0.05 הגענו ע"י הנתון כי : הסתברות של סטודנט מסוים לעבור בהצלחה **לפחות אחת מהבחינות** היא 0.95, ז"א שהמשלים מוגדר כהסתברות לא לעבור את הבחינה הראשונה וגם לא את השנייה. כיוון שזהו המשלים לנתון, נוכל לחשב ע"י ניצול העובדה כי מאורעות משלימים – משלימים ל-1 לכן $1-0.95=0.05$

	לעבור את הבדיקה הראשונה	לא לעבור את הבדיקה הראשונה	
לעבור את הבדיקה השנייה	$0.6-0.25=0.35$	$0.4-0.05=0.35$	0.7
לא לעבור את הבדיקה השנייה	$0.3-0.05=0.25$	0.05	$1-0.7=0.3$
	0.6	$1-0.6=0.4$	1

המאורע הנדרש בשאלה (סטודנט זה לא יקבל אישור מעבר) מתרחש כאשר אחת משלושת האפשרויות.

א. לא עבר את הבחינה הראשונה אך את השנייה עבר

ב. לא עבר את הבחינה השנייה אך את הראשונה עבר

ג. לא עבר את שתי הבחינות

כיוון שמדובר בשלושה אפשרויות, ז"א ש א' מתרחש **או** ש ב' מתרחש **או** ש ג' מתרחש, נצטרך לחבר בין ההסתברויות (אם היינו משתמשים במילה "וגם" היינו מכפילים בין האפשרויות)

$$p(no\ \text{approval})=\underbrace{0.35}_{\text{no 1}}+\underbrace{0.25}_{\text{no 2}}+\underbrace{0.05}_{\text{no (1\&2)}}=0.65$$

❖ נוכל ללמוד מדוגמא זו כי כאשר משתמשים במילה "או" בין אפשרויות אזי נחבר ביניהם

דוגמא: ההסתברות לעבור את הבחינה במתמטיקה היא 0.85 ובכלכלה היא 0.8. ההסתברות לעבור את שתיהן היא 0.75. מה הסיכוי להיכשל בשתיהן?

פתרון: הסיכוי להיכשל בשתי הבחינות, מסומן באליפסה:

	להיכשל מתמטיקה	לעבור מתמטיקה	
0.8	$0.8 - 0.75 = 0.05$	0.75	לעבור כלכלה
$1 - 0.8 = 0.2$	$0.2 - 0.1 = 0.1$	$0.85 - 0.75 = 0.1$	להיכשל כלכלה
1	$1 - 0.85 = 0.15$	0.85	

דוגמא: סטודנט מאחר לשיעור כאשר יורד גשם או כאשר יש עומס תנועה. ב-60% מהימים יש עומס תנועה. ב-10% מהימים יורד גשם. ב-60% מן הימים הסטודנט מאחר לשיעור. האם תמיד כשיורד גשם יש עומס? האם תמיד כשיש עומס יורד גשם?

פתרון: נתון כי סטודנט מאחר לשיעור כאשר יורד גשם או כאשר יש עומס תנועה, נתון גם, כי הסיכוי שסטודנט מאחר הינו 0.6, ז"א שהסיכוי 0.6 הינו שיש גשם או שיש עומס או שניהם, המשלים 1-0.6 הינו ההסתברות שאין גשם ואין עומס – לכן נוכל להסיק כי ההסתברות שאין גשם ואין עומס הינה $1 - 0.6 = 0.4$.

א. האם תמיד כשיורד גשם יש עומס? כן, כעת, נסביר מדוע, נתבונן בעמודה של "יש גשם" ונוכל לראות כי ההסתברות ש- "יש גשם" ו- "אין עומס" הינה אפס, זאת-אומרת שכאשר יש גשם לעולם לא יהיה עומס – זוהי אופציה שלא קיימת!

ב. האם תמיד כשיש עומס יורד גשם? לא, נתבונן בשורה של "עומס" ונוכל לראות כי ישנה הסתברות מסוימת (ערך מסוים) עבור -"יש גשם" ו- "עומס"- ועבור "אין גשם" ו- "עומס" לכן ישנה אופציה שכאשר יש עומס יש או אין גשם, מצב זה שונה מסעיף א' בו ההסתברות של "יש גשם" ו- "אין עומס" שווה ל-0

	אין גשם	יש גשם	
0.6	$0.9 - 0.4 = 0.5$	$0.1 - 0 = 0.1$	עומס
$1 - 0.6 = 0.4$	0.4	$0.4 - 0.4 = 0$	אין עומס
1	$1 - 0.1 = 0.9$	0.1	

כלל יסודי: אם הניסוי ראשון יש n תוצאות אפשריות ובניסוי שני יש m תוצאות אפשריות אזי מספר התוצאות האפשריות כאשר עושים את שני הניסויים הוא nm.

דוגמא: במספר רישוי של רכב 3 ספרות (מ-0 עד 9) ו3 אותיות לטיניות (A-Z), כמה מספרי רישוי אפשריים?

כל ספרה היא ניסוי בפני עצמו ולכן עבור המספרים: $10 \cdot 10 \cdot 10 = 10^3$

10 אפשרויות לבחירת המספר הראשון וגם 10 אפשרויות לבחירת המספר השני וגם אפשרויות לבחירת המספר השלישי.

לגבי האותיות: $26\cdot26\cdot26=26^3$.

בסך הכל נקבל: $10^3\cdot26^3$.

תמורות: Permutations

נתחיל בדוגמא: כאשר אנו מסדרים 5 אובייקטים ב-5 מקומות פנויים, ברור כי עבור הראשון יש 5 אופציות לבחור – כיוון שהרי הוא בוחר ראשון וכל המקומות פנויים עבורו.

עבור השני יש רק 4 מקומות פנויים לבחירה.

עבור השלישי יש רק 3 מקומות פנויים לבחירה.

עבור השני יש רק 2 מקומות פנויים לבחירה.

כל בחירה כזו היא ניסוי בפני עצמה, ולפי הכלל יסודי, אם נרצה שכל הניסויים יתקיימו נקבל: $5\cdot4\cdot3\cdot2\cdot1=5!$.

ולכן הכלל: מספר הדרכים לסדר n אובייקטים הוא $n!$.

דוגמא: כמה דרכים ניתן לסדר את האותיות במילה treat?

פתרון: בגלל שיש 5 אותיות אז $5!$, אבל אם נשים לב נוכל לראות כי האות t מופיעה פעמיים, מאחר ואין הבדל בין ה- t ראשונה ל- t השנייה, נצטרך להוריד את האפשרויות המיותרות, נעשה זאת ע"י חילוק ב$2!$. לכן בסה"כ נקבל: $\frac{5!}{2!}$.

דוגמא נוספת: כמה דרכים ניתן לסדר את האותיות במילה pepper?

פתרון: יש לנו 6 אותיות, שלוש פעמים מופיעה האות p, ופעמיים האות e. לכן: $\frac{6!}{2!\cdot3!}$

ביתר כלליות:

אם יש n איברים כך ש $n_1+n_2+n_3\ldots n_r=n$ ומתקיים כי

מסוג 1 יש n_1 איברים,

מסוג 2 יש n_2 איברים,

מסוג 3 יש n_3 איברים,

$\vdots$

מסוג r יש n_r איברים.

אזי, מספר הדרכים לסדר את n האיברים הינו: $\frac{n!}{n_1!\cdot n_2!\cdot n_3!\cdot\ldots\cdot n_r!}$

צירופים

נתחיל בדוגמא: נתון כי ישנם חמישה אובייקטים וברצוננו לבחור מתוכם שניים, בכמה צרופים ניתן לעשות זאת?

בתחילה יש לנו 5 אופציות לבחירה, לאחר שעשינו את בחירתנו, נותרו לנו 4 אופציות לבחירה, ז"א $5\cdot 4$, את הבחירה הספציפית שעשינו, יכולנו לבחור הפוך ז"א לבחור את האובייקט שבחרנו שני כראשון ואת האובייקט שבחרנו ראשון כשני. כמובן ששורה תחתונה, מבחינתנו, זה לא משנה כי האובייקטים שיש בידינו הם בדיוק אותם אובייקטים פשוט נבחרו בסדר שונה, אך מבחינה הסתברותית, צריך לקחת זאת בחשבון.

דוגמא: אם יש חמישה כדורים ממוספרים מ1-5 , בתחילה בחרתי את מספר 3 ואח"כ את מספר 1. אפשרות שנייה שבחרתי קודם את כדור הממוספר 1 ורק אח"כ את הכדור הממוספר 3.

בגלל שיש לנו שתי אופציות לבחירת שני הכדורים (הממוספרים 3 ו- 1) אזי מספר הדרכים לבחור שני כדורים מתוך חמשת הכדורים היא: $\frac{5\cdot 4}{2!}$ (אנו לא רוצים לספור פעמיים את אותה אפשרות, כי מבחינתינו 1,3 או 3,1 זאת אותה אפשרות)

כנוסחא מורחבת:

- מספר הדרכים בו ניתן לבחור r אובייקטים מתוך n ללא הקפדה על הסדר $\frac{n!}{r!(n-r)!}$. ומסמנים: $\binom{n}{r}=\frac{n!}{r!(n-r)!}$ ושמו: " מקדם בינומי r מתוך n ".
- מספר הדרכים בו ניתן לבחור r אובייקטים מתוך n כאשר ישנה הקפדה על הסדר $\frac{n!}{(n-r)!}$. ומסמנים: $(n)_r=\frac{n!}{(n-r)!}$
- נציין כי n הינו שלם ולא שלילי, r שלם ומקיים $0\le r\le n$. נוהגים להגדיר $0!=1$.

הערה: מספר האפשרויות לבחור r אובייקטים מתוך n: $\frac{n!}{r!(n-r)!}$.

דוגמא : מהו מספר האפשרויות לבחור שני אובייקטים משלושת האובייקטים המסומנים במספרים 1-3?

פתרון:

א. ללא הקפדה על הסדר: $3= \frac{3!}{2!(3-2)!}=\binom{3}{2}$ ואכן האפשרויות הן:

1,2. 1,3. 2,3.

ב. עם הקפדה על הסדר: $6= \frac{3!}{(3-2)!}=(3)_2$ ואכן האפשרויות הן:

1,2. 2,1. 3,1. 1,3. 2,3. 3,2.

דוגמא: בארגון מסוים יש 23 סטודנטים משנה א', 12 משנה ב' ו21 משנה ג'.

א. כמה ועדות ניתן להרכיב כך שיהיה נציג אחד מכל שנה?

ב. רוצים שהועדה תכיל שני סטודנטים משנה ב' ושלושה סטודנטים משנה א' או ג' (לא משנה)

פתרון: א. נרצה לבחור סטודנט משנה א' **וגם** סטודנט משנה ב' **וגם** סטודנט משנה ג. בגלל שאנו רוצים את כל זה בו זמנית – "**וגם**", נכפיל בין כל האפשרויות, ובסה"כ נקבל:

$$\binom{23}{1}\binom{12}{1}\binom{52}{1}$$

ב. $\binom{75}{3}\binom{12}{2}=\binom{23+52}{3}\binom{12}{2}$, צירפנו את 52 ל- 23 כיוון שלא משנה לנו איך נבחרים ומאיזה שנה א' **או** ג', בגלל ה- "**או**" חיברנו.

דוגמא:

אדם זוכר כי במספר הטלפון של חברו מופיעות הספרות:

3 – שלוש פעמים

4 – פעמיים

5 – פעמיים

כמה מספרים יוכל לקבל?

פתרון, בעצם נוכל להסתכל על השאלה כסידור של אובייקטים שיש בהם סוגים שונים, ולכן נשתמש בנוסחא: $\frac{n!}{n_1!\cdot n_2!\cdot n_3!\cdot\ldots\cdot n_r!}$ ונקבל $\frac{7!}{2!\cdot 2!\cdot 3!}$. אפשר להסתכל על זה באופן הבא: בהתחלה יש לנו 7 מקומות פנויים מתוכם נבחר 3 עבור המספר 3, אח"כ נשארו לנו 4 מקומות פנויים (כי כבר 3 תפסנו ע"י המספר 3) ונוכל לסדר את המספר 4 המופיע פעמיים בארבעת המקומות הפנויים,בסופו של דבר נשארו 2 מקומות פנויים עבור המספר 5 שגם מופיע פעמיים ובסה"כ: $\binom{2}{2}\binom{4}{2}\binom{7}{3}$. כמובן שהתשובות זהות עבור שתי הדרכים.

דוגמא: בכמה אופנים ניתן לסדר את המילה סטטיסטיקה?

פתרון:

ס – 2

ט – 3

י – 2

ק – 1

ה – 1

שוב נשתמש בנוסחא: $\frac{n!}{n_1!\cdot n_2!\cdot n_3!\cdot\ldots\cdot n_r!}$ ונקבל: $\frac{9!}{2!\cdot 2!\cdot 3!\cdot 1!\cdot 1!}$ או לפי הדרך השנייה נקבל: $\binom{1}{1}\binom{2}{1}\binom{4}{2}\binom{7}{3}\binom{9}{2}$

דוגמא: בכד יש 11 כדורים, 4 לבנים ו7 אדומים, מוציאים 5 כדורים. בכמה אופנים ניתן לעשות זאת?

א. ללא כל הגבלה

ב. שנים לבנים ושלושה אדומים

ג. לכל היותר 2 לבנים

ד. לפחות 2 לבנים

פתרון:

א. אם זה ללא כל הבדלה אז זה פשוט לבחור 5 מ- 11 $\binom{11}{5}$.

ב. נרצה לבחור 2 מ4 שהם סך כל הלבנים ו3 אדומים מ7 שהם סך כל האדומים $\binom{4}{2}\binom{7}{3}$

ג. במקרה הזה ישנן כמה אופציות, 0 לבנים **או** 1 לבן **או** 2 לבנים. בגלל ה "או" אנחנו נחבר בין האופציות:

0 לבנים – את כל ה-5 כדורים אנו בוחרים אדומים (יש 7 אדומים): $\binom{7}{5}$

1 לבן – יש 4 לבנים ומתוכם נבחר 1 וגם נבחר 4 אדומים מתוך ה7 שקיימים: $\binom{4}{1}\binom{7}{4}$

2 לבנים: $\binom{4}{2}\binom{7}{3}$

בסך הכל: $\binom{4}{2}\binom{7}{3}+\binom{4}{1}\binom{7}{4}+\binom{7}{5}$

ד. 2 לבנים **או** 3 לבנים **או** 4 לבנים:

2 לבנים – $\binom{4}{2}\binom{7}{3}$

3 לבנים – $\binom{4}{3}\binom{7}{2}$

4 לבנים – $\binom{4}{4}\binom{7}{1}$

בסך הכל: $\binom{4}{4}\binom{7}{1}+\binom{4}{3}\binom{7}{2}+\binom{4}{2}\binom{7}{3}$

דוגמא: בכמה אפשרויות ניתן לסדר 8 אנשים בשורה כך ששניים ספציפיים רוצים לשבת יחד.

פתרון: נצמצם את השניים שרוצים לשבת יחד לאחד ואז בעצם נסדר 7 אנשים בלבד ובין הצמד ניתן להחליף אז נכפיל ב $2!$. בסה"כ נקבל: $7!\cdot 2!$

בכלליות: מספר האפשרויות לסידור n אובייקטים בשורה כך ש r יושבים יחדיו: $r!(n-r+1)!$.

בכד שבו n כדורים ממוספרים מ1 עד n בוחרים k כדורים בתנאים הבאים:

	עם החזרה	בלי החזרה
עם חשיבות לסדר	n^r	$n_r=\dfrac{n!}{(n-r)!}$
ללא חשיבות לסדר	$\dbinom{n+r-1}{r}$	$\dbinom{n}{r}=\dfrac{n!}{r!(n-r)!}$

שאלה ממבחן: בוחרים באופן אקראי ועד בן 3 מתוך 6 עורכי דין ו5 מהנדסים. מה ההסתברות שהועד יהיה מורכב

1. משני עו"ד ומהנדס $\dfrac{\binom{6}{2}\binom{5}{1}}{\underbrace{\binom{11}{3}}_{\Omega}}$

2. מעו"ד אחד לפחות, אפשר להגיד בעצם שזה כל האפשרויות האפשריות שהן

ללא עו"ד: $1-\dfrac{\binom{5}{3}\binom{6}{0}}{\binom{11}{3}}$

3. רק עו"ד $\dfrac{\binom{5}{0}\binom{6}{3}}{\binom{11}{3}}$

אקסיומות של הסתברות:

קצת תורת הקבוצות – תזכורת: נתון כי $A_1,\dots A_n$ מאורעות, אזי נסמן:

$A_1\cap A_2$ - חיתוך, $A_1\cup A_2$ - איחוד, A_1^C - משלים ו- $\varnothing$ - קבוצה ריקה.

חילופיות, אסוציאטיביות, דיסטרבוטיביות (חוק הפילוג) וחוקי דמורגן

$1)\, p(A^C)=1-p(A)\ ,\quad p(A)=1-p(A^C)$

$2)\, p(A\cup A^C)=p(\Omega)=1$

$3)\, p(A\cup B)=p(A)+p(B)-p(A\cap B)$

$4)\, p(A\cap B)+p(A\cap B^C)=p(A)$

$5)\, p(A^C\cap B^C)=p(A\cup B)^C$

$A_1,\dots A_n$ מאורעות זרים $\Leftrightarrow$ $A_i\cap A_j=\varnothing \quad \forall\ i\neq j$

מידת ההסתברות: פונקציה שמתאימה לכל מאורע A מספר ממשי $p(A)$ כך שמתקיימים:

א. $0 \le p\left(A\right) \le 1$ לכל A

ב. $p\left(\Omega\right) = 1$

ג. לכל סדרה של מאורעות זרים זה לזה מתקיים $p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p\left(A_i\right)$

הסתברות מותנית: הסתברות על סמך מידע מקדים:

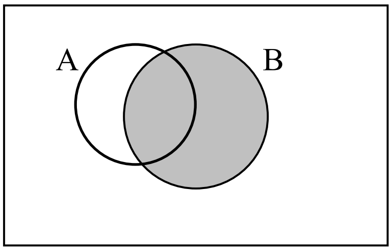
מבחינה רעיונית, אנו שואלים עבור הסתברות של מאורע מסוים אך לא מכלל מרחב המדגם אלא מתוך מאורע אחר. לשון אחר: מהי ההסתברות של המאורע A שנמצא בתוך החלק של המאורע B כאשר נתון כי המאורע B אינו קבוצה ריקה, ברור כי ההסתברות של החלק של A שלא נמצא בB הינה -0. הסתברות מותנית זו שתיארנו מסומנת $p(A \mid B)$ ובמילים "מצא את ההסתברות של A בהנתן B"

באופן פונקציונלי, נוכל להניח כי מרחב האפשרויות הצטמצם מ Ω ל- B. כעת, נוכל להגדיר פונקצית הסתברות חדשה על Ω:

$$P\left(x \mid B\right) = \begin{cases} \dfrac{P\left(x\right)}{P\left(B\right)} & x \in B \\ 0 & x \notin B \end{cases}$$

כעת, נרצה לדעת הסתברות זו (של x) מתוך הנתון שהוא ההסתברות של B. לשון אחר, אנו שואלים מה מהווה החלק של A שנמצא ב – B מתוך כל הקבוצה B. מתוך כך נקבל את הנוסחא הבאה:

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



ננסה להסביר אינטואיציה לנוסחא משני כיוונים:

ברור כי $p\left(B\right) = p\left(A \cap B\right) + p\left(A^c \cap B\right)$

לפי נוסחת הסתברות מותנית מתקיים: $P(B \mid A)P(A) = P(A \cap B)$ כעת, נוכל במקום A נציב A^c ונקבל $P(B \mid A^c)P(A^c) = P(A^c \cap B)$ ולכן:

$$p(B)=p(A\cap B)+p(A^c\cap B)\xrightarrow[P(B|A)P(A)=P(A\cap B)]{P(B|A^c)P(A^c)=P(A^c\cap B)}$$

$$p(B)=P(B\mid A)P(A)+P(B\mid A^c)P(A^c)$$

בנוסף, ברור כי $P(B\cap A)=P(A\cap B)$ ולכן נובע באופן ישיר כי $P(B\mid A)P(A)=P(A\mid B)P(B)$ ע"י העברת אגפים נקבל:

$P(B\mid A)=\dfrac{P(A\mid B)P(B)}{P(A)}$, לפי הנוסחא הקודמת:

$$P(B\mid A)=\frac{P(A\mid B)P(B)}{P(A)}\xrightarrow{p(A)=P(A|B)P(B)+P(A|B^c)P(B^c)}$$

$$P(B\mid A)=\frac{P(A\mid B)P(B)}{P(A\mid B)P(B)+P(A\mid B^c)P(B^c)}\xrightarrow{P(B^c)=1-P(B)}$$

$$P(B\mid A)=\frac{P(A\mid B)P(B)}{P(A\mid B)P(B)+P(A\mid B^c)\left(1-P(B)\right)}$$

וזהו **חוק בייס**.

הגדרה: חלוקה של מרחב מדגם הוא אוסף של מאורעות זרים זה לזה כך שהאיחוד שלהם הוא כל המרחב המדגם:

יהיו $F_1,F_2,F_3\ldots$ מאורעות זרים, $\forall\ i\neq j\quad F_i\cap F_j=\varnothing$ ומתקיים $\bigcup_i F_i=\Omega$. נתון מאורע מסוים A אזי בהכרח מתקיים: $A=(F_1\cap A)\cup(F_2\cap A)\cup(F_3\cap A)\ldots$.

חוק ההסתברות השלמה או חוק ההסתברות הכוללת:

אם Ω ניתן לפירוק של איחוד זר של F_i-ים (סופי או אינסופי) אזי

$$p(A)=\sum_i p(A\mid F_i)p(F_i)$$

$$p(F_i\mid A)=\frac{p(F_i\cap A)}{p(A)}=\frac{p(A\mid F_i)\,p(F_i)}{p(A)}$$

$$p(F_i\mid A)=\frac{p(A\mid F_i)\,p(F_i)}{p(A)}\xrightarrow{p(A)=\sum_i p(A|F_i)p(F_i)}=\frac{p(A\mid F_i)\,p(F_i)}{\sum_i p(A\mid F_i)p(F_i)}$$

דוגמא: חברת ביטוח מחלקת את עולם הנהגים ל-2 קבוצות – מסוכנים (30%) ובטיחותיים (70%), נהג מסוכן מעורב בתאונה בהסתברות 0.4 ונהג בטיחותי מעורב בהסתברות 0.2.

א. מה ההסתברות שאם בוחרים נהג באקראי הוא יעבור תאונה

ב. נהג מעורב בתאונה, מה ההסתברות שהוא מקבוצת המסוכנים

פתרון: נגדיר את המאורעות הבאים:

A – נהג מעורב בתאונה, A^c - נהג שלא מעורב בתאונה

B – נהג מסוכן, B^c - נהג בטיחותי.

לפי הנתונים: $p(B)=0.3\ ,\ \ p\left(B^c\right)=0.7\ \ ,\ \ p(A\,|\,B)=0.4\ \ \ ,\ \ p\left(A\,|\,B^c\right)=0.2$

א. $p(A)=p(A\,|\,B)\,p(B)+p\left(A\,|\,B^c\right)p\left(B^c\right)=0.4\cdot 0.3+0.2\cdot 0.7=0.26$

ב. $P(B\,|\,A)=\dfrac{P(A\,|\,B)P(B)}{P(A)}=\dfrac{0.4\cdot 0.3}{0.26}=\dfrac{6}{13}$

נוסחת הבינום של ניוטון עבור חזקה שלמה:

אם n מספר שלם. אזי עבור כל x ו-y מתקיים $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}$, כאשר

$\binom{n}{k}=\dfrac{n!}{k!(n-k)!}$ הם מקדמי הבינום (מבחינה קומבינטורית, מספר האפשרויות לבחור ללא חזרות וללא חשיבות לסדר).

ארבעת המקרים הראשונים של הנוסחא הם:

1. $(x+y)^1=x+y$
2. $(x+y)^2=x^2+2xy+y^2$
3. $(x+y)^3=x^3+3x^2y+3xy^2+y^3$
4. $(x+y)^4=x^4+4x^3y+6x^2y^2+3xy^3+y^4$

מבחינה קומבינטורית נוכל לראות זאת באופן הבא:

ישנן מכפלות של $(x+y)^3$ שלוש פעמים, ז"א $(x+y)(x+y)(x+y)$ והתוצאה היא בעצם סכום של כל האפשרויות שבהן נבחר איבר מכל מכפלה (מכל סוגריים).

למשל האפשרות הראשונה היא לבחור את x מכל סוגריים ולכן נקבל x^3, האפשרות השנייה היא לבחור את x מהסוגריים הראשונים, את x מהסוגריים השניים ואת y מהסוגריים השלישיים. את בחירה זו נוכל לעשות במספר ווריאציות למשל לבחור את y מהסוגריים הראשונים, את x מהסוגריים השניים ואת x מהסוגריים השלישיים

ובעצם יש לנו שלוש ווריאציות כאלו. ולכן בנוסחא יש את הגורם x^2y המוכפל ב-3

הוכחה: באינדוקציה.

עבור $n=1$: $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}\xrightarrow{n=1}\sum_{k=0}^{n=1}\binom{1}{k}x^k y^{1-k}=x+y$

הנחת האינדוקציה עבור $n=a$ $\quad(x+y)^a=\sum_{k=0}^{a}\binom{a}{k}x^k y^{a-k}$ ונוכיח עבור $n=a+1$:

$\sum_{k=0}^{a+1}\binom{a+1}{k}x^k y^{a-k+1}=(x+y)^{a+1}$. ראשית נפתח את אגף שמאל: $(x+y)^a(x+y)$ ולפי

הנחת האינדוקציה: $(x+y)^a(x+y)\to(x+y)\sum_{k=0}^{a}\binom{a}{k}x^k y^{a-k}$

נפתח סוגריים ונקבל: $x\sum_{k=0}^{a}\binom{a}{k}x^k y^{a-k}+y\sum_{k=0}^{a}\binom{a}{k}x^k y^{a-k}$ ונכניס את הקבועים x וy

בפנים: $\sum_{k=0}^{a}\binom{a}{k}x^{k+1}y^{a-k}+\sum_{k=0}^{a}\binom{a}{k}x^k y^{a-k+1}$. נשנה את הסכום הראשון באופן הבא:

$\sum_{k=1}^{a+1}\binom{a}{k-1}x^k y^{a-k+1}+\sum_{k=0}^{a}\binom{a}{k}x^k y^{a-k+1}$ נפתח את האיבר האחרון (ז"א האיבר ה- $a+1$)

בסכום הראשון ואת האיבר הראשון בסכון השני (ז"א עבור $k=0$):

$$\underbrace{\binom{a}{a}x^{a+1}y^{a-(a+1)+1}}_{last}+\sum_{k=1}^{a}\binom{a}{k-1}x^k y^{a-k+1}+\underbrace{\binom{a}{0}x^0 y^{a-0+1}}_{first}+\sum_{k=1}^{a}\binom{a}{k}x^k y^{a-k+1}$$ כעת קבלנו

בסה"כ: $x^{a+1}+y^{a+1}+\sum_{k=1}^{a}\binom{a}{k-1}x^k y^{a-k+1}+\sum_{k=1}^{a}\binom{a}{k}x^k y^{a-k+1}$ ועכשיו נוכל לאחד את

הסכומים: $x^{a+1}+y^{a+1}+\sum_{k=1}^{a}\left[\binom{a}{k-1}+\binom{a}{k}\right]x^k y^{a-k+1}$.

ידוע כי $\binom{a}{k-1}+\binom{a}{k}=\binom{a+1}{k}$

נוכיח:

$$\binom{a}{k-1}+\binom{a}{k}=\binom{a+1}{k}\xrightarrow{\binom{n}{k}=\frac{n!}{k!(n-k)!}}\frac{a!}{(k-1)!(a-k+1)!}+\frac{a!}{k!(a-k)!}=\frac{(a+1)!}{k!(a+1-k)!}$$

$$\frac{ka!}{(k)!(a-k)!(a-k+1)}+\frac{a!}{k!(a-k)!}=\frac{(a+1)(a)!}{k!(a-k)!(a-k+1)}$$

$$\frac{k}{(a-k+1)}+1=\frac{(a+1)}{(a-k+1)}\to\frac{k+a-k+1}{(a-k+1)}=\frac{(a+1)}{(a-k+1)}$$

$$\frac{(a+1)}{(a-k+1)}=\frac{(a+1)}{(a-k+1)}$$

נחזור להוכחה, $x^{a+1}+y^{a+1}+\sum_{k=1}^{a}\binom{a+1}{k}x^k y^{a-k+1}$ וכעת, נכניס את האיברים לתוך הסכום,

נשים לב כי אלו בעצם האיברים האחרונים של הסכום כך שעכשיו הסכום ירוץ עד $a+1$

לכן נקבל שאכן $\left(x+y\right)^{a+1}=\sum_{k=0}^{a+1}\binom{a+1}{k}x^k y^{a-k+1}$. משל.

דיאגראמת עץ

דיאגרמת עץ הינו כלי עזר שבאמצעותו ניתן לפתור את התרגילים שפתרנו עד עכשיו בעזרת כפל וחיבור הסתברויות. הדיאגרמה דומה לעץ עם ענפים ולכן נקראת דיאגרמת עץ. הדיאגרמה מפשטת את השאלה ומאפשרת לראות את כל המקרים האפשריים.

דוגמאות:

שאלה ממבחן: הלימודים לתואר ראשון מתחלקים בשיעורים הבאים: במדעי הרוח לומדים 10% מהסטודנטים, במדעי החברה לומדים 35%, במדעי הטבע לומדים 25%, ביהדות 20% ובמשפטים 10%. לאחר סיום התואר הראשון כל סטודנט מחליט אם להמשיך לתואר שני או לחפש עבודה. בכל פקולטה ממשיכים לתואר שני בשיעורים הבאים: מדעי הרוח 10%, מדעי החברה 50%, מדעי הטבע 50%, יהדות 5% ומשפטים 5%. (22 נקודות)

א. מה אחוז הסטודנטים הממשיכים לתואר שני? (6 נקודות)

ב. פגשת סטודנט לתואר שני, מה הסיכוי שהוא לומד מדעי הטבע? (8 נקודות)

ג. המכללה למנהל החליטה כי סטודנט שממשיך לתואר שני יוגדר מבחינתה כהצלחה, מצא את התוחלת ואת השונות של מספר ההצלחות עבור n סטודנטים מהמכללה. (8 נקודות)

פתרון:

א. נמצא את ההסתברות להיות סטודנט של תואר שני:

$$p(sheni)=0.1\cdot0.05+0.25\cdot0.5+0.1\cdot0.1+0.2\cdot0.05+0.35\cdot0.5=0.325$$

ב. בהנתן שסטודנט לומד תואר שני מה ההסתברות שלומד בפקולטה למדעי הטבע:

$$p(teva\mid sheni)=\frac{p(teva\cap sheni)}{p(sheni)}=\frac{0.25\cdot0.5}{0.325}=0.384$$

2. (25 נקודות) בסקר תעסוקה לבוגרי אוניברסיטה התברר כי מבין בוגרי חוג מחשבים 80% השיגו עבודה במשרה מלאה, 15% השיגו עבודה במשרה חלקית והיתר לא השיגו עבודה כלל. מבין שאר בוגרי האוניברסיטה, 60% השיגו עבודה במשרה מלאה, 25% השיגו עבודה במשרה חלקית והיתר לא השיגו עבודה כלל. ידוע כי אחוז הבוגרים בחוג מחשבים, מבין כלל הבוגרים, הוא 40%. (25 נקודות)

א. מה הסיכוי של בוגר אוניברסיטה להשיג עבודה במשרה מלאה? (4 נקודות)

ב. מה הסיכוי של בוגר אוניברסיטה להשיג עבודה במשרה בהיקף כלשהו? (4 נקודות)

ג. בוגר אוניברסיטה השיג עבודה במשרה מלאה, מה הסיכוי שהוא בוגר חוג מחשבים? (4 נקודות)

ד. ללא קשר לנתונים הקודמים, ידוע כי סטודנט שלמד מדעי המחשב מתקבל לעבודה לתקופת ניסיון של מספר חודשים. מספר חודשים זה מוגרל ע"י זריקת קובייה. אם התקבל מספר גדול או שווה ל -5 אזי תקופת הניסיון אורכת 5 חודשים, אם התקבל מספר קטן או

שווה ל-2 אזי תקופת הניסיון אורכת 3 חודשים. אם יתקבל מספר אחר תקופת הניסיון אורכת 4 חודשים.

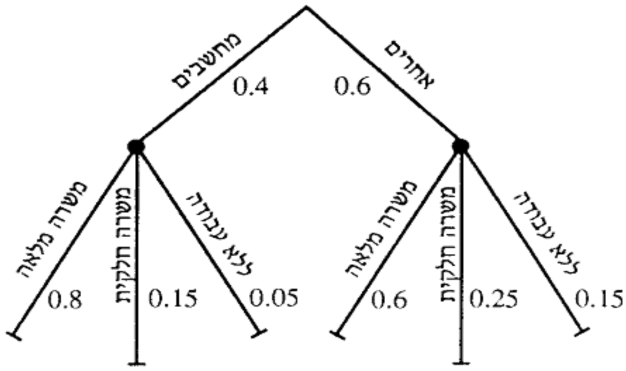
בנה טבלת התפלגות המתארת את הסיפור המוצג (5 נקודות)

ה. (בהמשך לסעיף ד׳) מה ההסתברות לתקופת ניסיון של 4 חודשים ומעלה? (3 נקודות)

ו. (בהמשך לסעיף ד׳) הסטודנטים נכנסים לחדר ההגרלות בממוצע של 10 לשעה. מה הסיכוי לפחות משני סטודנטים בחצי שעה? (5 נקודות)

<u>פתרון שאלה 2</u>

<u>דיאגרמת העץ המתאימה לנתוני השאלה :</u>



נגדיר : A – בוגר חוג מחשבים. B – השיג משרה מלאה. C – השיג משרה חלקית.

א. בעזרת נוסחת ההסתברות השלמה :

$$P(B)=P(A)\cdot P(B\mid A)+P(\bar{A})\cdot P(B\mid \bar{A})=0.4\cdot 0.8+0.6\cdot 0.6=0.68$$

ב. צריך למצוא את $P(B)+P(C)$. את $P(B)$ חישבנו בסעיף א׳. נחשב את $P(C)$:

$$P(C)=P(A)\cdot P(C\mid A)+P(\bar{A})\cdot P(C\mid \bar{A})=0.4\cdot 0.15+0.6\cdot 0.25=0.21$$

לכן : $P(B)+P(C)=0.68+0.21=0.89$

נשתמש בנוסחת בייס $P(A\mid B)=\dfrac{P(A)\cdot P(B\mid A)}{P(B)}$ כאשר נתון :

$P(A)=0.4$, $P(B\mid A)=0.8$, ובסעיף א׳ מצאנו $P(B)=0.68$. נציב בנוסחה :

$$P(A\mid B)=\frac{0.4\cdot 0.8}{0.68}=0.47$$

תלות ואי תלות

הגדרה : אם A ו- B בלתי תלויים (ב״ת) אזי מתקיים : $p(A\mid B)=p(A)$.

הסבר : האינפורמציה שיש ב -B אינה תורמת ומשפיעה על ההסתברות של A. לכן ההסתברות המותנית של A בהנתן B שווה להסתברות של A.

לשון אחר, אם ההסתברות של A בהינתן B שווה רק להסתברות של A , זה אומר ש- B לא נותן שום אינפורמציה חדשה על A . מצב כזה יכול להתרחש רק כאשר A ו- B ב״ת – כיוון שאין ב-B חידוש עבור A – אין זה אומר שהחיתוך ביניהם ריק, אז העובדה שכתוב $p(A\mid B)$ לא משפיעה על תוצאת ההסתברות ולכן ניתן לרשום $p(A)$.

דוגמא להמחשת הרעיון: נתון כד עם 5 כדורים, 3 לבנים ו2 שחורים. גיל מוציא 2 כדורים אחד אחר השני, אך מחזיר את הכדור לאחר הוצאתו, אזי ברור כי ההסתברות של ההוצאה השנייה אינה תלויה בתוצאת ההוצאה הראשונה. לעומת זאת, אם גיל אינו מחזיר את הכדור אזי ההסתברות של ההוצאה השנייה תלויה בתוצאת ההוצאה הראשונה.

אם A אינו תלוי ב – B , אזי השוויון הבא מתקיים: $p(A)\cdot p(B)=p(A\cap B)$. כעת, נסביר איך הגענו למסקנה זו: נתחיל מהנוסחה הבסיסית עבור הסתברות מותנית, ובמקום $p(A\mid B)$ נציב $p(A)$ כיוון ש- $p(A\mid B)=p(A)$ ומכאן נקבל כי מתקיים:

$$p(A|B)=\frac{p(A\cap B)}{p(B)}\xrightarrow{p(A|B)=p(A)}p(A)=\frac{p(A\cap B)}{p(B)}\rightarrow p(A)\cdot p(B)=p(A\cap B)$$

סיכום: - אם השוויון **מתקיים** אזי A ו- B **ב"ת**

- אם השוויון **לא מתקיים** אזי A ו- B **תלויים**

במקרה הכללי: המאורעות $A_1, A_2, \dots A_n$ ב"ת אם לכל תת קבוצה של אינדקסים $i_1, i_2, \dots i_m \in \{1, 2, \dots , n\}$ שכולם שונים, מתקיים:

$$P\left(\bigcap_{i=i_1,i_2,\dots i_m} A_i\right)=\prod_{i=i_1,i_2,\dots i_m} P(A_i)$$

דוגמא: הסיכוי לעבור את הבחינה במתמטיקה-0.9 , הסיכוי לעבור את הבחינה בכלכלה-0.6. הסיכוי לעבור את שתי הבחינות- 0.55. א. האם ישנה תלות בין מתמטיקה לכלכלה? ב. אם סטודנט עבר בכלכלה, מה הסיכוי שעבר במתמטיקה?

פתרון: א. נגדיר את המאורע A להיות הסיכוי לעבור את הבחינה במתמטיקה, ונגדיר את מאורע B להיות הסיכוי לעבור את הבחינה בכלכלה. בכדי לבדוק האם ישנה תלות בין המקצועות, נבדוק האם מתקיים השוויון הבא: $p(A)\,p(B)=p(A\cap B)$. אם השוויון מתקיים, אזי לא קיימת תלות בין A ל- B . אם השוויון לא מתקיים, אזי יש תלות בין A ל- B .

$$p(A)\,p(B)=p(A\cap B)$$
$$0.9\cdot 0.6\overset{?}{=}0.55\quad\rightarrow\quad 0.54\neq 0.55$$

נוכל לראות כי לא קיים שוויון בין המקצועות, ולכן המסקנה הינה שישנה תלות בין מתמטיקה לכלכלה.

ב. נגדיר את המאורע A להיות: סטודנט שעבר במתמטיקה, ונגדיר את מאורע B להיות סטודנט שעבר בכלכלה. לפי סעיף ב', נרצה למצוא את $p(A\mid B)$. נשתמש בנוסחה:

$$p(A\mid B)=\frac{p(A\cap B)}{p(B)}\quad\rightarrow\quad p(A\mid B)=\frac{0.55}{0.6}=0.9$$

דוגמא: הסתברות לגל חום הינה 0.9- וההסתברות לגל גובה-0.4 ההסתברות לגל חום או לגל גובה הינה 0.96. א. האם ישנה תלות בין גל חום לגל גובה? ב. היום יש גל חום מה ההסתברות לגל גובה?

פתרון: נגדיר את מאורע A להיות "ההסתברות לגל חום" ונגדיר את מאורע B להיות "ההסתברות לגל גובה". נסדר את הנתונים:

$$p(A)=0.9 \xrightarrow{p(A^C)=1-p(A)} p(A^C)=0.1$$
$$p(B)=0.4 \xrightarrow{p(B^C)=1-p(B)} p(B^C)=0.6$$
$$p(A\cup B)=0.96$$

על מנת לבדוק תלות נצטרך לבדוק האם מתקיים השוויון: $p(A)p(B)\overset{?}{=}p(A\cap B)$

כדי למצוא את החיתוך - $p(A\cap B)$ נפתח את $p(A\cup B)=0.96$:

$$p(A\cup B)=0.96 \xrightarrow{p(A\cup B)=p(A)+p(B)-p(A\cap B)} p(A)+p(B)-p(A\cap B)=0.96$$

$$0.9+0.4-p(A\cap B)=0.96 \quad \rightarrow \quad p(A\cap B)=0.34$$

נציב: $0.9\cdot 0.4 \neq 0.34 \xleftarrow{hatsava} p(A)p(B)\overset{?}{=}p(A\cap B)$ מכיוון שלא מתקיים השוויון אזי נסיק כי ישנה תלות בין A ל- B

נרצה למצוא את ההסתברות לגל גובה כאשר נתון כי כבר ישנו גל חום, לכן נשתמש בנוסחה להסתברות מותנית כיוון שהתנאי שלנו הינו שקיים היום גל חום, זאת לפי הנתון. $p(A|B)=\dfrac{p(A\cap B)}{p(B)} \xrightarrow{hatsava} \dfrac{0.34}{0.9}0.377$

דוגמא: הסתברות לגל חום הינה 0.9, כמו כן ההסתברות לגל גובה הינה 0.3
האם יש מצב של גל חום **שגם** גבוה?

דוגמא: 60% מהבוגרים מוצאים עבודה מיד עם סיום הלימודים. 35% מהבוגרים ממשיכים לתואר שני. 15% אינם עובדים ואינם לומדים בתום לימודיהם. האם מציאת עבודה והמשך לימודים הם מאורעות תלויים?

פתרון: בשאלה זו עדיף לעבוד עם המאורעות המשלימים כיוון שהנתונים לא מאפשרים אחרת... נגדיר את המאורע A להיות בוגר שמצא עבודה, את המאורע B להיות בוגר שהמשיך לתואר שני ולכן נקבל כי:

$$p(A)=0.6 \xrightarrow{p(A)=1-p(A^C)} p(A^C)=0.4$$
$$p(B)=0.35 \xrightarrow{p(B)=1-p(B^C)} p(B^C)=0.65$$

נתון כי 15% אינם עובדים ואינם לומדים בתום לימודיהם: לפי הצורה בה הגדרנו את המאורעות נקבל כי $p(A^C\cap B^C)=0.15$ כעת נבדוק תלות לפי ההגדרה ונקבל כי השוויון אינו מתקיים אזי יש תלות: $p(A^c)\cdot p(B^c)=p(A^c\cap B^c)\rightarrow 0.6\cdot 0.65\overset{?}{=}0.15$

משתנה מיקרי בדיד

תוצאה של **ניסוי מקרי**: כדי להגדיר זאת ניתן דוגמאות, תוצאה של ניסוי מקרי יכולות להיות:

- תנוחה מסוימת של קובייה,
- תנוחה מסוימת של מטבע,
- סוג דם של חולה (תוצאה זו אינה מספרית אך, ניתן לתת תוצאה מספרית למשל, נגדיר שסוג דם A הינו המספר 3 וסוג דם B הינו המספר 4 וכו').

אנו עורכים ניסוי כאשר התוצאות שלו מקריות, נסמן ב-X את המספר המצורף לתוצאה, או את התוצאה עצמה- אם היא מספרית.

הערך שקיבל- X יהיה תלוי בתוצאת הניסוי- שהוא **מקרי,** וכתוצאה מזה אנו מכנים את **X** – **משתנה מקרי, מ.מ**, למשל, אם זורקים קובייה ומקבלים את המספר 2 **אזי X- הינו התנוחה של הקובייה ו-2 זהו מספר <u>קבוע ולא מקרי</u>, אבל העובדה שהמספר -2 נבחר היא <u>מקרית</u>.**

דוגמא: מטילים קובייה פעמיים (ז"א מרחב המדגם הינו 36) נגדיר מ.מ – X : לכל זוג תוצאות , מתאימים את סכומן, ז"א X יכול לקבל: $\{2,3,\ldots,12\}$, בנו טבלת התפלגות של X.

פתרון: נבנה טבלת התפלגות של- X. כך שהטור x_i הינו הערכים ש-X יכול לקבל, הטור - $p(x_i)$ הינה ההסתברות שערכים אלו יתקבלו באופן מקרי.

i	x_i	$p(X = x_i)$
1	$2 \to (1,1)$	$\frac{1}{36}$
2	$3 \to (2,1), (1,2)$	$\frac{2}{36}$
3	$4 \to (2,2), (3,1), (1,3)$	$\frac{3}{36}$
4	5	$\frac{4}{36}$
5	6	$\frac{5}{36}$
6	7	$\frac{6}{36}$
7	8	$\frac{5}{36}$
8	9	$\frac{4}{36}$
9	10	$\frac{3}{36}$
10	11	$\frac{2}{36}$
11	12	$\frac{1}{36}$

נתון X מ.מ בדיד:

א. ישנה רשימה של כל ערכי X האפשריים

ב. לכל ערך של X מותאמת ההסתברות שמאורע זה יתרחש

ג. $\sum_{i=1}^{n} p\left(x_i\right) = 1$

אוסף של הערכים וההסתברויות נקראת **התפלגות.**

כאשר נתונה ההתפלגות ניתן לחשב הסתברות של כל מאורע.

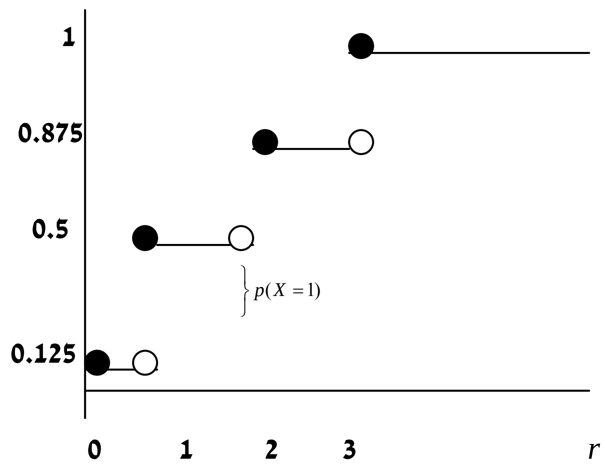
דוגמאות נוספות לבניית טבלת התפלגויות בסוף הספר

פונקצית ההצטברות:

נתון X מ.מ, ונתון a מספר ממשי כלשהו אזי $F(a)=p(X\le a)=\sum_{r\le a} p(X=r)$

דוגמא: נתונה ההתפלגות הבאה:

r	$p(X=r)$	$\sum_{r\le a} p(X=r)$
0	0.125	0.125
1	0.375	0.5
2	0.375	0.875
3	0.125	1



תכונות של פונקצית ההצטברות F

1. אם ישנו r כך שעבורו F אינה רציפה, אזי אותו ערך r שייך ל- X (אי רציפות ממין ראשון – ניתן לראות בציור)
2. $\lim_{a\to r^+} F(a)-\lim_{a\to r^-} F(a)=p(X=r)$

* ע"י שני סעיפים אלו ניתן לשחזר את ההתפלגות

3. $\lim_{a\to+\infty} F(a)=1$ וגם $\lim_{a\to-\infty} F(a)=0$

דוגמא: נתון X מ.מ בדיד עם התפלגות: $p(X=r)=cr$ כאשר $r=1,\dots,n$ ו - c קבוע כלשהו. בטאו את c ע"י n.

פתרון: ידוע כי נתונה התפלגות, ולכן נוכל להשתמש בתכונותיה, למשל שמתקיים

$\sum_{i=1}^{12} p(x_i)=1$ תמיד. ולכן בהתפלגות שלנו: $\sum_{i=1}^{n} p(x_i)=1 \rightarrow \sum_{i=1}^{n} cr=1$ נוכל את c

להוציא מחוץ לסכום כך שנשאר לנו בתוך הסכום רק r וזהו בעצם סכום סידרה חשבונית,

בסה"כ נקבל: $\sum_{i=1}^{n} cr=1 \rightarrow c\sum_{i=1}^{n} r=1 \xrightarrow{s_n=\left[2a_1+(n-1)d\right]\frac{n}{2}} c\frac{n(n+1)}{2}=1$ ולכן

$$c=\frac{2}{n(n+1)}$$

מדדים חשובים בהתפלגות:

שכיח: ערך r יהיה שכיח אם לכל ערך s מתקיים $p(X=s)\le p(X=r)$. לשון אחר: הערך שמופיע בהסתברות הגבוהה ביותר.

מבחינה סטטיסטית, שכיח (Mode): מסומן ב- MO. הערך ששכיחותו היא הגבוהה ביותר בהתפלגות. כאשר מדובר במשתנה כמותי רציף, אזי נמצא את השכיח על פי ערך הצפיפות הגבוה ביותר.

חציון: ערך r יהיה חציון אם $p(X\le r)\ge\frac{1}{2}$ וגם $p(X\ge r)\ge\frac{1}{2}$.

במילים פשוטות, משמאל ל- r יהיה חצי השטח וכך גם מימין.

מבחינה סטטיסטית, אותו ערך שמעליו יש 50% מהתצפיות ומתחתיו יש 50% מהתצפיות. מסומן ב- MD. לשון אחר, הערך שמחצית מהמקרים קטנים או שווים לו ומחצית מהמקרים גדולים או שווים לו. ערך זה מחלק את ההתפלגות לשני חלקים כך ששטחם (תחת פונקציית הצפיפות) שווים.

תוחלת: מסומנת $E(X)$, התוחלת מייצגת את הערך (ערך זה לא בהכרח נמצא בהתפלגות) אליו שואפת התוצאה הממוצעת של ניסוי, אם חוזרים עליו אינסוף פעמים – ממוצע משוקלל על פי ערכי ההתפלגות.

למשל, אם נטיל מטבע 3 פעמים יכולות לצאת מספר תוצאות (כמו בדוגמא הראשונה), אם נטיל את המטבע אינסוף פעמים אזי בממוצע נקבל ש50% מההטלות קבלנו פלי וב 50% מההטלות קבלנו עץ, ז"א שבאינסוף ניסויים נקבל מספרים שקרובים להסתברות התיאוריתית. התוחלת מחושבת: $E(X)=\sum_r r\cdot p(X=r)$

מבחינה סטטיסטית: ממוצע המדגם (Mean): מסומן ב- $\bar{X}$. סכום התצפיות מחולק במספרן - $\bar{X}=\frac{\sum_{i=1}^{n} X_i}{n}$. ניתן להציג את הממוצע גם כך $\bar{X}=\frac{\sum_{i=1}^{n} f(x_i)\cdot x_i}{n}$ נשתמש בנוסחא זו כאשר מדובר משתנה רציף.

שונות: מדד לפיזור הסטטיסטי של משתנה מקרי. הפיזור נמדד סביב התוחלת של המשתנה המקרי הנדון. השורש הריבועי של השונות נקרא "סטיית התקן" של המשתנה.

השונות מסומנת $V(X)$ או ע"י σ^2 ומחושבת באופן הבא: $V(X)=E(X^2)-(E(X))^2$.

סטיית התקן מסומנת σ ומחושבת כשורש של השונות.

כמו שכבר ציינו, סטיית תקן הינה המדד הנפוץ ביותר המורה על פיזור של משתני המדגם הינו סטיית התקן.

הגדרה: סכום הסטיות מהממוצע בריבוע חלקי גודל המדגם.

$$s=\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}{n}}$$

ניתן להציג את הנוסחא גם באופן הבא: $s=\sqrt{\frac{\sum_{i=1}^{n} f(x_i)\cdot x_i^2}{n}-\bar{X}^2}$

שונות: ערך סטיית התקן בריבוע. $s^2=\frac{\sum_{i=1}^{n} f(x_i)\cdot x_i^2}{n}-\bar{X}^2$

תכונות: תוחלת, שונות וטרנספורמציות ליניאריות.

תכונות התוחלת:

1. תוחלת של מספר קבוע הינה הקבוע עצמו: $E(\alpha)=\alpha$, למשל, $E(5)=5$
2. תוחלת של קבוע כפול מ.מ: $E(\alpha\cdot X)=\alpha\cdot E(X)$, למשל, $E(3\cdot x)=3\cdot E(X)$
3. טרנספורמציה ליניארית: $E(\alpha\cdot X+\beta)=E(\alpha\cdot X)+E(\beta)=\alpha\cdot E(X)+\beta$
4. תוחלת של סכום שווה סכום התוחלות: $E(X+Y)=E(X)+E(Y)$
5. אם X ו – Y ב"ת אזי $E(XY)=E(X)E(Y)$
6. הרחבה ל-5: אם X ו – Y ב"ת אזי $E(h(X)k(Y))=E(h(X))E(k(Y))$

תכונות השונות:

1. שונות של קבוע כפול מ.מ: $V(\alpha\cdot X)=\alpha^2\cdot V(X)$, למשל, $V(3\cdot X)=9\cdot V(X)$
2. שונות של מספר קבוע, שווה ל-0: $V(\alpha)=0$, למשל, $V(8)=0$
3. שונות של סכום מ.מ ו- קבוע: $V(X+\beta)=V(X)+V(\beta)=V(X)+0=V(X)$

למשל, $V(X+6)=V(X)+V(6)=V(X)+0=V(X)$

4. $V(\alpha\cdot X+\beta)=V(\alpha\cdot X)+V(\beta)=\alpha^2\cdot V(X)+0=\alpha^2\cdot V(X)$
5. $V(X)=E\left(\left(X-E(X)\right)^2\right)=E(X^2)-(E(X))^2$
6. $V(X+Y)=V(X)+V(Y)+2COV(X,Y)$ כאשר $COV(X,Y)=E(XY)-E(X)E(Y)$

7. אם X ו – Y ב"ת אזי $COV(X,Y)=0$

8. אם X ו – Y ב"ת אזי $V(X+Y)=V(X)+V(Y)$ כיוון ש $COV(X,Y)=0$

טרנספורמציות ליניאריות:

נבצע טרנספורמציה ליניארית כאשר נרצה להפעיל ארבע מפעולות החשבון הבסיסיות (חיבור, חיסור, כפל וחילוק) על מדדי המרכז ומדדי הפיזור.

דוגמאות לטרנספורמציות ליניאריות:

דוגמא: במפעל מסוים ידוע כי ממוצע המשכורות הינו 5000 ₪ מנהל המפעל החליט להעלות את המשכורות ב- 10% ולהפחית מכל משכורת סכום קבוע של 200 ₪ . מהו ממוצע המשכורות החדש?

דוגמא: מרצה החליט לתת פקטור למבחן באופן הבא: כל ציון יוכפל ב- 0.7 ובנוסף, יוסיף לו 13 נקודות. כיצד ישתנה ערך הממוצע?

דוגמא לטרנספורמציה לא ליניארית: מרצה החליט לתת פקטור למבחן באופן הבא: כל ציון נעלה בריבוע. כיצד ישתנה ערך הממוצע?

כיצד ישתנו מדדי המרכז כאשר x הינו ערך התצפית הישנה ו- y הינו הערך החדש כך שמתקיים: $y=bx+a$

הממוצע החדש: $\overline{Y}=b\overline{X}+a$

חציון חדש: $chetsyon\ \ chadash=b\cdot\left(chetsyon\ \ yashan\right)+a$

שכיח חדש: $shachiach\ \ chadash=b\cdot\left(shachiach\ \ yashan\right)+a$

כיצד ישתנו מדדיי הפיזור?

הוספת קבוע למדדיי הפיזור לא משנה את ערכם. חשוב להזכיר כי מדיי פיזור הינם ערכים אי שליליים. ההכפלה משנה את הערך באופן הבא:

טווח כולל חדש: $R_y=|b|\cdot R_x$

סטיית תקן חדשה: $s_y=|b|\cdot s_x$

שונות חדשה: $s^2{}_y=b^2\cdot s^2{}_x$

טרנספורמציה ליניארית ידועה ושימושית הינה "תקנון". אם ל-X יש תוחלת μ ושונות σ^2 אזי Y המשתנה המוגדר $Y=\dfrac{X-\mu}{\sigma}$ הינו בעל תוחלת השווה ל-0 ושונות השווה ל-1

נוכיח: $E\left(\dfrac{X-\mu}{\sigma}\right)=\dfrac{E(X)-\mu}{\sigma}=\dfrac{\mu-\mu}{\sigma}=0$, $V\left(\dfrac{X-\mu}{\sigma}\right)=\dfrac{V(X)-\overbrace{V(\mu)}^{=0}}{\sigma^2}=\dfrac{\sigma^2}{\sigma^2}=1$

הוכחה ל5 בתכונות שונות:

$$V(X)=E\left(\left(X-E(X)\right)^2\right)=E\left(X^2-2XE(X)+\left(E(X)\right)^2\right)=E\left(X^2\right)-2E(X)E(X)+E\left(\left(E(X)\right)^2\right)$$
$$=E\left(X^2\right)-2\left(E(X)\right)^2+\left(E(X)\right)^2=E\left(X^2\right)-\left(E(X)\right)^2$$

בדוגמא הנ"ל:

החציון: $p\left(X\le 1\right)=0.125+0.375\ge\frac{1}{2}$ ז"א ש $r=1$ הוא החציון

התוחלת: $E\left(X\right)=0.125\cdot 0+0.375\cdot 1+0.375\cdot 2+0.125\cdot 3=1.5$

שונות: על מנת לחשב את השונות נחשב תחילה את הביטוי $E\left(X^2\right)$

$E\left(X^2\right)=0.125\cdot 0+0.375\cdot 1+0.375\cdot 4+0.125\cdot 9=2.625$ ובסה"כ נקבל:

$V\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=2.625-1.5^2=0.375$

r	r^2	$p\left(X=r\right)$	$\sum_{r\le a} p\left(X=r\right)$
0	0	0.125	0.125
1	1	0.375	0.5
2	4	0.375	0.875
3	9	0.125	1

דוגמא: נתון X מ.מ בדיד עם התפלגות: $p\left(X=r\right)=cr$ כאשר $r=1,\ldots,n$ ו- c קבוע כלשהו, חשב את התוחלת, החציון ואת השכיח.

פתרון: כבר חישבנו את c ומצאנו כי $c=\frac{2}{n\left(n+1\right)}$. כעת, נחשב את התוחלת:

$$E\left(X\right)=\sum_{r=1}^{n} r\cdot p\left(X=r\right)\xrightarrow{p(X=r)=cr}\sum_{r=1}^{n} r\cdot r\,c=c\sum_{r=1}^{n} r^2$$

$$\xrightarrow{\sum_{r=1}^{n} r^2=\frac{n(n+1)(2n+1)}{6}\ ,\ c=\frac{2}{n(n+1)}}\frac{n\left(n+1\right)\left(2n+1\right)}{6}\cdot\frac{2}{n\left(n+1\right)}=\frac{2n+1}{3}$$

החציון:

$$F(a)=p\left(X\le a\right)\ge\frac{1}{2}\rightarrow\sum_{r=1}^{a} cr=c\frac{a\left(a+1\right)}{2}\xrightarrow{c=\frac{2}{n(n+1)}}\frac{2}{n\left(n+1\right)}\frac{a\left(a+1\right)}{2}\ge\frac{1}{2}$$

$$a\ge\sqrt{\frac{1}{2}n\left(n+1\right)+\frac{1}{4}}-\frac{1}{2}$$

החציון יהיה הערך השלם הראשון שמקיים את האי השוויון הראשון.

שכיח: בגלל ש $p\left(X=r\right)=\frac{2r}{n\left(n+1\right)}$ אזי ככל שנעלה בערך r כך תגדל ההסתברות ולכן השכיח הוא n.

פונקציה של משתנה מקרי:

יהי X מ.מ בדיד ויהי $Y = f(X)$ פונקציה כלשהי של X.

טענה: Y באופן טבעי גם הוא מ.מ בדיד וערכי y נוצרים ע"י הפעלת הפונקציה על ערכי x. לעיתים מקבלים אותו ערך של y מערכים שונים של x.

בהטלת קובייה ההסתברות לקבל מספר מסוים הינו $\frac{1}{6}$. נפעיל פונקציה על המ.מ שלנו

באופן הבא: $y = (r-2)^2$

r	1	2	3	4	5	6
$p(X=r)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$y=(r-2)^2$	1	0	1	4	9	16
$p(Y=y)$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

דוגמא: יהי X מ.מ המקבל ערכים שלמים. נגדיר את y להיות $y = \left|\sin\left(\frac{1}{2}\Pi x\right)\right|$. מהם ערכי y של ההתפלגות?

פתרון: נפריד מקרים, ניקח x זוגי, למשל $2n$ ונקבל: $y = |\sin(\Pi n)| = 0$. עבור x אי זוגי, למשל $2n+1$ נקבל: $y = \left|\sin\left(\Pi n + \frac{\Pi}{2}\right)\right| = 1$.

התפלגות משותפת (מתוך סיפרו של פרופ' עוזי וישנה):

אם נתונים שני מ.מ-ים $X,Y:\Omega \to \mathbb{R}$ אזי אפשר להגדיר את המאורעות $\{X=a, Y=b\}$ ולקבל פונקציית התפלגות משותפת המסומנת $(a,b) \to P(X=a, Y=b)$. כאשר $(X=a, Y=b)$ הינו החיתוך ביניהם. נרחיב נושא זה בהמשך.

הערה: מתקיים כי: $P(X=a) = \sum_b P(X=a, Y=b)$ ובאופן דומה $P(Y=b) = \sum_a P(X=a, Y=b)$

בנוסף, לכל ערך b של Y, $X|Y=b$ הוא מ.מ מותנה, וההתפלגות המותנית שלו היא:

$$P(X=a \mid Y=b) = \frac{P(X=a, Y=b)}{P(Y=b)}$$

טענה: אם X ו- Y בלתי תלויים, אזי לכל $f:\mathbb{R} \to \mathbb{R}$ גם Y ו- $f(X)$ בלתי תלויים. מכאן גם $g(Y)$ ו- $f(X)$ בלתי תלויים לכל שתי פונקציות $f,g:\mathbb{R} \to \mathbb{R}$.

כעת, ננסה להוכיח עמה תכונות עבור התוחלות והשוניות (מהסעיף הקודם):

1. אדיטיביות של התוחלת: לכל שני מ.מ –ים , $E(X+Y)=E(X)+E(Y)$ (סעיף 4 בתכונות תוחלת).

הוכחה:

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

עבור כל ערך של a סוכמים על כל החיתוכים של הערך a עם כל ערכי b.

$$\begin{aligned}
&\sum_{a}\left(\underbrace{\sum_{b} P\left(X=a, Y=b\right)}_{P(X=a)}\right)a + \sum_{b}\left(\underbrace{\sum_{a} P\left(X=a, Y=b\right)}_{P(Y=b)}\right)b \\
&= \sum_{a} P\left(X=a\right)a + \sum_{b} P\left(Y=b\right)b = E\left(X\right)E\left(Y\right)
\end{aligned}$$

2. טענה: אם $\Omega = A_1 \cup \ldots \cup A_n$ הוא פירוק של המרחב לקבוצות זרות, אז $E\left(X\right)=\sum_i P\left(A_i\right)E\left(X \mid A_i\right)$.

הוכחה: נתחיל לפי הגדרה, $E\left(X\right)=\sum_a P\left(X=a\right)a$ כעת, נפתח את ההסתברות $P\left(X=a\right)$ ע"י נוסחת ההסתברות השלמה $P\left(X=a\right)=\sum_i P\left(X=a \mid A_i\right)P\left(A_i\right)$. ולכן:

$$\begin{aligned}
E\left(X\right) &= \sum_a P\left(X=a\right)a \xrightarrow{P(X=a)=\sum_i P(X=a|A_i)P(A_i)} \sum_a \sum_i P\left(X=a \mid A_i\right)P\left(A_i\right)a \\
&= \sum_i P\left(A_i\right)\underbrace{\sum_a P\left(X=a \mid A_i\right)a}_{E(X|A_i)} = \sum_i P\left(A_i\right)E\left(X \mid A_i\right)
\end{aligned}$$

3. חוק התוחלת החוזרת או משפט ההחלקה או משפט התוחלת השלמה:

$$E\left(X\right)=E\left(E\left(X \mid Y\right)\right).$$

הוכחה: לפי הטענה הקודמת מתקיים כי

$$E\left(X\right)=\sum_b P\left(Y=b\right)\left(E\left(X \mid Y=b\right)\right)=E\left(E\left(X \mid Y\right)\right)$$

השלב האחרון נובע מהגדרת תוחלת (עדיף להסתכל מימין לשמאל)

הערה: $E\left(X\right)=E\left(E\left(x\right)|Y\right)$, התוחלת $E\left(X\right)$ הינה ערך מספרי, ולכם המ.מ המותנה $E\left(x\right)|Y$ הוא קבוע ושווה תמיד ל $E\left(X\right)$ (ה מ.מ Y אינו משפיע על מספר קבוע ולכן אין להתניה זו משמעות).

פונקציה יוצרת מומנטים

פונקציה יוצרת מומנטים של משתנה מקרי X , כשמה כן היא, פונקציה שיוצרת לנו את המומנטים, משמע שממנה נוכל להניב את המומנטים, זאת על-ידי מציאת התוחלות: המומנט הראשון הינו התוחלת, $E(X)$, המומנט השני הינו $E(X^2)$ וכך הלאה. חשוב להדגיש כי המומנטים הם עבור המשתנה X. נסמן את המומנט ה-i-י באופן הבא: μ_i והוא מוגדר כך: $\mu_i = E(X^i) = \sum_r r^i \cdot p(X=r)$ למשל: $\mu_1 = E(X) = \sum_r r \cdot p(X=r)$ באותו אופן $\mu_2 = E(X^2) = \sum_r r^2 \cdot p(X=r)$ וכך הלאה.

חשיבותה התאורטית של פונקציה יוצרת מומנטים היא בכך שבתנאים מסוימים אפשר לשחזר ממנה את ההתפלגות של המשתנה (ע"י היפוך טרנספורם לפלס), והיא מאפשרת לבנות התפלגות מתוך המומנטים בלבד. למשל אם אין בידינו את בפונקצית ההתפלגות או הצפיפות אזי נוכל להסיק עבור ההתפלגות בעזרת המומנטים, תוחלת, נוכל לבנות את השונות, הקורטוזיס (גבנוניות) והסקיונס (צידוד).

הגדרתה של פונקציה יוצרת מומנטים: פונקציה יוצרת מומנטים של משתנה X היא **פונקציה של משתנה ממשי t** המוגדרת כתוחלת של e^{tX}, נסמנה $M_X(t)$, כאשר היא קיימת (יש פונקציות עבורן לא מוגדרת פונקציה יוצרת מומנטים). ומוגדרת באופן הבא:

$$M_X(t) = E(e^{tX}) = \sum_r e^{tr} \cdot p(X=r) \ .$$

אם הפונקציה יוצרת המומנטים גזירה n פעמים בקטע הכולל את הנקודה $t=0$,אז המומנט ה-n-י של המשתנה הוא הנגזרת ה-n-ית של הפונקציה בנקודה זו (את הנגזרת עושים לפי המשתנה t), כלומר $E(X^n) = M_X^{(n)}(0)$. כעת, קל לחשב את התוחלת או השונות של מ.מ מסוים: דוגמאות:

$$E(X) = M'_X(t=0)$$
$$E(X^2) = M''_X(t=0)$$
$$E(X^3) = M'''_X(t=0)$$
$$E(X^4) = M^{(4)}_X(t=0)$$

באופן הזה קל למצוא את השונות כיוון ש $V(X) = E(X^2) - (E(X))^2$

התפלגויות בדידות ידועות:

ישנן שש התפלגויות בדידות מיוחדות:

1. התפלגות אחידה.
2. התפלגות בינומית.
3. התפלגות גיאומטרית.
4. התפלגות פואסון.
5. התפלגות ברנולי
6. התפלגות היפרגאומטרית
7. התפלגות בינומית שלילית

כעת, נפרט על כל אחת מההתפלגויות ונלמד כיצד לזהות שמדובר בהתפלגות מסוימת.

התפלגות אחידה

נהוג לסמן מ.מ בעל התפלגות אחידה כך: $X \sim U(a,b)$ - כלומר, X מתפלג - ~ בהתפלגות אחידה (Uniform) עם הפרמטרים a ו- b , ההסתברות האחידה שווה ל- $\frac{1}{b-a+1}$

הסבר: בהתפלגות אחידה לכל x_i ישנה בדיוק אותה הסתברות להתרחש, דוגמא: תוצאה בהטלת קובייה תקינה: נרשום את טבלת ההתפלגות:

(x_i)	1	2	3	4	5	6
$p(x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

נוכל לראות לפי הטבלה, שלא משנה על איזה צד נפלה הקובייה ומה היא מראה, כיוון שההסתברות להראות כל מספר מ-1 עד- 6 היא - $\frac{1}{6}$ זוהי הסתברות אחידה, לכל מספר 1,5,2, או כל מספר אחר בקובייה יש בדיוק אותה הסתברות, נוכל לרשום זאת כך: $X \sim U(1,6)$ זאת כיוון שהערכים שלנו נעים מהמספר -1 עד למספר- 6 ולפי הנוסחה, ההסתברות האחידה של כל ה- x_i -ים הינה, $\frac{1}{6} = \frac{1}{6-1+1} = \frac{1}{b-a+1}$.

התפלגות בינומית

נהוג לסמן מ.מ בעל התפלגות בינומית כך: $X \sim B(n, p)$ כאשר:

n - הינו מספר הניסויים.

p - ההסתברות להצלחה.

$q = 1 - p$ - הסתברות לכישלון.

r - מספר הצלחות רצוי.

ההסתברות הבינומית הינה: $p_r = p(X = r) = \binom{n}{r} p^r \cdot q^{n-r}$

$\binom{n}{r}$ - בחירה של- r איברים מתוך n ניסויים **כאשר אין חשיבות לסדר**

<u>סימני זיהוי להתפלגות הבינומית:</u>

א. סדרת n ניסויים בלתי תלויים (ב"ת –למשל, אם מדובר על השאלה הנפוצה עם הוצאת הכדורים מספר פעמים מתוך כד, אז מדובר על האופציה שהיא עם החזרת הכדורים לכד).

ב. בכל ניסוי קיימת הצלחה בסיכוי P וכישלון בסיכוי q=1-p.

ג. המשתנה המקרי X סופר את מספר ההצלחות מתוך n הניסויים הבלתי תלויים .

כאשר n מייצג מספר ניסויים מקריים בלתי תלויים זה בזה. כך שלכל ניסוי מקרי יש שתי תוצאות אפשריות "הצלחה" בהסתברות p ו- "כשלון" בהסתברות $1 - p$.

$\binom{n}{r}$ - זהו סימון של מספר האפשרויות לקבל r מתוך n פריטים ללא החזרה, בתנאי שהסדר אינו חשוב. $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

תכונות ההתפלגות: לא נוכיח אך ברור כי מתקיים: $\sum_{r=0}^{n} \binom{n}{r} p^r \cdot q^{n-r} = 1$.

נמצא את התוחלת: $E(X) = \sum_{r=0}^{n} r \binom{n}{r} p^r \cdot q^{n-r} = E(X) = \sum_{r=1}^{n} r \binom{n}{r} p^r \cdot q^{n-r}$

השוויון נכון כיוון שהמחובר הראשון בסכום שווה ל-0.

$$\sum_{r=1}^{n} r \frac{n!}{r!(n-r)!} p^r \cdot q^{n-r} = \sum_{r=1}^{n} \frac{n(n-1)!}{(r-1)!(n-r)!} p \cdot p^{r-1} \cdot q^{n-r}$$

$$np \sum_{r=1}^{n} \frac{(n-1)!}{(r-1)!(n-r)!} p^{r-1} \cdot q^{n-r}$$

כעת נבצע הצבה: $s=r-1$ ונקבל:

$$np\sum_{r=1}^{n}\frac{(n-1)!}{(r-1)!(n-r)!}p^{r-1}\cdot q^{n-r}\xrightarrow{s=r-1}np\sum_{s=0}^{n-1}\frac{(n-1)!}{s!(n-s-1)!}p^{s}\cdot q^{n-s-1}$$

נסמן את $n-1=N$ ונקבל: $\underbrace{np\sum_{s=0}^{N}\frac{N!}{s!(N-s)!}p^{s}\cdot q^{N-s}}_{1}$ ובסה״כ: $E(X)=np$

פונקציה יוצרת מומנטים: $M_X(t)=\left(q+p\cdot e^t\right)^n$

נמצא את השונות: תחילה נמצא את $E\left(X^2\right)$, נעשה זאת ע״י פונקציה יוצרת מומנטים:

$$\left(M_X(t)\right)'=n\left(q+p\cdot e^t\right)^{n-1}p\cdot e^t$$
$$E(X)=\left(M_X(0)\right)'=n(q+p)^{n-1}p=np$$
$$\left(M_X(t)\right)''=n(n-1)\left(q+p\cdot e^t\right)^{n-2}p^2\cdot e^{2t}+p\cdot e^t n\left(q+p\cdot e^t\right)^{n-1}$$
$$E\left(X^2\right)=\left(M_X(0)\right)''=n(n-1)(q+p)^{n-2}p^2+pn(q+p)^{n-1}=n(n-1)p^2+pn$$
$$V(X)=E\left(X^2\right)-\left(E(X)\right)^2=n(n-1)p^2+pn-n^2p^2$$
$$=n^2p^2-np^2+pn-n^2p^2=-np^2+pn=np(-p+1)=npq$$

באופן טכני: איך נוכל לזהות כי מדובר בהתפלגות בינומית? סוג של שאלות שיש הסתברות להצלחה וכישלון, למשל:

*לידת בן או בת כאשר לידת בן נחשבת כהצלחה לידת בת כישלון

*הטלת מטבע כאשר למשל עץ נחשב הצלחה ופלי כישלון

*הטלת קובייה והתוצאה, למשל, 4 נחשבת הצלחה ושאר המספרים נחשבים כישלון.

*בד״כ השאלות עם כד הכדורים הידועות, כאשר ישנה חשיבות להוצאת כדורים עם צבעים מסוימים וישנה חזרה (ב״ת).

*דגימת אדם מאוכלוסיה ורישום האם משתמש בסמים (כישלון) או לא משתמש (הצלחה)

דוגמא: בתהליך יצור נורות, קימת הסתברות של 3% שתתקבל נורה פגומה.

א. מה ההסתברות שבמשלוח של 60 נורות תימצאנה בדיוק 4 נורות פגומות?

פתרון: נגדיר ״הצלחה״ – התקבלות נורה פגומה.

א. אם ישנם 60 נורות במשלוח אזי יש 60 ניסויים, ואנו רוצים ״להצליח״ בדיוק 4 פעמים, כיוון שאנו רוצים 4 נורות פגומות, לכן

n - מספר הניסויים- 60.

p - ההסתברות להצלחה- 0.03.

$q=1-p$ - 0.97.

r - מספר הצלחות רצוי- 4.

לכן, $p_4=p(X=4)=\binom{60}{4}0.03^4\cdot 0.97^{60-4}=\frac{60!}{4!(56)!}0.03^4\cdot 0.97^{46}\approx 0.097$

$\binom{60}{4}=\frac{60!}{4!(56)!}$ - בחירה של 4 הצלחות מתוך 60 ניסויים עם כל הוריאציות- חשיבות לסדר

0.03^4 - סיכוי להצלחה בחזקת 4, מכיוון שרוצים 4 הצלחות.

0.97^{56} - סיכוי לכישלון בחזקת 46 מכיוון שרוצים 46 כישלונות.

נסכם: בסה"כ 60 ניסויים מתוכם 4 הצלחות ו- 56 כישלונות.

ב. ב. נשנה את נתוני השאלה להלן, נתון כי בכל משלוח ישנן 50 נורות והסתברות לנורה פגומה הינו 2%. נשאלת השאלה מה ההסתברות שבמשלוח תהיינה פחות מ-4 נורות פגומות ?

פחות מ- 4 נורות פגומות, ז"א, 0 פגומות או 1 פגומה או 2 פגומות או 3 פגומות, נרשום זאת כנוסחה:

$$p(X<4)=p(X=0)+p(X=1)+p(X=2)+p(X=3)=$$

$$\binom{50}{0}0.02^0\cdot0.98^{50}+\binom{50}{1}0.02^1\cdot0.98^{49}+\binom{50}{2}0.02^2\cdot0.98^{48}+\binom{50}{3}0.02^3\cdot0.98^{47}\approx0.982$$

דוגמא: נסמן ב-X את מספר התוצאות- עץ ב-3 הטלות מטבע תקינה.

א. מהי ההסתברות לא לקבל עץ בכלל ?

ב. מהי ההסתברות לקבל עץ פעם אחת ?

ג. מהי ההסתברות לקבל פעמיים עץ ?

ד. מהי ההסתברות לקבל 3 פעמים עץ ?

פתרון: נגדיר את ההצלחה: לקבל עץ, ז"א שכישלון הינו קבלת פלי.

n - מספר הניסויים- 3.

p - ההסתברות להצלחה $-\frac{1}{2}$

$q=1-p$ - $\frac{1}{2}$.

א. לא לקבל עץ בכלל, ז"א, r- מספר הצלחות רצוי- 0. לכן, ההסתברות היא:

$$p_0=p(X=0)=\binom{3}{0}\left(\frac{1}{2}\right)^0\cdot\left(\frac{1}{2}\right)^3=\frac{3!}{0!(3)!}\left(\frac{1}{2}\right)^0\cdot\left(\frac{1}{2}\right)^3=\frac{1}{8}\cdot$$

ב. לקבל עץ פעם אחת מוגדר כהצלחה פעם אחת, ולקבל פלי פעמיים משמעותו להיכשל פעמיים

$$p_1=p(X=1)=\binom{3}{1}\left(\frac{1}{2}\right)^1\cdot\left(\frac{1}{2}\right)^2=\frac{3!}{1!(2)!}\left(\frac{1}{2}\right)^1\cdot\left(\frac{1}{2}\right)^2=\frac{3}{8}\cdot$$

ג. לקבל עץ פעמיים מוגדר כהצלחה פעמיים, ולקבל פלי פעם אחת משמעותה להיכשל פעם אחת.

$$p_2=p(X=2)=\binom{3}{2}\left(\frac{1}{2}\right)^2\cdot\left(\frac{1}{2}\right)^1=\frac{3!}{2!(1)!}\left(\frac{1}{2}\right)^2\cdot\left(\frac{1}{2}\right)^1=\frac{3}{8}\cdot$$

ד. לקבל עץ 3 פעמים מוגדר להצליח 3 פעמים, ולא לקבל פלי בכלל נחשב לא להיכשל.

$$p_3 = p(X=3) = \binom{3}{3}\left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^0 = \frac{3!}{3!(0)!}\left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{2}\right)^0 = \frac{1}{8}$$

התפלגות גיאומטרית

בהתפלגות גיאומטרית **לא** נאמר לנו כמה ניסויים עושים ושואלים אותנו "**באיזה ניסוי תחול ההצלחה הראשונה ?**"

בהתפלגות בינומית, היו נתונים מספר הניסויים, מספר זה היה **קבוע**, ונשאלנו לגבי על **מספר ההצלחות**.

<u>סימני זיהוי להתפלגות הגיאומטרית :</u>

- סדרת ניסויים בלתי תלויים (ב"ת).
- בכל ניסוי קיימת הצלחה בסיכוי P וכישלון בסיכוי q=1-p.
- המשתנה המקרי X סופר את מספר הניסויים הדרושים עד להצלחה ראשונה כולל הניסוי המצליח.

נסמן ב- X את מספר הניסוי שבו חלה הצלחה ראשונה, כך שיהיו $x-1$ כישלונות ואז באה ההצלחה הראשונה!

נהוג לסמן מ.מ בעל התפלגות גיאומטרית כך : $X \sim G(p)$ - כלומר, X מתפלג בהתפלגות גיאומטרית עם פרמטר p , כאשר p הינו הסיכוי להצלחה.

אנו קוראים להתפלגות זו – גיאומטרית בגלל שסדרת ההסתברויות שלה מהווה טור הנדסי\גיאומטרי אינסופי- $p,\ p\cdot q,\ \ p\cdot q^2,\ \ p\cdot q^3 \ldots$

הנוסחה להסתברות הגיאומטרית : $p_r = p(X=r) = p\cdot(1-p)^{r-1}$

כאשר :

p - ההסתברות להצלחה.

$q=1-p$ - הסתברות לכישלון.

r - מיקום ההצלחה הראשונה.

תחילה, נוכיח שהסכום בכל התחום של ההתפלגות שווה ל-1 :

$\sum_{r=1}^{\infty} p\cdot(1-p)^{r-1} = p\sum_{r=1}^{\infty}(1-p)^{r-1}$ ביטוי הסכום הינו סידרה הנדסית אינסופית מתכנסת

שסכומה $\sum_{r=1}^{\infty}(1-p)^{r-1} = \frac{1}{1-(1-p)}$ ולכן נקבל כי $p\sum_{r=1}^{\infty}(1-p)^{r-1} \xrightarrow{\sum_{r=1}^{\infty}(1-p)^{r-1}=\frac{1}{1-(1-p)}} p\frac{1}{1-(1-p)} = 1$.

תוחלת: $E(X) = \sum_{r=1}^{\infty} p\cdot r\cdot(1-p)^{r-1} \xrightarrow{1-p=q} p\sum_{r=1}^{\infty} r\cdot(q)^{r-1}$ נציב $1-p=q$

$$p\sum_{r=1}^{\infty} r\cdot(q)^{r-1} = p\sum_{r=0}^{\infty} r\cdot q^{r-1}$$

כעת, נעצור לרגע, ונתבונן בעובדה הידועה הבאה: ידוע לנו כי $\sum_{r=0}^{\infty} a^r = \frac{1}{1-a}$, אם נגזור את שני האגפים לפי a נקבל: $\sum_{r=0}^{\infty} ra^{r-1} = \frac{1}{(1-a)^2}$. במקרה זהו הסכום שקבלנו ולכן:

$$p\sum_{r=0}^{\infty} r\cdot q^{r-1} \xrightarrow{\sum_{r=0}^{\infty} ra^{r-1}=\frac{1}{(1-a)^2}} \frac{p}{(1-q)^2} \xrightarrow{1-a=p} \frac{p}{p^2} = \frac{1}{p}$$

פונקציה יוצרת מומנטים: $M_X(t) = \frac{pe^t}{1-qe^t}$

שונות: תחילה נמצא את $E\left(X^2\right)$, נעשה זאת ע"י פונקציה יוצרת מומנטים:

$$M_X(t) = \frac{pe^t}{1-qe^t}$$

$$\left(M_X(t)\right)' = \frac{pe^t\left(1-qe^t\right)-\left(-qe^t\right)pe^t}{\left(1-qe^t\right)^2} = \frac{pe^t\left(1-qe^t+qe^t\right)}{\left(1-qe^t\right)^2} = \frac{pe^t}{\left(1-qe^t\right)^2}$$

$$E(X) = \left(M_X(0)\right)' = \frac{p}{(1-q)^2} = \frac{p}{(p)^2} = \frac{1}{p}$$

$$\left(M_X(t)\right)'' = \frac{pe^t\left(1-qe^t\right)^2 - 2\left(1-qe^t\right)\left(-qe^t\right)pe^t}{\left(1-qe^t\right)^4}$$

$$E\left(X^2\right) = \left(M_X(0)\right)'' = \frac{p(1-q)^2-2(1-q)(-q)p}{(1-q)^4} = \frac{pp^2-2p(-q)p}{(1-q)^4} = \frac{p^3+2qp^2}{(p)^4} = \frac{p+2q}{(p)^2}$$

$$V(X) = E\left(X^2\right)-\left(E(X)\right)^2 = \frac{p+2q}{(p)^2}-\frac{1}{p^2} = \frac{p+2q-1}{(p)^2} \xrightarrow{p-1=-q} \frac{-q+2q}{p^2} = \frac{q}{p^2}$$

דוגמא: בתהליך יצור של בובות, ההסתברות של בובה פגומה הינה 1%.

א. מה ההסתברות שבבדיקת איכות של בובות בזו אחר זו תמצאנה שש בובות תקינות ושביעית פגומה?

ב. מהי ההסתברות שבביקורת שגרתית בה נבדקות 5 בובות בזו אחר זו לא נגלה אף בובה פגומה?

פתרון: נגדיר את ההצלחה - גילוי בובה פגומה.

א. 6 בובות ראשונות תקינות ושביעית פגומה, לפי הנוסחה נקבל:

p - ההסתברות להצלחה: 0.01.

$q=1-p$ - הסתברות לכישלון: 0.99.

r - מיקום ההצלחה הראשונה: 7.

$p_7 = p(X=7) = 0.01\cdot(0.99)^{7-1} = 0.01\cdot(0.99)^6 \approx 0.0094$.

ב. כאן אנו נשאלים בעצם, על הסתברות המאורע, שהבובה הפגומה הראשונה לא תהיה ב- 5 המקומות הראשונים, ז"א שהבובה הפגומה הראשונה יכולה להיות במקום השישי או השביעי או השמיני או התשיעי וכך הלאה.... כנוסחה נוכל לרשום:

$$p(X>5) = p(X=6) + p(X=7) + p(X=8) + p(x=9) + \ldots$$

טור מסוג זה הינו טור הנדסי אינסופי, ניתן למצוא את סכומו, ישנה נוסחה וניתן להשתמש בה, אנו נפתור את השאלה זו המשלים, ז"א, ניתן לומר כי:

$p(x>5) = 1 - p(x\le 5)$ ואת $p(x\le 5)$ נוכל לחשב, מכיוון שזהו סכום סופי!

$$p(X>5) = 1 - p(X\le 5) =$$
$$1 - \left[p(X=1) + p(X=2) + p(X=3) + p(X=4) + p(X=5)\right] =$$
$$1 - \left[0.01 + 0.01\cdot(0.99) + 0.01\cdot(0.99)^2 + 0.01\cdot(0.99)^3 + 0.01\cdot(0.99)^4\right] \approx 0.95$$

דוגמא: בכד 5 כדורים לבנים, 3 שחורים ו- 2 אדומים. מוציאים כדורים בזה אחר זה ומחזירים.

א. מה הסיכוי ששלושת הראשונים לבנים או שחורים והרביעי אדום ?

ב. מה הסיכוי שב- 3 הוצאות לא יהיה אפילו כדור אדום אחד ?

פתרון: א. נגדיר הצלחה: הוצאת כדור אדום וכישלון הוצאת כדור שהוא לא אדום, ז"א לבן או שחור- לא תהיה הבדלה בין שחור ללבן אלא בין אדום, למה שלא אדום.

p - ההסתברות להצלחה $\frac{2}{10} = \frac{1}{5}$.

$q = 1-p$ - הסתברות לכישלון הינה $1-\frac{1}{5} = \frac{4}{5}$ או ניתן לראות זאת אחרת, $\frac{8}{10} = \frac{4}{5}$

r - מיקום ההצלחה הראשונה הינה 4.

לכן, $p_4 = p(x=4) = \frac{1}{5}\cdot\left(\frac{4}{5}\right)^{4-1} = \frac{1}{5}\cdot\left(\frac{4}{5}\right)^3 \approx 0.1024$

ב. באופן דומה לסעיף ב' של הדוגמא הקודמת נצטרך לבצע את החישוב כך:

$$p(X>3) = 1 - p(X\le 3) = 1 - \left[p(X=1) + p(X=2) + p(X=3)\right] =$$
$$1 - \left[\frac{1}{5}\cdot\left(\frac{4}{5}\right)^0 + \frac{1}{5}\cdot\left(\frac{4}{5}\right)^1 + \frac{1}{5}\cdot\left(\frac{4}{5}\right)^2\right] = \frac{64}{125}$$

התפלגות פואסון

התפלגות פואסונית מופיעה בבעיות כגון: מספר תקלות במרכזית טלפונים ביום מסוים, מספר התפרקויות גרעיניות של חומר רדיואקטיבי ביחידות זמן, מספר האנשים הפונים לפקיד מסוים בין השעות 8:00 עד 12:00, וכדו'. נוכל לסכם באופן כללי, כי אם X מתפלג פואסונית אזי הוא סופר את מספר האירועים/המופעים ביחידה קנה מידה נתונה(יחידת זמן.

נהוג לסמן מ.מ בעל התפלגות פואסונית כך: $X \sim P(\lambda)$ - כלומר, X מתפלג בהתפלגות פואסונית עם פרמטר λ, כאשר $\lambda > 0$ ומוגדר כממוצע מס' האירועים בזמן המוגדר מראש.

הנוסחה להסתברות פואסונית: $p_k = p(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}$

X – מס' האירועים ביחידות זמן, יכול לקבל את הערכים $\{0,1,2,3,\ldots,\infty\}$.

תחילה, נוכיח שהסכום בכל התחום של ההתפלגות שווה ל-1:

$$\sum_{r=0}^{\infty} p(X=r) = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} = e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}$$

נזכר כי אם עושים פיתוח טור טיילור סביב 0 – טור מקלורן ל- e^x מקבלים:

$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ לכן נוכל לומר כי $e^{\lambda} = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}$ והסה"כ נקבל

כי $$\sum_{r=0}^{\infty} p(X=r) = e^{-\lambda} \sum_{r=0}^{\infty} \frac{\lambda^r}{r!} \xrightarrow{e^{\lambda} = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}} e^{-\lambda} \cdot e^{\lambda} = 1$$

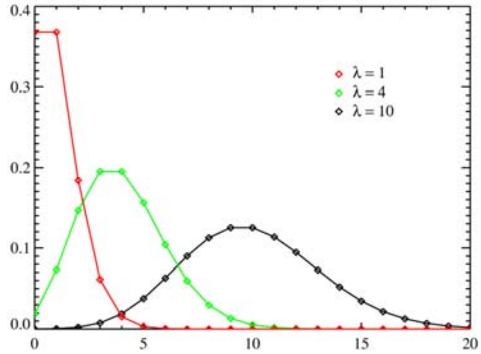
התוחלת שווה ל- λ, $E(X) = \lambda$:

$$E(X) = \sum_{r=0}^{\infty} r \cdot \frac{\lambda^r}{r!} e^{-\lambda} \xrightarrow{if\ r=0 \ \to\ r \cdot \frac{\lambda^r}{r!} e^{-\lambda} = 0} \sum_{r=1}^{\infty} r \cdot \frac{\lambda^r}{r!} e^{-\lambda} = \sum_{r=1}^{\infty} r \frac{\lambda^r}{r(r-1)!} e^{-\lambda} = \sum_{r=1}^{\infty} \frac{\lambda^r}{(r-1)!} e^{-\lambda}$$

נעשה הצבה: $r - 1 = s$ ונקבל:

$$\sum_{r=1}^{\infty} \frac{\lambda \lambda^{r-1}}{(r-1)!} e^{-\lambda} \xrightarrow{r-1=s} \lambda e^{-\lambda} \sum_{s=0}^{\infty} \frac{\lambda^s}{s!} \xrightarrow{e^{\lambda} = \sum_{r=0}^{\infty} \frac{\lambda^r}{r!}} \lambda e^{-\lambda} e^{\lambda} = \lambda$$

בסה"כ קבלנו כי תוחלת התפלגות פואסון שווה לפרמטר ההתפלגות. נוכל לראות דוגמאות בגרפים הבאים:



פונקציה יוצרת מומנטים : $M_X\left(t\right)=e^{\lambda\left(e^t-1\right)}$

שונות : תחילה נמצא את $E\left(X^2\right)$, נעשה זאת ע"י פונקציה יוצרת מומנטים :

$$M_X\left(t\right)=M_X\left(t\right)=e^{\lambda\left(e^t-1\right)}$$
$$\left(M_X\left(t\right)\right)'=e^{\lambda\left(e^t-1\right)}\lambda e^t$$
$$E\left(X\right)=\left(M_X\left(0\right)\right)'=\lambda$$
$$\left(M_X\left(t\right)\right)''=e^{\lambda\left(e^t-1\right)}\lambda^2e^{2t}+\lambda e^te^{\lambda\left(e^t-1\right)}$$
$$E\left(X^2\right)=\left(M_X\left(0\right)\right)''=\lambda^2+\lambda$$
$$V\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=\lambda^2+\lambda-\lambda^2=\lambda$$

דוגמא : במרכזית קלדניות, ממוצע הטעיות בהקלדת נתונים הוא 10 ליום.

מהי ההסתברות שמספר הטעויות יהיה קטן או שווה ל-12?

פתרון : לפי נתוני השאלה $\lambda=10$ בכדי לדעת מהי ההסתברות שמספר הטעויות יהיה קטן או שווה ל-12, נצטרך למצוא את ההסתברות הבאה : $p(X\le 12)$. לכן,

$$p(x\le 12)=p(X=0)+p(X=1)+p(X=2)+\ldots+p(X=12)\ =$$

$$\frac{10^0}{0!}e^{-10}+\frac{10^1}{1!}e^{-10}+\frac{10^2}{2!}e^{-10}+\frac{10^3}{3!}e^{-10}+\frac{10^4}{4!}e^{-10}+\frac{10^5}{5!}e^{-10}+\frac{10^7}{7!}e^{-10}+\frac{10^8}{8!}e^{-10}+\frac{10^9}{9!}e^{-10}+\frac{10^{10}}{10!}e^{-10}+\frac{10^{11}}{11!}e^{-10}+\frac{10^{12}}{12!}e^{-10}$$

דוגמא : ממוצע מס' טעויות בהעברת פס תקשורת ביממה הינו 6.

א. מהי ההסתברות ליממה ללא טעויות?

ב. מהו הסיכוי לפחות מ-2 טעויות בשעה?

ג. מה הסיכוי ליותר מטעות אחת לשעה?

פתרון : א. לפי נתוני השאלה $\lambda=6$. בכדי לדעת מהי ההסתברות ליממה ללא טעויות (0 טעויות) נציב בנוסחה : $p_0=p(X=0)=\frac{6^0}{0!}e^{-6}$.

ג. בסעיף זה נשאלנו לגבי הסיכוי לפחות מ-2 טעויות בשעה אחת, ולא ביממה! כאמור, ה- λ בשאלה שווה ל-6 פר יממה, כלומר, עבור שעה בודדת נקבל λ שונה. בכדי לדעת מהי ה- λ , נלמד <u>**ערך משולש**</u> :

מס' שעות	λ
24	6
1	?

את "?" נוכל למצוא ע"י הפעולה הבאה:

$\frac{1\cdot 6}{24}=? \quad \rightarrow \quad ?=\frac{1}{4}$, ולכן ה- λ החדשה שלנו שווה $\frac{1}{4}$. כעת נמצא את ההסתברות,

כאשר נתונה לנו λ החדשה. נסמן את λ החדשה כ- λ_{new}:

$$p(X<2)=p(X=0)+p(X=1)=\frac{{\lambda_{new}}^{0}}{0!}e^{-\lambda_{new}}+\frac{{\lambda_{new}}^{1}}{1!}e^{-\lambda_{new}}=$$

$$=\frac{\left(\frac{1}{4}\right)^{0}}{0!}e^{\left(-\frac{1}{4}\right)}+\frac{\left(\frac{1}{4}\right)^{1}}{1!}e^{\left(-\frac{1}{4}\right)}\approx 0.973$$

ג. עדיין מדובר על $\lambda=\frac{1}{4}$ כיון שעדיין מדובר על פר שעה. ההסתברות שאנו צריכים

למצוא היא: $p(X>1)=1-p(X\le 1)=1-\left[p(X=0)+p(X=1)\right]$
$=1-0.973=0.027$

המעבר האחרון נובע על סמך הסעיף הקודם בו מצאנו כי:
$p(X=0)+p(X=1)=0.973$

שאלה 10: מספר הטעויות בכתיבת תוכנית ב C++ בממוצע ביממה הוא 3, הטעות יכולה לקרות בכל חלק של היממה בהסתברות שווה.

א. מה הסיכוי ל-2 טעויות ביממה

ב. הסיכוי ליותר מ- 2 טעויות ביממה

ג. הסיכוי שלא התרחשה אף טעות בשעה מסוימת

ד. הסיכוי שהתרחשה טעות או יותר בין השעות 10 ל- 12

פתרון: א. מהנתון נוכל לומר כי $\lambda=3$ ביממה – 24 שעות.

הסיכוי ל – 2 טעויות ביממה, בגלל שמדובר ביממה אז אנו יכולים להישאר עם ה- λ שלנו, ז"א $\lambda=3$ בסעיף זה $p(X=2)=\frac{3^{2}}{2!}e^{-3}\approx 0.22$

ב. הסיכוי ליותר מ- 2 טעויות ביממה, שוב מדובר על יממה לכן אנו יכולים להישאר עם ה- λ שלנו ,ז"א $\lambda=3$.

$$p(x>2)=1-p(X\le 2)=$$

$$1-\left[p(X=0)+p(X=1)+p(X=2)\right]=1-\left[\frac{\lambda^{0}}{0!}e^{-\lambda}+\frac{\lambda^{1}}{1!}e^{-\lambda}+\frac{\lambda^{2}}{2!}e^{-2}\right]=$$

$$1-\left[\frac{(3)^{0}}{0!}e^{(-3)}+\frac{(3)^{1}}{1!}e^{(-3)}+\frac{(3)^{2}}{2!}e^{(-3)}\right]\approx 0.568$$

ג. הסיכוי שלא התרחשה אף טעות בשעה מסוימת- ז"א 0 טעויות, כאשר מדובר על שעה ולא על יממה, כמו בנתון המקורי, נצטרך לעשות ערך משולש.

שעות	λ
24	3
1	?

$\frac{1\cdot 3}{24}=? \rightarrow ?=\frac{1}{8}$ לכן, ה- λ החדשה שלנו שווה $\frac{1}{8}$. כעת נמצא את ההסתברות, כאשר נתונה לנו λ החדשה. נסמן את λ החדשה כ- λ_{new} :

$$p(X=0)=\frac{{\lambda_{new}}^{0}}{0!}e^{-\lambda_{new}} \rightarrow \frac{\left(\frac{1}{8}\right)^{0}}{0!}e^{\left(-\frac{1}{8}\right)}=\frac{1}{\sqrt[8]{e}}$$

ד. הסיכוי שהתרחשה טעות או יותר בין השעות 10 ל- 12, גם פה מדובר על שעתיים ולא על יממה, כמו בנתון המקורי, לכן שוב נצטרך לעשות ערך משולש.

שעות	λ
24	3
2	?

$\frac{2\cdot 3}{24}=? \rightarrow ?=\frac{1}{4}$ לכן, ה- λ החדשה שלנו שווה $\frac{1}{4}$. כעת נמצא את ההסתברות, כאשר נתונה לנו λ החדשה. נסמן את λ החדשה כ- λ_{new} :

$$p(X\geq 1)=1-p(X<1)$$
$$=1-\left[p(X=0)\right]=1-\left[\frac{{\lambda_{new}}^{0}}{0!}e^{-\lambda_{new}}\right]$$
$$=1-\left[\frac{\left(\frac{1}{4}\right)^{0}}{0!}e^{-\frac{1}{4}}\right]=1-e^{-\frac{1}{4}}=1-\frac{1}{\sqrt[4]{e}}$$

התפלגות ברנולי

בהתפלגות ברנולי, המשתנה המקרי X מקבל שני ערכים, 0 ו-1. התפלגות בינומית עם $n=1$ הינה התפלגות ברנולי. בדומה להתפלגות בינומית, התפלגות זו מתאימה לתיאור מצבים של הצלחה או כישלון. למשל, בהטלת קובייה תקינה תסומן התוצאה 3 כהצלחה וכל שאר התוצאות האחרות ככישלון. ההסתברות לנפילה על 3 בקובייה תקינה היא 1/6, ולפיכך ההסתברות המשלימה המתייחסת לכל שאר התוצאות (1,2,4,5,6) היא 5/6. בדוגמה זו המשתנה המציין את המאורע המתאים הוא בעל התפלגות ברנולי עם פרמטר $p=\frac{1}{6}$. נהוג לסמן מ.מ בעל התפלגות ברנולי כך: $X \sim Ber(p)$ כאשר:

p - ההסתברות להצלחה.

$q=1-p$ - הסתברות לכישלון.

הסתברות ברנולי הינה: $p\left(X=1\right)=1-p\left(X=0\right)=1-q=p$

פונקציה יוצרת מומנטים: $M_X\left(t\right)=q+pe^t$

התוחלת שווה ל $E\left(X\right)=p$ נראה על-ידי פונקציה יוצרת מומנטים:

$$E\left(X\right)=\left(M_X\left(t\right)\right)'=pe^t \xrightarrow{t=o} p$$

השונות: $V\left(X\right)=pq$. נראה על-ידי פונקציה יוצרת מומנטים

$$M_X\left(t\right)=q+pe^t$$
$$E\left(X\right)=\left(M_X\left(t\right)\right)'=pe^t \xrightarrow{t=o} p$$
$$E\left(X^2\right)=\left(M_X\left(t\right)\right)''=pe^t \xrightarrow{t=o} p$$
$$V\left(X\right)=E\left(X^2\right)-\left(E\left(X\right)\right)^2=p-p^2=p\left(1-p\right)=pq$$

נזכור כי n ניסויי ברנולי הוא ניסוי בינומי.

התפלגות היפרגאומטרית

נתונה אוכלוסיה בת N פרטים, שמהם R פרטים בעלי תכונה מסוימת שייקראו "מיוחדים". שאר N-R פרטים הם "רגילים". מוציאים מהאוכלוסייה ללא החזרה מדגם של n פריטים. יהי X משתנה מקרי הסופר את מספר הפרטים ה"מיוחדים" שהתקבלו במדגם.

כך לדוגמה, התפלגות זו מתארת מספר הכדורים הלבנים שמתקבלים (ללא החזרה) כאשר מוציאים n כדורים מכד שיש בו N כדורים, ומתוכם יש r כדורים לבנים.

הגדרת ההתפלגות:

$$P(X=k)=\frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} \qquad k=0,1,2....n \qquad \text{if} \quad \mathrm{n \le R \le N}$$

$$E(X)=n\frac{R}{N} \qquad\qquad V(X)=n\frac{R}{N}(1-\frac{R}{N})(\frac{N-n}{N-1})$$

התפלגות בינומית שלילית

התפלגות בינומית שלילית מתארת את מספר ההצלחות בסדרת ניסויי ברנולי בלתי תלויים לפני שמתרחשים מספר קבוע נתון מראש, r, של הצלחות. **לדוגמא**, אם נטיל מטבע שוב ושוב ונעצור כאשר נקבל 1 בפעם השלישית, אז מספר האפסים שנראה יתפלג באופן בינומי שלילי. אם ההסתברות להצלחה בכל ניסוי היא p וההסתברות לכישלון היא $1-p$ אז המספר האקראי של ההצלחות שנראה, X, יהיה בעל התפלגות בינומית שלילית עם הפרמטרים (r,p) ונסמן זאת כך: $X \sim NB(r,p)$

$$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r} \qquad ,k=r,r+1....$$

פונקציה יוצרת מומנטים: $M_X(t)=\left(\frac{p}{1-qe^t}\right)^r$

תוחלת: $E(X)=\frac{r}{p}$

שונות: $V(X)=\frac{rq}{p^2}$

נוכיח לפי פונקציה יוצרת מומנטים:

$$M_X(t) = M_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r$$

$$E(X) = \left(M_X(t)\right)' = r\left(\frac{pe^t}{1-qe^t}\right)^{r-1} \frac{pe^t\left(1-qe^t\right)-\left(-qe^t\left(pe^t\right)\right)}{\left(1-qe^t\right)^2} \xrightarrow{t=0} r\frac{p^2+pq}{p^2} = r\frac{p\left(p+p\right)}{p^2} = \frac{r}{p}$$

$$\left(M_X(t)\right)' = r\left(\frac{pe^t}{1-qe^t}\right)^{r-1} \frac{pe^t - qpe^{2t} + qpe^{2t}}{\left(1-qe^t\right)^2} = r\left(\frac{pe^t}{1-qe^t}\right)^{r-1} \frac{pe^t}{\left(1-qe^t\right)^2}$$

$$\left(M_X(t)\right)'' = r\left(r-1\right)\left(\frac{pe^t}{1-qe^t}\right)^{r-2}\left(\frac{pe^t}{\left(1-qe^t\right)^2}\right)^2 + \frac{pe^t\left(1-qe^t\right)^2 - 2\left(1-qe^t\right)\left(-qe^t\right)pe^t}{\left(1-qe^t\right)^4}\cdot r\left(\frac{pe^t}{1-qe^t}\right)^{r-1}$$

$$\xrightarrow{t=0} r\left(r-1\right)\left(\frac{1}{p}\right)^2 + \frac{p^3+2p^2q}{p^4}\cdot r = \frac{r\left(r-1\right)}{p^2} + \frac{r\left(1+q\right)}{p^2} = \frac{r^2+qr}{p^2}$$

$$V(X) = E\left(X^2\right) - \left(E(X)\right)^2 = \frac{r^2+qr}{p^2} - \frac{r^2}{p^2} = \frac{r^2+qr-r^2}{p^2} = \frac{qr}{p^2}$$

התפלגויות רציפות

מ.מ רציף – משתנה המקבל טווח ערכים המהווה קטע בציר המספרים הממשיים, המשתנה מקבל ערכים רציפים כגון משקל, גובה, כל זווית אפשרית במעגל, זמן חיים ועוד.

חשוב לציין כי לכל תוצאה אפשרית של מ.מ רציף יש הסתברות השווה ל-0, לכן לא נשאל מה ההסתברות ש- X יקבל ערך מסוים (הרי זה שווה ל-0) אלא מה ההסתברות ש X יקבל ערך מתוך רווח מסוים: למשל, $p\left(X<a\right)$ או $p\left(a<X<b\right)$. ברור כי $p\left(X=a\right)=0$ וכך גם עבור b, לכן אין הבדל בין $p\left(X<a\right)$ ל- $p\left(X\le a\right)$.

תהי $f\left(x\right)$ פונקצית צפיפות (PDF - Probability Density Function) המוגדרת על הממשיים $\mathbb{R}$ ומקיימת את התנאים הבאים:

1. $f\left(x\right)\ge 0$
2. $f\left(x\right)$ רציפה פרט אולי למספר סופי של נקודות
3. $\int\limits_{-\infty}^{\infty} f\left(x\right)dx=1$

הגדרה פורמאלית: יהי X מ.מ רציף עם פונקצית צפיפות $f\left(x\right)$ אם $p\left(X\in A\right)=\int\limits_{A} f\left(x\right)dx$ כך שאם $A=\left(a,b\right)$ אזי $\int\limits_{A} f\left(x\right)dx=\int\limits_{a}^{b} f\left(x\right)dx$ ובנוסף מתקיים: $p\left(X=a\right)=0$.

דוגמאות:

נתונה הפונקציה
$$f\left(x\right)=\begin{cases} c\left(4x-2x^2\right) & 0\le x\le 2 \\ 0 & \text{otherwise} \end{cases}$$

א. עבור איזה ערך של c הפונקציה תהיה פונקצית צפיפות?

ב. מצאו את $p\left(X>1\right)$

פתרון: א. נרצה שהפונקציה הנתונה תקיים את שלושת התנאים:

1. $f\left(x\right)\ge 0$ - עבור כל ערך של x. בתחום בין $0\le x\le 2$ הפונקציה $4x-2x^2$ חיובית, נשאר לנו לקבוע את c כך שלא יקלקל ולכן נדרוש כי $c>0$.
2. הפונקציה רציפה – גם בנקודות הפיצול (ניתן לבדוק ☺)
3. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int\limits_{-\infty}^{\infty} f\left(x\right)dx=1$

כעת, $\int_{-\infty}^{\infty} f(x)dx=\int_0^2 c\left(4x-2x^2\right)dx=c\left(2x^2-\frac{2x^3}{3}\right)_0^2=c\frac{8}{3}$ נשווה ל-1 ונקבל כי

$$c\frac{8}{3}=1\rightarrow c=\frac{3}{8}$$

ב. $p(X>1)=\int_1^2 \frac{3}{8}\left(4x-2x^2\right)dx=\frac{3}{8}\left(2x^2-\frac{2x^3}{3}\right)_1^2=\frac{3}{8}\left(\frac{8}{3}-\frac{4}{3}\right)=\frac{1}{2}$

דוגמא: נתונה הפונקציה $f(x)=\lambda e^{\frac{-x}{100}}$ כאשר $x>0$.

א. מצאו עבור איזה ערך של λ הפונקציה תהווה פונקצית צפיפות

ב. מצאו $p(50\le X\le 150)$

פתרון: א. נרצה שהפונקציה הנתונה תקיים את שלושת התנאים:

1. $f(x)\ge 0$ - עבור כל ערך של x כאשר מתקיים $\lambda>0$

2. ברור כי הפונקציה רציפה עבור $x>0$

3. נבדוק עבור איזה ערך של λ מתקיים התנאי השלישי: $\int_{-\infty}^{\infty} f(x)dx=1$

$\int_{-\infty}^{\infty} f(x)dx=\int_0^{\infty}\lambda e^{\frac{-x}{100}}dx=\left(\frac{\lambda e^{\frac{-x}{100}}}{\frac{-1}{100}}\right)_0^{\infty}$ נזכור כי $\lim_{x\to-\infty} e^x=0$ ולכן נקבל כי

$\left(-100\lambda e^{\frac{-x}{100}}\right)_0^{\infty}=0-(-100\lambda)=100\lambda$ ביטוי זה שווה ל 1- ולכן $\lambda=\frac{1}{100}$.

ב. $p(50\le X\le 150)=\int_{50}^{150}\frac{1}{100}e^{\frac{-x}{100}}dx=\left(\frac{1}{100}(-100)e^{\frac{-x}{100}}\right)_0^{\infty}=e^{\frac{-2}{3}}-e^{\frac{-1}{2}}$

כמה דברים שניתן להסיק: אם האינטגרל של $f(x)$ מוגדר ותמיד חיובי, אזי נוכל להגדיר את $F(x)$ כגזירה, רציפה ולא יורדת. $F(x)$ מוגדרת כפונקצית ההתפלגות ודרכה נוכל לחשב התפלגויות וגם להגדיר את פונקצית הצפיפות: $F(a)$ המסומנת גם $F_x(a)$ מקיימת $F(a)=F_x(a)=p(X\le a)$ (השטח משמאל). באופן דומה: $F(a\le X\le b)=p(X\le b)-p(X\le a)=F(b)-F(a)$.

הגדרה פורמאלית: תהי $f(x)$ פונקציה רציפה כך שעבור כל מתקיים $\int_a^b f(x)dx=F(b)-F(a)$ אזי קוראים ל $f(x)$ פונקצית צפיפות של שמ.מ רציף X

ו- $F(x)$ נקראת פונקצית ההצטברות או ההתפלגות.

דוגמא: נתונה הפונקציה $f\left(x\right)=\frac{c}{x^2}$ כאשר $1\le x\le 2$.

א. מצא את c על מנת שהפונקציה תהיה פונקצית צפיפות

ב. מצא את פונקצית ההסתברות

פתרון: א. נרצה שהפונקציה הנתונה תקיים את שלושת התנאים:

1. $f\left(x\right)\ge 0$ - מתקיים, עבור כל ערך של x (כיוון שמופיע בפונקציה- x^2), כאשר מתקיים $c>0$

2. ברור כי הפונקציה רציפה בתחום ההגדרה: $1\le x\le 2$

3. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int_{-\infty}^{\infty} f\left(x\right)dx=1$

$$\int_{-\infty}^{\infty} f\left(x\right)dx=1 \xrightarrow{f(x)=\frac{c}{x^2}} \int_{1}^{2}\frac{c}{x^2}dx=\left(\frac{-c}{x}\right)_1^2=\frac{-c}{2}-\frac{-c}{1}=1\rightarrow c=2$$

ב. לפי הגדרה $F\left(a\right)=p\left(X\le a\right)$ ולכן נרצה למצוא את

$$p\left(X\le a\right)=\int_{1}^{a} f\left(x\right)dx \xrightarrow{f(x)=\frac{c}{x^2}} \int_{1}^{a}\frac{2}{x^2}dx=\left(\frac{-2}{x}\right)_1^a=\frac{-2}{a}-\frac{-2}{1}\rightarrow 2\left(1-\frac{1}{a}\right)$$

כעת נוכל להגדיר את הפונקציה:

$$F\left(a\right)=\begin{cases}0 & a<1\\ 2\left(1-\frac{1}{a}\right) & 1\le a\le 2\\ 1 & a>2\end{cases}$$

נצטרך לדאוג שהפונקציה רציפה ועולה:

עבור הרציפות: בנקודה 2 הגבולות החד צדדיים קיימים ושווים, ושווים לערך הפונקציה בנקודה ולכן הפונקציה רציפה

$$\lim_{a\to 2^-}F\left(a\right)=\lim_{a\to 2^-}2\left(1-\frac{1}{a}\right)=1$$

$$\lim_{a\to 2^+}F\left(a\right)=1$$

גם בנקודה 1 הגבולות החד צדדיים קיימים ושווים, ושווים לערך הפונקציה בנקודה ולכן הפונקציה רציפה

$$\lim_{a\to 1^+}F\left(a\right)=\lim_{a\to 1^+}2\left(1-\frac{1}{a}\right)=0$$

$$\lim_{a\to 1^-}F\left(a\right)=0$$

בנוסף, הפונקציה עולה כי נגזרתה ($f\left(x\right)=\frac{2}{x^2}$) תמיד חיובית.

תוחלת: הגדרת תוחלת עבור משתנה רציף: $E(X)=\int_{-\infty}^{\infty} x\cdot f(x)dx$ בהנתן שהאינטגרל קיים.

דוגמא: מצא את התוחלת של ההתפלגות עבורה נתונה פונקצית הצפיפות הבאה:

$f(x)=\frac{3}{8}\left(4x-2x^2\right)$ בתחום $0\le x\le 2$.

פתרון, נעשה זאת לפי הגדרת התוחלת:

$$E(X)=\int_{-\infty}^{\infty} x\cdot f(x)dx=\int_{0}^{2}\frac{3}{8}x\left(4x-2x^2\right)dx=\frac{3}{8}\left(\frac{2x^3}{3}-\frac{2x^4}{4}\right)_0^2=1$$

משמעות התוצאה של התוחלת: הניסוי נותן תוצאה בין 0 ל- 2, אם נחזור מספר רב של פעמים יהיו מספר גדול של תוצאות שהממוצע שלהם יהיה קרוב ל-1.

חציון: יהי a חציון אם $p(X\le a)=\frac{1}{2}\Leftrightarrow p(X\ge a)=\frac{1}{2}$ (לא בהכרח מספר יחיד, ז"א יכולים להיות מספר ערכים המקימים זאת)

שכיח: יהי a שכיח אם עבור כל ערך של b מתקיים $f(a)\ge f(b)$.

פונקציה של משתנה מקרי: יהי X מ.מ בעל פונקצית צפיפות $f(x)$ ופונקצית הסתברות $F(x)$. ויהי y פונקציה של x אזי גם y הינו מ.מ.

דוגמא: נתון כי $y=x^2$ בהנתן שיודעים כיצד X מתפלג, מצאו כיצד Y מתפלג.

פתרון: במקרים אלו, שנרצה לדעת התפלגות של מ.מ שהוא פונקציה של משתנה מקרי ידוע בהתפלגותו, נעבוד בד"כ עם פונקצית ההתפלגות:

$$F_y(a)=p(Y\le a)\xrightarrow{y=x^2}p\left(X^2\le a\right)=p\left(-\sqrt{a}\le X\le\sqrt{a}\right)=F_x\left(\sqrt{a}\right)-F_x\left(-\sqrt{a}\right)$$

כעת, נעבור לפונקצית צפיפות, נעשה זאת ע"י נגזרת:

$$f_x\left(\sqrt{a}\right)\frac{1}{2\sqrt{a}}-f_x\left(-\sqrt{a}\right)\frac{-1}{2\sqrt{a}}=\frac{1}{2\sqrt{a}}\left(f_x\left(\sqrt{a}\right)+f_x\left(-\sqrt{a}\right)\right)$$

דוגמא: נתון כי $y=x^3$ בהנתן שיודעים כיצד X מתפלג, מצאו כיצד Y מתפלג.

פתרון: $F_y(a)=p(Y\le a)\xrightarrow{y=x^3}p\left(X^3\le a\right)=p\left(X\le\sqrt[3]{a}\right)=F_x\left(\sqrt[3]{a}\right)$

דוגמא: נתונה הפונקציה הבאה:

$$f(x)=\begin{cases}0 & x<0\\ c\left(2x-x^2\right) & 0\le x\le 2\\ 0 & x>2\end{cases}$$

א. מצאו את ערכו של c על מנת שתהווה פונקצית צפיפות לגיטימית

ב. חשב את $p(x>1.5)$

ג. הגדר ומצא את פונקצית ההתפלגות

פתרון: נרצה שהפונקציה הנתונה תקיים את שלושת התנאים:

1. $f(x)\ge 0$ - מתקיים, עבור כל ערך של x בתחום $0\le x\le 2$ כאשר מתקיים $c>0$
2. ברור כי הפונקציה רציפה בתחום ההגדרה- פולינום
3. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int_{-\infty}^{\infty} f(x)dx=1$

$$\int_{-\infty}^{\infty} f(x)dx=1\to\int_0^2 c\left(2x-x^2\right)dx=c\left(x^2-\frac{x^3}{3}\right)_0^2=c\left(4-\frac{8}{3}\right)=1\to c=\frac{3}{4}$$

ב. לפי הגדרה $F(a)=p(X\le a)$ ולכן נרצה למצוא את

$$p(X<1.5)=\frac{3}{4}\int_{1.5}^{2}\left(2x-x^2\right)dx=\frac{3}{4}\left(x^2-\frac{x^3}{3}\right)_{1.5}^{2}=0.156$$

ג. נגדיר את פונקציית ההתפלגות באופן הבא:

$F(x)=(X\le x)=\int_{-\infty}^{x} f(x)dx=\frac{3}{4}\int_0^x\left(2x-x^2\right)dx=\frac{3}{4}\left(x^2-\frac{x^3}{3}\right)_0^x=\frac{3}{4}\left(x^2-\frac{x^3}{3}\right)$ ולכן:

$$F(x)=\begin{cases}0 & x<0\\ \frac{3}{4}\left(x^2-\frac{x^3}{3}\right) & 0\le x\le 2\\ 1 & x>2\end{cases}$$

נצטרך לדאוג שהפונקציה רציפה ועולה:

עבור הרציפות: בנקודה 2 הגבולות החד צדדיים קיימים ושווים, ושווים לערך הפונקציה בנקודה ולכן הפונקציה רציפה

$$\lim_{x\to 2^-}F(x)=\lim_{x\to 2^-}\frac{3}{4}\left(x^2-\frac{x^3}{3}\right)=1$$

$$\lim_{x\to 2^+}F(x)=1$$

גם בנקודה 1 הגבולות החד צדדיים קיימים ושווים, ושווים לערך הפונקציה בנקודה ולכן הפונקציה רציפה

$$\lim_{x\to 0^+}F(x)=\lim_{x\to 0^+}\frac{3}{4}\left(x^2-\frac{x^3}{3}\right)=0$$

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

בנוסף, הפונקציה עולה כי נגזרתה ($f(x)=\frac{3}{4}\left(2x-x^2\right)$) תמיד חיובית.

דוגמא: נתונה פונקציית ההתפלגות הבאה:

$$F(x)=\begin{cases}0 & x<0\\ \frac{x^2}{4} & 0\le x\le 2\\ 1 & x>2\end{cases}$$

א. מצאו את פונקציית הצפיפות של X

ב. מצאו את $p(x<1.5)$, $p(1\le x\le 2)$ ואת $p(0.5\le x\le 1)$

פתרון: א. על מנת לחזור אחורה לפונקצית הצפיפות נצטרך לגזור:

$$F(x)=\begin{cases}0 & x<0\\ \frac{x^2}{4} & 0\le x\le 2\\ 1 & x>2\end{cases}\xrightarrow{\text{derivative}} f(x)=\begin{cases}0 & x<0\\ \frac{x}{2} & 0\le x\le 2\\ 0 & x>2\end{cases}$$

ג. בגלל שנתונה פונקצית ההתפלגות אנו יכולים לחשב בשתי שיטות:

$$p(x<1.5)=F(1.5)=\frac{1.5^2}{4}=\frac{9}{16} \text{ או } p(x<1.5)=\int_0^{1.5}\frac{x}{2}dx=\frac{9}{16}$$

$$\cdot\ p(1\le x\le 2)=F(2)-F(1)=\frac{2^2}{4}-\frac{1}{4}=\frac{3}{4}$$

$$\cdot\ p(0.5\le x\le 1)=F(1)-F(0.5)=\frac{2^2}{4}-\frac{1}{4}=\frac{3}{4}$$

דוגמא- שאלה ממבחן:

יהי מ.מ רציף בעל פונקצית צפיפות הבאה:

$$f(x)=\begin{cases}c\left(4x-2x^2\right) & 0\le x\le 2\\ 0 & \text{otherwise}\end{cases}$$

א. מצא את c

ב. מצא את פונקצית ההתפלגות

ג. מצא את $p(x<1.5\,|\,x>1)$

פתרון: א. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int_{-\infty}^{\infty} f(x)dx=1$

$$\int_{-\infty}^{\infty} f(x)dx=1\rightarrow\int_0^2 c\left(4x-2x^2\right)dx=\rightarrow c=\frac{3}{4}$$

ב. $F(x)=(X\le x)=\int_{-\infty}^{x} f(x)dx=\frac{3}{4}\int_0^x\left(4x-2x^2\right)dx=\frac{3}{4}\left(2x^2-\frac{2x^3}{3}\right)_0^x=\frac{3}{4}x^2-\frac{x^3}{4}$

$$F(x)=\begin{cases}0 & x<0\\ \frac{3}{4}x^2-\frac{x^3}{4} & 0\le x\le 2\\ 1 & x>2\end{cases}$$

ג.

$$p\left(x>1.5\,|\,x>1\right)=\frac{p\left(x>1.5\cap x>1\right)}{p\left(x>1\right)}=\frac{p\left(x>1.5\right)}{p\left(x>1\right)}=\frac{1-p\left(x\le 1.5\right)}{1-p\left(x\le 1\right)}=\frac{1-F\left(1.5\right)}{1-F\left(1\right)}$$

$$=\frac{1-\frac{3}{8}\left(2\cdot 1.5^2-\frac{2}{3}\cdot 1.5^3\right)}{1-\frac{3}{8}\left(2\cdot 1^2-\frac{2}{3}\cdot 1^3\right)}=0.32$$

שאלה ממבחן: יהי מ.מ רציף בעל פונקצית צפיפות הבאה:

$$f\left(x\right)=\begin{cases} c\left(\cos x\right) & |x|\le\frac{\pi}{2} \\ 0 & \text{otherwise}\end{cases}$$

א. מצא את c

ב. מצא את פונקצית ההתפלגות

ג. מצא את $p\left(|X|<\frac{\pi}{4}\right)$

פתרון: א. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int_{-\infty}^{\infty} f\left(x\right)dx=1$

$$\int_{-\infty}^{\infty} f\left(x\right)dx=1\rightarrow\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} c\left(\cos x\right)dx=\rightarrow\left(c\left(\sin x\right)\right)_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=c\left(\underbrace{\sin\left(\frac{\pi}{2}\right)}_{1}-\underbrace{\sin\left(-\frac{\pi}{2}\right)}_{-1}\right)=1\rightarrow c=\frac{1}{2}$$

ב.

$$F\left(x\right)=\left(X\le x\right)=\int_{-\infty}^{x} f\left(x\right)dx=\frac{1}{2}\int_{-\frac{\pi}{2}}^{x}\cos x dx=\left(\frac{1}{2}\left(\sin x\right)\right)_{-\frac{\pi}{2}}^{x}=\frac{1}{2}\left(\sin x-\underbrace{\sin\left(-\frac{\pi}{2}\right)}_{-1}\right)=\frac{1}{2}\left(\sin x+1\right)$$

ולכן נוכל להגדיר את פונקצית ההתפלגות באופן הבא:

$$F\left(x\right)=\begin{cases} 1 & x>\frac{\pi}{2} \\ \frac{1}{2}\left(\sin x+1\right) & |x|\le\frac{\pi}{2} \\ 0 & x<-\frac{\pi}{2}\end{cases}$$

ג.

$$p\left(|X|<\frac{\pi}{4}\right)=p\left(-\frac{\pi}{4}<X<\frac{\pi}{4}\right)=F\left(\frac{\pi}{4}\right)-F\left(-\frac{\pi}{4}\right)\xrightarrow{F\left(-\frac{\pi}{4}\right)=1-F\left(\frac{\pi}{4}\right)}F\left(\frac{\pi}{4}\right)-\left(1-F\left(\frac{\pi}{4}\right)\right)=$$

$$2F\left(\frac{\pi}{4}\right)-1=2\left(\frac{1}{2}\left(\sin\frac{\pi}{4}+1\right)\right)-1=\frac{\sqrt{2}}{2}$$

ניתן כמובן לחשב ע"י פונקציית הצפיפות: $\frac{1}{2}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}}\cos xdx=\left(\frac{1}{2}(\sin x)\right)_{-\frac{\pi}{4}}^{\frac{\pi}{4}}=\frac{\sqrt{2}}{2}$

דוגמא ממבחן: למ.מ X יש פונקציית צפיפות $f(x)=D\sin x$ כאשר $0\le x\le\frac{\pi}{2}$. מצא את D ואת התוחלת.

פתרון: על מנת למצוא את D נשתמש בתכונה כי $\int_{-\infty}^{\infty}f(x)dx=1$:

$$\int_0^{\frac{\pi}{2}}D\sin xdx=1\rightarrow(-D\cos x)_0^{\frac{\pi}{2}}=-D\left(\underbrace{\cos\frac{\pi}{2}}_{0}-\underbrace{\cos 0}_{1}\right)\rightarrow D=1$$

נמצא את התוחלת, נעשה זאת לפי הגדרת בסיסית של תוחלת: $E(X)=\int_{-\infty}^{\infty}xf(x)dx$

ונקבל:

$$E(X)=\int_{-\infty}^{\infty}x\sin xdx=\int_0^{\frac{\pi}{2}}x\sin xdx\xrightarrow{\text{Integration by parts}}\begin{cases}u=x & u'=1\\ v'=\sin x & v=-\cos x\end{cases}$$

$$-x\cos x-\int-\cos xdx=(-x\cos x+\sin x)_0^{\frac{\pi}{2}}=-\frac{\pi}{2}\underbrace{\cos\frac{\pi}{2}}_{0}+\underbrace{\sin\frac{\pi}{2}}_{1}-\left(-0\underbrace{\cos 0}_{1}+\underbrace{\sin 0}_{0}\right)=1$$

דוגמא מבחינה: יהי מ.מ רציף בעל פונקציית צפיפות הבאה:

$$f(x)=\begin{cases}\frac{c}{\sqrt{a^2-x^2}} & -a\le x\le a\\ 0 & \text{otherwise}\end{cases}$$

א. מצא את c

ב. מצא את פונקציית ההתפלגות

ג. מצא את $p(0\le X\le a)$ ואת $p\left(-\frac{a}{2}\le X\le\frac{a}{2}\right)$

פתרון: א. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int_{-\infty}^{\infty}f(x)dx=1$

$$\int_{-\infty}^{\infty}f(x)dx=1\rightarrow\int_{-a}^{a}\frac{c}{\sqrt{a^2-x^2}}dx\rightarrow\left(c\arcsin\left(\frac{x}{a}\right)\right)_{-a}^{a}=c(\arcsin 1-\arcsin(-1))$$

$$\xrightarrow{\arcsin(-x)=-\arcsin x}2c\underbrace{\arcsin 1}_{\frac{\pi}{2}}=1\rightarrow c=\frac{1}{\pi}$$

ב.

$$F\left(x\right)=\left(X\le x\right)=\frac{1}{\pi}\int\limits_{-a}^{x}\frac{1}{\sqrt{a^2-x^2}}dx=\frac{1}{\pi}\int\limits_{-a}^{x}\left(\arcsin\left(\frac{x}{a}\right)\right)_{-a}^{x}dx=\frac{1}{\pi}\left(\arcsin\left(\frac{x}{a}\right)\underbrace{-\arcsin\left(\frac{-a}{a}\right)}_{+\frac{\pi}{2}}\right)$$

$$=\frac{1}{\pi}\arcsin\left(\frac{x}{a}\right)+\frac{1}{2}$$

$$F\left(x\right)=\begin{cases}0 & x>a\\ \frac{1}{\pi}\arcsin\left(\frac{x}{a}\right)+\frac{1}{2} & -a\le x\le a\\ 1 & x<-a\end{cases}$$

ג.

$$p\left(0\le X\le a\right)=\frac{1}{\pi}\int\limits_{0}^{a}\frac{1}{\sqrt{a^2-x^2}}dx\to\left(\frac{1}{\pi}\arcsin\left(\frac{x}{a}\right)\right)_{0}^{a}=\frac{1}{2}$$

$$p\left(-\frac{a}{2}\le X\le\frac{a}{2}\right)=\frac{1}{\pi}\int\limits_{-\frac{a}{2}}^{\frac{a}{2}}\frac{1}{\sqrt{a^2-x^2}}dx\to\left(\frac{1}{\pi}\arcsin\left(\frac{x}{a}\right)\right)_{-\frac{a}{2}}^{\frac{a}{2}}=\frac{1}{3}$$

פונקציה יוצרת מומנטים – עבור התפלגויות רציפות

נסמן את המומנט ה-i-י באופן הבא: μ_i והוא מוגדר כך:

$\mu_i = E\left(X^i\right) = \int_{-\infty}^{\infty} x^i f_X(x)dx$ למשל: $\mu_1 = E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$ באותו אופן $\mu_2 = E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f_X(x)dx$ וכך הלאה.

חשיבותה התיאורטית של פונקציה יוצרת מומנטים היא בכך שבתנאים מסוימים אפשר לשחזר ממנה את ההתפלגות של המשתנה, והיא מאפשרת לבנות התפלגות מתוך המומנטים בלבד.

הגדרתה של פונקציה יוצרת מומנטים: פונקציה יוצרת מומנטים של משתנה X היא **פונקציה של משתנה ממשי** t המוגדרת כתוחלת של e^{tX}, נסמנה $M_X(t)$, כאשר היא קיימת. ומוגדרת באופן הבא:

$$M_X(t) = E\left(e^{tX}\right) = \int_{-\infty}^{\infty} e^{tX} f_X(x)dx \cdot$$

אם הפונקציה יוצרת המומנטים גזירה n פעמים בקטע הכולל את הנקודה $t=0$, אז המומנט ה-n-י של המשתנה הוא הנגזרת ה-n-ית של הפונקציה בנקודה זו, כלומר $E\left(X^n\right) = M_X^{(n)}(0)$. כעת, קל לחשב את התוחלת או השונות של מ.מ מסויים:

דוגמאות:

$$E(X) = M'_X(0)$$
$$E\left(X^2\right) = M''_X(0)$$
$$E\left(X^3\right) = M'''_X(0)$$

התפלגויות רציפות ידועות:

1. התפלגות אחידה רציפה- Uniform distribution- Continuous
2. התפלגות מעריכית או התפלגות אקספוננציאלית
3. התפלגות נורמאלית
4. התפלגות לוגנורמאלית
5. התפלגות ביתא
6. התפלגות גמא
7. התפלגות χ^2 (חי בריבוע)
8. התפלגות T
9. התפלגות F
10. התפלגות וייבל
11. התפלגות פארטו - power law

התפלגות אחידה רציפה- Uniform distribution- Continuous

יהי X מ.מ המתפלג אחידה (יוניפומית) המקבל ערכים בקטע הסגור $[a,b]$ כאשר פונקצית הצפיפות היא:

$$f(x)=\begin{cases}0 & x<a \quad or \quad x>b\\ \dfrac{1}{b-a} & a\le x\le b\end{cases}$$

ופונקצית ההתפלגות היא

$$F_X(x)=\begin{cases}0 & x<a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & x>b\end{cases}$$

נמצא את התוחלת של התפלגות זו:

$$E(X)=\int_{-\infty}^{\infty}xf(x)dx=\int_a^b x\frac{1}{b-a}dx=\frac{1}{b-a}\left(\frac{1}{2}\left(b^2-a^2\right)\right)=\frac{b+a}{2}$$

נמצא את השונות של התפלגות זו:

$$E\left(X^2\right)=\int_{-\infty}^{\infty}x^2f(x)dx=\int_a^b x^2\frac{1}{b-a}dx=\frac{b^3-a^3}{3(b-a)}$$

$$V(X)=E\left(X^2\right)-\left(E(X)\right)^2=\frac{(b-a)^2}{12}$$

פונקציה יוצרת מומנטים: $M_X(t)=\dfrac{e^{bt}-e^{at}}{t(b-a)}$ ניתן להראות דרך גבול (פעמיים לופיטל).

תרגיל: נתון כי ניתן להגריל מספר תוך 5 עד 10 שנים בהתפלגות אחידה.

א. מה ההסתברות לחכות בין 4 ל- 7 שנים?

ב. כעבור 6 שנים לא הוגרל המספר, מה ההסתברות לחכות יותר משלוש שנים נוספות?

פתרון: נגדיר כי X מתפלג אחידה $X\sim U(5,10)$.

א. אנו בעצם נשאלים מהי ההסתברות: $P(4\le X\le 7)$ לכן, נחשב:

$$P(4\le X\le 7)=F_X(7)-F_X(4)\xrightarrow{X\sim U(5,10)\ ,\ F_X(x)=\frac{x-5}{10-5}}\frac{7-5}{10-5}-\frac{4-5}{10-5}=0.6$$

ב. $$P(X>9\,|\,X>6)=\frac{P(X>9\cap X>6)}{P(X>6)}=\frac{P(X>9)}{P(X>6)}=\frac{1-F_X(9)}{1-F_X(6)}\xrightarrow{X\sim U(5,10)\ ,\ F_X(x)=\frac{x-4}{10-5}}\frac{1-\frac{9-5}{10-5}}{1-\frac{6-5}{10-5}}=\frac{1}{4}$$

תרגיל: נגדיר את X מ.מ, כאשר $X\sim U\left(0,1\right)$, נגדיר את Y מ.מ באופן הבא $y=\sin\left(\frac{\pi x}{2}\right)$.

מצא את פונקצית הצפיפות עבור Y כך שערכיו נעים בין 0 ל- 1.

פתרון:

$$F_Y(a)=p\left(Y\le a\right)=p\left(\sin\left(\frac{\pi x}{2}\right)\le a\right)=p\left(\sin\left(\frac{\pi x}{2}\right)\le a\right)=p\left(\frac{\pi x}{2}\le\arcsin a\right)=p\left(X\le\frac{2}{\pi}\arcsin a\right)$$

כעת, בגלל שאנו יודעים כיצד מתפלג X נוכל לחשב את פונקצית ההתפלגות:

$$F_X(a)=p\left(X\le\frac{2}{\pi}\arcsin a\right)=\int_0^{\frac{2}{\pi}\arcsin a}\frac{1}{1-0}dx=\frac{2}{\pi}\arcsin a$$

נוכל לחשב את פונקצית הצפיפות, ע"י גזירה: $f\left(a\right)=\left(\frac{2}{\pi}\arcsin a\right)'=\frac{2}{\pi\sqrt{1-a^2}}$ כאשר $0\le a<1$.

התפלגות מעריכית או התפלגות אקספוננציאלית:

אנחנו נשתמש בהתפלגות מעריכית כאשר מדובר במשתנים כגון זמן המתנה לאוטובוס, זמן המתנה בבנק, זמן עד קלקול מחשב\נורה וכו'. משתנה שמתפלג מעריכית מסומן: $X\sim\exp\left(\lambda\right)$. התפלגות זו היא רציפה על המספרים האי שליליים

λ - הקצב בו המאורע מתרחש

$E\left(X\right)=\frac{1}{\lambda}$ - התוחלת. ננסה להסביר את תוצאת התוחלת: הזמן שאנו מצפים להמתין גדל ככל שהקצב יותר איטי, אפשר לראות זאת באופן הבא: ככל שהקצב גדול אנו ממתינים פחות.

פונקצית הצפיפות של התפלגות מעריכית היא: $f\left(x\right)=\lambda e^{-\lambda x}$ כאשר $\lambda>0$ ו- $0<X<\infty$. ניתן לכתוב: $f\left(x\right)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0 & x<0\end{cases}$

- נראה כי מתקיים $\int_{-\infty}^{\infty}f\left(x\right)dx=1$:

$$\int_{-\infty}^{\infty}f\left(x\right)dx=\int_0^{\infty}\lambda e^{-\lambda x}=\left(-e^{-\lambda x}\right)_0^{\infty}=-e^{-\infty}-\left(-e^{-0}\right)\xrightarrow{e^{-\infty}\to 0}1$$

- נראה כי התוחלת שווה $\frac{1}{\lambda}$:

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_{0}^{\infty}\lambda xe^{-\lambda x}\xrightarrow{\text{Integration by parts}}\begin{cases}u=x & u'=1\\ v'=\lambda e^{-\lambda x} & v=-e^{-\lambda x}\end{cases}$$

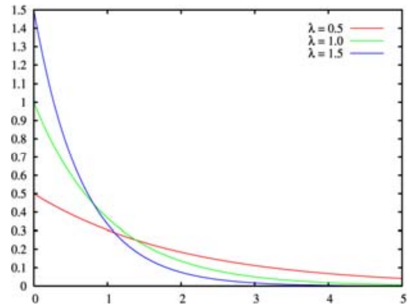
$$-xe^{-\lambda x}-\int -e^{-\lambda x}dx=\left(-xe^{-\lambda x}-\frac{1}{\lambda}e^{-\lambda x}\right)_0^{\infty}\rightarrow \lim_{x\to\infty}\frac{x}{e^{\lambda x}}\xrightarrow{lopital}\lim_{x\to\infty}\frac{1}{\lambda e^{\lambda x}}\rightarrow 0$$

$$0-0-\left(0-\frac{1}{\lambda}\right)=\frac{1}{\lambda}$$

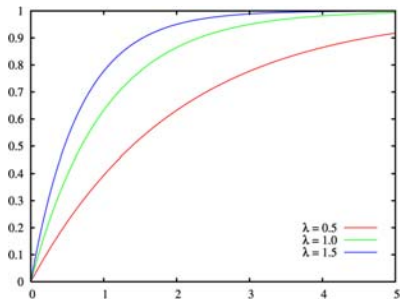
- נמצא את פונקצית ההתפלגות : $F(x)=\int_{-\infty}^{x}\lambda e^{-\lambda x}dx=\left(-e^{-\lambda x}\right)_0^x=-e^{-\lambda x}-\left(-e^{-0}\right)=1-e^{-\lambda x}$

נוכל להגדיר : $F(x)\begin{cases}1-e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$

פונקצית הצפיפות



פונקצית הצטברות



פונקציה יוצרת מומנטים $M_X(t)=\dfrac{\lambda}{\lambda-t}$

השונות שווה : $V(X)=\dfrac{1}{\lambda^2}$.

נוכיח לפי פונקציה יוצרת מומנטים : נמצא את המומנט השני :

$$E(X)=\left(M_X(t)\right)'=\frac{\lambda}{(\lambda-t)^2}\xrightarrow{t=0}\frac{1}{\lambda}$$

$$E\left(X^2\right)=\left(M_X(t)\right)''=\frac{-2(\lambda-t)(-1)\lambda}{(\lambda-t)^4}=\frac{2\lambda}{(\lambda-t)^3}\xrightarrow{t=0}\frac{2\lambda}{(\lambda)^3}=\frac{2}{\lambda^2}$$

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

• **להתפלגות מעריכית אין זיכרון**: משמע שמתקיים השוויון הבא: $p(X>s+t\,|\,x>s)=p(X>t)$. העובדה שהמתנו s זמן לא משפיעה על ההסתברות להמתין עוד t זמן.

הוכחה לתכונת אי זיכרון:

$$p(X>s+t\,|\,X>s)=\frac{p((X>s+t)\cap(X>s))}{p(X>s)}=\frac{p(X>s+t)}{p(X>s)}=\frac{1-p(X\le s+t)}{1-p(X\le s)}$$

$$=\frac{1-F(s+t)}{1-F(s)}=\frac{1-\left(1-e^{-\lambda(s+t)}\right)}{1-\left(1-e^{-\lambda s}\right)}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda(s+t)-(-\lambda s)}=e^{-\lambda t}=1-F(t)=1-p(X\le t)=p(X>t)$$

דוגמא: אורך של שיחת טלפון היא 10 דקות בממוצע, בהנחה שלשיחת טלפון יש התפלגות מעריכית, מה ההסתברות לשיחה בין 10 דקות ל20 דקות.

פתרון: לפי נתוני השאלה, נגדיר את X להיות אורך זמן שיחה המתפלג מעריכית: $X\sim P(\lambda)$. בנוסף, ישנו נתון על ממוצע אורך זמן השיחה, ז"א שהתוחלת ידועה: $E(X)=\frac{1}{\lambda}=10\rightarrow\lambda=\frac{1}{10}$. ומכאן $X\sim P\left(\frac{1}{10}\right)$.

נרצה לחשב את ההסתברות לשיחה בין 10 דקות ל- 20 דקות:

$$p(10<X<20)=\int_{10}^{20}\frac{1}{10}e^{-\frac{1}{10}x}dx=\frac{1}{10}\left[\frac{e^{-\frac{1}{10}x}}{-\frac{1}{10}}\right]_{10}^{20}=-\left[\left(e^{-\frac{1}{10}20}\right)-\left(e^{-\frac{1}{10}10}\right)\right]=-e^{-2}+e^{-1}$$

דוגמא: מ.מ מתפלג אקספוננציאלית עם פרמטר $\lambda=0.1$ מצאו את $p(5<X<15)$, $V(X)$ ואת $E(X)$.

פתרון: נתון $X\sim\exp(0.1)$ נחשב: $p(5<X<15)=\int_{5}^{15}0.1e^{-0.1x}dx=\left(-e^{-0.1x}\right)_{5}^{15}=-e^{-0.1\cdot15}-\left(-e^{-0.1\cdot5}\right)$

$$E(X)=\frac{1}{\lambda}=\frac{1}{0.1}=10$$

$$V(X)=\frac{1}{\lambda^2}=\frac{1}{0.01}=100$$

דוגמא: הזמן הדרוש לתיקון מכונה הוא מ.מ בעל התפלגות אקפוננציאלית עם פרמטר $\lambda=2$.

א. מה ההסתברות שזמן התיקון יעלה על 0.5 שעות?

ב. מה ההסתברות שתיקון ייקח לפחות 12.5 שעות בהנחה שכבר נמשך יותר מ12 שעות.

פתרון: א. $p(X>0.5)=\int_{0.5}^{\infty}2e^{-2x}dx=\left(-e^{-2x}\right)_{0.5}^{\infty}=e^{-1}$

ב. $p(X>12.5\,|\,X>12)=p(X>0.5)$

התפלגות נורמאלית

ההתפלגות מאופיינת על-ידי שני פרמטרים: התוחלת המסומנת μ והשונות המסומנת σ^2. מ.מ המתפלג נורמאלית עם תוחלת ושונות, נהוג לסמן באופן הבא: $X \sim N\left(\mu,\sigma^2\right)$.

פונקצית הצפיפות מוגדרת: $f\left(x\right)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right\}$

נוכיח כי $\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right\}dx=1$: על מנת לפתור את האינטגרל נציב $\frac{x-\mu}{\sigma}=y\rightarrow\frac{1}{\sigma}=\frac{dy}{dx}\rightarrow dx=\sigma dy$. לכן, נקבל את האינטגרל הבא: $\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{y^2}{2}\right\}dy$. נסמן את האינטגרל $\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{y^2}{2}\right\}dy$ ב- I ולכן I^2 יהיה: $I^2=\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{y^2}{2}\right\}dy\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{z^2}{2}\right\}dz$ ששווה ל - $I^2=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{y^2+z^2}{2}\right\}dydz$.

נעבור לאורדינאטות פולריות $y=r\sin\theta$ $z=r\cos\theta$ ומתקיים כי

$$dydz=\begin{vmatrix}\frac{dy}{d\theta} & \frac{dy}{dr}\\ \frac{dz}{d\theta} & \frac{dz}{dr}\end{vmatrix}drd\theta\ =\begin{vmatrix}r\cos\theta & \sin\theta\\ -r\sin\theta & \cos\theta\end{vmatrix}drd\theta=r\left(\cos^2\theta+\sin^2\theta\right)drd\theta=rdrd\theta$$

בסה"כ נקבל:

$$I^2=\int\limits_0^{\infty}\int\limits_0^{2\pi}r\exp\left\{-\frac{r^2}{2}\right\}d\theta dr=\int\limits_0^{\infty}2\pi r\exp\left\{-\frac{r^2}{2}\right\}dr\xrightarrow{\frac{r^2}{2}=t\quad\rightarrow dr=\frac{dt}{r}}2\pi\left(-e^{-\frac{r^2}{2}}\right)_0^{\infty}=2\pi$$

ולכן $\mathrm{I}=\sqrt{2\pi}$ ובסה"כ נקבל כי $\frac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{\infty}\exp\left\{-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right\}dx=1$.

מקרה מיוחד ונפוץ בשימושו הוא כאשר $X\sim N\left(0,1\right)$ ז"א שהתוחלת שווה ל-0 והשונות שווה ל-1, עבור מקרה זה נוכל להגדיר את פונקצית ההצטברות באופן הבא: $\Phi\left(a\right)=\int\limits_{-\infty}^{a}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$. ההתפלגות במקרה זה נקראת התפלגות נורמאלית סטנדרטית.

הפונקציה שהתקבלה $\Phi\left(x\right)$ איננה פונקציה אלמנטארית כלומר, היא אינה מתקבלת מהרכבה סופית של פולינומים ,פונקצית האקספוננט והפונקציות הטריגונומטריות , והפונקציות ההפוכות להם. משום כך, ישנה טבלה המכילה את הערכים המקורבים להתפלגות הנורמאלית הסטנדרטית, שחושבו בשיטות נומריות. הקירוב הבא שימושי

למדי כאשר הערך שברצוננו למצוא עבורו את השטח משמאל גדול (כשאנו אומרים גדול אנו מתכוונים לערך הגדול מ3.5):

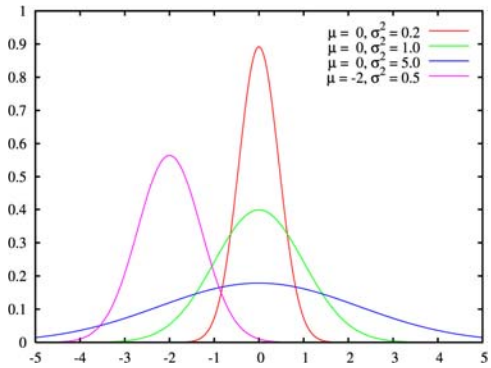
$$\Phi(a) \cong 1 - \frac{1}{a\sqrt{2\pi}} e^{-\frac{a^2}{2}}.$$

עקומת הנורמאליות הינה העקומה הרציפה השימושית ביותר. היא סימטרית וידועה בשם עקומת הפעמון או עקומת גאוס.

הסיבה לחשיבותה הרבה הינה העובדה כי למשתנים רבים יש התפלגות נורמאלית. כגון: גובה, משקל, ציוני משכל ועוד. חשוב לזכור:

א. בהתפלגות נורמאלית מתקיים כי הממוצע, החציון והשכיח מתלכדים.

ב. קיימות אינסוף עקומות נורמאליות. דוגמאות:



טענה: יהי מ.מ X המתפלג נורמאלית עם תוחלת μ ושונות σ^2 אזי Z המוגדר $Z = \frac{X-\mu}{\sigma}$ מתפלג נורמאלית סטנדרטית – עם תוחלת שווה ל-0 ושונות שווה ל-1.

הוכחה: נמצא את פונקציית ההצטברות של Z:

$$F(a) = p(Z \le a) = p\left(\frac{X-\mu}{\sigma} \le a\right) \rightarrow p(X \le \sigma a + \mu)$$

$$\xrightarrow{f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}} \int_{-\infty}^{\sigma a+\mu} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} dx \xrightarrow{\frac{x-\mu}{\sigma}=y \rightarrow \frac{1}{\sigma}=\frac{dy}{dx} \rightarrow dx=\sigma dy}$$

נצטרך לשנות גם את הגבולות ולכן כאשר X שואף לגבולות האינטגרל אזי

$$X \rightarrow -\infty \xrightarrow{y=\frac{x-\mu}{\sigma}} y = -\infty$$

$$X = \sigma a + \mu \xrightarrow{y=\frac{x-\mu}{\sigma}=\frac{\sigma a+\mu-\mu}{\sigma}} y = a$$

כעת, נקבל את האינטגרל הבא $\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} dx$ שהוא בדיוק $F(a)$ המסומן $\Phi(a)$. קבלנו שההתפלגות של Z הינה התפלגות של משתנה המתפלג נורמאלית סטנדרטית.

פונקציה יוצרת מומנטים: $M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2 t^2\right]$

נוכיח: $M_X(t)=E\left(e^{tx}\right)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{tx}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$ כעת, ע״י השלמה לריבוע נקבל:

$\left(x-\mu-\sigma^2 t\right)^2=(x-\mu)^2-2(x-\mu)\sigma^2 t+\sigma^4 t^2=\underbrace{(x-\mu)^2-2x\sigma^2 t}+2\mu\sigma^2 t+\sigma^4 t^2$ נרצה לבטא את הביטוי המסומן, ולכן, ע״י העברת אגפים נקבל $(x-\mu)^2-2x\sigma^2 t=\left(x-\mu-\sigma^2 t\right)^2-\left(2\mu\sigma^2 t+\sigma^4 t\right)$. נציב, ונוציא כל ביטוי שאינו כולל x מחוץ לאינטגרל ואז נקבל:

$$M_X(t)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{\left(x-\mu-\sigma^2 t\right)^2-\sigma^4 t^2-2\mu\sigma^2 t}{2\sigma^2}\right\}dx\rightarrow\frac{\exp\left(\frac{\sigma^4 t^2+2\mu\sigma^2 t}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left\{-\frac{\left(x-\mu-\sigma^2 t\right)^2}{2\sigma^2}\right\}dx$$

$$\xrightarrow{\frac{x-\mu-\sigma^2 t}{\sigma^2}=z}\left[\frac{x-\mu-\sigma^2 t}{\sigma}=z\rightarrow\ dx=\sigma dz\right]\rightarrow\frac{\exp\left(\frac{\sigma^4 t^2+2\mu\sigma^2 t}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}\sigma\sqrt{2\pi}\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{(z)^2}{2}\right\}dz}_{1}$$

$$M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2 t^2\right]$$

השונות: את השונות נקבל ע״י פונקציה יוצאת מומנטים:

$$M'_t(t)=\left(\mu+\sigma^2 t\right)\exp\left[\mu t+\frac{1}{2}\sigma^2 t^2\right]\xrightarrow{t=0}\mu=E(X)$$

$$M''_t(t)=\sigma^2\exp\left[\mu t+\frac{1}{2}\sigma^2 t^2\right]+\left(\mu+\sigma^2 t\right)^2\exp\left[\mu t+\frac{1}{2}\sigma^2 t^2\right]\xrightarrow{t=0}\sigma^2+\mu^2=E\left(X^2\right)$$

$$V(X)=E\left(X^2\right)-\left(E(X)\right)^2=\sigma^2+\mu^2-\mu^2=\sigma^2$$

השטח מתחת לעקומה:

כמו שאמרנו והוכחנו, השטח מתחת לעקומה הנורמאלית סטנדרטית שווה ל- 1 בדיוק כמו סכום הסתברויות במקרה נתון כלשהו. לכן, אם נרצה למצוא את **השטח** עבורו מתקיים $Z\le 1.5$ אז נוכל לומר כי ברצוננו למצוא את **השכיחות היחסית המצטברת** עבור $Z\le 1.5$ וזה שווה בדיוק ל**הסתברות** עבורה מתקיים $Z\le 1.5$.

נסכם ונאמר כי מטרתנו היא אחת וישנן כמה צורות לבקשה: שטח, השכיחות היחסית המצטברת, אחוזים והסתברות עבור $Z\le 1.5$.

ישנה טבלה מיוחדת הנקראת טבלת התפלגות מצטברת. טבלה זו נותנת לנו את **השטח משמאל לציון התקן החיובי.**

חשוב לדעת כי:

א. השטח מתחת העקומה כולה שווה ל- 1

ב. השטח משאל ל- $z=0$ שווה ל- 0.5 וכך גם השטח מימין ל $z=0$ כיוון שהעקומה סימטרית

ג. עיקר השטח שבין העקומה הנורמאלית לציר ה- Z מרוכז בין $z=3$ ל- $z=-3$ השטחים שמעבר לערכים אלו ניתנים להזנחה

ד. היות וההתפלגות הנורמאלית הינה רציפה אזי השטח עבור $Z\le 1.5$ שווה לשטח $Z<1.5$

ה. אם Z הוא חיובי אז השטח משמאלו גדול מ-0.5 ואם Z שלילי אז השטח משמאלו קטן מ-0.5. גם ההפך נכון:לשטח שגדול מ-0.5 מתאים Z חיובי ולשטח קטן מ-0.5 מתאים Z שלילי.

מציאת שטחים:

נעזר בטבלה הבאה: טבלת התפלגות נורמאלית סטנדרטית המצטברת:

	0.00	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	**0.5319**	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6103	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	**0.8133**
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	**0.8413**	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	**0.9332**	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	**0.9884**	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	**0.9984**	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

1. חשב את השטח משמאל ל- $z=1.5$ ניתן גם לומר מצא את השטח עבורו $Z\le 1.5$.

תשובה: **0.9332**

2. מצא את $p\left(Z \le 2.27\right)$. תשובה: **0.9884**

3. מצא את $p\left(Z \le 1\right)$. תשובה: **0.8413**

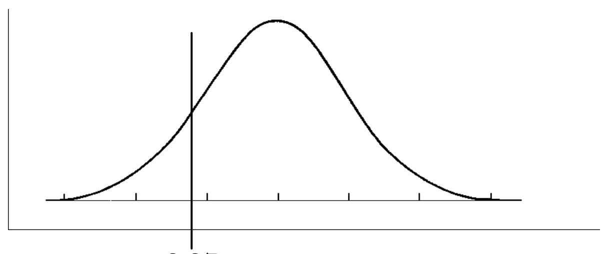
4. מצא את $p\left(Z \le 0.08\right)$. תשובה: **0.5319**

5. מצא את $p\left(Z \le 2.95\right)$. תשובה: **0.9984**

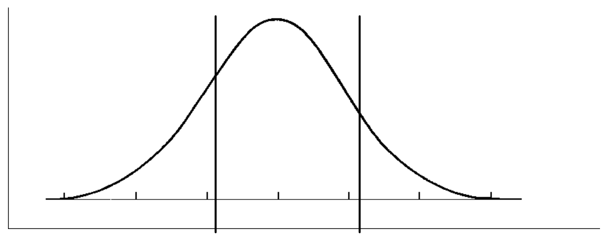
6. מצא את $p\left(Z \ge 2.27\right)$. תשובה: שטח זה אינו מופיע בטבלה כיוון שזה שטח מימין ולא משמאל. ידוע כי שטח תחת כל העקומה שווה ל-1. בנוסף, ידוע כי השטח משמאל לערך זה שווה ל- **0.9884** ולכן השטח המבוקש הינו השטח מימין שווה ל- **1- 0.9884** =0.0116

7. מצא את $p\left(Z \ge 0.08\right)$. תשובה: **1- 0.5319** =0.4681

8. מצא את $p\left(Z \le -2.27\right)$. תשובה: במקרים כאלו רצוי לצייר. לפי הציור נוכל לראות כי התחום המבוקש הינו **1-0.9884**=0.0116



9. חשב את השטח בין $z_1 = -1.32$ ל- $z_2 = 0.89$. או במילים אחרות $p\left(-1.32 \le Z \le 0.89\right)$ תשובה: ראשית נצייר



נוכל לראות כי על מנת למצוא את השטח המבוקש נצטרך תחילה למצוא את השטח משמאל ל- $z_2 = 0.89$ ונוריד ממנו את השטח מימין של $z_1 = -1.32$. השטח משמאל ל- $z_2 = 0.89$ הינו 0.8133. השטח מימין של $z_1 = -1.32$ הינו 0.9066 -1 = 0.0934. כעת, נחסיר: 0.7199=0.8133-0.0934

ההתפלגות הנורמאלית הכללית

היות וקיימת טבלה רק עבור התפלגות נורמאלית סטנדרטית אזי נצטרך להיעזר בנוסחא לציוני תקן על מנת שנוכל לעבוד עם הטבלה הקיימת. נוסחא זו מקשרת בין התפלגות הנורמאלית סטנדרטית להתפלגות נורמאלית כלשהי.

דוגמא: משקל של קבוצת סטודנטים מתפלגים באופן הבא: $X \sim N(73,225)$

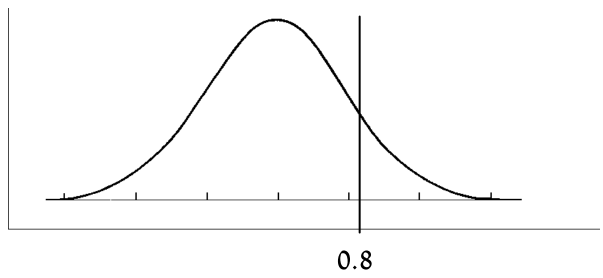
א. מצא את ההסתברות שמשקל של אחד הסטודנטים הינו נמוך מ-85

ב. מצא את ההסתברות שמשקל של אחד הסטודנטים הינו בין 60 ל-85

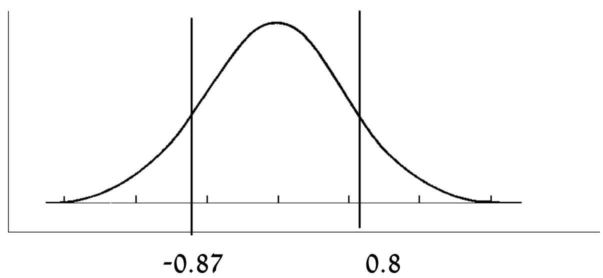
מהנתון אנו רואים שהמשקל מתפלג נורמאלית עם הפרמטרים הבאים $\overline{X}=73$ ו-$s^2=225 \;\rightarrow\; s=15$.

ג. מהו המשקל שרבע מהסטודנטים גבוהים ממנו?

א. נמצא כעת את ציון התקן שמתאים לציון גלם 85 – על מנת להעביר להתפלגות נורמאלית סטנדרטית זאת כדי להשתמש בטבלה. $z_{85}=\frac{85-73}{15}=\frac{12}{15}=0.8$. נרצה הסתברות לציון נמוך מזה ולכן נרצה את השטח משמאל ל-0.8 שטח זה הינו – 0.7881. מבחינת הגרף:



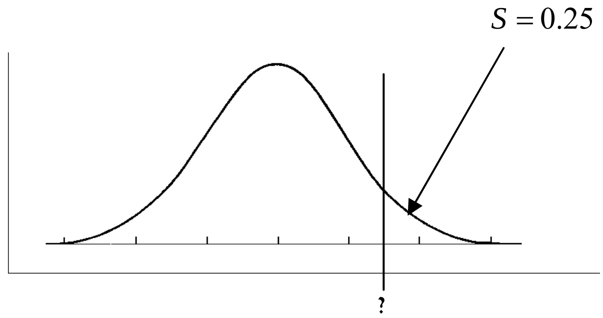
ב. נמצא את ציון התקן של הציון גלם 60: $z_{60}=\frac{60-73}{15}=\frac{-13}{15}=-0.867\approx-0.87$



נחסר מהשטח משמאל של ציון התקן 0.8 את השטח משמאל של ‎-0.87‎. השטח משמאל ל-0.8 הינו – 0.7881. השטח משמאל ל- ‎-0.87‎ הינו : 0.1922=1-0.8078. כעת נחסר : 0.5959 =0.7881-0.1922

ג. במקרה זה מחפשים ציון תקן ולא שטח. השטח נתון בשאלה, מחפשים ציון תקן כך שהשטח משמאל הינו 0.75. שטח זה גדול מ-0.5 ולכן אנו מחפשים ציון תקן חיובי.

נחפש את הערך כך ששטח מימין לערך זה הינו 0.25



כעת, נחפש בתוך הטבלה את הערך הקרוב ביותר ל-0.75 ערך זה הינו 0.68 וזה ציון התקן. עכשיו נוכל למצוא את ציון הגלם המתאים לו :

$z = \frac{x-73}{15} = 0.68 \to = 83.2$. וזהו בדיוק המשקל שמעליו רבע מהמשקלים ומתחתיו שלושת רבעים. במילים אחרות זהו הרבעון העליון.

שאלה ממבחן : נתון כי קבוצת תצפיות מתפלגת נורמאלית עם ממוצע m וסטית תקן s , מצא קבוע a כך שאחוז התצפיות,

א. בתוך הטווח $m \pm a \cdot s$ היא 75%

ב. הקטנה מ $m - a \cdot s$ היא 22%

פתרון : נרצה לחשב, $p\left(m - as \le X \le m + as\right) = 0.75$ אך תחילה נעשה תקנון $Z = \frac{X-m}{s}$ אנו עושים זאת כי אנו יודעים לעבוד רק עם התפלגות נורמאלית סטנדרטית.

$$p\left(\frac{m-as-m}{s} \le \frac{X-m}{s} \le \frac{m+as-m}{s}\right) = p\left(-a \le Z \le a\right) \to \Phi(a) - \Phi(-a) =$$

$$\Phi(a) - \left(\underbrace{1-\Phi(a)}_{\Phi(-a)}\right) = 2\Phi(a) - 1 \to 2\Phi(a) - 1 = 0.75 \to \Phi(a) = 0.875 \to a = 1.15$$

ב.

$$p\left(X \le m-as\right)=0.22 \xrightarrow{Z=\frac{X-m}{s}} p\left(\frac{X-m}{s} \le \frac{m-as-m}{s}\right)=0.22$$

$$p\left(Z \le -a\right)=\Phi\left(-a\right)=1-\Phi\left(a\right) \to \Phi\left(-a\right)=0.78 \to a=0.77$$

תוחלת: נוכיח כי התוחלת שווה ל- μ (לא ע"י פונקציה יוצאת מומנטים):

$$E\left(X\right)=\int_{-\infty}^{\infty} xf\left(x\right)dx=\int_{-\infty}^{\infty}\frac{x}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right\}dx \xrightarrow[x=\sigma y+\mu]{\frac{x-\mu}{\sigma}=y \to \frac{1}{\sigma}=\frac{dy}{dx} \to dx=\sigma dy}$$

$$\int_{-\infty}^{\infty}\frac{\sigma y+\mu}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}dy \to \underbrace{\int_{-\infty}^{\infty}\frac{\sigma y}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}dy}_{A}+\underbrace{\int_{-\infty}^{\infty}\frac{\mu}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}dy}_{B}$$

$$A:\frac{\sigma}{\sqrt{2\pi}}\int_{-\infty}^{\infty} y\exp\left\{-\frac{y^2}{2}\right\}dy \xrightarrow{\frac{y^2}{2}=t \quad \to dy=\frac{dt}{y}} \frac{\sigma}{\sqrt{2\pi}}\int e^{-t}dt=\frac{\sigma}{\sqrt{2\pi}}e^{-t} \to \left(\frac{\sigma}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}\right)_{-\infty}^{\infty}=0$$

$$B=\int_{-\infty}^{\infty}\frac{\mu}{\sqrt{2\pi}}\exp\left\{-\frac{y^2}{2}\right\}dy=\frac{\mu}{\sqrt{2\pi}}\underbrace{\int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}dy}_{\sqrt{2\pi}}=\mu$$

$$A+B=\mu$$

שאלה ממבחן: יהיו $Y \sim N\left(\mu_Y,\sigma_Y^2\right)$ ו $X \sim N\left(\mu_X,\sigma_X^2\right)$ מ.מ בלתי תלויים. הוכח כי $X+Y$ גם מתפלג נורמאלית.

פתרון: נבדוק האם הפונקציה יוצרת מומנטים של $X+Y$ היא פונקציה יוצרת של ההתפלגות הנורמאלית: $E\left(e^{t(X+Y)}\right)=E\left(e^{tX}e^{tY}\right)$ בגלל שנתון כי המ.מ-ים בלתי תלויים ניתן להפריד $E\left(e^{t(X+Y)}\right)=E\left(e^{tX}\right)E\left(e^{tY}\right)$. כעת, כל מ.מ מתפלג נורמאלית ולכן ניתן להציב את הפונקציה היוצרת המתאימה לו:

$$E\left(e^{t(X+Y)}\right)=E\left(e^{tX}\right)E\left(e^{tY}\right)\xrightarrow{M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2t^2\right]}\exp\left[\mu_X t+\frac{1}{2}\sigma_X^2t^2\right]\exp\left[\mu_Y t+\frac{1}{2}\sigma_Y^2t^2\right]$$

$$\exp\left[\left(\mu_X+\mu_Y\right)t+\frac{1}{2}\left(\sigma_X^2+\sigma_Y^2\right)t^2\right]$$

קבלנו התפלגות נורמאלית עם תוחלת $\mu_X+\mu_Y$ ושונות $\sigma_X^2+\sigma_Y^2$. ז"א בסה"כ $X+Y \sim N\left(\mu_X+\mu_Y,\sigma_X^2+\sigma_Y^2\right)$.

התפלגות וייבול – הרחבה של התפלגות מעריכית

מ.מ X המתפלג וייבול מסומן $X \sim \text{Weibull}(a,b)$. פונקציית הצפיפות מוגדרת באופן הבא: $f(x) = abx^{b-1}e^{-ax^b}$ כאשר $a>0, b>0$ ו $0<x<\infty$.

התפלגות מעריכית היא מקרה פרטי של התפלגות ויבול, מ.מ שמתפלג וייבול עם הפרמטרים $X \sim Weibull(\lambda,1)$ הינו מ.מ $X \sim \exp(\lambda)$.

תוחלת: $E(X) = a^{\frac{-1}{b}}\Gamma\left(1+b^{-1}\right)$

שונות: $V(X) = a^{\frac{-2}{b}}\left[\Gamma\left(1+2b^{-1}\right) - \Gamma^2\left(1+b^{-1}\right)\right]$

התפלגות גמא:

יהי מ.מ X המתפלג גמא נסמנו $X \sim Gamma(r,\lambda)$. פונקצית הצפיפות הינה:

$\lambda>0, r>0$ $f_X(x) = \frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$ כאשר $x>0$ ו $\Gamma(r) = \int_0^\infty x^{r-1}e^{-x}dx$. ומתקיים עבור r שלם $\Gamma(r) = (r-1)!$ ו $\Gamma(r) = \Gamma(r-1)\cdot(r-1)$

פונקציה יוצרת מומנטים: $M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r \quad for \quad t<\lambda$

התוחלת: $E(X) = \frac{r}{\lambda}$

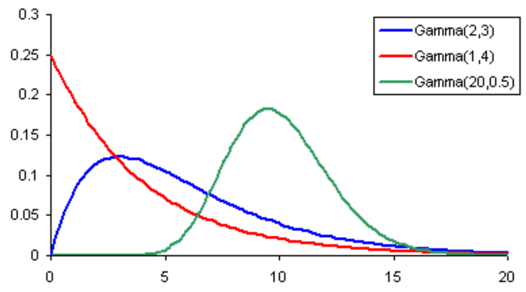
הוכחה: $M'_X(t) = E(X) = r\left(\frac{\lambda}{\lambda-t}\right)^{r-1}\frac{\lambda}{(\lambda-t)^2} = \frac{r\lambda^r}{(\lambda-t)^{r+1}} \xrightarrow{t=0} \frac{r}{\lambda}$

השונות: $V(X) = \frac{1}{\lambda^2}$

הוכחה: תחילה נמצא את המומנט השני $M''_X(t) = E\left(X^2\right) = \frac{r\lambda^r(r+1)}{(\lambda-t)^{r+2}} \xrightarrow{t=0} \frac{r(r+1)}{\lambda^2}$

ולכן, $V(X) = E\left(X^2\right) - \left(E(X)\right)^2 = \frac{r(r+1)}{\lambda^2} - \left(\frac{r}{\lambda}\right)^2 = \frac{r}{\lambda^2}$

חשוב להדגיש כי עבור $r=1$ אז התפלגות גמא שווה להתפלגות מעריכית.



טענה: יהיו $X_1, X_2, X_3, \dots X_r \sim Gamma(1,\lambda)$ ב"ת אזי $X_1+X_2+X_3+\dots+X_r \sim Gamma(r,\lambda)$.

הוכחה: ע"י פונקציה יוצרת מומנטים $M_{\sum_i^r X_i^2}(t) = E\left(e^{t(X_1+X_2+X_3+\dots+X_r)}\right)$ בגלל שהם ב"ת אזי $E\left(e^{t(X_1+X_2+X_3+\dots+X_r)}\right) = \prod_{i=1}^{r} E\left(e^{tX_i}\right)$ ובגלל שנתון שהם שווי התפלגות אזי $\prod_{i=1}^{r} E\left(e^{tX_i}\right) = \prod_{i=1}^{r} \frac{\lambda}{\lambda - t} = \left(\frac{\lambda}{\lambda - t}\right)^r$. בסה"כ קבלנו פונקציה יוצרת מומנטים של התפלגות גמא, משמע, שאם סוכמים מ.מ-ים המתפלגים מעריכית עם פרמטר λ אזי התפלגותם היא גמא עם פרמטרים r ו- λ.

התפלגות ביתא:

יהי מ.מ X המתפלג ביתא נסמנו $X \sim Beta(a,b)$. פונקציית הצפיפות הינה:

$f_X(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}$ כאשר $a,b>0$ $\quad 0 \le x \le 1$ ו $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$.

ומתקיים $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

פונקציה יוצרת מומנטים: אין שימוש

התוחלת: $E(X) = \frac{a}{a+b}$

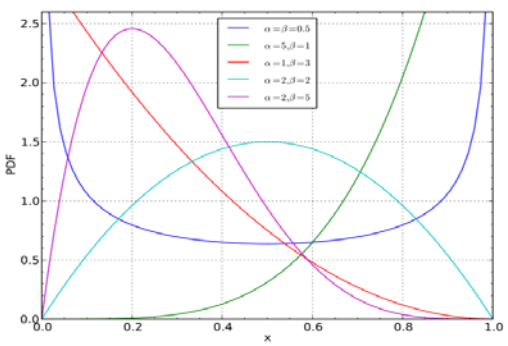
הוכחה:

$$E(X) = \int_0^1 \frac{x}{B(a,b)} x^{a-1}(1-x)^{b-1} = \frac{1}{B(a,b)} \underbrace{\int_0^1 x^a (1-x)^{b-1}}_{B(a+1,b)} = \frac{B(a+1,b)}{B(a,b)}$$

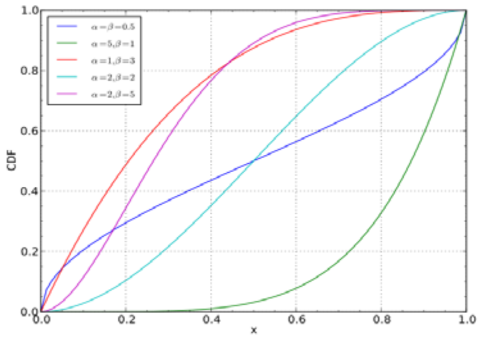
$$= \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} \Big/ \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \frac{a}{a+b}$$

השונות: $V(X) = \frac{ab}{(a+b+1)(a+b)^2}$

פונקציית הצפיפות



פונקצית ההצטברות



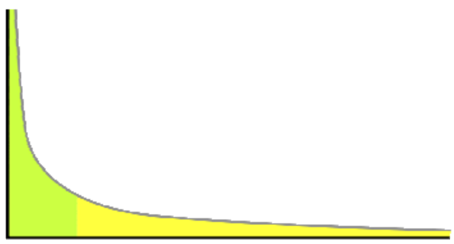
התפלגות פארטו

עקרון פארטו, הידוע גם בשם כלל 80-20, הוא כלל שנוסח על ידי הכלכלן האיטלקי וילפרדו פארטו, וגורס כי בתופעות רבות 80% מהפעילות מקורם ב-20% מהגורמים הפעילים. הכלל מאפשר הבחנה בין הגורמים העיקריים (המהווים 20% מכלל הגורמים) לגורמים הטפלים (יתר 80% מהגורמים). פארטו, שעסק בחקר פערים כלכליים, שם לב שכ-80% מהעושר באיטליה נמצאים בבעלות כ-20% מהאוכלוסייה.

מ.מ X המתפלג פארטו מסומן $\mathrm{X} \sim \mathrm{pareto}\left(\mathrm{x}_0, \theta\right)$. פונקציית הצפיפות מוגדרת באופן הבא: $f\left(x\right)=\dfrac{\theta x_0^{\theta}}{x^{\theta+1}}$ כאשר $x_0 > 0, \theta > 0$ ו $x_0 < x < \infty$.

תוחלת: $E\left(X\right)=\dfrac{\theta x_0^{\theta}}{\theta-1} \quad for \quad \theta > 1$

שונות: $V\left(X\right)=\dfrac{\theta x_0^{2}}{\left(\theta-1\right)^2\left(\theta-2\right)} \quad for \quad \theta > 2$



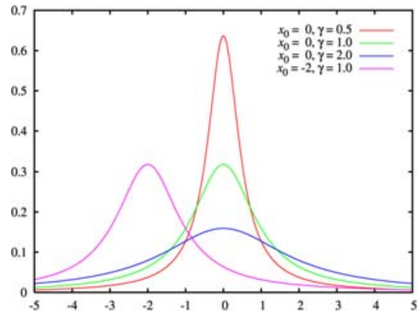
התפלגות קושי

התפלגות קושי מוגדרת כהתפלגות רציפה בעלת פונקציית צפיפות :

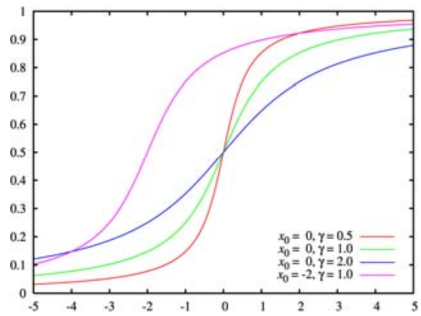
$f(x)=\dfrac{1}{\pi\gamma\left[1+\left(\dfrac{x-x_0}{\gamma}\right)^2\right]}$ כאשר $\gamma>0$, $-\infty<x_0<\infty$. x_0 הינו פרמטר מיקום אשר קובע את החציון של ההתפלגות. γ הינו פרמטר הקובע את רוחב ההתפלגות.

להתפלגות זו לא קיימים התוחלת והשונות.

פונקצית הצפיפות :



פונקצית ההצטברות



אי שוויון צ'בישב וחוק המספרים הגדולים:

ראשית נסביר את הרעיון הכללי של אי שוויון זה : במילים פשוטות, ככל שסטיית תקן של מ.מ X קטנה כל תגדל ההסתברות כי הערכים שהמשתנה המקרי זה יקבל הינם קרובים לתוחלת. כמובן שאת המרחק של התוחלת מהערך של המ.מ אנו מודדים בסטיות תקן.

משפט – אי שוויון צ׳בישב:

לכל מ.מ X בעל תוחלת μ ושונות σ^2 ולכל מספר חיובי k מתקיים

$$p\left(\mu-k\sigma<X<\mu+k\sigma\right)\geq 1-\frac{1}{k^2}$$

ברור כי ככל ש k גדל כך גדל הרווח, עוד ניתן לומר כי ההסתברות שימצא ערך x מסוים גדלה כיוון שהאורך רווח גדל.

דוגמא בסיסית: נניח כי $k=1.5$ אזי נקבל $p\left(\mu-1.5\sigma<X<\mu+1.5\sigma\right)\geq 1-\frac{1}{1.5^2}\approx 0.56$. משמע שהסיכוי ש- X יקבל ערך שריחוקו מהתוחלת ב-1.5 סטיות תקן, יהיה לפחות 0.56. ניתן גם לראות זאת באופן גרפי על ידי כך שהשטח (השטח הינו ההסתברות) בתחום $\left(\mu-1.5\sigma,\mu+1.5\sigma\right)$ יהי לפחות 0.56. ברור כי ככל שניקח k גדול אזי ההסתברות תגדל.

אי שוויון צ׳בישב נכון לכל מ.מ, בדיד ורציף. נוכיח עבור מ.מ בדיד:

ראשית נכתוב את האי שיווין באופן הבא

$$p(|X-\mu|<k\sigma)\geq 1-\frac{1}{k^2}\ \rightarrow\ p\left(\left(X-\mu\right)^2<k^2\sigma^2\right)\geq 1-\frac{1}{k^2}\xrightarrow{\text{complete}}p\left(\left(X-\mu\right)^2\geq k^2\sigma^2\right)\leq\frac{1}{k^2}$$

מספיק להוכיח את האי שיווין האחרון ובכך סיימנו.

במקום σ^2 נוכל לכתוב $\sigma^2=E\left(\left(X-\mu\right)^2\right)\rightarrow\sigma^2=\sum_i p_i\cdot\left(x_i-\mu\right)^2$ כעת, ננסה להקטין את הסכום $\sum_i p_i\cdot\left(x_i-\mu\right)^2$ על ידי הפחתה של כל התצפיות עבורם מתקיים $\left(x_i-\mu\right)^2<k^2\sigma^2$ ולכן נקבל $\sigma^2\geq\sum_i{}^{*}p_i\cdot\left(x_i-\mu\right)^2$ כאשר האיברים $\left(x_i-\mu\right)^2$ מקיימים $\left(x_i-\mu\right)^2\geq k^2\sigma^2$: (אם השמטנו את כל המקיימים $\left(x_i-\mu\right)^2<k^2\sigma^2$ אזי נשארנו עם האיברים המקיימים $\left(x_i-\mu\right)^2\geq k^2\sigma^2$) אם ננסה לסכם מה קבלנו:

$\sigma^2\geq\sum_i{}^{*}p_i\cdot\left(x_i-\mu\right)^2\xrightarrow{\left(x_i-\mu\right)^2\geq k^2\sigma^2}\geq\sum_i{}^{*}p_i\cdot k^2\sigma^2=k^2\sigma^2\sum_i{}^{*}p_i$, כאשר מתקיים $\sum_i{}^{*}p_i=p\left(\left(X-\mu\right)^2\geq k^2\sigma^2\right)$, כעת נציב ב $\sigma^2\geq k^2\sigma^2\sum_i{}^{*}p_i$ ונקבל:

$$\sigma^2\geq k^2\sigma^2\sum_i{}^{*}p_i\xrightarrow{\sum_i{}^{*}p_i=p\left(\left(X-\mu\right)^2\geq k^2\sigma^2\right)}\sigma^2\geq k^2\sigma^2 p\left(\left(X-\mu\right)^2\geq k^2\sigma^2\right)$$

$$\frac{\sigma^2}{k^2\sigma^2}\geq p\left(\left(X-\mu\right)^2\geq k^2\sigma^2\right)\rightarrow\frac{1}{k^2}\geq p\left(\left(X-\mu\right)^2\geq k^2\sigma^2\right)$$

התפלגות דגימה:

הקדמה: תהליך הדגימה הינו תהליך מקרי וכל אחת מהתוצאות הינה משתנה מקרי. כמו כן ברור כי אם נבצע דגימות נוספות נקבל תוצאות מקריות אחרות. למשל, כאשר נחזור על פעולת הדגימה נקבל מדגמים שונים: אם נדגום באופן מקרי 5 ציונים של סטודנטים מהאוניברסיטה נקבל מדגם מסוים וכאשר נבצע דגימה חדשה של 5 ציוניי סטודנטים נקבל ערכים אחרים וכך הלאה.

כמה הגדרות:

1. סטטיסטי: מדד במדגם נקרא סטטיסטי, כל סטטיסטי הינו פונקציה של תצפיות המדגם ולכן מעצם הגדרתו ברור כי אם נבצע דגימות שונות אזי הסטטיסטיים יהיו שונים בערכם ממדגם למדגם. למשל: ממוצע הינו סטטיסטי כיוון שהוא פונקציה של התצפיות. נחזור לרגע לדוגמא עבור דגימת ציונים של 5 סטודנטים. ברור כי עבור כל מדגם נקבל ממוצע שונה. באותו אופן נקבל גם סטיית תקן שונה, ערך מינימאלי או מקסימאלי שונים וכו.
2. פרמטר: מדד **באוכלוסיה** נקרא פרמטר. פרמטר הינו קבוע ולא משתנה כמו סטטיסטי. למשל: תוחלת- ממוצע האוכלוסיה מסומנת ב- μ, שונות האוכלוסיה מסומנת ב- σ^2 ועוד.

מטרתנו העיקרית: להסיק ממדגם עבור האוכלוסיה כולה. אנו עושים זאת בעזרת הסטטיסטיים ובעזרתם ללמוד על הפרמטרים.

התפלגות דגימה של סטטיסטי מסוים היא פונקצית התפלגות שלו. התפלגות זו תלויה בצורתו המתמטית של הסטטיסטי, בגודל המדגם ובתכונות האוכלוסיה ממנה נדגם.

נחזור לדוגמא עבור דגימת ציונים של 5 סטודנטים. ברור כי אם נדגום 100 פעמים חמישה סטודנטים מתוך האוכלוסיה נקבל 100 ממוצעים (סביר להניח שהם שונים בערכם) ממוצעים אלו הם גם משתנים מקריים. נוכל להסתכל על ההתפלגות של ממוצעים אלו.

עבור הממוצע נקבל כי תוחלת ממוצע המדגם שווה לתוחלת המשתנה המקרי שממנו דוגמים.

נוכיח: $E\left(\overline{X}\right)=E\left(\frac{\sum_{i=1}^{n}X_i}{n}\right)=\frac{1}{n}E\left(\sum_{i=1}^{n}X_i\right)$ ובגלל שהם שווי התפלגות:

$$\frac{1}{n}E\left(\sum_{i=1}^{n}X_i\right)=\frac{1}{n}n\cdot E\left(X\right)=E\left(X\right)$$

במילים פשוטות יותר נוכל לומר כי ממוצע של הממוצעים של כל המדגמים האפשריים הינו בדיוק ממוצע האוכלוסיה זאת ללא תלות בגודל המדגם.

עבור סטיית התקן של ממוצע המדגמים: בפועל אין אנו דוגמים הרבה פעמים, אלא אנו מעוניינים לדעת עד כמה הממוצע שלנו במדגם שלנו הספציפי רחוק ממוצע

האוכלוסיה כולה או במילים אחרות אנו רוצים לדעת את מידת הפיזור של הממוצעים ממוצע הממוצעים. פיזור זה שווה : $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$.

נוכיח : $V\left(\bar{X}\right)=\sigma_{\bar{X}}^2 = V\left(\frac{\sum_{i=1}^{n}\left(X_i\right)}{n}\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n}X_i\right)$, בגלל שהמ.מ-ים שווי התפלגות אזי $\sigma_{\bar{X}}^2 = V\left(\bar{X}\right) = \frac{1}{n^2}\left(\sum_{i=1}^{n}V\left(X\right)\right) = \frac{1}{n^2}n\cdot V\left(X\right) = \frac{V\left(X\right)}{n} = \frac{\sigma_X^2}{n}$

משמעות הדבר כי שונות זאת שווה לשונות האוכלוסייה מחולקת בגודל המדגם (מספר הממוצעים שקיימים). נוכל להסיק כי שונות המדגם קטנה יותר ככל ש- n גדול יותר.

משפט הגבול המרכזי – דגימה מתוך התפלגות כללית:

עד כה דברנו על התפלגות דגימה של $\bar{X}$ והסקנו כמה דברים :

1. תוחלת התפלגות הדגימה שווה לתוחלת של האוכלוסייה (מסמנים זאת כך : $E\left(\bar{X}\right)=E\left(X\right)$)
2. שונות התפלגות הדגימה של הממוצע שווה לשונות האוכלוסייה חלקי גודל המדגם : $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$
3. אם האוכלוסייה מתפלגת נורמאלית אז גם התפלגות הדגימה עבור הממוצע גם מתפלגת נורמאלית באופן הבא : אם $X \sim N\left(\mu,\sigma^2\right)$ אזי $\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right)$ ההתפלגות הנורמאלית של הממוצע נובעת מהעובדה כי דגמנו מאוכלוסייה שמתפלגת נורמאלית,אך נוכל להרחיב ולומר דבר חשוב המוביל למשפט הגבול המרכזי.

ניסוחו המתמטי של משפט הגבול המרכזי:

יהי X משתנה מקרי כלשהו בעל תוחלת μ וסטיית תקן σ , ויהיו $X_1, X_2, \dots X_n$ משתנים מקריים בלתי תלויים שלכל אחד מהם התפלגות כמו זו של X אזי **ההתפלגות של הממוצע** $\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}$ שואפת להתפלגות נורמאלית עם תוחלת μ וסטיית תקן $\frac{\sigma}{\sqrt{n}}$. זאת **כאשר** n **שואף ל-** ∞.

נוכיח: נעשה זאת ע"י פונקציה יוצרת מומנטים: נגדיר את המשתנים Z ו- Y להיות

$Y=\frac{X-\mu}{\sigma}$ ואת $Z=\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}}$. ברור כי למ.מ Y יש תוחלת השווה ל-0 ושונות השווה ל-1.

נראה זאת: לפי המשפט, X משתנה מקרי כלשהו בעל תוחלת μ וסטיית תקן σ , ויהיו

$X_1, X_2, \ldots X_n$ משתנים מקריים בלתי תלויים שלכל אחד מהם התפלגות כמו זו של X

לכן Y הינו המשתנה המתוקנן (כידוע טרנספורמציה ליניארית של תקנון מניבה תוחלת השווה ל-0 ושונות השווה ל-1 – הוכחנו זאת בתחילת הסמסטר).

נשים לב כי איננו יודעים את התפלגות המשתנים!!

$E\left(e^{tZ}\right)=E\left(e^{t\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}}}\right)=E\left(e^{\frac{t(Y_1+Y_2+\ldots+Y_n)}{\sqrt{n}}}\right)=E\left[\prod_{i=1}^{n}\left(e^{\frac{ty_i}{\sqrt{n}}}\right)\right]$ בגלל שהמ.מ-ים ב"ת אזי,

$E\left(e^{tZ}\right)=E\left[\prod_{i=1}^{n}\left(e^{\frac{tY_i}{\sqrt{n}}}\right)\right]=\prod_{i=1}^{n} E\left(e^{\frac{tY_i}{\sqrt{n}}}\right)$ בגלל שהמ.מ-ים שווי התפלגות אזי

$E\left(e^{tZ}\right)=\prod_{i=1}^{n} E\left(e^{\frac{tY_i}{\sqrt{n}}}\right)=\left[E\left(e^{\frac{tY}{\sqrt{n}}}\right)\right]^n$ את הביטוי $e^{\frac{tX}{\sqrt{n}}}$ נפתח לפי טור מקלורן עד האיבר

השלישי כיוון שכאשר n שואף לאינסוף אזי שאר האיברים שואפים לאפס.

$$E\left(e^{tZ}\right)=\left[E\left(e^{\frac{tY}{\sqrt{n}}}\right)\right]^n=\left(1+\frac{tE(Y)}{\sqrt{n}}+\frac{t^2}{2n}E\left(Y^2\right)\right)^n \xrightarrow{E(Y)=0\ ,\ \ V(Y)=1\ \ ,\ \ E\left(Y^2\right)=V(Y)+(E(Y))^2=1}$$

$$\left[1+0+\frac{t^2}{2n}\cdot 1\right]^n \xrightarrow{e^{\ln a}=a\ \ ,\ \ \ln a^n=n\ln a} e^{n\ln\left[1+\frac{t^2}{2n}\right]}$$

כעת, נבדוק למה שואפת החזקה כאשר נשאיף את n לאינסוף:

$$\lim_{n\to\infty} n\ln\left[1+\frac{t^2}{2n}\right]=\lim_{n\to\infty}\frac{\ln\left[1+\frac{t^2}{2n}\right]}{\frac{1}{n}} \xrightarrow{lopital} \lim_{n\to\infty}\frac{\frac{1}{\left[1+\frac{t^2}{2n}\right]}\left(-\frac{t^2}{2n^2}\right)}{\frac{-1}{n^2}}=\frac{t^2}{2}$$

ובסה"כ: $E\left(e^{tZ}\right)=e^{\frac{t^2}{2}}$ שזה פונקציה יוצרת מומנטים של התפלגות נורמאלית סטנדרטית.

מכאן נוכל להסיק כי לפי הגדרת Z:

$$Z=\frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}} \xrightarrow{Y=\frac{X-\mu}{\sigma}} Z=\frac{\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma}}{\sqrt{n}}=\frac{\sum_{i=1}^{n} X_i-n\mu}{\sqrt{n}\sigma}=\frac{n\left(\frac{\sum_{i=1}^{n} X_i}{n}-\mu\right)}{\sqrt{n}\sigma}=\frac{n\left(\bar{X}-\mu\right)}{\sqrt{n}\sigma}=\frac{\left(\bar{X}-\mu\right)}{\sigma/\sqrt{n}}\ .$$

נעביר אגפים ונקבל כי $Z=\frac{\left(\bar{X}-\mu\right)}{\sigma/\sqrt{n}} \to \bar{X}=\mu+Z\cdot\sigma/\sqrt{n}$ כעת, בגלל שהסקנו כי ל Z יש

התפלגות נורמאלית סטנדרטית נוכל לומר כי $\bar{X}\sim N\left(\mu,\sigma^2/n\right) \rightarrow \bar{X}\sim \mu+N(0,1)\cdot\sigma/\sqrt{n} \rightarrow \bar{X}=\mu+Z\cdot\sigma/\sqrt{n}$ המעברים האחרונים נכונים כי מדובר בטרנספורמציה ליניארית, כאשר מבצעים טרנספורמציה ליניארית על משתנה, אזי התוחלת שלו משתנה בהתאם. אך השונות מושפעת רק מפעולת כפל.

בסה"כ, מתברר כי גם אם לא נדגום מאוכלוסייה נורמאלית, התפלגות הדגימה של הממוצע תשאף לנורמאליות, **אם** נבחר גודל מדגם גדול במידה מספקת. פירוש הדבר כי אם נדגום מאוכלוסייה שמתפלגת בצורה כלשהי, המון פעמים ובכל פעם נחשב את $\bar{X}$, הרי שכאשר נערוך את כל התוצאות של $\bar{X}$ בהסטוגרמה אזי צורתה תהיה קרובה מאוד להתפלגות נורמאלית.

דגימה מהתפלגות נורמאלית:

כשאנו אומרים דגימה מתפלגות נורמאלית, אנו מתכוונים כי ידוע שהאוכלוסייה מתפלגת נורמאלית עם תוחלת μ ושונות σ^2. במקרה כזה ידוע כי אם נבצע התפלגות דגימה ונחשב כל פעם את הממוצע של המדגם אזי התפלגות הממוצע תהיה גם נורמאלית כאשר התוחלת שווה μ ושונות $\frac{\sigma^2}{n}$. נסכם:

* אם $X\sim N\left(\mu,\sigma^2\right)$ אזי $\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$

* אם $X\sim N\left(\mu,\sigma^2\right)$ אזי $Z_x\sim N(0,1)$ כאשר $Z_x=\frac{X-\mu}{\sigma}\sim N(0,1)$ (טרנספורמצית תקנון)

באופן דומה נוכל לומר כי אם $\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$ אזי $Z_{\bar{X}}=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$.

תרגילים כלליים:

1. יהי X מ.מ בעל פונקצית צפיפות $f(x)=c\left(1-x^2\right),\ 0<x<1$

א. מצא את c

ב. מצא את פונקצית ההצטברות

ג. מצא את התוחלת של X

ד. מצאו את ההסתברות ש $p(X\le 0.5)$

ה. האם החציון של X גדול או קטן מ 0.5?

פתרון: א. בגלל שמדובר בפונקצית צפיפות אזי האינטגרל על כל התחום שווה ל-1:

$$\int_0^1 c\left(1-x^2\right)dx=1 \rightarrow c\left(x-\frac{x^3}{3}\right)_0^1=\frac{2c}{3} \rightarrow \frac{2c}{3}=1 \rightarrow c=\frac{3}{2}$$

ב. $F(x)=\int_0^x \frac{3}{2}\left(1-x^2\right)dx \rightarrow \frac{3}{2}\left(x-\frac{x^3}{3}\right)_0^x=\frac{3x}{2}-\frac{x^3}{2}$ ולכן נגדירה באופן מצא:

$$F(x)=\begin{cases}0 & x<0\\ \dfrac{3x}{2}-\dfrac{x^3}{2} & 0\le x\le 1\\ 1 & x>1\end{cases}$$

ג. התוחלת של X : $E(X)=\int_0^1 \frac{3}{2}x\left(1-x^2\right)dx=1 \rightarrow \frac{3}{2}\left(\frac{x^2}{2}-\frac{x^4}{4}\right)_0^1=\frac{3}{8}$

ד. $p(X\le 0.5)\rightarrow F(0.5)=\frac{3\cdot 0.5}{2}-\frac{(0.5)^3}{2}=\frac{11}{16}=0.6875$

ה. לפי הסעיף הקודם $\frac{11}{16}>\frac{1}{2}$ ולכן ההסתברות ש $X\le 0.5$ גדולה מחצי ולכם החציון קטן מ0.5.

2. יהי X מ.מ בעל פונקצית צפיפות $|x|<a$, $f(x)=\frac{c}{\sqrt{a^2-x^2}}$ כאשר a הינו פרמטר חיובי.

א. מצא את הקבוע c

ב. מצא את פונקצית ההצטברות של X

ג. חשב את ההסתברויות $p(0<X<a)$ ו- $p\left(|X|\le\frac{a}{2}\right)$

פתרון : א. בגלל שמדובר בפונקצית צפיפות אזי האינטגרל על כל התחום שווה ל-1 :

$$\int_{-a}^{a}\frac{c}{\sqrt{a^2-x^2}}dx=1 \xrightarrow{\int\frac{1}{\sqrt{a^2-x^2}}dx=\arcsin\frac{x}{a}+c} c\left(\arcsin\frac{x}{a}\right)_{-a}^{a}=c\left(\arcsin 1-\arcsin(-1)\right)$$

$$\xrightarrow{\arcsin(-1)=-\arcsin 1} c\left(2\arcsin 1\right)\xrightarrow{\arcsin 1=\frac{\pi}{2}}=c\pi=1\rightarrow c=\frac{1}{\pi}$$

ב. פונקצית ההצטברות :

$$F(x)=\frac{1}{\pi}\int_{-a}^{x}\frac{1}{\sqrt{a^2-x^2}}dx=\frac{1}{\pi}\left(\arcsin\frac{x}{a}\right)_{-a}^{x}=\frac{1}{\pi}\left(\arcsin\frac{x}{a}-\arcsin(-1)\right)$$

$$\xrightarrow{\arcsin(-1)=-\arcsin 1\ ,\ \arcsin 1=\frac{\pi}{2}}\frac{1}{\pi}\left(\arcsin\frac{x}{a}+\frac{\pi}{2}\right)=\frac{1}{\pi}\arcsin\frac{x}{a}+\frac{1}{2}$$

ובסה"כ נקבל : $$F(x)=\begin{cases}0 & x<-a\\ \dfrac{1}{\pi}\arcsin\dfrac{x}{a}+\dfrac{1}{2} & -a\le x\le a\\ 1 & x>a\end{cases}$$

ג. $p\left(0<X<a\right)=F\left(a\right)-F\left(0\right)=\frac{1}{\pi}\underbrace{\arcsin\frac{a}{a}}_{\frac{\pi}{2}}+\frac{1}{2}-\left(\frac{1}{\pi}\underbrace{\arcsin\frac{0}{a}}_{0}+\frac{1}{2}\right)=\frac{1}{2}$

$$p\left(|X|\le\frac{a}{2}\right)=p\left(-\frac{a}{2}\le X\le\frac{a}{2}\right)=F\left(-\frac{a}{2}\right)-F\left(\frac{a}{2}\right)=\frac{1}{\pi}\underbrace{\arcsin\frac{-1}{2}}_{\frac{\pi}{6}}+\frac{1}{2}-\left(\frac{1}{\pi}\underbrace{\arcsin\frac{1}{2}}_{\frac{\pi}{6}}+\frac{1}{2}\right)=\frac{1}{3}$$

3. התפלגותו של מ.מ מסוים היא בעלת תוחלת השווה ל 90 ושונות 0.25. השתמש באי שוויון צ'בישב על מנת למצוא את התחום בו נמצאים לפחות 0.7 מהתצפיות

פתרון: אנו מעוניינים לפחות בהסתברות של 0.7 ולכן $1-\frac{1}{k^2}=0.7\rightarrow k=1.83$ לפי אי שיוון צ'בישב $p\left(\mu-k\sigma<X<\mu+k\sigma\right)\ge1-\frac{1}{k^2}\xrightarrow{k=1.83}p\left(\mu-1.83\sigma<X<\mu+1.83\sigma\right)\ge0.7$ כעת, נציב את הנתונים ונקבל: $p\left(89.1<X<91\right)\ge0.7$ על-ידי החישוב הבא:

$$p\left(\mu-1.83\sigma<X<\mu+1.83\sigma\right)\ge0.7\xrightarrow{\mu=90\ ,\ \sigma=0.5}p\left(90-1.83\cdot0.5<X<90+1.83\cdot0.5\right)\ge0.7$$

משמע, לפחות 0.7 מהתצפיות נמצאות בתחום שבין 89.1 ל-91.

4. יהי X מ.מ בעל פונקצית ההצטברות

$$F\left(x\right)=\begin{cases}0 & x\le c\\ 1-\left(\frac{c}{x}\right)^a & x>c\end{cases}$$

כאשר a ו- c הם פרמטרים חיוביים.

א. מצא את פונקצית הצפיפות של X.

ב. מצאת את התוחלת בהנחה ש $a>1$

ג. מצאת את התוחלת בהנחה ש $a<1$

פתרון: א: נגזור, $\left(F\left(x\right)\right)'=f\left(x\right)=\begin{cases}0 & x\le c\\ ac^ax^{-a-1} & x>c\end{cases}$

ב. $E\left(X\right)=\int_{-\infty}^{\infty}xf\left(x\right)=\int_{c}^{\infty}xac^ax^{-a-1}dx=\int_{c}^{\infty}ac^ax^{-a}dx=ac^a\left(\frac{x^{-a+1}}{-a+1}\right)_c^{\infty}$. כעת, נפריד מקרים:

$$\xrightarrow{a>1} ac^a\left[\left(\underbrace{\lim_{x\to\infty}\frac{x^{\overbrace{-a+1}^{<0}}}{-a+1}}_{0}\right)-\frac{c^{-a+1}}{-a+1}\right]=ac^a\frac{c^{-a+1}}{a-1}=\frac{ac}{a-1}$$

$$\xrightarrow{0<a<1} ac^a\left[\left(\underbrace{\lim_{x\to\infty}\frac{x^{\overbrace{-a+1}^{>0}}}{-a+1}}_{\infty}\right)-\frac{c^{-a+1}}{-a+1}\right]\to\infty$$

נשים לב כי עבור $0<a<1$ נקבל תוחלת אינסופית.

5. יהי X מ.מ בעל התפלגות קושי, פונקצית ההצטברות הינה $x\in\mathbb{R}$, $F(x)=b+c\cdot\arctan\left(\frac{x}{a}\right)$

כאשר a הינו פרמטר חיובי.

א. מצא את הקבועים b ו- c

ב. מצא את פונקצית הצפיפות של X

ג. מצא את ההסתברות $p(-a\le X\le a)$

פתרון :א

אנחנו צריכים לחשוב על שתי משוואות כיוון שיש לנו שני נעלמים. ברור כי $F(\infty)=1$ וגם $F(-\infty)=0$.

$$\left.\begin{array}{l} F(\infty)=1\to\ \lim\limits_{x\to\infty}\left[b+c\cdot\arctan\left(\frac{x}{a}\right)\right]\to b+c\cdot\frac{\pi}{2}\ =1 \\ F(-\infty)=1\to\ \lim\limits_{x\to\infty}\left[b+c\cdot\arctan\left(\frac{x}{a}\right)\right]\to b-c\cdot\frac{\pi}{2}=0 \end{array}\right\}\to c=\frac{1}{\pi}\quad,\quad b=\frac{1}{2}$$

ב. נגזור, $f(x)=c\frac{a}{a^2+x^2}\to{}'\left(b+c\cdot\arctan\left(\frac{x}{a}\right)\right)='(F(x))$, לכן נקבל בסה"כ

$$f(x)=\begin{cases}\frac{1}{\pi}\frac{a}{a^2+x^2} & x\in\mathbb{R}\\ 0 & otherwise\end{cases}$$

ג. $p(-a\le X\le a)=F(-a)-F(a)=\frac{1}{2}+\frac{1}{\pi}\cdot\arctan\left(\frac{-a}{a}\right)-\left(\frac{1}{2}+\frac{1}{\pi}\cdot\arctan\left(\frac{a}{a}\right)\right)=2\frac{1}{\pi}\cdot\underbrace{\arctan 1}_{\frac{\pi}{4}}=\frac{1}{2}$

6. יהי X מ.מ בעל פונקצית צפיפות $f(x)=\frac{3}{2}\left(1-x^2\right),\ 0<x<1$

א. מהו התחום של הערכים של המ.מ $Y=\frac{1}{1+X}$

ב. מצא את הצפיפות של Y

ג. מצא את התוחלת של Y

חזור על שאלה זו למקרה ש Y מוגדר באופן הבא $Y=\left(X-\frac{1}{2}\right)^2$.

פתרון: א. נסתכל על התחום של X ונסיק לגבי Y (הפונקציה המתוארת יורדת תמיד ולכן ערכי הקיצון עבור התחום הסגור מתקבל בקצוות): אם X=0 אזי נקבל $Y=\frac{1}{1+X}\xrightarrow{x=0}1$. עבור X=1 נקבל: $Y=\frac{1}{1+X}\xrightarrow{x=1}\frac{1}{2}$ ולכן: $\frac{1}{2}<Y<1$.

ב. הצפיפות של Y: ראשית נמצא את ההתפלגות, כיוון שלא ידועה את התפלגותו של Y.

$$F_Y(a)=p(Y\le a)\xrightarrow{Y=\frac{1}{1+X}}p\left(\frac{1}{1+X}\le a\right)=p\left(X\ge\frac{1-a}{a}\right)=1-p\left(X<\frac{1-a}{a}\right)=1-F_X\left(\frac{1-a}{a}\right)$$

כעת, נגזור על מנת למצוא את הצפיפות:

$$\left(F_Y(a)\right)'=f(y)=\left(1-F_X\left(\frac{1-a}{a}\right)\right)'=-f_x\left(\frac{1-a}{a}\right)\cdot\left(-\frac{1}{a^2}\right)\xrightarrow{f_x\left(\frac{1-a}{a}\right)=\frac{3}{2}\left(1-\left(\frac{1-a}{a}\right)^2\right)}\frac{3}{a^3}-\frac{3}{2a^4}$$

ג. תוחלת: $E(x)=\int_{\frac{1}{2}}^{1}a\left(\frac{3}{a^2}-\frac{3}{2a^3}\right)da=\left(\frac{-3}{a}+\frac{3}{4a^2}\right)\Big|_{\frac{1}{2}}^{1}=\frac{3}{4}$.

נחזור על השאלה כאשר $Y=\left(X-\frac{1}{2}\right)^2$ (בבית)

א. נסתכל על התחום של X ונסיק לגבי Y: אם X=0 אזי נקבל $Y=\left(X-\frac{1}{2}\right)^2\xrightarrow{x=0}\frac{1}{4}$.

הגבול התחתון של התחום מתקבל בנקודת המינימום שהיא X=0.5 שיעור הY בנקודה זו היא אפס, ולכן: $0<Y<\frac{1}{4}$.

ד. הצפיפות של Y: ראשית נמצא את ההתפלגות, כיוון שלא ידועה את התפלגותו של Y.

$$F_Y(a)=p(Y\le a)\xrightarrow{Y=\frac{1}{1+X}}p\left(\left(X-\frac{1}{2}\right)^2\le a\right)=p\left(\frac{1}{2}-\sqrt{a}\le X\le\frac{1}{2}+\sqrt{a}\right)=F_X\left(\frac{1}{2}+\sqrt{a}\right)-F_X\left(\frac{1}{2}-\sqrt{a}\right)$$

כעת, נגזור על מנת למצוא את הצפיפות:

$$\left(F_Y(a)\right)'=f(y)=\left(F_X\left(\frac{1}{2}+\sqrt{a}\right)-F_X\left(\frac{1}{2}-\sqrt{a}\right)\right)'=f_X\left(\frac{1}{2}+\sqrt{a}\right)\cdot\frac{1}{2\sqrt{a}}-f_X\left(\frac{1}{2}-\sqrt{a}\right)\cdot\frac{-1}{2\sqrt{a}}$$

נציב בפונקציה, $f(x)=\frac{3}{2}\left(1-x^2\right)$ ונקבל $\frac{1}{2\sqrt{a}}\left(\frac{9}{4}-3a\right)$ בתחום $0<Y<\frac{1}{4}$

ה. תוחלת: $E(x)=\int_{0}^{\frac{1}{4}}a\frac{1}{2\sqrt{a}}\left(\frac{9}{4}-3a\right)da=\frac{3}{40}$.

7. ידוע כי אוכלוסיה מתפלגת נורמאלית עם תוחלת 84 וסטיית תקן 7

א. מהי התפלגות הדגימה עבור הממוצע אם לוקחים את כל המדגמים הקיימים בגודל 124

ב. אם התפלגות האוכלוסיה אינה ידועה, האם ניתן להסיק עבור התפלגות הדגימה של הממוצע המתוארת הסעיף א'?

ג. מה ההסתברות למצוא ממוצע מדגם מעל 87 בהינתן גודל מדגם של 30?

פתרון:

א. ההתפלגות שתתקבל עבור התפלגות דגימה של הממוצע היא התפלגות נורמאלית כיוון שמדובר באוכלוסייה שמתפלגת נורמאלית

ב. ההתפלגות שתתקבל עבור התפלגות דגימה של הממוצע היא התפלגות נורמאלית כיוון שמדובר במדגם גדול מספיק.

ד. תחילה נמצא את התפלגות דגימת הממוצע: כידוע אם $X \sim N\left(\mu,\sigma^2\right)$ אזי $\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right)$ במקרה שלנו: $X \sim N(84,49)$ אזי

$$\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right) \xrightarrow{X\sim N(84,49)\,,\,n=30} \bar{X} \sim N\left(84,\frac{49}{30}\right) \to \bar{X} \sim N(84,1.633)$$

כעת, נמצא את ציון התקן של הציון גלם 87: $Z_{87} = \frac{87-84}{\sqrt{1.633}} = 2.347$ הסתברות זאת (לפי הטבלה) שווה: 0.9906-1 = 0.0094

8. לאוכלוסיה מסוימת ישנה התפלגות א-סימטרית עם ממוצע 100 וסטיית תקן 15.

א. מהי התפלגות הדגימה של הממוצע עבור מדגם בן 64 תצפיות?

ב. מה אחוז המדגמים בגודל 64 בהם הממוצע קטן מ- 110?

ג. מה התחום הסימטרי סביב הממוצע בו מרוכזים 95% ממוצעי כל המדגמים האפשריים בגודל 64?

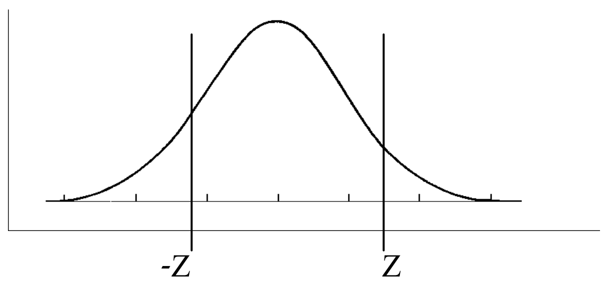
פתרון:

א. התפלגות הדגימה עבור הממוצע הינה נורמאלית כיוון שמדובר במדגם גדול מספיק. התפלגות היא:

$$\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right) \xrightarrow{X\sim N(100,225)\,,\,n=64} \bar{X} \sim N\left(100,\frac{225}{64}\right) \to \bar{X} \sim N(100,3.515)$$

ב. נמצא את ציון התקן של הציון גלם 110: $Z_{110} = \frac{110-100}{\sqrt{3.515}} = 5.33$ השטח שמהווה את האחוז (בטבלה) הינו שואף ל-1 (ערך שגדול מ-3)

ג. אנו מחפשים את ציוני התקן העונים על בקשה שהשטח ביניהם יהיה 95% מכלל השטח וגם ששטח זה יהיה סביב הממוצע. נדגים בציור:



בזכות העובדה שהתפלגות זו סימטרית, אנו יכולים לומר כי מימין ל- Z ומשמאל ל $-Z$ יש 2.5%. ז"א שנוכל לחפש בטבלה את השטח משמאל ל-Z.

משמאל ל-Z יש 0.975=0.95+0.025 שטח. לכן נחפש בתוך הטבלה את מספר זה ונקבל כי Z עבור שטח זה משמאלו הוא: 1.96. באותו אופן ה Z – שווה ל -1.96. כעת נרצה למצוא את ציון הגלם המתאימים: $Z_{\bar{X}}=\frac{x-100}{\sqrt{3.515}}=1.96\rightarrow 103.67$ כעת נמצא את $-Z$: $Z_{\bar{X}}=\frac{x-100}{\sqrt{3.515}}=-1.96\rightarrow 96.3$. ז"א ש95% מהממוצעים נמצאים בין הערכים 103.67 ל- 96.3.

התפלגויות המבוססות על ההתפלגות הנורמאלית:

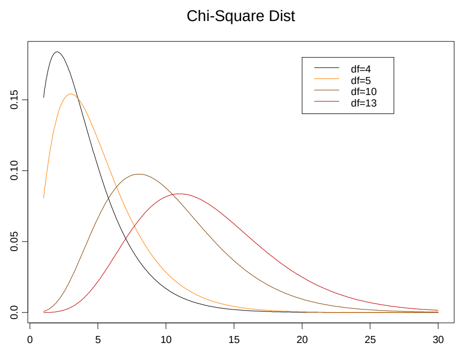
12. התפלגות χ^2 (חי בריבוע)
13. התפלגות T
14. התפלגות F

התפלגות χ^2 Chi-square distribution

יהי X מ.מ המתפלג חי בריבוע עם k דרגות חופש . אזי נסמן $X\sim\chi_k^2$ כך שפונקצית הצפיפות היא: $x^{(k/2)-1}e^{-x/2}$ $f_X(x)=\frac{1}{\Gamma(k/2)}\left(\frac{1}{2}\right)^{(k/2)}$ כאשר $x\leq 0$.

דרגות חופש - Degrees of freedom

מספר דרגות החופש של מודל מתמטי הוא מספר הערכים הבלתי-תלויים המאפשר לתאר את המצב באופן חד משמעי. באופן טכני יותר, מספר הפרמטרים המעורבים בקביעת התפלגות, או מספר המשתנים החופשיים בחישוב (למשל אם יש לנו n מחוברים אז דרגות החופש יהיו n-1 כיוון שאיבר אחד ניתן להציג ע"י כל השאר).



התוחלת: $E(x)=k$

שונות: $V(x)=2k$

פונקציה יוצרת מומנטים: $M_X(t)=\left(\frac{1}{1-2t}\right)^{\frac{k}{2}}$

הקשר בין התפלגות χ^2 להתפלגות הנורמאלית.

טענה: יהי $Z\sim N(0,1)$ אזי $Z^2\sim\chi_1^2$.

הוכחה: נפתח את הפונקציה יוצרת מומנטים של Z^2 ונבדוק אם נגיע לפונקציה יוצרת של מ.מ המתפלג χ_1^2, ז"א נרצה להגיע בסוף הפיתוח שמתקיים $M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$.

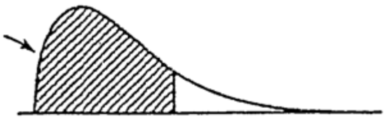
$$M_{Z^2}(t)=\int_{-\infty}^{\infty}e^{tz^2}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\}dz=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2-2tz^2}{2}\right\}dz=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2(1-2t)}{2}\right\}dz$$

$$\xrightarrow{\frac{1}{1-2t}=y^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2y^2}\right\}dz\xrightarrow{\frac{z}{y}=w\ ,\ dz=ydw}y\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{w^2}{2}\right\}dw}_{1}=y$$

$$M_{Z^2}(t)=\sqrt{\frac{1}{1-2t}}\quad\Rightarrow\quad Z^2\sim\chi_1^2$$

דוגמא: הוכח כי אם $Z_1,Z_2,Z_3,\ldots Z_k$ מ.מ-ים ב"ת כך ש $\forall i=1\ldots k\quad Z_i\sim N(0,1)$ אזי $\sum_i^k Z_i^2\sim\chi_k^2$.

הוכחה: שוב נתחיל מפונקציה יוצרת מומנטים, $M_{\sum_i^k Z_i^2}(t)=E\left(e^{t\sum_i^k Z_i^2}\right)$ בגלל שנתון שהמ.מ-ים ב"ת אזי מתקיים $E\left(e^{t\sum_i^k Z_i^2}\right)=E\left(e^{t\left(Z^2_1+Z^2_2+\ldots+Z^2_k\right)}\right)=E\left(e^{tZ^2_1}\right)\cdot E\left(e^{tZ^2_2}\right)\cdot\ldots\cdot E\left(e^{tZ^2_k}\right)$ ובגלל שהמ.מ-ים שווי התפלגות (בעלי התפלגות זהה לחלוטין) לפי הנתון $\forall i=1\ldots k\quad Z_i\sim N(0,1)$ אזי $E\left(e^{tZ^2_1}\right)\cdot E\left(e^{tZ^2_2}\right)\cdot\ldots\cdot E\left(e^{tZ^2_k}\right)=\left[M_{Z^2}(t)\right]^k$. כעת מההוכחה הקודמת ראינו כי $M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$ ולכן $\left[M_{Z^2}(t)\right]^k\xrightarrow{M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}}\left(\frac{1}{1-2t}\right)^{\frac{k}{2}}$. *משמעות הדבר היא כי סכום של מ.מ-ים ב"ת המתפלגים נורמאלית סטנדרטית מתפלג חי בריבוע עם מספר הדרגות החופש השווה למספר המחוברים.*



χ^2_p

טבלת התפלגות חי־בריבוע – ערכי החלוקה χ^2_p

	p												
דרגות חופש	0.005	0.010	0.025	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.975	0.990	0.995
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7

התפלגות F:

יהי מ.מ U המתפלג χ^2_m ויהי מ.מ V המתפלג χ^2_n . מ.מ אלו הם ב״ת. אזי למ.מ X

המוגדר $\frac{U/m}{V/n}$ יש התפלגות F עם m,n דרגות חופש. נסמן $X \sim F_{m,n}$

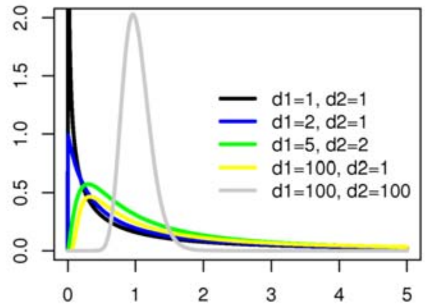
פונקצית הצפיפות הינה: $f_X(x) = \frac{\Gamma[(m+n)/2]}{\Gamma[m/2]\Gamma[n/2]}\left(\frac{m}{n}\right)^{m/2}\frac{x^{(m-2)/2}}{[1+(m/n)x]^{(m+n)/2}}$ כאשר

$m,n = 1,2\ldots$

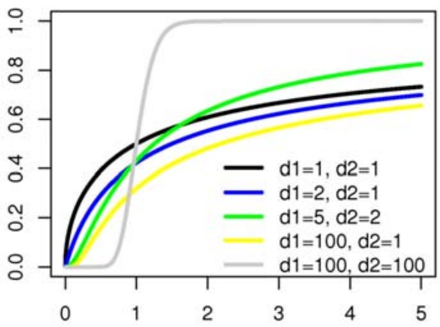
פונקציה יוצרת מומנטים: לא קיימת

התוחלת: $E(X) = \frac{n}{n-2}$ $for\ \ n>2$ השונות: $V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$

פונקצית הצפיפות



פונקצית הצטברות



טבלת התפלגות F עבור $\alpha = 0.05$

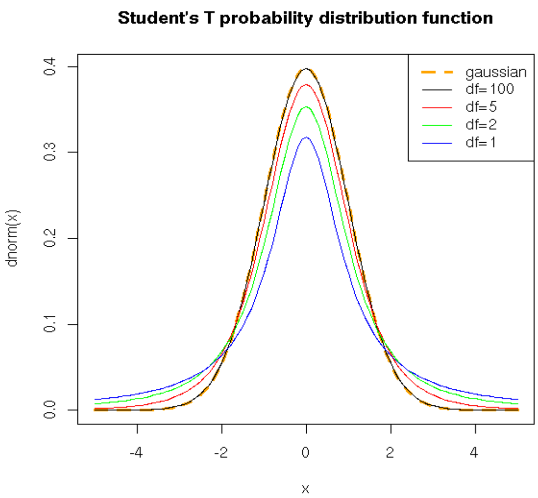
df2/df1	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**20**	**24**	**30**	**40**	**60**	**120**	**INF**
1	161.448	199.5	215.707	224.583	230.162	233.99	236.768	238.883	240.543	241.882	243.91	245.95	248.013	249.052	250.095	251.143	252.196	253.253	254.314
2	18.5128	19	19.1643	19.2468	19.2964	19.33	19.3532	19.371	19.3848	19.3959	19.413	19.4291	19.4458	19.4541	19.4624	19.4707	19.4791	19.4874	19.4957
3	10.128	9.552	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855	8.7446	8.7029	8.6602	8.6385	8.6166	8.5944	8.572	8.5494	8.5264
4	7.7086	6.944	6.5914	6.3882	6.2561	6.1631	6.0942	6.041	5.9988	5.9644	5.9117	5.8578	5.8025	5.7744	5.7459	5.717	5.6877	5.6581	5.6281
5	6.6079	5.786	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351	4.6777	4.6188	4.5581	4.5272	4.4957	4.4638	4.4314	4.3985	4.365
6	5.9874	5.143	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.099	4.06	3.9999	3.9381	3.8742	3.8415	3.8082	3.7743	3.7398	3.7047	3.6689
7	5.5914	4.737	4.3468	4.1203	3.9715	3.866	3.787	3.7257	3.6767	3.6365	3.5747	3.5107	3.4445	3.4105	3.3758	3.3404	3.3043	3.2674	3.2298
8	5.3177	4.459	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472	3.2839	3.2184	3.1503	3.1152	3.0794	3.0428	3.0053	2.9669	2.9276
9	5.1174	4.257	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373	3.0729	3.0061	2.9365	2.9005	2.8637	2.8259	2.7872	2.7475	2.7067
10	4.9646	4.103	3.7083	3.478	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782	2.913	2.845	2.774	2.7372	2.6996	2.6609	2.6211	2.5801	2.5379
11	4.8443	3.982	3.5874	3.3567	3.2039	3.0946	3.0123	2.948	2.8962	2.8536	2.7876	2.7186	2.6464	2.609	2.5705	2.5309	2.4901	2.448	2.4045
12	4.7472	3.885	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534	2.6866	2.6169	2.5436	2.5055	2.4663	2.4259	2.3842	2.341	2.2962
13	4.6672	3.806	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.671	2.6037	2.5331	2.4589	2.4202	2.3803	2.3392	2.2966	2.2524	2.2064
14	4.6001	3.739	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022	2.5342	2.463	2.3879	2.3487	2.3082	2.2664	2.2229	2.1778	2.1307
15	4.5431	3.682	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437	2.4753	2.4034	2.3275	2.2878	2.2468	2.2043	2.1601	2.1141	2.0658
16	4.494	3.634	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935	2.4247	2.3522	2.2756	2.2354	2.1938	2.1507	2.1058	2.0589	2.0096
17	4.4513	3.592	3.1968	2.9647	2.81	2.6987	2.6143	2.548	2.4943	2.4499	2.3807	2.3077	2.2304	2.1898	2.1477	2.104	2.0584	2.0107	1.9604
18	4.4139	3.555	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117	2.3421	2.2686	2.1906	2.1497	2.1071	2.0629	2.0166	1.9681	1.9168
19	4.3807	3.522	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779	2.308	2.2341	2.1555	2.1141	2.0712	2.0264	1.9795	1.9302	1.878
20	4.3512	3.493	3.0984	2.8661	2.7109	2.599	2.514	2.4471	2.3928	2.3479	2.2776	2.2033	2.1242	2.0825	2.0391	1.9938	1.9464	1.8963	1.8432
21	4.3248	3.467	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.366	2.321	2.2504	2.1757	2.096	2.054	2.0102	1.9645	1.9165	1.8657	1.8117
22	4.3009	3.443	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967	2.2258	2.1508	2.0707	2.0283	1.9842	1.938	1.8894	1.838	1.7831
23	4.2793	3.422	3.028	2.7955	2.64	2.5277	2.4422	2.3748	2.3201	2.2747	2.2036	2.1282	2.0476	2.005	1.9605	1.9139	1.8648	1.8128	1.757
24	4.2597	3.403	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547	2.1834	2.1077	2.0267	1.9838	1.939	1.892	1.8424	1.7896	1.733
25	4.2417	3.385	2.9912	2.7587	2.603	2.4904	2.4047	2.3371	2.2821	2.2365	2.1649	2.0889	2.0075	1.9643	1.9192	1.8718	1.8217	1.7684	1.711
26	4.2252	3.369	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197	2.1479	2.0716	1.9898	1.9464	1.901	1.8533	1.8027	1.7488	1.6906
27	4.21	3.354	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043	2.1323	2.0558	1.9736	1.9299	1.8842	1.8361	1.7851	1.7306	1.6717
28	4.196	3.34	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.236	2.19	2.1179	2.0411	1.9586	1.9147	1.8687	1.8203	1.7689	1.7138	1.6541
29	4.183	3.328	2.934	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768	2.1045	2.0275	1.9446	1.9005	1.8543	1.8055	1.7537	1.6981	1.6376
30	4.1709	3.316	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646	2.0921	2.0148	1.9317	1.8874	1.8409	1.7918	1.7396	1.6835	1.6223
40	4.0847	3.232	2.8387	2.606	2.4495	2.3359	2.249	2.1802	2.124	2.0772	2.0035	1.9245	1.8389	1.7929	1.7444	1.6928	1.6373	1.5766	1.5089
60	4.0012	3.15	2.7581	2.5252	2.3683	2.2541	2.1665	2.097	2.0401	1.9926	1.9174	1.8364	1.748	1.7001	1.6491	1.5943	1.5343	1.4673	1.3893
120	3.9201	3.072	2.6802	2.4472	2.2899	2.175	2.0868	2.0164	1.9588	1.9105	1.8337	1.7505	1.6587	1.6084	1.5543	1.4952	1.429	1.3519	1.2539
inf	3.8415	2.996	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307	1.7522	1.6664	1.5705	1.5173	1.4591	1.394	1.318	1.2214	1

התפלגות T :

התפלגות t היא התפלגות המתארת את הערכים הצפויים למדגם מתוך אוכלוסייה המקיימת התפלגות נורמאלית כאשר גודלו של המדגם קטן. הצורה הכללית של התפלגות ה-t דומה לזו של ההתפלגות הנורמאלית וכאשר מספר הפרטים במדגם גדול היא אף זהה לה. התפלגות ה-t היא הבסיס למבחן t, המשמש לבדיקת מובהקות ההבדל בין שני מדגמים של אוכלוסיות.

מאפייני ההתפלגות:

- סימטריות
- התפלגות בעלת ממוצע השווה ל-0
- צורתה של ההתפלגות תלויה בגודל המדגם : עבור מדגם גדול, נקבל התפלגות המתקרבת להתפלגות נורמאלית סטנדרטית. עבור מדגם קטן, נקבל התפלגות שטוחה יותר מהתפלגות נורמאלית סטנדרטית.
- התפלגות זו בעלת השם הידוע "**התפלגות T בעלת n-1 דרגות חופש**" דרגות החופש הן גודל המדגם פחות 1 . כך שעבור מדגמים בגדלים שונים נקבל צורות התפלגות T שונה. ככל שערך דרגות החופש גדול כל מתקרבים לנורמאליות. כפי שניתן לראות בגרף, צורת הפעמון משתנה עם שינוי של דרגות החופש. דרגות החופש מסומנות df (degrees of freedom).



יהי X מ.מ המתפלג T עם m דרגות חופש . אזי נסמן $X \sim t_m$ כך שפונקציית הצפיפות היא :

$$f_X(x)=\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma(m/2)}\frac{1}{\sqrt{m\pi}}\frac{1}{\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}$$ כאשר $x \le 0$.

השימוש בהתפלגות T בא לידי ביטוי כאשר אין מידע על שונות האוכלוסיה.

התוחלת: $E(x)=0$

שונות: $V(x)=\dfrac{m}{m-2} \quad for \;\; m>2$

משפט: יהיו Z ו U מ.מ-ים בלתי תלויים המתפלגים באופן הבא $Z\sim N(0,1)$ ו-

$U\sim\chi_m^2$ אזי X המוגדר $X=\dfrac{Z}{\sqrt{U/m}}$ הינו מ.מ המתפלג T עם m דרגות חופש.

הוכחה: נמצא את ההתפלגות של X לפי הגדרה:

$F_X(x)=P(X<x)\xrightarrow{X=\frac{Z}{\sqrt{U/m}}}P\left(\dfrac{Z}{\sqrt{U/m}}<x\right)=P\left(Z<x\sqrt{U/m}\right)$ כעת, בגלל שמעורבים שני משתנים מקריים, נעשה זאת קצת אחרת, נרצה תחילה "לנטרל" אחד מהמשתנים המקריים הכי קל "לנטרל" את U , נעשה זאת כך: נטגרל על כל המרחב התפלגות של U ומתוכה נרצה את החלק שמקיים את ההסתברות שאנו רוצים לחשב: $P\left(Z<x\sqrt{U/m}\right)=\int_0^\infty P\left(Z<x\sqrt{u/m}\right)f_U(u)\,du$ עכשיו נוכל לחשב את ההסתברות $P\left(Z<x\sqrt{u/m}\right)$ (ה U הפך לu ועכשיו כבר אינו מ.מ) ולהציבה בתוך האינטגרל: $P\left(Z<x\sqrt{u/m}\right)=\Phi\left(x\sqrt{u/m}\right)$ ובסה"כ נקבל $\int_0^\infty \Phi\left(x\sqrt{u/m}\right)f_U(u)\,du$. אם אנו רוצים את פונקצית הצפיפות עלינו לגזור לפי המשתנה x: $\left(F_X(x)\right)_x'=\left(\int_0^\infty \Phi\left(x\sqrt{u/m}\right)f_U(u)\,du\right)_x'$ וביטוי המתקבל הינו:

$$f_X(x)=\int_0^\infty \sqrt{u/m}\,\phi\left(x\sqrt{u/m}\right)f_U(u)\,du$$

נציב את פונקציות הצפיפות עצמן:

$$f_X(x)=\int_0^\infty \sqrt{u/m}\,\phi\left(x\sqrt{u/m}\right)f_U(u)\,du=\int_0^\infty \sqrt{u/m}\underbrace{\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}\frac{u}{m}}}_{\phi\left(x\sqrt{u/m}\right)}\underbrace{\frac{1}{\Gamma(m/2)}\left(\frac{1}{2}\right)^{(m/2)}u^{(m/2)-1}e^{-u/2}}_{f_U(u)}du$$

נשים לב כי הביטוי שקבלנו ניתן לפירוק באופן הבא:

$$\sqrt{1/m}\frac{1}{\sqrt{2\pi}}\left(\frac{1}{2}\right)^{(m/2)}\frac{1}{\Gamma(m/2)}\int_0^\infty\sqrt{u}\;e^{\frac{-x^2}{2}\frac{u}{m}}\;u^{(m/2)-1}\;e^{-u/2}du=\sqrt{1/m}\frac{1}{\sqrt{2\pi}}\left(\frac{1}{2}\right)^{(m/2)}\frac{1}{\Gamma(m/2)}\int_0^\infty e^{-u\left(\frac{1+\frac{x^2}{m}}{2}\right)}u^{\frac{m+1}{2}-1}du$$

נסתכל על האינטגרל:

$$\int_0^\infty e^{-u\left(\frac{1+\frac{x^2}{m}}{2}\right)} u^{\frac{m+1}{2}-1}\,du = \frac{\Gamma\left(\frac{m+1}{2}\right)}{\frac{1}{2}\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}} \int_0^\infty \frac{\frac{1}{2}\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}{\Gamma\left(\frac{m+1}{2}\right)} e^{-u\left(\frac{1+\frac{x^2}{m}}{2}\right)} u^{\frac{m+1}{2}-1}\,du$$

כאשר מכפילים ומחלקים בביטוי $\frac{\frac{1}{2}\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}{\Gamma\left(\frac{m+1}{2}\right)}$ (ז"א שהכפלנו ב-1) נשים לב כי הפונקציה המתקבלת בתוך האינטגרל הינה בדיוק פונקציה צפיפות של התפלגות גמא ($\lambda>0, r>0 \quad f_X(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$) עם הפרמטרים הבאים $\lambda=\frac{1+\frac{x^2}{m}}{2}$ ו $r=\frac{m+1}{2}$. אם כן אינטגרל זה על כל המרחב שווה ל-1. מכאן נקבל שפונקצית הצפיפות שמתקבלת היא בדיוק פונקציה צפיפות של התפלגות T עם m דרגות חופש.

$$\sqrt{1/m}\,\frac{1}{\sqrt{2\pi}}\left(\frac{1}{2}\right)^{(m/2)}\frac{1}{\Gamma(m/2)}\frac{\Gamma\left(\frac{m+1}{2}\right)}{\left(\frac{1+\frac{x^2}{m}}{2}\right)^{\frac{m+1}{2}}}\underbrace{\int_0^\infty \frac{\left(\frac{1+\frac{x^2}{m}}{2}\right)^{\frac{m+1}{2}}}{\Gamma\left(\frac{m+1}{2}\right)} e^{-u\left(\frac{1+\frac{x^2}{m}}{2}\right)} u^{\frac{m+1}{2}-1}\,du}_{=1} =$$

$$=\sqrt{1/m}\,\frac{1}{\sqrt{\pi}}\frac{1}{\Gamma(m/2)}\frac{\Gamma\left(\frac{m+1}{2}\right)}{\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}$$

טענה: יהי X מ.מ המתפלג T עם m דרגות חופש, אזי

1. $E(X)=0$
2. $V(X)=\frac{m}{m-2}$

הוכחה 1: נוכל לבטא את המ.מ באופן הבא $X=\frac{Z}{\sqrt{U/m}}$, כעת, נמצא את התוחלת שלו:

$$E(X)=E\left(\frac{Z}{\sqrt{U/m}}\right)\xrightarrow{ID}E(Z)E\left(\frac{1}{\sqrt{U/m}}\right)\xrightarrow{E(Z)=0}0$$

הוכחה 2:

$$V(X)=E(X^2)-(E(X))^2\xrightarrow{E(X)=0}E(X^2)$$

$$X^2=\left(\frac{Z}{\sqrt{U/m}}\right)^2=\frac{Z^2}{U/m}$$

כבר הוכחנו בעבר כי אם Z מתפלג נורמאלית סטנדרטית אזי Z^2 מתפלג חי עם דרגת חופש אחת.בנוסף, ידוע כי U מתפלג חי עם m דרגות חופש. לכן בסה"כ נקבל כי $X^2\sim\frac{\chi_1^2/1}{\chi_m^2/m}=F_{1,m}$. ותוחלת של מ.מ המתפלג F עם דרגות החופש הנ"ל היא בדיוק $\frac{m}{m-2}$

טבלת התפלגות T

Significance level = α

Degrees of Freedom	.005 (1-tail) .01 (2-tails)	.01 (1-tail) .02 (2-tails)	.025 (1-tail) .05 (2-tails)	.05 (1-tail) .10 (2-tails)	.10 (1-tail) .20 (2-tails)	.25 (1-tail) .50 (2-tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	.765
4	4.604	3.747	2.776	2.132	1.533	.741
5	4.032	3.365	2.571	2.015	1.476	.727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.878	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
Large	2.575	2.327	1.960	1.645	1.282	.675

ב. יהיו משתנים מקריים Z_i כך שלכל $i=1\ldots n$ מתקיים $Z_i \sim N(0,1)$ מצא כיצד מתפלג

X אם מוגדר באופן הבא: $X=\frac{42\sum_{i=1}^{52}(Z_i)^2}{52\sum_{i=1}^{42}(Z_i)^2}$

פתרון :נשתמש בשני משפטים:

1. אם $Z_1,Z_2,Z_3,\ldots Z_k$ מ.מ-ים ב"ת כך ש $\forall i=1\ldots k$ $Z_i \sim N(0,1)$ אזי $\sum_i^k Z_i^2 \sim \chi_k^2$. לכן כל סכום המוצג במשתנה המקרי מתפלג חי בריבוע עם מספר דרגות חופש כמספר המחוברים: $\sum_{i=1}^{52}(Z_i)^2 \sim \chi_{52}^2$ ו $\sum_{i=1}^{42}(Z_i)^2 \sim \chi_{42}^2$

2. יהי מ.מ U המתפלג χ_m^2 ויהי מ.מ V המתפלג χ_n^2 . מ.מ אלו הם ב"ת. אזי למ.מ X המוגדר $\frac{U/m}{V/n}$ יש התפלגות F עם m,n דרגות חופש. נסמן $X \sim F_{m,n}$

עם קצת אלגברה בסיסית... נסיק כי הביטוי המוצג בתרגיל מתפלג $X \sim F_{52,42}$ נסו!!

4. א. הוכח את הטענה הבאה: יהיו $X_1 \sim\sim Poisson(\lambda_1),\ X_2 \sim Poisson(\lambda_2),\ldots,X_n \sim Poisson(\lambda_n)$ ב"ת כיצד יתפלג $\sum_{i=1}^n X_i$ ומהם הפרמטרים המתאימים?

פתרון: נשים לב כי לא מדובר במשתנים שווי התפלגות!!!!

$$E\left(e^{t\sum_{i=1}^n X_i}\right)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right)=E\left(\prod_{i=1}^n e^{tx_i}\right)\xrightarrow{I.D}\prod_{i=1}^n E\left(e^{tx_i}\right)$$

$$\xrightarrow{X_1\sim\sim Poisson(\lambda_1),\ X_2\sim Poisson(\lambda_2),\ldots,X_n\sim Poisson(\lambda_n)}\prod_{i=1}^n E\left(e^{tx_i}\right)$$

$$\xrightarrow{M_X(t)=E\left(e^{tx}\right)=e^{\lambda(e^t-1)}} e^{\lambda_1(e^t-1)}\cdot e^{\lambda_2(e^t-1)}\cdots e^{\lambda_n(e^t-1)}=e^{(e^t-1)\sum_{i=1}^n \lambda_i}$$

קבלנו התפלגות פואסון עם פרמטר: $\sum_{i=1}^n \lambda_i$ ז"א $\sum_{i=1}^n X_i \sim Poisson\left(\sum_{i=1}^n \lambda_i\right)$

שאלה ממבחן: יהיו $Y \sim N(2,25)$ ו $X \sim N(4,16)$ מ.מ בלתי תלויים. (24 נקודות)

א. כיצד מתפלג $2Y,\ 3X$? (4 נקודות)

ב. יהיו $K \sim N\left(\mu_K, \sigma_K^2\right)$ ו $W \sim N\left(\mu_W, \sigma_W^2\right)$ מ.מ בלתי תלויים. הוכח כי $W-K$ גם מתפלג נורמאלית ומצא את הפרמטרים המתאימים. (5 נקודות)

ג. כיצד מתפלג $Y-X$? (2 נקודות)

ד. מצאו $p\left(2Y+3X+40<10\right)$. (5 נקודות)

ה. יהיו משתנים מקריים Z_i כך שלכל $i=1\ldots n$ מתקיים $Z_i \sim N\left(0,1\right)$ מצא כיצד מתפלג X אם מוגדר באופן הבא: $X=\dfrac{Z_{32}\sqrt{20}}{\sqrt{\sum_{i=1}^{20}\left(Z_i\right)^2}}$ (8 נקודות)

א.
$$X \sim N\left(4,16\right) \xrightarrow{3X} 3X \sim N\left(3\cdot 4, 9\cdot 16\right) \rightarrow 3X \sim N\left(12,144\right)$$
$$Y \sim N\left(2,25\right) \xrightarrow{2Y} 2Y \sim N\left(2\cdot 2, 4\cdot 25\right) \rightarrow 2Y \sim N\left(4,100\right)$$

ב. פתרון: נבדוק האם הפונקציה יוצרת מומנטים של $X+Y$ היא פונקציה יוצרת של ההתפלגות הנורמאלית: $E\left(e^{t(W-K)}\right)=E\left(e^{tW}e^{-tK}\right)$ בגלל שנתון כי המ.מ-ים בלתי תלויים ניתן להפריד $E\left(e^{t(W-K)}\right)=E\left(e^{tW}\right)E\left(e^{-tK}\right)$. כעת, כל מ.מ מתפלג נורמאלית ולכן ניתן להציב את הפונקציה היוצרת המתאימה לו:

$$E\left(e^{t(W-Y)}\right)=E\left(e^{tW}\right)E\left(e^{-tK}\right)\xrightarrow{M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2t^2\right]}\exp\left[\mu_W t+\frac{1}{2}\sigma_W^2t^2\right]\exp\left[-\mu_K t+\frac{1}{2}\sigma_K^2t^2\right]$$
$$\exp\left[\left(\mu_W-\mu_K\right)t+\frac{1}{2}\left(\sigma_W^2+\sigma_K^2\right)t^2\right]$$

קבלנו התפלגות נורמאלית עם תוחלת $\mu_W-\mu_K$ ושונות $\sigma_W^2+\sigma_K^2$. ז"א בסה"כ $X+Y \sim N\left(\mu_W-\mu_K, \sigma_W^2+\sigma_K^2\right)$.

ג. $Y \sim N\left(2,25\right), \quad X \sim N\left(4,16\right) \rightarrow Y-X \sim N\left(-2,\ 41\right)$

ד.
$$p\left(2Y+3X+40<10\right)\longrightarrow p\left(2Y+3X<-30\right)$$
$$3X \sim N\left(12,144\right),\quad 2Y \sim N\left(4,100\right)\rightarrow 2Y+3X \sim N\left(16,244\right)$$
$$\xrightarrow{tiknun} p\left(Z<\frac{-30-16}{\sqrt{244}}\right)=p\left(Z<-2.9\right)=1-0.9981=0.0019$$

הערה, ניתן לפתור גם כך, ברור שזה נכון ומתקבל ☺

$$p\left(2Y+3X+40<10\right)$$
$$3X \sim N\left(12,144\right),\quad 2Y \sim N\left(4,100\right)\rightarrow 2Y+3X+40 \sim N\left(56,244\right)$$
$$\xrightarrow{tiknun} p\left(Z<\frac{10-56}{\sqrt{244}}\right)=p\left(Z<-2.9\right)=1-0.9981=0.0019$$

ה. מבחינה אלגברית הביטוי $X=\frac{Z_{32}\sqrt{20}}{\sqrt{\sum_{i=1}^{20}(Z_i)^2}}$ שווה בדיוק ל $X=\frac{Z_{32}}{\sqrt{\frac{\sum_{i=1}^{20}(Z_i)^2}{20}}}$ לפי

המשפט בעמוד 129 " משפט: יהיו Z ו U מ.מ-ים בלתי תלויים המתפלגים באופן הבא $Z\sim N(0,1)$ ו- $U\sim\chi_m^2$ אזי X המוגדר $X=\frac{Z}{\sqrt{U/m}}$ הינו מ.מ המתפלג T עם m דרגות חופש"

ההתפלגות היא T עם 20 דרגות חופש.

שאלה ממבחן: יהיו $Y\sim N(1,0.125)$ ו $X\sim N(2,0.5)$ מ.מ בלתי תלויים.

א. כיצד מתפלג $W=X-2Y$?

ב. כיצד מתפלג W^2?

ג. יהיו משתנים מקריים Z_i כך שלכל $i=1\ldots n$ מתקיים $Z_i\sim N(0,1)$ מצא כיצד מתפלג A אם מוגדר באופן הבא: $A=\frac{W\sqrt{15}}{\sqrt{\sum_{i=1}^{15}(Z_i)^2}}$

פתרון: א. $W=X-2Y\sim N(0,1)$

ב. לפי טענה בעמוד 124 $W^2\sim\chi_1^2$

ג. מבחינה אלגברית הביטוי $A=\frac{W\sqrt{15}}{\sqrt{\sum_{i=1}^{15}(Z_i)^2}}$ שווה בדיוק ל $A=\frac{W}{\sqrt{\frac{\sum_{i=1}^{15}(Z_i)^2}{15}}}$ לפי

המשפט בעמוד 129 " משפט: יהיו Z ו U מ.מ-ים בלתי תלויים המתפלגים באופן הבא $Z\sim N(0,1)$ ו- $U\sim\chi_m^2$ אזי X המוגדר $X=\frac{Z}{\sqrt{U/m}}$ הינו מ.מ המתפלג T עם m דרגות חופש"

ההתפלגות היא T עם 15 דרגות חופש.

התפלגות לוגנורמאלית:

מ.מ X נקרא מ.מ לוג נורמאלי אם $\ln X\sim N(\mu,\sigma^2)$

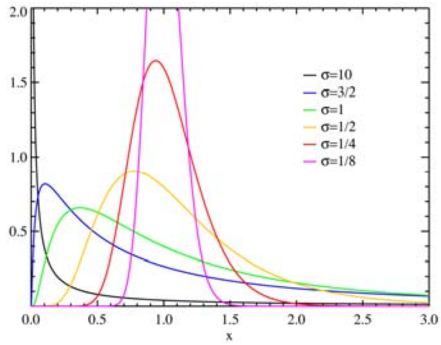
פונקצית הצפיפות הינה:

$$f_X(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}\qquad x>0$$

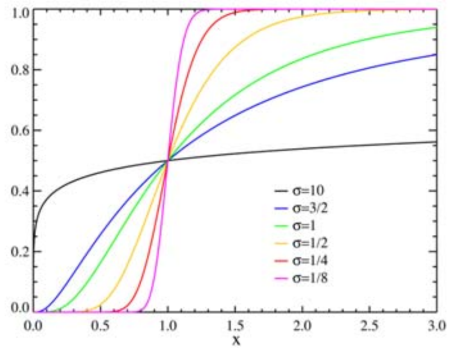
פונקציה יוצרת מומנטים: אין שימוש

התוחלת: $E(X)=\exp\left(\mu+\frac{1}{2}\sigma^2\right)$ השונות: $V(X)=(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$

פונקצית הצפיפות



פונקצית הצטברות



משתנה מקרי דו מימדי

הגדרה: יהיו X ו- Y שני מ.מ-ים המוגדרים על אותו מרחב הסתברותי, אזי הזוג (X,Y) יקרא משתנה מקרי דו מימדי Bivariate random variable

הסתברות של תוצאה מסוימת (X_i,Y_j) מסומנת ע"י $p_{XY}(X_i,Y_j)=p(X=x_i,Y=y_j)$ ונקראת פונקצית ההסתברות המשותפת The joint probability function $p_{XY}(X,Y)$ של המשתנה הדו מימדי (X,Y).

תכונות של פונקצית ההסתברות המשותפת.

1. לכל זוג ערכים (x_i,y_j) מתקיים $0\le p_{XY}(x_i,y_j)\le 1$
2. סכום ההסתברויות שווה ל-1: $\sum_{x_i}\sum_{y_j} p_{XY}(x_i,y_j)=1$

פונקצית הסתברות שולית Marginal probability function

פונקצית ההסתברות השולית של מ.מ X מוגדרת $p_X(x_i)=\sum_{y_j} p_{XY}(x_i,y_j)$, באותו אופן, פונקצית ההסתברות השולית של מ.מ Y מוגדרת $p_Y(y_i)=\sum_{x_i} p_{XY}(x_i,y_j)$.

נסביר: $p_X(x_i)$ הינה ההסתברות לערך x_i כאשר אין חשיבות מהו הערך של המשתנה השני.

תלות ומתאם:

- שני מ.מ-ים בעלי התפלגות משותפת p_{XY} יקראו בלתי תלויים (Independent random variable) אם לכל (X_i,Y_j) מתקיים $p_{XY}(X_i,Y_j)=p_X(X_i)p_Y(Y_j)$
- שני מ.מ-ים בעלי התפלגות משותפת p_{XY} יקראו בלתי מתואמים (Uncorrelated) אם מתקיים $E(XY)=E(X)\cdot E(Y)$ כאשר $E(XY)=\sum_{i,j}x_iy_jp_{XY}(x_i,y_j)$

אם X ו Y מ.מ-ים בלתי תלויים אזי הם גם בלתי מתואמים.

דוגמא: בכד נמצאים 4 כדורים, שנים מסומנים בספרה 1, אחד מהם בספרה 2 ואחד בספרה 3. מוציאים שני כדורים מהכד ללא החזרה. יהי X המספר הרשום על הכדור הראשון שהוצא, ו Y המספר על הכדור השני שהוצא.

א. בנה טבלת התפלגות משותפת

ב. מצא את $E(X)$ ואת $E(Y)$

ג. מצא את $E(X+Y)$

ד. מצא את $V(X)$ ואת $V(Y)$

ה. מצא את $V(X+Y)$

פתרון: א.

Y \ X	1	2	3	$p_Y(y_i)$
1	$1/6$	$1/6$	$1/6$	$1/6+1/6+1/6=0.5$
2	$1/6$	0	$1/12$	$1/6+0+1/12=0.25$
3	$1/6$	$1/12$	0	$1/6+1/12+0=0.25$
$p_X(x_i)$	$1/6+1/6+1/6=0.5$	$1/6+0+1/12=0.25$	$1/6+1/12+0=0.25$	1

נסביר כיצד הגענו להסתברויות, למשל $(X=2,Y=1)=1/4\cdot 2/3=1/6$ וכך הלאה.

ב. $E(X)=1\cdot 0.5+2\cdot 0.25+3\cdot 0.25=1.75$ באותו אופן $E(Y)=1\cdot 0.5+2\cdot 0.25+3\cdot 0.25=1.75$

ג. על מנת למצוא את $E(X+Y)$ נצטרך, תחילה למצוא את ההסתברויות $p(X+Y)$ ולכן:

$(X+Y)$	2	3	4	5	6	
$p(X+Y)$	$1/6$	$1/3$	$1/3$	$1/6$	0	1

נסביר כיצד הגענו להסתברויות, למשל על מנת להגיע לערך 3 נצטרך $(X=1,Y=2)\ or\ (X=2,Y=1)$ ז״א $p(X=1,Y=2)+p(X=2,Y=1)=1/6+1/6=1/3$

וכך הלאה. לכן התוחלת המבוקשת : $E(X+Y)=2\cdot 1/6+3\cdot 1/3+4\cdot 1/3+5\cdot 1/6=3.5$

ד. על מנת למצוא את השונויות, נמצא תחילה את $E(X^2)$ ואת $E(Y^2)$:

$E(X^2)=1^2\cdot 0.5+2^2\cdot 0.25+3^2\cdot 0.25=3.75$ ומכאן : $V(X)=E(X^2)+(E(X))^2=3.75+1.75^2=0.6875$

$E(Y^2)=1^2\cdot 0.5+2^2\cdot 0.25+3^3\cdot 0.25=3.75$ ומכאן $V(Y)=E(Y^2)+(E(y))^2=3.75+1.75^2=0.6875$

ו. לפי הנוסחא $V(X+Y)=V(X)+V(Y)+2(E(XY)-E(X)E(Y))$

נצטרך למצוא את $E(XY)$

(XY)	1	2	3	4	6	9	
$p(XY)$	$1/6$	$1/3$	$1/3$	0	$1/6$	0	1

$E(XY)=1\cdot 1/6+2\cdot 1/3+3\cdot 1/3+6\cdot 1/6=2\frac{5}{6}$ ולכן : $V(X+Y)=2\frac{5}{6}\cdot 2-4.75=0.9167$

דוגמא : חברת אינטרנט מספקת ללקוחותיה גישה למספר מאגרי מידע. זמן חיפוש במאגר נתון ע״י הנוסחא $T=\frac{N_1N_2}{S}$, כאשר N_1 הוא מספר המילים בשאילתא של הלקוח, N_2 הוא גודל המאגר בו הוא מחפש, ו S מודד את מהירות המחשב בו נערך החיפוש. אם לN_1 יש התפלגות פואסון עם פרמטר λ ו N_2 ו S מתפלגים לפי הטבלאות. מצא את התוחלת והשונות של T בהנחה ש N_1 , N_2 ו S כולם ב״ת

r	5	10	100
$p(N_2=r)$	0.2	0.7	0.1

S	30	60
$p(S=s)$	0.9	0.1

פתרון : $E(T)=E\left(\frac{N_1N_2}{S}\right)$ בגלל שהם ב״ת ו N_1 מתפלג פואסון עם פרמטר λ אזי מתקיים

$$E\left(\frac{N_1N_2}{S}\right)=E(N_1)E(N_2)E\left(\frac{1}{S}\right)\xrightarrow{E(N_1)=\lambda}\lambda E(N_2)E\left(\frac{1}{S}\right)$$
$$E(N_2)=5\cdot 0.2+10\cdot 0.7+100\cdot 0.1=18$$
$$E\left(\frac{1}{S}\right)=\frac{1}{30}0.9+\frac{1}{60}0.1=0.0316$$
$$E(T)=\lambda E(N_2)E\left(\frac{1}{S}\right)=18\cdot 0.0316\lambda=0.57\lambda$$

עבור השונות:

$$V(T)=E\left(T^2\right)-\left(E(T)\right)^2=$$
$$E\left(T^2\right)=E\left(\frac{N_1^2N_2^2}{S^2}\right)=E\left(N_1^2\right)E\left(N_2^2\right)E\left(\left(\frac{1}{S}\right)^2\right)$$
$$E\left(N_2^2\right)=5^2\cdot 0.2+10^2\cdot 0.7+100^2\cdot 0.1=1075$$
$$E\left(\frac{1}{S^2}\right)=\frac{1}{30^2}0.9+\frac{1}{60^2}0.1=0.00102$$
$$E\left(N_1^2\right)=V(N_1)+\left(E(N_1)\right)^2\xrightarrow{N_1\sim P(\lambda)}\lambda+\lambda^2$$
$$E\left(T^2\right)=1075\cdot 0.00102\left(\lambda+\lambda^2\right)=1.0965\left(\lambda+\lambda^2\right)$$
$$V(T)=1.0965\left(\lambda+\lambda^2\right)-\left(0.57\lambda\right)^2$$

דוגמא: הוכח כי אם X ו- Y ב"ת אזי מתקיים $V(X+Y)=V(X)+V(Y)$ (תכונה 7).

פתרון: לפי תכונה 5 של שונות: $V(X)=E\left(X^2\right)+\left(E(X)\right)^2$ מתקיים,

$V(X+Y)=E\left(\left(X+Y\right)^2\right)+\left(E(X+Y)\right)^2$ כעת, בגלל שנתון שהם ב"ת אזי:

$$V(X+Y)=E\left(X^2+2XY+Y^2\right)-\left(E(X)+E(Y)\right)^2$$
$$=\underbrace{E\left(X^2\right)}+E(2XY)+E\left(Y^2\right)-\underbrace{\left(E(X)\right)^2}-2E(X)E(Y)-\left(E(Y)\right)^2$$
$$=V(X)+E(2XY)+\underbrace{E\left(Y^2\right)}-2E(X)E(Y)-\underbrace{\left(E(Y)\right)^2}$$
$$=V(X)+V(Y)+E(2XY)-2E(X)E(Y)\rightarrow V(X)+V(Y)+2E(XY)-2E(X)E(Y)$$
$$V(X)+V(Y)+2\left[E(XY)-E(X)E(Y)\right]\xrightarrow{E(XY)=E(X)E(Y)}V(X)+V(Y)+2\cdot 0=V(X)+V(Y)$$

המעבר האחרון נובע התכונה 5 עבור התוחלת (עמוד 31) - עבור משתנים ב"ת.

דוגמא: הוכח כי אם X ו – Y ב"ת אזי $E\left(h(X)k(Y)\right)=E\left(h(X)\right)E\left(k(Y)\right)$ (תכונה 6 עבור תוחלת)

פתרון: לפי הגדרת התוחלת, נקבל $E\left(h(X)k(Y)\right)=\sum_r\sum_s h(r)k(s)\,p(X=r,Y=s)$ בגלל שנתון כי X ו – Y ב"ת אזי נוכל להפריד את $p(X=r,Y=s)$ ל- $p(X=r)\,p(Y=s)$

$$E\left(h\left(X\right)k\left(Y\right)\right)=\sum_{r}\sum_{s}h\left(r\right)k\left(s\right)p\left(X=r\right)p\left(Y=s\right)$$

$$=\underbrace{\sum_{r}h\left(r\right)p\left(X=r\right)}_{E\left(h\left(X\right)\right)}\underbrace{\sum_{s}k\left(s\right)p\left(Y=s\right)}_{E\left(k\left(Y\right)\right)}$$

הסקה סטטיסטית:

ישנם שלושה סוגים של הסקות סטטיסטיות.

1. **אמידה נקודתית** – משערכים עבור הפרמטר ערך מסוים
2. **אמידת רווח** – רווח סמך בו נמצא הפרמטר בביטחון מסוים המוגדר מראש
3. **בדיקת השערות** – מכריעים בין השערות על הערך של הפרמטר

הגדרה פורמאלית: פרמטר של פונקצית התפלגות של מ.מ (או של אוכלוסיה) הוא מספר הנקבע על-ידי התפלגות זו, ומהווה אחד המאפיינים שלה.

כאשר אנו אומרים פרמטר אנו מתכוונים לתוחלת, לשונות, ל- P כאשר מדובר בהתפלגות בינומית, ל- λ כאשר מדובר בהתפלגות אקספוננציאלית וכו.

נציין כי פרמטר אינו משתנה מקרי, אלא מספר קבוע. למשל, תוחלת האוכלוסייה לא משתנה ממדגם למדגם בניגוד לממוצע שאכן משתנה ממדגם למדגם כיוון שתלוי במדגם עצמו, התוחלת הינה ממוצע של כל האוכלוסייה עצמה ולכן אין היא משתנה!

כיצד נאמוד פרמטר? נעשה זאת בעזרת סטטיסטי.

סטטיסטי - הינו פונקציה של תצפיות המדגם. אם את הפרמטר נסמן ב- θ אזי הסטטיסטי שאמור לאמוד אותו – לשערך אותו נסמן ב- $\hat{\theta}$.

כמו שאמרנו, הסטטיסטי הינו משתנה מקרי – המשתנה ממדגם למדגם, באופן תיאוריתי, נוכל לדגם מספר רב של מדגמים ובכל מדגם לחשב את הסטטיסטי – למשל נוכל לחשב בכל מדגם את הממוצע. כתוצאה מכך, נקבל ממוצעים רבים שהם בעצמם מהווים התפלגות, כלומר, אם ניקח את כל הממוצעים שקבלנו ונצייר עבורם מצולע שכיחות נקבל את התפלגותם של ממוצעים אלו. התפלגות מיוחדת זו נקראת **התפלגות דגימה** של הממוצע. ברור שביכולתנו ליצור התפלגות דגימה של כל סטטיסטי שנרצה. ברור כי עבור כל מדגם שבו נאמוד את θ על-ידי $\hat{\theta}$ יש שגיאה של: $|\hat{\theta}-\theta|$.

1. אמידה נקודתית

ברצוננו למצוא סטטיסטי המשערך בצורה הטובה ביותר את הפרמטר. נשאלות השאלות, כיצד נמצא אומדנים סבירים? איך נגדיר אומדן טוב?

עבור השאלה הראשונה: ישנן שתי שיטות העוזרות למציאת מועמדים לאומדנים:

1. שיטת שחזור המומנטים
2. שיטת מירב הנראות

עבור נושא זה לא נפרט במסגרת הקורס.

עבור השאלה השנייה: מה מוגדר כאומדן טוב?

1. **קריטריון ראשון חוסר הטיה** - נדרוש כי האומדן יהיה חסר הטיה – אנו רוצים שבממוצע, האומדן יהיה שווה לפרמטר. במילים אחרות שההטיה תהיה שווה ל-0 .

כעת, נתן הגדרה מתמטית:

אומד חסר הטיה: הגדרה פורמאלית: סטטיסטי $\hat{\theta}$ ייקרא אומד חסר הטיה לפרמטר θ אם $E\left(\hat{\theta}\right)=\theta$.

אומדן מסוים יקרא אומד חסר הטיה, אם תוחלת התפלגות הדגימה שלו מתלכדת עם הפרמטר הנאמד. נוכל גם לומר כי אם אומד מסוים חסר הטיה אזי תוחלת השגיאה שווה תמיד ל-0. הטיה מוגדרת באופן הבא: $|E\left(\hat{\theta}\right)-\theta|=b\left(\hat{\theta}\right)$

1.א. אומד חסר הטיה לתוחלת:

אומד חסר הטיה ידוע לתוחלת הינו הממוצע. נוכיח.

נרצה להראות כי מתקיים $E\left(\bar{X}\right)=\mu$ עבור כל x בעל התפלגות כלשהי ו- n כלשהו.

$$E\left(\bar{X}\right)\xrightarrow{\bar{X}=\frac{\sum x_i}{n}}E\left(\frac{\sum x_i}{n}\right)=\frac{1}{n}\sum\left(E\left(x\right)\right)\xrightarrow{E(x)=\mu}\frac{1}{n}\sum\mu=\frac{1}{n}\cdot n\mu=\mu$$

2. **קריטריון שני – תוחלת שגיאה ריבועית קטן כמה שניתן**

MSE- תוחלת השגיאה הריבועית של האומדן $\hat{\theta}$ עבור θ הוא: $MSE\left(\hat{\theta}\right)=E\left[\left(\hat{\theta}-\theta\right)^2\right]$. במילים קצת יותר פשוטות, אנו שואלים כמה רחוק הסטטיסטי שבחרנו מהפרמטר האמיתי – מהי הסטייה\שגיאה. אנו מעלים בריבוע על מנת לקבל ערך חיובי.

2. רווח סמך

כמו שאמרנו המטרה העיקרית הינה לאמוד את הפרמטר שאינו ידוע, למשל אמידת התוחלת, אומדן הגיוני לכך יהיה הממוצע. אך נרצה לדעת מהו טיב אומדן זה, ז"א עד כמה הוא קרוב לתוחלת האמיתית, עד כמה הוא אומדן טוב משמע נרצה לדעת את טיב הדיוק שלו. שאלה זו קשה, אך נוכל למצוא את ההסתברות לכך שסטיית האומדן מן הפרמטר גדולה ממספר יחידות מסוים שנחליט עליו.

2.א. רווח סמך לתוחלת כאשר השונות ידועה

מהסעיף הקודם הגענו למסקנה כי $\bar{X}$ הנו אומד לפרמטר התוחלת μ. על פי התהליך שתיארנו, תהליך התפלגות הדגימה, ישנו משתנה מקרי $\bar{X}$ הלא הוא וקטור של הרבה ערכי $\bar{x}$ שיתקבלו מאינסוף מדגמים. ברור כי אם נתבונן בערך מסוים $\bar{x}$ - ז"א תוצאת הממוצע במדגם אחד מכלל אינסוף המדגמים, סביר להניח כי ערך זה יהיה שונה מהפרמטר μ הלא הוא התוחלת האמיתית אותה אנו רוצים לאמוד. משמעות הדבר כי ההפרש בין $\bar{x}$ לבין μ יהיה שונה מ-0 או במילים אחרות שגיאת

האמידה שונה מ-0. ממשפט הגבול המרכזי קבלנו כי התפלגות הממוצע, $\bar{X}$, הינה נורמאלית בקירוב ובעלת ממוצע μ וסטיית תקן $\frac{\sigma}{\sqrt{n}}$.

כאשר אנו מעוניינים למצוא רווח בר סמך, אנו בעצם מחפשים את ההסתברות לכך שממוצע המדגם שהתקבל בניסוי שלנו יהיה שונה מהפרמטר האמיתי בלא יותר ממספר יחידות המוגדר מראש.

עוד הסבר, בהנתן שקבלנו $\bar{x}$ מסוים מתוך מדגם שדגמנו אזי μ שהיא התוחלת האמיתית, תמצא ברווח בר הסמך בביטחון של 95%. כל זאת כמובן כי אנו יודעים את התפלגותו של $\bar{X}$ מתוך משפט הגבול המרכזי.

נוכל לדמיין כי אם נבצע מספר רב של דגימות ועבור כל דגימה נחשב את הממוצע (ניצור התפלגות דגימה של הממוצע) אזי סביר להניח כי מספר רב של ממוצעים יהיו קרובים לתוחלת האמיתית μ

נדגיש כי רווח הסמך הינו רווח מקרי כיוון שתלוי בממוצע שהתקבל (כמו שנאמר, עבור כל מדגם נקבע ערך אחר של ממוצע)

כעת נציג את הנוסחא לרווח הסמך : $\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$.

כיצד נקבע רווח זה?

לפי משפט הגבול המרכזי ומהדיון עד כה, הסקנו כי התפלגות דגימת הממוצע הינה $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ כעת, אם ניקח שטח סימטרי מסוים סביב הממוצע נקבל רווח סמך ברמת ביטחון של השטח שנלקח. למשל אם נרצה רווח סמך ברמת ביטחון של 95%, אזי ניקח שטח סימטרי של 0.95 סביב הממוצע שמתקבל במדגם. באופן זה נהיה בטוחים ב95% כי התוחלת נמצאת ברווח זה, וזאת בהנתן ממוצע מסוים שהתקבל במדגם שלנו, ברור כי מדגם זה צריך להיות גדול מספיק כך שהתוצאה שנציג תהיה טובה. נציג דיון והסבר למציאת דוגל מדגם בהמשך. כעת נרצה להסביר כיצד הגענו לנוסחא המוצגת לרווח סמך

$$P\left(-z_{\frac{\alpha}{2}} < z_{\bar{X}} < z_{\frac{\alpha}{2}}\right) \xrightarrow{z_{\bar{X}} = \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}} P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

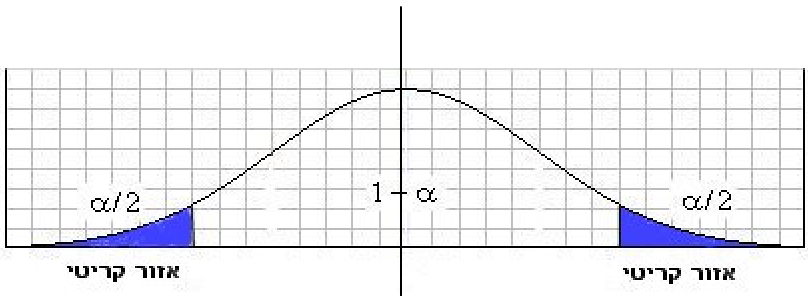
כאשר $1-\alpha = 0.95 \rightarrow \alpha = 0.05$

$z_{\bar{X}} = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ התקבל כתוצאה מהתקנון

$\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right) \rightarrow \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$. נוכל להסתכל על הנוסחא באופן הבא

$\bar{X} = \sigma/\sqrt{n} \cdot z_{\bar{X}} + \mu$ ולהסיק מידית כי כאשר גודל המדגם גדל כך השבר $\sigma/\sqrt{n}$ קטן ולכן $\bar{X}$ מתקרב יותר ויותר ל- μ.

μ



$\downarrow$ $\downarrow$

$\dfrac{\alpha}{2} = 0.025$ $\underbrace{\qquad\qquad\qquad}_{0.95}$ $\dfrac{\alpha}{2} = 0.025$

האזור הכחול מתקבל בשטח של $\dfrac{\alpha}{2} = 0.025$ כך שנשאר לנו סביב תוחלת הממוצע שטח של 0.95.

אם נרצה באופן קבוע את רווח הסמך סביב 95% אזי נוכל לכתוב אותו כך :

$$P\left(-z_{0.025} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{0.025}\right) = \underbrace{1-0.05}_{0.95} \xrightarrow{z_{0.025}=1.96} \bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}$$

נסכם ונאמר כי רמת הדיוק באמידת התוחלת באמצעות הממוצע נמדדת בעזרת רווח סמך. ומלכתחילה נצטרך להניח שתי הנחות :

1. סטיית התקן σ של ה- מ.מ או של האוכלוסייה ידועה
2. המדגם גדול (מעל 30) או שהמשתנה הנבדק בעל התפלגות נורמאלית

נוכל להסיק כי כאשר יש בידינו מידע על התפלגות הסטטיסטי נוכל בקלות לבנות רווח סמך לפרמטר. כמובן שהבעיה מתחילה כאשר אין אנו יודעים את התפלגות הפרמטר.

2.א.1. קביעת גודל המדגם – שגיאת האמידה

בדיון הקודם נאמר כי רווח המסך הינו רווח מקרי כיון שערך הממוצע משתנה מניסוי לניסוי, בניגוד לכך ברור כי אורך רווח הסמך אינו משתנה מניסוי לניסוי (ניתן להסיק זאת, בקלות, מהנוסחא לרווח סמך) :

אורך רווח הסמך:

בהינתן רווח הסמך: $\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$. אורכו שווה:

$$L = \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} - \left(\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) \rightarrow \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} - \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

כעת ברור כי האורך עצמו, $L = 2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$, אינו תלוי ב $\bar{X}$ ולכן אינו מקרי. אם כך, ניתן לקבוע מראש את אורכו של רווח זה וזאת עבור רמת ביטחון שאנו קובעים מראה.

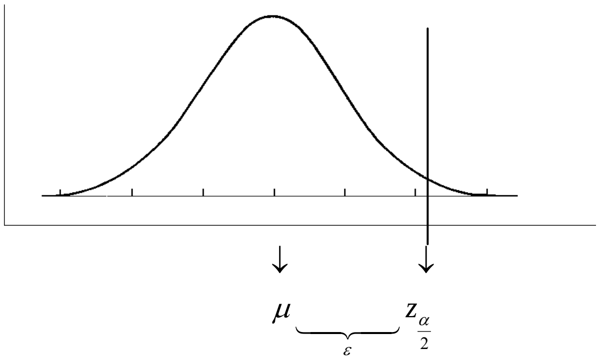
כעת, נבנה נוסחא לגודל מדגם רצוי: נתחיל מכך שאנו רוצים להבטיח כי שגיאת האמידה ז"א $|\bar{X} - \mu|$ לא תעלה על ערך מסוים שאנו קובעים מראש, נסמן ערך זה באות ε. נוכל לומר כי המרחק בין האומד $\bar{X}$ מן הפרמטר μ לא יעלה על הערך מסוים ε. ובאופן הסתברותי, נרצה שהמרחק בין האומד $\bar{X}$ מן הפרמטר μ יהיה קטן מ ε, וזאת בהסתברות של $1-\alpha$. נראה זאת בנוסחאות:

$P\left(|\bar{X} - \mu| < \varepsilon\right) = 1-\alpha \rightarrow P\left(\bar{X} - \varepsilon < \mu < \bar{X} + \varepsilon\right) = 1-\alpha$ תחילה נחלק בסטיית התקן על מנת לקבל התפלגות נורמאלית סטנדרטית

$$P\left(|\bar{X} - \mu| < \varepsilon\right) = 1-\alpha \xrightarrow{:\sigma/\sqrt{n}} P\left(\frac{-\varepsilon}{\sigma/\sqrt{n}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = 1-\alpha,$$

ביטוי זה שווה $\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) - \left(1 - \Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right)\right) = 2\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) - 1$ כאשר Φ היא ההתפלגות הנורמאלית הסטנדרטית מכאן, $2\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) - 1 = 1-\alpha$ נעביר אגפים ונקבל, $\Phi\left(\frac{\varepsilon}{\sigma/\sqrt{n}}\right) = 1 - \frac{\alpha}{2}$ ולכן: $\frac{\sqrt{n}\varepsilon}{\sigma} = z_{1-\alpha/2}$, כעת, נוכל לחלץ את n הלא הוא גודל המדגם:

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2$$



ראשית, נצטרך לבחור כרצוננו את הריחוק של התוחלת מהממוצע. כמובן שנרצה שריחוק זה יהיה קטן. לאחר מכן נבחר את רמת הביטחון.

דוגמאות:

דוגמא 1: נתון כי סטיית התקן שווה ל- $\sigma = 0.45$. במדגם בגודל 50 התקבל ממוצע של 1.2 חשב רווח סמך לתוחלת:

א. ברמת ביטחון של 95%

ב. ברמת ביטחון של 99%

ג. איזה רווח סמך רחב יותר ומדוע?

פתרון: גודל המדגם גדול מספיק כדי שנוכל לומר כי מדובר בהתפלגות דגימה עבור נורמאלית עבור הממוצע. התפלגות האוכלוסייה: $X \sim N\left(?, 0.2025\right)$ ולכן

$$\bar{X} \sim N\left(?, \frac{0.2025}{50}\right)$$

א. נשתמש בנוסחא לרווח סמך:

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \xrightarrow{n=50,\ \sigma=0.45,\ \alpha=0.05,\ \bar{X}=1.2} 1.2 - z_{\frac{0.05}{2}} \frac{0.45}{\sqrt{50}} < \mu < 1.2 + z_{\frac{0.05}{2}} \frac{0.45}{\sqrt{50}}$$

$$1.2 - z_{0.025} \frac{0.45}{\sqrt{50}} < \mu < 1.2 + z_{0.025} \frac{0.45}{\sqrt{50}} \xrightarrow{z_{0.025}=1.96} 1.2 - 1.96 \frac{0.45}{\sqrt{50}} < \mu < 1.2 + 1.96 \frac{0.45}{\sqrt{50}}$$

$$\xrightarrow{1.96\frac{0.45}{\sqrt{50}}=0.1247} 1.2 - 0.1247 < \mu < 1.2 + 0.1247 \rightarrow 1.075 < \mu < 1.324$$

ז"א שבמדגם בעל גודל של 50 מתוך משתנה מקרי זה התפלגות הדגימה היא נורמאלית שסטיית התקן שלו שווה 0.45 ותוחלתו אינה ידועה, ההסתברות שהתוחלת תהיה שונה ממוצע המדגם בלא יותר מ 0.1247 יחידות היא 0.95.

ב.

$$\bar{X}-z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\bar{X}+z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\xrightarrow{n=50,\ \sigma=0.45,\ \alpha=0.01,\ \bar{X}=1.2}1.2-z_{\frac{0.01}{2}}\frac{0.45}{\sqrt{50}}<\mu<1.2+z_{\frac{0.01}{2}}\frac{0.45}{\sqrt{50}}$$

$$1.2-z_{0.005}\frac{0.45}{\sqrt{50}}<\mu<1.2+z_{0.005}\frac{0.45}{\sqrt{50}}\xrightarrow{z_{0.005}=2.58}1.2-2.58\frac{0.45}{\sqrt{50}}<\mu<1.2+2.58\frac{0.45}{\sqrt{50}}$$

$$\xrightarrow{2.58\frac{0.45}{\sqrt{50}}=0.164}1.2-0.164<\mu<1.2+0.164\rightarrow 1.035<\mu<1.364$$

ג. נוכל לראות כי ככל שרמת הסמך עולה כך מתרחב רווח הסמך. ככל שנרצה להגדיל את הביטחון ($1-\alpha$) בהימצאות הערך האמיתי של התוחלת, μ, בתוך הרווח, נצטרך להגדיל את הרווח. באותו אופן נוכל לומר כי אם הרווח צר יותר כן גדלה ההסתברות שהוא לא יכיל את התוחלת, μ, כלומר רמת הסמך פוחתת.

דוגמא 2: על מנת לאמוד רווח סמך לפרמטר μ ברמת ביטחון של 95%, נדגמו 169 תצפיות. הממוצע שהתקבל הוא 24. עוד ידוע כי שונות האוכלוסייה היא 12.25. חשב את אורך רווח הסמך.

פתרון: על מנת לחשב את אורך רווח הסמך נשתמש בנוסחא : $L=2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$.

נדגיש כי מותר לנו להשתמש בנוסחא זו כיוון שאנו מניחים **התפלגות נורמאלית להתפלגות דגימה** עבור התוחלת בגלל גודל המדגם הגדול, 169.

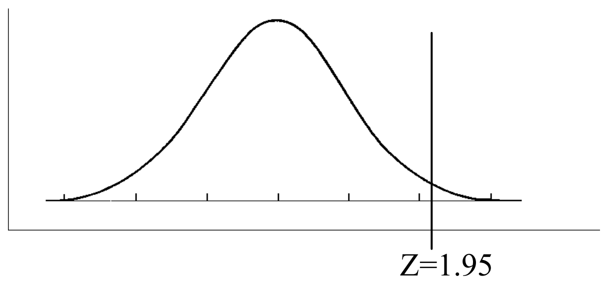
$$L=2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\xrightarrow{z_{\frac{0.05}{2}}=1.96\ ,\ \sigma=3.5,\ n=169}L=2\cdot 1.96\frac{3.5}{\sqrt{169}}=1.055$$

דוגמא 2: באוכלוסיה מסוימת ידוע כי סטיית התקן היא 28. על מנת לאמוד רווח סמך נדגם עבור הפרמטר μ מדגם בגודל 121. מהי רמת הביטחון- רמת הסמך $(1-\alpha)$ אם ידוע כי רווח הסמך שהתקבל הינו: $83.036<\mu<92.963$?

פתרון: כמו בשאלה הקודמת, גם פה נוכל להניח התפלגות נורמאלית. נתון לנו רווח הסמך ולכן נוכל למצוא את אורכו: $83.036<\mu<92.963\xrightarrow{orech=L}L=92.963-83.036=9.927$. כעת, אנו יודעים לפי הנוסחא כי אורך מקיים: $L=2z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\xrightarrow{L=9.927\ ,\ \sigma=28\ ,\ n=121}9.927=2z_{\frac{\alpha}{2}}\frac{28}{\sqrt{121}}\rightarrow z_{\frac{\alpha}{2}}\approx 1.95$. כעת, נשאר לחפש את השטח המתאים ל- Z הנ"ל. ז"א שנמצא את השטח משמאל. לפי הטבלה ל-Z זה מתאים השטח: 0.9744 ולכן ה- $\frac{\alpha}{2}$ (השטח מימין) שווה ל-

$\frac{\alpha}{2}=1-0.9744=0.0256$. ולכן $\alpha=0.0512$. על מנת לדעת את רמת הביטחון נרצה

למצוא את הביטוי $1-\alpha \xrightarrow{\alpha=0.0512} 1-0.0512=0.9488$.



דוגמא 1: אם נרצה לדעת כמה תצפיות עלינו לדגום עבור אוכלוסייה בעלת שונות השווה ל- 100 ברמת ביטחון 0.9 כך שהממוצע שקיבלנו במדגם לא יהיה רחוק מהתוחלת ביחידה אחת בדיוק. נקבל כי:

תמיד $n \ge \left(\frac{z_{\frac{\alpha}{2}}\cdot\sigma}{\varepsilon}\right)^2$. $n \ge \left(\frac{z_{\frac{\alpha}{2}}\cdot\sigma}{\varepsilon}\right)^2 \xrightarrow{\varepsilon=1,\ z_{\frac{0.1}{2}}=1.645,\ \sigma=10} n \ge \left(\frac{1.645\cdot 10}{1}\right)^2 \to n \ge 270.6025$

מעגלים כלפי מעלה – במקרה שלנו: 271.

ניתן לראות כי ככל שנגדיל את גודל המדגם כך נקבל אורך קצר יותר. נובע ישירות מהנוסחא $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}=\varepsilon$ אם נגדיל את n השבר $z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ יקטן ולכן האורך

$\frac{L}{2}=\varepsilon=\bar{X}-\mu$ יקטן. כך שיש לנו יותר מידע על הפרמטר μ באוכלוסיה.

דוגמא 2: מהו גודל המדגם שיש לדגום על מנת להבטיח ששגיאת האמידה לא תעלה על 0.001 כאשר נתון כי $\sigma=0.5$

1. ברמת סמך של 95%
2. ברמת סמך של 99%
3. מה יקרה לאורך הרווח במידה ונגדיל את שגיאת האמידה.

פתרון:

.1

$$n \ge \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2 \xrightarrow{\varepsilon=0.001,\ z_{0.025}=1.96\ ,\ \sigma=0.5} n \ge \left(\frac{1.96 \cdot 0.5}{0.001}\right)^2 \to n \ge 960400$$

.2 $$n \ge \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon}\right)^2 \xrightarrow{\varepsilon=0.001,\ z_{0.005}=2.575\ ,\ \sigma=0.5} n \ge \left(\frac{2.575 \cdot 0.5}{0.001}\right)^2 \to n \ge 1657.6$$

3. גודל המדגם יקטן

מקדם המתאם של פירסון:

מדדי קשר:כאשר אנו עורכי מחקר, אחת השאלות שבדרך כלל עולות היא מציאת קשר בין משתנים. למשל, נרצה למצוא האם יש קשר בין גובה האב לגובה הבן, נרצה למצוא קשר בין ציוני הבגרות לציון הפסיכומטרי, נרצה למצוא קשר בין הכנסתו הכלכלית של אדם לבין הסיכוי שלא יוכל לעמוד בהלוואתו בבנק, נרצה למצוא קשר בין ציון בקורס סטטיסטיקה 1 לציון בסטטיסטיקה 2 ועוד.

מציאת קשר זה הינה חשובה ביותר, למשל בדוגמה האחרונה, אם נדע כי קיים קשר בין רמת הכנסה מסוימת להחזרת הלוואה אז נוכל לנבא, באמצעות ההכנסה החודשית את הסיכוי שאדם מסוים יוכל להחזיר את הלוואתו לבנק. כך יוכל הבנק למנוע בעיות רבות.

ישנם הרבה סוגי קשר: ליניארי, פרבולי, מעריכי, לוגריתמי ועוד. אנו מתמקדים בקשר הליניארי.

כיצד נוכל למצוא קשר וניבוי?

ראשית אנו מתבססים על ידע העבר, למשל בדוגמת הקשר בין ציון בקורס סטטיסטיקה 1 לציון בסטטיסטיקה 2. המדד שבאמצעותו נחשב קשר נקרא : מקדם המתאם של פירסון וסימונו - r

על מנת לחשב את r נצטרך לדעת לחשב את השונות המשותפת של שני משתנים.

שונות משותפת מסומנת : $\text{cov}(x,y)$ או S_{xy} מוגדרת באופן הבא: $\text{cov}(x,y) = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)}{n}$

ניתן להשתמש בנוסחא חלופית : $\text{cov}(x,y) = \frac{\sum_{i=1}^{n}\left(x_i\right)\left(y_i\right)}{n} - \bar{X} \cdot \bar{Y}$.

r - מוגדר להיות כיחס בין השונות המשותפת למכפלת סטיות התקן של X ו- Y. נכתוב זאת כנוסחא:

$$r = \frac{\text{cov}(x,y)}{s_x \cdot s_y} = \frac{\sum_{i=1}^{n}\left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)}{n \cdot s_x \cdot s_y}$$

טענה: מקדם המתאם של פירסון נע בין 1 ל- -1.

הוכחה: כאשר הקשר בין X ל- Y חיובי ומושלם, הרי שמתקיים $y_i = b \cdot x_i + a$ לכל x_i ואנו כבר יודעים כי כאשר ישנה טרנספורמציה ליניארית אזי מתקיים: $s_y = |b| \cdot s_x$ וגם $s_y^2 = b^2 \cdot s_x^2$. כמו כן אנו יודעים כי $\overline{Y} = b\overline{X} + a$

ואז נוכל לרשום את $r = \frac{\mathrm{cov}(x,y)}{s_x \cdot s_y} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{n \cdot s_x \cdot s_y}$ באופן הבא:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{n \cdot s_x \cdot s_y} \xrightarrow{y_i = b\cdot x_i + a\ ,\ \overline{Y} = b\overline{X} + a} r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left[b \cdot x_i + a - \left(b\overline{X} + a\right)\right]}{n \cdot s_x \cdot s_y}$$

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left[b \cdot x_i + a - b\overline{X} - a\right]}{n \cdot s_x \cdot s_y} \rightarrow r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left[b \cdot x_i - b\overline{X}\right]}{n \cdot s_x \cdot s_y}$$

$$\xrightarrow{gorem\ \ meshutaf\ \ b} r = \frac{b\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left[x_i - \overline{X}\right]}{n \cdot s_x \cdot s_y} \xrightarrow{s_y = |b| \cdot s_x\ ,\ s_y^2 = b^2 \cdot s_x^2} r = \frac{b\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n \cdot s_x \cdot |b| \cdot s_x}$$

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^2}{n \cdot s_x^2} = \frac{s_x^2}{s_x^2} = 1$$

כאשר הקשר מושלם אך שלילי כל ההבדל בהוכחה הוא בעובדה כי צריך להציב $-b$ ובאותו אופן נקבל תשובה סופית של 1-.

משפט 1: טרנספורמציות ליניארית חיוביות על המשתנים אינן משנות את ערכו של מקדם המתאם של פירסון.

הוכחה:

$$r=\frac{\operatorname{cov}(x',y')}{s'_x\cdot s'_y}=\frac{\sum_{i=1}^{n}\left(x'_i-\bar{X}'\right)\left(y'_i-\bar{Y}'\right)}{s'_x\cdot s'_y\, n}\xrightarrow{x'=b\cdot x_i+a\quad ,\quad y'=c\cdot y_i+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\bar{X}'\right)\left[c\cdot y_i+d-\left(\bar{Y}'\right)\right]}{s'_x\cdot s'_y\, n}$$

$$\xrightarrow{\bar{X}'=b\cdot\bar{X}+a\ ,\ \bar{Y}'=c\cdot\bar{Y}+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\left(b\cdot\bar{X}+a\right)\right)\left[c\cdot y_i+d-\left(c\cdot\bar{Y}+d\right)\right]}{s'_x\cdot s'_y\, n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-b\cdot\bar{X}-a\right)\left[c\cdot y_i+d-c\cdot\bar{Y}-d\right]}{s'_x\cdot s'_y\, n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i-b\cdot\bar{X}\right)\left[c\cdot y_i-c\cdot\bar{Y}\right]}{s'_x\cdot s'_y\, n}\xrightarrow{gorem\ meshutaf\ \ b\ \&\ c}\frac{bc\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[y_i-\bar{Y}\right]}{bcs_x\cdot s_y n}=\frac{\operatorname{cov}(x,y)}{s_x\cdot s_y}=r$$

משפט 2 : טרנספורמציות ליניאריות על שני המשתנים משנות את ערך השונות המשותפת באופן הבא : $x'=b\cdot x_i+a\quad ,\quad y'=c\cdot y_i+d\quad\rightarrow\operatorname{cov}(x',y')=b\cdot c\cdot\operatorname{cov}(x,y)$ כאשר b ו- c חיובים.

הוכחה: נתון כי ישנן טרנספורמציות ליניאריות על המשתנים : x ו- y באופן הבא : $x'=b\cdot x_i+a\quad ,\quad y'=c\cdot y_i+d$ ולכן מתקיים :

$$\operatorname{cov}(x',y')=\frac{\sum_{i=1}^{n}\left(x'_i-\bar{X}'\right)\left(y'_i-\bar{Y}'\right)}{n}\xrightarrow{x'=b\cdot x_i+a\quad ,\quad y'=c\cdot y_i+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\bar{X}'\right)\left[c\cdot y_i+d-\left(\bar{Y}'\right)\right]}{n}$$

$$\xrightarrow{\bar{X}'=b\cdot\bar{X}+a\ ,\ \bar{Y}'=c\cdot\bar{Y}+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\left(b\cdot\bar{X}+a\right)\right)\left[c\cdot y_i+d-\left(c\cdot\bar{Y}+d\right)\right]}{n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-b\cdot\bar{X}-a\right)\left[c\cdot y_i+d-c\cdot\bar{Y}-d\right]}{n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i-b\cdot\bar{X}\right)\left[c\cdot y_i-c\cdot\bar{Y}\right]}{n}\xrightarrow{gorem\ meshutaf\ \ b\ \&\ c}\frac{bc\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[y_i-\bar{Y}\right]}{n}=b\cdot c\cdot\operatorname{cov}(x,y)$$

משפט 3 : שונות משותפת של משתנה עם עצמו שווה לשונות שלו.

הוכחה : $s_x^{\,2} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left(x_i-\bar{X}\right)}{n} = \operatorname{cov}(x,x)\cdot$

משפט 4 : כאשר נרצה למצוא מקדם מתאם של פירסון של משתנה עם עצמו נמצא כי $r=1$

הוכחה : $1 = \frac{s_x^{\,2}}{s_x^{\,2}} \xleftarrow{\frac{\sum_{i=1}^{n}(x_i-\bar{X})^2}{n}=s_x^{\,2}} \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n\cdot s_x^{\,2}} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left(x_i-\bar{X}\right)}{n\cdot s_x\cdot s_x} = \frac{\operatorname{cov}(x,x)}{s_x\cdot s_x} = r$

דוגמא חישובית : נתונות 10 תצפיות עבור 10 סטודנטים, כך שעבור כל סטודנט ידועים ציוני הקורס בסטטיסטיקה 1 וציוני הקורס בסטטיסטיקה 2. מנתונים אלו נרצה לדעת האם יש ביניהם קשר. מהסתכלות על הנתונים ניתן לראות כי קיים קשר (סטודנט שקיבל ציון גבוה בסטטיסטיקה 1 גם קיבל ציון גבוה בסטטיסטיקה 2).

i	Stat1	Stat2
1	65.00	68.00
2	85.00	88.00
3	84.00	95.00
4	72.00	78.00
5	71.00	75.00
6	69.00	72.00
7	62.00	76.00
8	54.00	66.00
9	81.00	91.00
10	77.00	100.00

1. מצא את הממוצע וסטיית התקן של כל אחד מהמשתנים Stat1 ו- Stat2

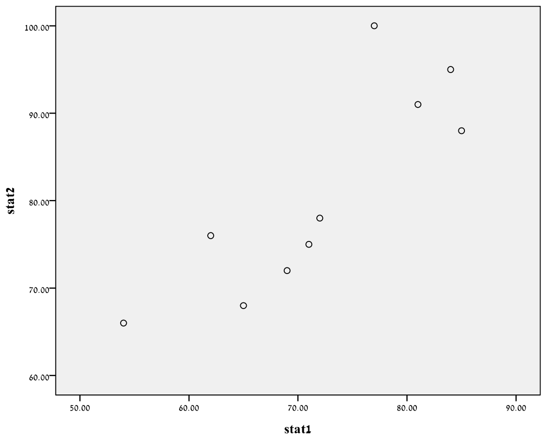
סטטיסטיקה תיאורית בסיסית :

Descriptive Statistics

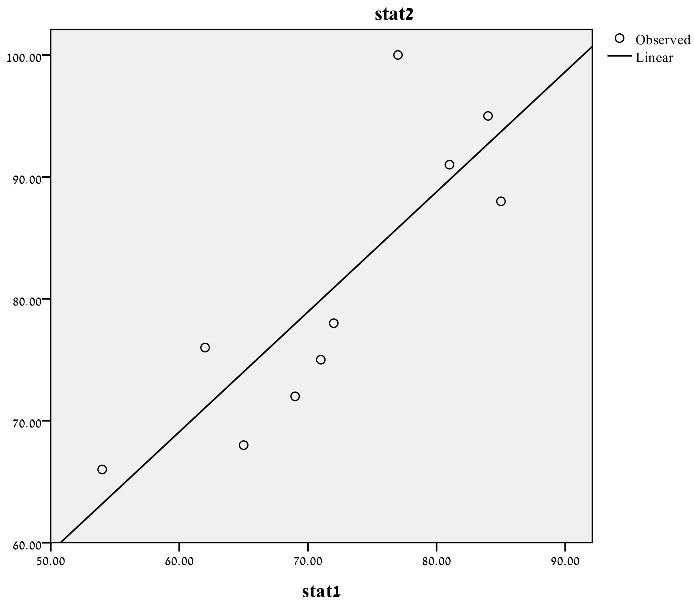
	N	Minimum	Maximum	Mean	Std. Deviation
stat1	10	54.00	85.00	72.0000	10.01110
stat2	10	66.00	100.00	80.9000	11.78935
Valid N (listwise)	10				

2. נוכל לראות על–ידי דיאגרמת פיזור כי ישנו קשר חיובי.

צייר דיאגרמת פיזור :



נוכל לראות כי ניתן להעביר קו ישר, באופן הבא :



3. חשב את מקדם המתאם של פירסון :

Correlations

		stat1	stat2
stat1	Pearson Correlation	1	.836**
	Sig. (2-tailed)		.003
	N	10	10
stat2	Pearson Correlation	.836**	1
	Sig. (2-tailed)	.003	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

4. מצא את המשתנים המתוקננים :

i	Stat1- מתוקנן	Stat2- מתוקנן
1	-0.7	-1.09
2	1.30	0.6
3	1.20	1.2
4	0	-0.25
5	-0.1	-0.5
6	-0.3	-0.75
7	-1.00	-0.42
8	-1.80	-1.26
9	0.9	0.86
10	0.5	1.62

5. מצא את הממוצע וסטיית התקן של המשתנים המתוקננים :

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
metuknan1	10	-1.80	1.30	.0000	1.00000
metuknan2	10	-1.26	1.62	.0000	1.00000
Valid N (listwise)	10				

6. מצא את מקדם המתאם של פירסון עבור המשתנים המתוקננים :

Correlations

		metuknan1	metuknan2
metuknan1	Pearson Correlation	1	.836**
	Sig. (2-tailed)		.003
	N	10	10
metuknan2	Pearson Correlation	.836**	1
	Sig. (2-tailed)	.003	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

7. מדוע ערכי המתאם יצאו שווים בערכם?

משתנים מתוקננים, הם בעצם המשתנים המקוריים עם טרנספורמציה ליניארית. לפי משפט 1 אנו יודעים כי מקדם המתאם של פירסון אינו משתנה בערכו כאשר מופעלת על המשתנים טרנספורמציה ליניארית.

8. מהו המתאם בין ציוני סטטיסטיקה 1 לציוני סטטיסטיקה 1 המתוקננים? האם תוכל לחזות זאת (לא על ידי חישוב)?

Correlations

		metuknan1	stat1
metuknan1	Pearson Correlation	1	1.000**
	Sig. (2-tailed)		.000
	N	10	10
stat1	Pearson Correlation	1.000**	1
	Sig. (2-tailed)	.000	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

9. מהו המתאם בין ציוני סטטיסטיקה 2 לציוני סטטיסטיקה 1 המתוקננים? האם תוכל לחזות זאת (לא על ידי חישוב)?

Correlations

		metuknan1	stat2
metuknan1	Pearson Correlation	1	.836**
	Sig. (2-tailed)		.003
	N	10	10
stat2	Pearson Correlation	.836**	1
	Sig. (2-tailed)	.003	
	N	10	10

**. Correlation is significant at the 0.01 level (2-tailed).

10. לציוני סטטיסטיקה 2 נעשתה טרנספורמציה הבאה: $stat2_{new} = 0.75 \cdot stat2 + 15$.

א. חשב את ערכו של r בין המשתנים $stat2_{new}$ ו- $stat2$. תשובה: 1

ב. חשב את היחס : $\frac{\text{cov}(stat2_{new}, stat1)}{\text{cov}(stat2, stat1)}$. תשובה: 0.75

ג. מה המתאם בין $stat2_{new}$ ל- $stat1$. תשובה: 0.836

עוצמת הקשר: עד כה אמרנו כי r נע בין המספרים 1 ל- -1 וכאשר r מקבל את הערכים האלו אזי יש קשר מושלם: כאשר $r=1$ אזי הקשר מושלם חיובי וכאשר $r=-1$ אזי יש קשר מושלם שלילי. במציאות כמעט ואין מצב של קשרים מושלמים.

נהוג להגדיר את עוצמות הקשר באופן הבא:

כאשר $0<r<0.4$ אז נאמר כי הקשר חיובי חלש

כאשר $0.4<r<0.7$ אז נאמר כי הקשר חיובי בינוני

כאשר $0.7<r<1$ אז נאמר כי הקשר חיובי חזק

כאשר $-0.4<r<0$ אז נאמר כי הקשר שלילי חלש

כאשר $-0.7<r<-0.4$ אז נאמר כי הקשר שלילי בינוני

כאשר $-1<r<-0.7$ אז נאמר כי הקשר שלילי חזק

רגרסיה ליניארית פשוטה

המטרה העיקרית שעומדת לפנינו כעת הינה מבנית מודל ליניארי המנבא את ערכי Y באמצעות X.

כיצד נוכל לנבא משהו שאינו ידוע?

אכן קשה ובעייתי לנבא משהו שאינו ידוע, אך אם זאת נוכל לעשות את הדבר הבא: נוכל ללמוד מן העבר ולנסות לנבא את העתיד, כך שאם לעבר ישנה התנהגות עקבית אזי אנו עשויים לחשוב שכך תמשיך התופעה להתנהל.

למשל, בהנתן כי אנו מעוניינים לנבא האם לקוח חדש בבנק יוכל או לא יוכל להחזיר הלוואה שברצונו לקחת – האם הבנק יאפשר לו מתן של הלוואה?, מה עלינו לעשות? נוכל לקחת נתוני עבר רבים ולנסות ללמוד את דפוס ההתנהגות של נתונים אלו, למשל, נוכל ליצור מטריצה בה מרוכזים מספר רב של תצפיות – אנשים שלקחו בעבר הלוואה ואינפורמציה עבורם, למשל:

קידוד: עבור לקוח שהחזיר הלוואה -1, עבור לקוח שלא החזיר הלוואה -0.

ותק: נספר בשנים

מין: זכר – 0, נקבה – 1.

מין	ותק	מספר נכסים	משכורת ממוצעת	החזיר\לא החזיר	i
0	10	2	15000	1	1
0	9	3	17000	1	2
0	2	1	6800	0	3
0	5	0	7420	0	4
1	4	1	16000	1	$\vdots$ n

דבר ראשון נרצה לבדוק האם ישנו קשר בין המשתנה "החזיר/לא החזיר" לשאר המשתנים: "משכורת ממוצעת", "מספר נכסים", "ותק" ו "מין". כמו שהגדרנו מראש אנו מדברים על קשר ליניארי שזהו הנושא שלנו, אם ישנו קשר אחר אזי נצטרך לבנות מודל מתאים אחר.

נמשיך בדוגמא, אם למשל מצאנו כי ישנו קשר ליניארי חזק בין המשתנה "החזיר/לא החזיר" למשתנה "משכורת ממוצעת" (כמובן שיכול להיות קשר חזק גם עם שאר המשתנים אך לצורך הדוגמא והנוחות ניקח רק משתנה אחד) אז ננסה לבנות מודל מהצורה הבאה: $y = \beta x + \alpha$ כאשר x הינו המשתנה "משכורת ממוצעת" ו- y הינו המשתנה "החזיר/לא החזיר" באופן זה כאשר יגיע לקוח חדש, נכניס את הנתון של המשכורת הממוצעת שלו וכך ננבא עוד לפני שניתן לו הלוואה, האם יוכל להחזירה או לאו. כמובן שהקושי בכל העניין הזה הוא מציאת המקדם β, אותו ערך מיוחד שמביע את הקשר בין שני המשתנים.

סימונים:

משתנה מנבא, נקרא גם משתנה בלתי תלוי: X

משתנה מנובא, נקרא גם משתנה תלוי: Y

ניתן לראות את המטרה שלנו גם באופן הבא: אנו מנסים להסביר את התופעה Y באמצעות התופעה X, אנו מנסים לברר מה גורם להשתנות מסוימת של הגורם Y ומנסים לתת נימוק לשינוי זה ע"י משתנים המשפיעים עליו. מאחר שאנו בודקים קשר ליניארי חזק בין X ל-Y אזי נצפה כי משוואת קו ישר תתאר בצורה הטובה ביותר את הקשר בין X ל- Y. השתמשנו במושג "הטובה ביותר" כיוון שקשה לנבא את Y באמצעות X ללא טעויות, אנו טוענים לנכונות משוואת הישר כאשר כמות הטעויות תהיה מינימאלית

המשוואה הליניארית המשוערת המקשרת בין X ל- Y היא: $y_i = \alpha + \beta x_i + \varepsilon_i$ כאשר:

x_i - הערך המנבא

y_i - הערך המנבא

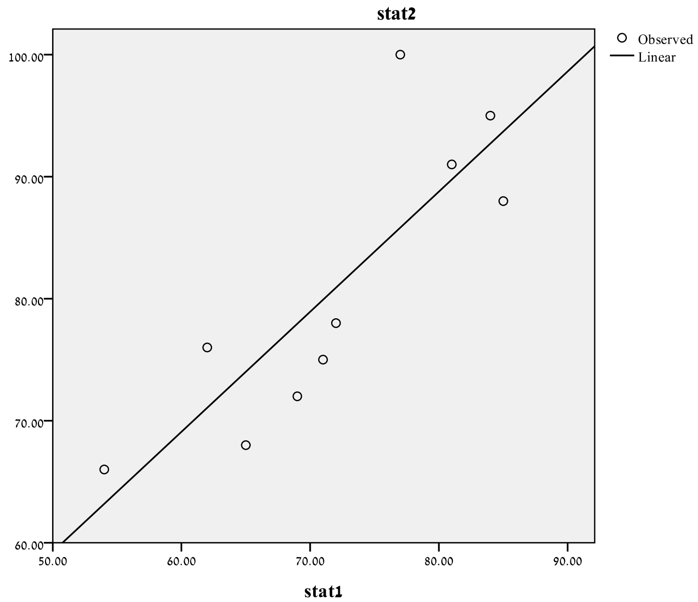
ε_i - טעות הניבוי עבור התצפית ה - i

α - קבוע הרגרסיה (החותך של ציר ה- Y)

β - שיפוע הרגרסיה

כעת, נרצה להתמקד בטעות הניבוי, ε_i, ונעמיק בהסבר:

במציאות, הקשר הליניארי אינו מושלם, למשל, נזכר בדוגמא על ציוני הסטודנטים בסטטיסטיקה 1 ובסטטיסטיקה 2:



נוכל לראות כי הזוגות X וY שהתקבלו מהמדגם נופלות סביב הישר, נוכל לומר שהמרחקים מהישר מהווים את השונות סביב הישר. ישר זה הינו קו הרגרסיה שלנו. כעת כאשר יגיע סטודנט חדש ובידיו ציונו בסטטיסטיקה 1 נוכל מיד לנבא את ציונו בסטטיסטיקה 2 על הצבת ציונו בסטטיסטיקה 1 במשוואת הישר, כמובן שעשוי להיות פער מהמציאות, בגלל העובדה כי ישר זה נבנה על סמך העיקרון של מינימום שגיאות (של תצפיות רבות מן העבר), אך עדיין הן קיימות. לטעות זאת או פער זה אנו קוראים טעות הניבוי ומסמנים אותו ב- ε_i.

מהו קו הרגרסיה הטוב ביותר?

נסמן את הערך המנובא לפי קו הרגרסיה $\widehat{y_i}$ והערך האמיתי שהתצפית קיבלה מסומן - y_i כמו שאמרנו, נרצה להביא למינימום את גודל הטעויות ועל פי עקרון זה נבנה את קו הרגרסיה שלנו. ננסה כעת, להפוך את המילים לנוסחא: הגדרנו כבר את גודל הטעויות באופן הבא: $\varepsilon_i = y_i - \widehat{y_i}$. אנו רוצים לדבר על כל הטעויות ולא רק על טעות אחת ולכן נרצה לסכום את הטעויות $\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} \left(y_i - \widehat{y_i} \right)$ הבעיה בסכום זה היא שיכול להיווצר מצב בו סוכמים גורם שלילי עם גורם חיובי. על מנת להתגבר על בעיה זו נעלה בריבוע כל טעות ונקבל: $\sum_{i=1}^{n} \varepsilon^2{}_i = \sum_{i=1}^{n} \left(y_i - \widehat{y_i} \right)^2$. קו רגרסיה שנבנה על-ידי עקרון בניה של הבאה למינימום את סכום הטעויות בריבוע נקרא: "**קו הריבועים הפחותים**".

אם כך, כיצד נמצא אומדנים עבור קו הרגרסיה על פי עקרון קו הריבועים הפחותים?

אז כמו שאמרנו נרצה להביא למינימום את סכום הטעויות בריבוע: $\min\sum_{i=1}^{n}\left(y_i-\widehat{y_i}\right)^2$:

על מנת להביא למינימום נצטרך לגזור נגזרות חלקיות לפי α ולפי β. להשוואת ל-0 ולוודא כי האומדנים שקיבלנו הינם מינימאליים.

נבצע את התהליך, תחילה נפתח את הביטוי בתוך הסוגריים:

$$\min\sum_{i=1}^{n}\left(y_i-\widehat{y_i}\right)^2 \xrightarrow{\widehat{y_i}=\alpha+\beta x_i} \sum_{i=1}^{n}\left(y_i-\underbrace{\left(\alpha+\beta x_i\right)}_{\widehat{y_i}}\right)^2$$

$$\sum_{i=1}^{n}\left(y_i-\alpha-\beta x_i\right)^2=\sum_{i=1}^{n}\left[\left(y_i-\alpha-\beta x_i\right)\left(y_i-\alpha-\beta x_i\right)\right]$$

$$\sum_{i=1}^{n}\left[y^2{}_i-y_i\alpha-y_i\beta x_i-y_i\alpha+\alpha^2+\alpha\beta x_i-y_i\beta x_i+\alpha\beta x_i+\beta^2x^2{}_i\right]$$

$$\sum_{i=1}^{n}\left[y^2{}_i-2y_i\alpha-2y_i\beta x_i+\alpha^2+2\alpha\beta x_i+\beta^2x^2{}_i\right]$$

$$\sum_{i=1}^{n}y^2{}_i-2\alpha\sum_{i=1}^{n}y_i-2\beta\sum_{i=1}^{n}y_ix_i+\sum_{i=1}^{n}\alpha^2+2\beta\alpha\sum_{i=1}^{n}x_i+\beta^2\sum_{i=1}^{n}x^2{}_i$$

$$\sum_{i=1}^{n}y^2{}_i-2\alpha\sum_{i=1}^{n}y_i-2\beta\sum_{i=1}^{n}y_ix_i+n\alpha^2+2\beta\alpha\sum_{i=1}^{n}x_i+\beta^2\sum_{i=1}^{n}x^2{}_i$$

כעת, נגזור לפי α :

$$\left[\min\sum_{i=1}^{n}\left(y_i-\widehat{y_i}\right)^2\right]' = \left[\ \sum_{i=1}^{n}y^2{}_i-2\alpha\sum_{i=1}^{n}y_i-2\beta\sum_{i=1}^{n}y_ix_i+n\alpha^2+2\beta\alpha\sum_{i=1}^{n}x_i+\beta^2\sum_{i=1}^{n}x^2{}_i\right]'$$

$$=-2\sum_{i=1}^{n}y_i+2n\alpha+2\beta\sum_{i=1}^{n}x_i\ =\ -2\underbrace{\frac{\sum_{i=1}^{n}y_i}{n}}_{\bar{Y}}n+2n\alpha+2\beta\underbrace{\frac{\sum_{i=1}^{n}x_i}{n}}_{\bar{X}}n\xrightarrow{nashve\ \ le\ \ 0}$$

$$-2\bar{Y}n+2n\alpha+2\beta\bar{X}n=0\xrightarrow{nechalek\ \ be\ \ 2n}-\bar{Y}+\alpha+\beta\bar{X}=0$$

$$\alpha=\bar{Y}-\beta\bar{X}$$

כעת, נגזור לפי β :

$$\left[\min\sum_{i=1}^{n}\left(y_i-\widehat{y_i}\right)^2\right]' =$$

$$\left[\ \sum_{i=1}^{n}y^2{}_i-2\alpha\sum_{i=1}^{n}y_i-2\beta\sum_{i=1}^{n}y_ix_i+n\alpha^2+2\beta\alpha\sum_{i=1}^{n}x_i+\beta^2\sum_{i=1}^{n}x^2{}_i\right]'$$

$$=-2\sum_{i=1}^{n}y_ix_i+2\alpha\sum_{i=1}^{n}x_i+2\beta\sum_{i=1}^{n}x^2{}_i\ =-2\frac{\sum_{i=1}^{n}y_ix_i}{n}n+2\alpha\underbrace{\frac{\sum_{i=1}^{n}x_i}{n}}_{\overline{X}}n+2\beta\sum_{i=1}^{n}x^2{}_i$$

$$\xrightarrow{nashve\ \ le\ \ 0\ \ ve\ \ netsamtsem\ \ be2}\ -\frac{\sum_{i=1}^{n}y_ix_i}{n}n+\alpha\underbrace{\frac{\sum_{i=1}^{n}x_i}{n}}_{\overline{X}}n+\beta\frac{\sum_{i=1}^{n}x^2{}_i}{n}n=0$$

$$\xrightarrow{\operatorname{cov}(x,y)=\frac{\sum_{i=1}^{n}y_ix_i}{n}-\overline{X}\cdot\overline{Y}\ \ \rightarrow\ \ \operatorname{cov}(x,y)+\overline{X}\cdot\overline{Y}=\frac{\sum_{i=1}^{n}y_ix_i}{n}}$$

$$\xrightarrow{r=\frac{\operatorname{cov}(x,y)}{s_x\cdot s_y}\rightarrow r\cdot s_x\cdot s_y=\operatorname{cov}(x,y)\ \rightarrow\ r\cdot s_x\cdot s_y+\overline{X}\cdot\overline{Y}=\frac{\sum_{i=1}^{n}y_ix_i}{n}}\ \xrightarrow{s_x{}^2=\frac{\sum_{i=1}^{n}x_i^2}{n}-\overline{X}^2\ \rightarrow\ s_x{}^2+\overline{X}^2=\frac{\sum_{i=1}^{n}x_i^2}{n}}$$

$$-r\cdot s_x\cdot s_y-\overline{X}\cdot\overline{Y}+\alpha\overline{X}+\beta\left(\ s_x{}^2+\overline{X}^2\right)=0\ \xrightarrow{\alpha=\overline{Y}-\beta\overline{X}}$$

$$-r\cdot s_x\cdot s_y-\overline{X}\cdot\overline{Y}+\left(\underbrace{\overline{Y}-\beta\overline{X}}_{\alpha}\right)\overline{X}+\beta\left(\ s_x{}^2+\overline{X}^2\right)=0$$

$$-r\cdot s_x\cdot s_y-\overline{X}\cdot\overline{Y}+\overline{X}\cdot\overline{Y}-\beta\overline{X}^2+\beta s_x{}^2+\beta\overline{X}^2=0$$

$$-r\cdot s_x\cdot s_y+\beta s_x{}^2=0\quad\rightarrow\quad\beta=\frac{r\cdot s_x\cdot s_y}{s_x{}^2}\ \rightarrow\beta=\frac{r\cdot s_y}{s_x}$$

ניתן בקלות להראות שאומדנים אלו מינימאליים.

תכונות קו הריבועים הפחותים:

1. ממוצע הטעויות שווה ל-0
2. ממוצע האומדנים שווה ל- $\bar{Y}$ ($\bar{y} = \bar{\hat{y}}$)
3. הקו עובר דרך נקודת הממוצעים
4. המתאם בין הטעויות ל Y המנובא $\hat{y}$ שווה ל-0 ($r_{\hat{y}\hat{\varepsilon}} = 0$)
5. נוכל לפרק את השונות של Y לשונות מוסברת ולשונות לא מוסברת
6. פרופורציות השונות המוסברת מחושבת ע"י r^2 באופן הבא:

$$r^2 = \frac{\sum_{i=1}^{n}\left(\widehat{y_i} - \bar{Y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2} \rightarrow \frac{shonut\ \ shel\ \ reg}{shonut\ \ shel\ \ y}$$

ובמילים: אחוז השונות של הניבוי מתוך כל השונות של Y

7. פרופורציות השונות הלא מוסברת מחושבת ע"י $1-r^2$ באופן הבא:

$$1-r^2 = \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y_i}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{Y}\right)^2}$$

ובמילים: סכום ריבועי הטעויות מתוך כל השונות של Y.

דוגמא 26: בעקבות עליית המחירים במוצר מסוים הייתה ירידה בכמות המכירות והתקבלו הממצאים הבאים: מקדם המתאם של פירסון = 0.67- . ממוצע המחירים 521 ליום מכירות וסטיית התקן =2.54. וממוצע כמות המכירות ביום 76 פריטים וסטיית התקן שווה ל-6.5.

א. מה עוצמת הקשר בין המחירים לכמות המכירות?

ב. מצא את משוואת הרגרסיה המנבא את כמות המכירות על ידי המחירים.

ג. כאשר המחיר הינו 277 ₪ איזה כמות תמכר?

פתרון :א. עוצמת הקשר בינונית.

ב. על מנת למצוא את משוואת הרגרסיה, עלינו למצוא את $\beta = \frac{r \cdot s_y}{s_x}$, $\alpha = \bar{Y} - \beta\bar{X}$. נציב את הנתונים:

$$\beta = \frac{r \cdot s_y}{s_x} \xrightarrow{r=-0.67\ ,\ s_y=6.5\ ,\ s_x=2.54} \beta = \frac{-0.67 \cdot 6.5}{2.54} = -1.714$$

$$\alpha = \bar{Y} - \beta\bar{X} \xrightarrow{\bar{Y}=76\ \ \bar{X}=521\ \ \beta=-1.714} \alpha = 76 - (-1.714) \cdot 521 = 968.99$$

ולכן משוואת הרגרסיה היא: $\hat{y} = -1.714x + 968.99$

ג. $y=-1.714x+968.99 \xrightarrow{x=277} y_{277}=-1.714\cdot 277+968.99=494.212$

דוגמא 27 חוקר השתמש במדגם של 10 תצפיות לחישוב רגרסיה ליניארית פשוטה לניבוי Y לפי X. הנתונים אבדו לו, אך זכר כי סכום כל ה-X-ים שווה ל-100 וסכום כל ה-Y-ים שווה ל- 250. בנוסף זכר כי, כאשר $x=0$ אז $\hat{y}=0$. מצא את קו הרגרסיה.

פתרון:

$$\sum_{i=1}^{n} X_i = 100 \xrightarrow{:10} \frac{\sum_{i=1}^{n} X_i}{10} = \frac{100}{10} \rightarrow \overline{X} = 10$$

$$\sum_{i=1}^{n} Y_i = 100 \xrightarrow{:10} \frac{\sum_{i=1}^{n} Y_i}{10} = \frac{250}{10} \rightarrow \overline{Y} = 25$$

$$\alpha = \overline{Y} - \beta\overline{X} \xrightarrow{\overline{X}=10\ ,\ \overline{Y}=25} \alpha = 25-10\beta \quad \rightarrow \hat{y} = \beta x + \alpha$$

$$\xrightarrow{\alpha=25-10\beta\ ,} \hat{y} = \beta x + \left(25-10\beta\right) \xrightarrow{\hat{y}=0 \leftrightarrow x=0} 0 = \beta\cdot 0 + \left(25-10\beta\right) \rightarrow \beta = 2.5$$

$$\hat{y} = \beta x + \alpha \xrightarrow{\alpha=25-10\beta\ ,\ \beta=2.5\ \rightarrow\ \alpha=0} \hat{y} = 2.5x + 0 \rightarrow \hat{y} = 2.5x$$

מבחנים לדוגמא:

מבחן לדוגמא 1 – הסתברות – עשינו יחד בכיתה – מצ"ב פתרונות מלאים

1. יהי X מ.מ בעל התפלגות קושי, פונקצית ההצטברות הינה $x \in \mathbb{R} \quad , \quad F(x) = b + c \cdot \arctan\left(\frac{x}{a}\right)$ כאשר a הינו פרמטר חיובי.

א. מצא את הקבועים b ו- c

ב. מצא את פונקצית הצפיפות של X

ג. מצא את ההסתברות $p(-a \le X \le a)$

ד. ללא קשר לסעיפים הקודמים או להגדרת X בראשית השאלה – סעיף נפרד לחלוטין:

נגדיר את X מ.מ, כאשר $X \sim U(0,1)$, נגדיר את Y מ.מ באופן הבא $y = \sin\left(\frac{\pi X}{2}\right)$.

מצא את פונקצית הצפיפות וההצטברות עבור Y כך שערכיו נעים בין 0 ל- 1.

פתרון: א

אנחנו צריכים לחשוב על שתי משוואות כיוון שיש לנו שני נעלמים. ברור כי $F(\infty) = 1$ וגם $F(-\infty) = 0$.

$$\left.\begin{array}{l} F(\infty) = 1 \rightarrow \lim\limits_{x\to\infty}\left[b + c\cdot\arctan\left(\frac{x}{a}\right)\right] \rightarrow b + c\cdot\frac{\pi}{2} = 1 \\ F(-\infty) = 1 \rightarrow \lim\limits_{x\to\infty}\left[b + c\cdot\arctan\left(\frac{x}{a}\right)\right] \rightarrow b - c\cdot\frac{\pi}{2} = 0 \end{array}\right\} \rightarrow c = \frac{1}{\pi} \quad , \quad b = \frac{1}{2}$$

ב. נגזור, $\left(F(x)\right)' = \left(b + c\cdot\arctan\left(\frac{x}{a}\right)\right)' \rightarrow f(x) = c\frac{a}{a^2 + x^2}$, לכן נקבל בסה"כ

$$f(x) = \begin{cases} \frac{1}{\pi}\frac{a}{a^2 + x^2} & x \in \mathbb{R} \\ 0 & otherwise \end{cases}$$

ג. $p(-a \le X \le a) = F(-a) - F(a) = \frac{1}{2} + \frac{1}{\pi}\cdot\arctan\left(\frac{-a}{a}\right) - \left(\frac{1}{2} + \frac{1}{\pi}\cdot\arctan\left(\frac{a}{a}\right)\right) = 2\frac{1}{\pi}\cdot\underbrace{\arctan 1}_{\frac{\pi}{4}} = \frac{1}{2}$

ד. $F_Y(a) = p(Y \le a) = p\left(\sin\left(\frac{\pi x}{2}\right) \le a\right) = p\left(\sin\left(\frac{\pi x}{2}\right) \le a\right) = p\left(\frac{\pi x}{2} \le \arcsin a\right) = p\left(X \le \frac{2}{\pi}\arcsin a\right)$

כעת, בגלל שאנו יודעים כיצד מתפלג X נוכל לחשב את פונקצית ההתפלגות:

$$F_X(a) = p\left(X \le \frac{2}{\pi}\arcsin a\right) = \int_0^{\frac{2}{\pi}\arcsin a} \frac{1}{1-0}dx = \frac{2}{\pi}\arcsin a$$

נוכל לחשב את פונקצית הצפיפות, ע"י גזירה: $f(a) = \left(\frac{2}{\pi}\arcsin a\right)' = \frac{2}{\pi\sqrt{1-a^2}}$ כאשר $0 \le a < 1$.

2. חברת ביטוח מחלקת את עולם הנהגים ל-2 קבוצות – מסוכנים (30%) ובטיחותיים (70%), נהג מסוכן מעורב בתאונה בהסתברות 0.4 ונהג בטיחותי מעורב בהסתברות 0.2.

א. מה ההסתברות שאם בוחרים נהג באקראי הוא יעבור תאונה

ב. נהג מעורב בתאונה, מה ההסתברות שהוא מקבוצת המסוכנים

פתרון: נגדיר את המאורעות הבאים:

A – נהג מעורב בתאונה, A^c - נהג שלא מעורב בתאונה

B – נהג מסוכן, B^c - נהג בטיחותי.

לפי הנתונים: $p(A|B^c)=0.2$, $p(A|B)=0.4$, $p(B^c)=0.7$, $p(B)=0.3$

א. $p(A)=p(A|B)p(B)+p(A|B^c)p(B^c)=0.4\cdot 0.3+0.2\cdot 0.7=0.26$

ב. $P(B|A)=\dfrac{P(A|B)P(B)}{P(A)}=\dfrac{0.4\cdot 0.3}{0.26}=\dfrac{6}{13}$

3. יהיו $Y\sim N(10,144)$ ו $X\sim N(20,169)$ מ.מ בלתי תלויים.

א. כיצד מתפלג $2Y,\ 3X$?

ב. כיצד מתפלג $2Y+3X$?

ג. כיצד מתפלג $2Y+3X-70$?

ד. מצאו $p(2Y+3X-70<10)$.

ה. מצאו $p\left(X(X-Y)\le XY-Y^2+100\right)$.

פתרון:

$$X\sim N(20,169)$$
$$Y\sim N(10,144)$$
$$2Y\sim N(2\cdot 10,4\cdot 144)\to 2Y\sim N(20,576)$$
$$3X\sim N(3\cdot 20,9\cdot 169)\to 3X\sim N(60,1521)$$
$$2Y+3X\sim N(20+60,1521+576)\to N(80,2097)$$
$$2Y+3X-70\sim N(80-70,2097)\to N(10,2097)$$
$$p\left(X(X-Y)\le XY-Y^2+100\right)\to p\left((X-Y)^2\le 100\right)$$
$$W=X-Y\sim N(20-10,169+144)\to W\sim N(10,313)$$
$$p(-10\le W\le 10)$$

כעת על מנת למצוא את ההסתברות המבוקשת, נהפוך להתפלגות נורמאלית סטנדרטית (כיוון שW מתפלג נורמאלית אך לא סטנדרטית ועבור זאת אין טבלת קירובים).

$$W\sim N(10,313)\to p(-10\le W\le 10)$$
$$Z\sim N(0,1)\xrightarrow{tiknun\ to\ w\to\frac{x-\mu}{\sigma}}p(-1.13\le W\le 0)$$

4. א. הוכח את הטענה הבאה: יהיו $X_1, X_2, X_3, \dots X_n \sim NB(n,p)$ ב"ת אזי $\sum_{i=1}^{n} X_i \sim NB(n^2,p)$.

פתרון:

$$E\left(e^{t\sum_{i=1}^{n} X_i}\right) = E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right) = E\left(\prod_{i=1}^{n} e^{tX_i}\right) \xrightarrow{I.D} \prod_{i=1}^{n} E\left(e^{tX_i}\right)$$

$$\xrightarrow{X_1,X_2,X_3,\ldots X_n \sim NB(n,p)} \prod_{i=1}^{n} E\left(e^{tx}\right) \xrightarrow{M_X(t)=E\left(e^{tx}\right)=\left(\frac{p}{1-pe^t}\right)^n} \prod_{i=1}^{n}\left(\frac{p}{1-pe^t}\right)^n = \left(\left(\frac{p}{1-pe^t}\right)^n\right)^n$$

$$= \left(\frac{p}{1-pe^t}\right)^{n^2}$$

קבלנו פונקציה יוצרת מומנטים של התפלגות בינומית שלילית עם הפרמטרים המבוקשים ולכן $\sum_{i=1}^{n} X_i \sim NB(n^2,p)$

ב. טענה: מקדם המתאם של פירסון נע בין 1 ל- 1-.

פתרון:

הוכחה: כאשר הקשר בין X ל- Y חיובי ומושלם, הרי שמתקיים $y_i = b\cdot x_i + a$ לכל x_i ואנו כבר יודעים כי כאשר ישנה טרנספורמציה ליניארית אזי מתקיים: $s_y = |b|\cdot s_x$ וגם $s_y{}^2 = b^2\cdot s_x{}^2$. כמו כן אנו יודעים כי $\bar{Y} = b\bar{X} + a$

ואז נוכל לרשום את $r = \frac{\text{cov}(x,y)}{s_x\cdot s_y} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left(y_i-\bar{Y}\right)}{n\cdot s_x\cdot s_y}$ באופן הבא:

$$r=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left(y_i-\overline{Y}\right)}{n\cdot s_x\cdot s_y}\xrightarrow{y_i=b\cdot x_i+a\ ,\ \overline{Y}=b\overline{X}+a} r=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left[b\cdot x_i+a-\left(b\overline{X}+a\right)\right]}{n\cdot s_x\cdot s_y}$$

$$r=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left[b\cdot x_i+a-b\overline{X}-a\right]}{n\cdot s_x\cdot s_y}\rightarrow r=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left[b\cdot x_i-b\overline{X}\right]}{n\cdot s_x\cdot s_y}$$

$$\xrightarrow{gorem\ \ meshutaf\ \ b} r=\frac{b\sum_{i=1}^{n}\left(x_i-\overline{X}\right)\left[x_i-\overline{X}\right]}{n\cdot s_x\cdot s_y}\xrightarrow{s_y=|b|\cdot s_x\ \ ,\ \ {s_y}^2=b^2\cdot {s_x}^2} r=\frac{b\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n\cdot s_x\cdot |b|\cdot s_x}$$

$$r=\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n\cdot {s_x}^2}=\frac{{s_x}^2}{{s_x}^2}=1$$

כאשר הקשר מושלם אך שלילי כל ההבדל בהוכחה הוא בעובדה כי צריך להציב $-b$ ובאותו אופן נקבל תשובה סופית של 1-.

ג. הוכח כי תוחלת של מ.מ המתפלג בינומית הינה $E\left(X\right)=np$

פתרון: נמצא את התוחלת: $E\left(X\right)=\sum_{r=0}^{n}r\binom{n}{r}p^r\cdot q^{n-r}=E\left(X\right)=\sum_{r=1}^{n}r\binom{n}{r}p^r\cdot q^{n-r}$

השוויון נכון כיוון שהמחובר הראשון בסכום שווה ל-0.

$$\sum_{r=1}^{n}r\frac{n!}{r!\left(n-r\right)!}p^r\cdot q^{n-r}\ =\sum_{r=1}^{n}\frac{n\left(n-1\right)!}{\left(r-1\right)!\left(n-r\right)!}p\cdot p^{r-1}\cdot q^{n-r}$$

$$np\sum_{r=1}^{n}\frac{\left(n-1\right)!}{\left(r-1\right)!\left(n-r\right)!}p^{r-1}\cdot q^{n-r}$$

כעת נבצע הצבה: $s=r-1$ ונקבל:

$$np\sum_{r=1}^{n}\frac{\left(n-1\right)!}{\left(r-1\right)!\left(n-r\right)!}p^{r-1}\cdot q^{n-r}\xrightarrow{s=r-1}np\sum_{s=0}^{n-1}\frac{\left(n-1\right)!}{s!\left(n-s-1\right)!}p^s\cdot q^{n-s-1}$$

נסמן את $n-1=N$ ונקבל: $np\underbrace{\sum_{s=0}^{N}\frac{N!}{s!\left(N-s\right)!}p^s\cdot q^{N-s}}_{1}$ ובסה"כ: $E\left(X\right)=np$

מבחן לדוגמא 2 – הסתברות

1. נתון כי קבוצת תצפיות מתפלגת נורמאלית על ממוצע m וסטית תקן s , מצא קבוע a כך שאחוז התצפיות,

א. בתוך הטווח $m \pm a \cdot s$ היא 75%

ב. הקטנה מ $m - a \cdot s$ היא 22%

פתרון: נרצה לחשב, $p\left(m - as \le X \le m + as\right) = 0.75$ אך תחילה נעשה תקנון $Z = \dfrac{X - m}{s}$ אנו עושים זאת כי אנו יודעים לעבוד רק עם התפלגות נורמאלית סטנדרטית.

$$p\left(\frac{m - as - m}{s} \le \frac{X - m}{s} \le \frac{m + as - m}{s}\right) = p\left(-a \le Z \le a\right) \to \Phi\left(a\right) - \Phi\left(-a\right) =$$

$$\Phi\left(a\right) - \left(\underbrace{1 - \Phi\left(a\right)}_{\Phi(-a)}\right) = 2\Phi\left(a\right) - 1 \ \to 2\Phi\left(a\right) - 1 = 0.75 \ \to \Phi\left(a\right) = 0.875 \ \to \ a = 1.15$$

ב.

$$p\left(X \le m - as\right) = 0.22 \xrightarrow{Z = \frac{X - m}{s}} p\left(\frac{X - m}{s} \le \frac{m - as - m}{s}\right) = 0.22$$

$$p\left(Z \le -a\right) = \Phi\left(-a\right) = 1 - \Phi\left(a\right) \to \Phi\left(-a\right) = 0.78 \to a = 0.77$$

2. 13. נתונה הפונקציה $f\left(x\right) = \dfrac{a}{2}\left(1 - x^2\right), \ 0 < x < 1$

א. מצא את a על מנת שהפונקציה תהיה פונקצית צפיפות לגיטימית.

ב. מצא את פונקצית ההצטברות

ג. נתון התחום של המ.מ Y אם הוא מוגדר באופן הבא: $Y = \dfrac{1}{1 + X}$

ב. מצא את הצפיפות של Y

ג. מצא את התוחלת של Y

פתרון: א. נסתכל על התחום של X ונסיק לגבי Y (הפונקציה המתוארת יורדת תמיד ולכן ערכי הקיצון עבור התחום הסגור מתקבל בקצוות): אם X=0 אזי נקבל $Y = \dfrac{1}{1 + X} \xrightarrow{x=0} 1$. עבור X=1 נקבל: $Y = \dfrac{1}{1 + X} \xrightarrow{x=1} \dfrac{1}{2}$ ולכן: $\dfrac{1}{2} < Y < 1$.

ו. הצפיפות של Y: ראשית נמצא את ההתפלגות, כיוון שלא ידועה את התפלגותו של Y.

$$F_Y(a)=p(Y\le a)\xrightarrow{Y=\frac{1}{1+X}}p\left(\frac{1}{1+X}\le a\right)=p\left(X\ge\frac{1-a}{a}\right)=1-p\left(X<\frac{1-a}{a}\right)=1-F_X\left(\frac{1-a}{a}\right)$$

כעת, נגזור על מנת למצוא את הצפיפות:

$$\left(F_Y(a)\right)'=f(y)=\left(1-F_X\left(\frac{1-a}{a}\right)\right)'=-f_x\left(\frac{1-a}{a}\right)\cdot\left(-\frac{1}{a^2}\right)\xrightarrow{f_x\left(\frac{1-a}{a}\right)=\frac{3}{2}\left(1-\left(\frac{1-a}{a}\right)^2\right)}\frac{3}{a^3}-\frac{3}{2a^4}$$

ז. תוחלת: $E(x)=\int_{\frac{1}{2}}^{1}a\left(\frac{3}{a^2}-\frac{3}{2a^3}\right)da=\left(\frac{-3}{a}+\frac{3}{4a^2}\right)_{\frac{1}{2}}^{1}=\frac{3}{4}$.

ח. על מנת למצוא את השונות נמצא את

$$E\left(x^2\right)=\int_{\frac{1}{2}}^{1}a^2\left(\frac{3}{a^2}-\frac{3}{2a^3}\right)da=\int_{\frac{1}{2}}^{1}3-\frac{3}{2a}da=\left(3a-\frac{3\ln(2a)}{2}\right)_{\frac{1}{2}}^{1}$$

כעת, נציב בנוסחא $V(X)=E\left(X^2\right)-\left(E(X)\right)^2$

3. מנהל חברה מעוניין לשכור 5 מתכנתים. הוא מראיין מועמד אחר מועמד עד שיאייש את חמש המשרות. ההסתברות של כל מועמד להתקבל היא $p=0.3$.

א. מהי התוחלת והשונות של מספר המועמדים שיראיין?

ב. מה ההסתברות שיראיין בדיוק 10 מועמדים?

פתרון:

א. התפלגות בינומית שלילית:

$$p(n)=\binom{n-1}{r-1}p^r(1-p)^{n-r}\xrightarrow{r=5,\quad p=0.3}E(X)=\frac{r}{p}=\frac{5}{0.3}=16.666$$

$$V(X)=\frac{r(1-p)}{p^2}\xrightarrow{r=5,\quad p=0.3}\approx 35$$

ב.מציבים בהתפלגות ומקבלים0.051

4. א. הוכח את הטענה הבאה: יהיו $X_1,X_2,X_3,\ldots X_n\sim\exp(\lambda)$ ב"ת אזי $\sum_{i=1}^{n}X_i\sim Gamma(n,\lambda)$.

פתרון:

$$E\left(e^{t\sum_{i=1}^{n}X_i}\right)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right)=E\left(\prod_{i=1}^{n}e^{tx_i}\right)\xrightarrow{I.D}\prod_{i=1}^{n}E\left(e^{tx_i}\right)$$

$$\xrightarrow{X_1,X_2,X_3,\ldots X_n\sim\exp(\lambda)}\prod_{i=1}^{n}E\left(e^{tx}\right)\xrightarrow{M_X(t)=E\left(e^{tx}\right)=\frac{\lambda}{\lambda-t}}\prod_{i=1}^{n}\frac{\lambda}{\lambda-t}=\left(\frac{\lambda}{\lambda-t}\right)^n$$

קבלנו פונקציה יוצרת מומנטים של התפלגות גמא ולכן $\sum_{i=1}^{n} X_i \sim Gamma(n,\lambda)$

ב. הוכח: $V(X)=E\left(\left(X-E(X)\right)^2\right)=E\left(X^2\right)-\left(E(X)\right)^2$

פתרון:

$$V(X)=E\left(\left(X-E(X)\right)^2\right)=E\left(X^2-2XE(X)+\left(E(X)\right)^2\right)=E\left(X^2\right)-2E(X)E(X)+E\left(\left(E(X)\right)^2\right)$$
$$=E\left(X^2\right)-2\left(E(X)\right)^2+\left(E(X)\right)^2=E\left(X^2\right)-\left(E(X)\right)^2$$

ט. הוכח כי תוחלת של מ.מ המתפלג פואסון עם פרמטר λ הינה $E(X)=\lambda$

פתרון:

$$E(X)=\sum_{r=0}^{\infty} r\cdot\frac{\lambda^r}{r!}e^{-\lambda}\xrightarrow{if\ r=0\ \to\ r\cdot\frac{\lambda^r}{r!}e^{-\lambda}=0}\sum_{r=1}^{\infty} r\cdot\frac{\lambda^r}{r!}e^{-\lambda}=\sum_{r=1}^{\infty} r\frac{\lambda^r}{r(r-1)!}e^{-\lambda}=\sum_{r=1}^{\infty}\frac{\lambda^r}{(r-1)!}e^{-\lambda}$$

נעשה הצבה: $r-1=s$ ונקבל:

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

מבחן לדוגמא 3

1. יהי מ.מ רציף בעל פונקצית צפיפות הבאה:

$$f(x)=\begin{cases} c(\cos x) & |x|\le\frac{\pi}{2} \\ 0 & \text{otherwise}\end{cases}$$

ד. מצא את c

ה. מצא את פונקצית ההתפלגות

ו. מצא את $p\left(|X|<\frac{\pi}{4}\right)$

פתרון:

א. נבדוק עבור איזה ערך של c מתקיים התנאי השלישי: $\int_{-\infty}^{\infty} f(x)dx=1$

$$\int_{-\infty}^{\infty} f(x)dx=1\to\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} c(\cos x)dx=\to\left(c(\sin x)\right)_{-\frac{\pi}{2}}^{\frac{\pi}{2}}=c\left(\underbrace{\sin\left(\frac{\pi}{2}\right)}_{1}-\underbrace{\sin\left(-\frac{\pi}{2}\right)}_{-1}\right)=1\to c=\frac{1}{2}$$

ב.

$$F(x)=(X\le x)=\int_{-\infty}^{x} f(x)\,dx=\frac{1}{2}\int_{-\frac{\pi}{2}}^{x}\cos x\,dx=\left(\frac{1}{2}(\sin x)\right)_{-\frac{\pi}{2}}^{x}=\frac{1}{2}\left(\sin x-\underbrace{\sin\left(-\frac{\pi}{2}\right)}_{-1}\right)=\frac{1}{2}(\sin x+1)$$

ולכן נוכל להגדיר את פונקצית ההתפלגות באופן הבא:

$$F(x)=\begin{cases}1 & x>\frac{\pi}{2}\\ \frac{1}{2}(\sin x+1) & |x|\le\frac{\pi}{2}\\ 0 & x<-\frac{\pi}{2}\end{cases}$$

ג.

$$p\left(|X|<\frac{\pi}{4}\right)=p\left(-\frac{\pi}{4}<X<\frac{\pi}{4}\right)=F\left(\frac{\pi}{4}\right)-F\left(-\frac{\pi}{4}\right)\xrightarrow{F\left(-\frac{\pi}{4}\right)=1-F\left(\frac{\pi}{4}\right)}F\left(\frac{\pi}{4}\right)-\left(1-F\left(\frac{\pi}{4}\right)\right)=$$

$$2F\left(\frac{\pi}{4}\right)-1=2\left(\frac{1}{2}\left(\sin\frac{\pi}{4}+1\right)\right)-1=\frac{\sqrt{2}}{2}$$

ניתן כמובן לחשב ע"י פונקצית הצפיפות: $\frac{1}{2}\int_{-\frac{\pi}{4}}^{\frac{\pi}{4}}\cos x\,dx=\left(\frac{1}{2}(\sin x)\right)_{-\frac{\pi}{4}}^{\frac{\pi}{4}}=\frac{\sqrt{2}}{2}$

2 בקניה בסופר קונים לחם או פיתות. ההסתברות לקנית לחם היא 0.6. אם קונים לחם אז מצרפים לו חמאה או גבינה כאשר ההסתברות לקנית חמאה היא 0.2. אם קונים פיתות אז מוסיפים חמאה, גבינה או חומוס. ההסתברות לקנית חמאה היא 0.3, לקנית גבינה היא 0.4 ולקנית חומוס היא 0.3. בקנית גבינה יש הסתברות של 0.6 לקנית גבינה לבנה ו 0.4 לקנית גבינה צהובה. בקנית חמאה יש הסתברות של 0.7 לחמאה בלי מלח ו - 0.3 לחמאה עם מלח. ענו על השאלות הבאות: (א) מה ההסתברות לקנית חמאה? (ב) מה ההסתברות לקנית חמאה ללא מלח? (ג) אם ידוע שנקנתה גבינה צהובה מה ההסתברות שנקנה לחם? (ד) אם ידוע שנקנה חומוס מה ההסתברות שנקנו פיתות?

פתרון: א. ההסתברות לקניית חמאה הינה $p(butter)=0.6\cdot0.2+0.4\cdot0.3=0.24$

ב. ההסתברות לקניית חמאה ללא מלח

$p(butter\quad w.o.\quad salt)=0.6\cdot0.2\cdot0.7+0.4\cdot0.3\cdot0.7=0.168$

ג. אם ידוע שנקנתה גבינה צהובה, מהי ההסתברות שנקנה לחם?

$$p(bread \mid yellow\,chease)=\frac{p(bread\cap yellow\,chease)}{p(yellow\,chease)}=\frac{0.192}{0.256}=0.75$$

$$p(yellow\,chease)=0.4\cdot0.4\cdot0.4+0.6\cdot0.8\cdot0.4=0.256$$

$$p(bread\cap yellow\,chease)=0.6\cdot0.8\cdot0.4=0.192$$

ד. אם ידוע שנקנה חומוס, מהי ההסתברות שנקנו פיתות? ברור כי התשובה היא 1, כיוון שחומוס ניתן לאכול רק כאשר קונים איתו פיתה.

$$p(pita \mid hoomoos)=\frac{p(pita\cap hoomoos)}{p(hoomoos)}=\frac{0.12}{0.12}=1$$

$$p(hoomoos)=0.4\cdot0.3=0.12$$

$$p(pita\cap hoomoos)=0.4\cdot0.3=0.12$$

3. א. הוכח את משפט – אי שוויון צ'בישב.

לכל מ.מ X בעל תוחלת μ ושונות σ^2 ולכל מספר חיובי k מתקיים

$$p\left(\mu-k\sigma<X<\mu+k\sigma\right)\ge1-\frac{1}{k^2}$$

נוכיח עבור מ.מ בדיד:

ראשית נכתוב את האי שיוויון באופן הבא

$$p\left(|X-\mu|<k\sigma\right)\ge1-\frac{1}{k^2}\;\rightarrow\;p\left(\left(X-\mu\right)^2<k^2\sigma^2\right)\ge1-\frac{1}{k^2}\xrightarrow{\text{complete}}p\left(\left(X-\mu\right)^2\ge k^2\sigma^2\right)\le\frac{1}{k^2}$$

מספיק להוכיח את האי שיוויון האחרון ובכך סיימנו.

במקום σ^2 נוכל לכתוב $\sigma^2=E\left(\left(X-\mu\right)^2\right)\rightarrow\sigma^2=\sum_i p_i\cdot\left(x_i-\mu\right)^2$ כעת, ננסה להקטין את הסכום $\sum_i p_i\cdot\left(x_i-\mu\right)^2$ על ידי הפחתה של כל התצפיות עבורם מתקיים $\left(x_i-\mu\right)^2<k^2\sigma^2$ ולכן נקבל $\sigma^2\ge\sum_i{}^{*}p_i\cdot\left(x_i-\mu\right)^2$ כאשר האיברים $\left(x_i-\mu\right)^2$ מקיימים $\left(x_i-\mu\right)^2\ge k^2\sigma^2$: (אם השמטנו את כל המקיימים $\left(x_i-\mu\right)^2<k^2\sigma^2$ אזי נשארנו עם האיברים המקיימים $\left(x_i-\mu\right)^2\ge k^2\sigma^2$) אם ננסה לסכם מה קבלנו:

$\sigma^2\ge\sum_i{}^{*}p_i\cdot\left(x_i-\mu\right)^2\xrightarrow{\left(x_i-\mu\right)^2\ge k^2\sigma^2}\ge\sum_i{}^{*}p_i\cdot k^2\sigma^2=k^2\sigma^2\sum_i{}^{*}p_i$, כאשר מתקיים

$\sum_i^* p_i = p\left(\left(X-\mu\right)^2 \ge k^2\sigma^2\right)$, כעת נציב ב $\sigma^2 \ge k^2\sigma^2 \sum_i^* p_i$ ונקבל:

$$\sigma^2 \ge k^2\sigma^2 \sum_i^* p_i \xrightarrow{\sum_i^* p_i = p\left(\left(X-\mu\right)^2 \ge k^2\sigma^2\right)} \sigma^2 \ge k^2\sigma^2 p\left(\left(X-\mu\right)^2 \ge k^2\sigma^2\right)$$

$$\frac{\sigma^2}{k^2\sigma^2} \ge p\left(\left(X-\mu\right)^2 \ge k^2\sigma^2\right) \to \frac{1}{k^2} \ge p\left(\left(X-\mu\right)^2 \ge k^2\sigma^2\right)$$

ב. יהיו משתנים מקריים Z_i כך שלכל $i=1\ldots n$ מתקיים $Z_i \sim N(0,1)$ מצא כיצד מתפלג

X אם מוגדר באופן הבא: $X = \dfrac{7\sum_{i=8}^{20}(Z_i)^2}{13\sum_{i=1}^{7}(Z_i)^2}$

פתרון :נשתמש בשני משפטים:

3. אם $Z_1, Z_2, Z_3, \ldots Z_k$ מ.מ-ים ב"ת כך ש $\forall i = 1\ldots k$ $Z_i \sim N(0,1)$ אזי $\sum_i^k Z_i^2 \sim \chi_k^2$. לכן כל סכום המוצג במשתנה המקרי מתפלג חי בריבוע עם מספר דרגות חופש כמספר המחוברים: $\sum_{i=8}^{20}(Z_i)^2 \sim \chi_{13}^2$ ו $\sum_{i=1}^{7}(Z_i)^2 \sim \chi_7^2$

4. יהי מ.מ U המתפלג χ_m^2 ויהי מ.מ V המתפלג χ_n^2. מ.מ אלו הם ב"ת. אזי למ.מ X המוגדר $\frac{U/m}{V/n}$ יש התפלגות F עם m,n דרגות חופש. נסמן $X \sim F_{m,n}$

עם קצת אלגברה בסיסית... נסיק כי הביטוי המוצג בתרגיל מתפלג $X \sim F_{13,7}$ נסו!!

4. א. הוכח את הטענה הבאה: יהיו $X_1 \sim\sim Poisson(\lambda_1),\ X_2 \sim Poisson(\lambda_2), \ldots, X_n \sim Poisson(\lambda_n)$ ב"ת כיצד יתפלג $\sum_{i=1}^{n} X_i$ ומהם הפרמטרים המתאימים?

פתרון: נשים לב כי לא מדובר במשתנים שווי התפלגות!!!!

$$E\left(e^{t\sum_{i=1}^{n}X_i}\right)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right)=E\left(\prod_{i=1}^{n}e^{tX_i}\right)\xrightarrow{I.D}\prod_{i=1}^{n}E\left(e^{tX_i}\right)$$

$$\xrightarrow{X_1\sim\sim Poisson(\lambda_1),\ X_2\sim Poisson(\lambda_2),\ldots,X_n\sim Poisson(\lambda_n)}\prod_{i=1}^{n}E\left(e^{tX_i}\right)$$

$$\xrightarrow{M_X(t)=E\left(e^{tx}\right)=e^{\lambda(e^t-1)}}e^{\lambda_1(e^t-1)}\cdot e^{\lambda_2(e^t-1)}\ldots\cdot e^{\lambda_n(e^t-1)}=e^{(e^t-1)\sum_{i=1}^{n}\lambda_i}$$

קבלנו התפלגות פואסון עם פרמטר : $\sum_{i=1}^{n}\lambda_i$ ז״א $\sum_{i=1}^{n}X_i\sim Poisson\left(\sum_{i=1}^{n}\lambda_i\right)$

ב. חוקר השתמש במדגם של 10 תצפיות לחישוב רגרסיה ליניארית פשוטה לניבוי Y לפי X. הנתונים אבדו לו, אך זכר כי סכום כל ה-X-ים שווה ל-100 וסכום כל ה-Y-ים שווה ל- 250. בנוסף זכר כי, כאשר $x=0$ אז $\hat{y}=0$. מצא את קו הרגרסיה.

פתרון :

$$\sum_{i=1}^{n}X_i=100\xrightarrow{:10}\frac{\sum_{i=1}^{n}X_i}{10}=\frac{100}{10}\rightarrow\bar{X}=10$$

$$\sum_{i=1}^{n}Y_i=100\xrightarrow{:10}\frac{\sum_{i=1}^{n}Y_i}{10}=\frac{250}{10}\rightarrow\bar{Y}=25$$

$$\alpha=\bar{Y}-\beta\bar{X}\xrightarrow{\bar{X}=10\ ,\ \bar{Y}=25}\alpha=25-10\beta\quad\rightarrow\hat{y}=\beta x+\alpha$$

$$\xrightarrow{\alpha=25-10\beta\ ,}\hat{y}=\beta x+(25-10\beta)\xrightarrow{\hat{y}=0\leftrightarrow x=0}0=\beta\cdot 0+(25-10\beta)\rightarrow\beta=2.5$$

$$\hat{y}=\beta x+\alpha\xrightarrow{\alpha=25-10\beta\ ,\ \beta=2.5\ \rightarrow\ \alpha=0}\hat{y}=2.5x+0\rightarrow\hat{y}=2.5x$$

ז. הוכח כי תוחלת של מ.מ המתפלג פואסונית הינה $E(X)=\lambda$

$$E(X)=\sum_{r=0}^{\infty}r\cdot\frac{\lambda^r}{r!}e^{-\lambda}\xrightarrow{if\ r=0\ \rightarrow\ r\cdot\frac{\lambda^r}{r!}e^{-\lambda}=0}\sum_{r=1}^{\infty}r\cdot\frac{\lambda^r}{r!}e^{-\lambda}=\sum_{r=1}^{\infty}r\frac{\lambda^r}{r(r-1)!}e^{-\lambda}=\sum_{r=1}^{\infty}\frac{\lambda^r}{(r-1)!}e^{-\lambda}$$

נעשה הצבה : $r-1=s$ ונקבל :

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

מבחן לדוגמא 4

1. יהי X מ.מ בעל פונקצית צפיפות $|x|<a$, $f(x)=\dfrac{c}{\sqrt{a^2-x^2}}$ כאשר a הינו פרמטר חיובי.

ד. מצא את הקבוע c

ה. מצא את פונקצית ההצטברות של X

ו. חשב את ההסתברויות $p(0<X<a)$ ו- $p\left(|X|\le\dfrac{a}{2}\right)$

פתרון: א. בגלל שמדובר בפונקצית צפיפות אזי האינטגרל על כל התחום שווה ל-1:

$$\int_{-a}^{a}\frac{c}{\sqrt{a^2-x^2}}dx=1\xrightarrow{\int\frac{1}{\sqrt{a^2-x^2}}dx=\arcsin\frac{x}{a}+c}c\left(\arcsin\frac{x}{a}\right)_{-a}^{a}=c\left(\arcsin 1-\arcsin(-1)\right)$$

$$\xrightarrow{\arcsin(-1)=-\arcsin 1}c\left(2\arcsin 1\right)\xrightarrow{\arcsin 1=\frac{\pi}{2}}=c\pi=1\rightarrow c=\frac{1}{\pi}$$

ב. פונקצית ההצטברות:

$$F(x)=\frac{1}{\pi}\int_{-a}^{x}\frac{1}{\sqrt{a^2-x^2}}dx=\frac{1}{\pi}\left(\arcsin\frac{x}{a}\right)_{-a}^{x}=\frac{1}{\pi}\left(\arcsin\frac{x}{a}-\arcsin(-1)\right)$$

$$\xrightarrow{\arcsin(-1)=-\arcsin 1\ ,\ \ \arcsin 1=\frac{\pi}{2}}\frac{1}{\pi}\left(\arcsin\frac{x}{a}+\frac{\pi}{2}\right)=\frac{1}{\pi}\arcsin\frac{x}{a}+\frac{1}{2}$$

ובסה"כ נקבל:
$$F(x)=\begin{cases}0 & x<-a\\ \dfrac{1}{\pi}\arcsin\dfrac{x}{a}+\dfrac{1}{2} & -a\le x\le a\\ 1 & x>a\end{cases}$$

ג.
$$p(0<X<a)=F(a)-F(0)=\frac{1}{\pi}\underbrace{\arcsin\frac{a}{a}}_{\frac{\pi}{2}}+\frac{1}{2}-\left(\frac{1}{\pi}\underbrace{\arcsin\frac{0}{a}}_{0}+\frac{1}{2}\right)=\frac{1}{2}$$

$$p\left(|X|\le\frac{a}{2}\right)=p\left(-\frac{a}{2}\le X\le\frac{a}{2}\right)=F\left(-\frac{a}{2}\right)-F\left(\frac{a}{2}\right)=\frac{1}{\pi}\underbrace{\arcsin\frac{-1}{2}}_{\frac{\pi}{6}}+\frac{1}{2}-\left(\frac{1}{\pi}\underbrace{\arcsin\frac{1}{2}}_{\frac{\pi}{6}}+\frac{1}{2}\right)=\frac{1}{3}$$

2. בוחרים באופן אקראי ועד בן 3 מתוך 6 עורכי דין ו5 מהנדסים. מה ההסתברות שהועד יהיה מורכב באופנים הבאים:

א. משני עו"ד ומהנדס

ב. מעו"ד אחד לפחות

ג. רק עו"ד

פתרון: א. משני עו"ד ומהנדס $\dfrac{\binom{6}{2}\binom{5}{1}}{\underbrace{\binom{11}{3}}_{\Omega}}$

ב. מעו"ד אחד לפחות, אפשר להגיד בעצם שזה כל האפשרויות האפשריות שהן

ללא עו"ד: $1-\dfrac{\binom{5}{3}\binom{6}{0}}{\binom{11}{3}}$

ג. רק עו"ד $\dfrac{\binom{5}{0}\binom{6}{3}}{\binom{11}{3}}$

3. הוכח את חוק ההסתברות השלמה: אם Ω ניתן לפירוק של איחוד זר של F_i-ים (סופי או אינסופי) אזי $p(A)=\dfrac{p(A|F_i)\,p(F_i)}{\sum_i p(A|F_i)p(F_i)}$

הוכחה: ידוע כי אם יהיו $F_1,F_2,F_3\ldots$ מאורעות זרים, $\forall\ i\neq j$ $F_i\cap F_j=\varnothing$ ומתקיים $\bigcup_i F_i=\Omega$. נתון מאורע מסוים A אזי בהכרח מתקיים:

$A=(F_1\cap A)\cup(F_2\cap A)\cup(F_3\cap A)\ldots$.

לפי הנתון: אם Ω ניתן לפירוק של איחוד זר של F_i-ים (סופי או אינסופי) אזי

$$p(A)=\sum_i p(A|F_i)p(F_i)$$

$$p(F_i\,|\,A)=\frac{p(F_i\cap A)}{p(A)}=\frac{p(A|F_i)\,p(F_i)}{p(A)}$$

$$p(F_i\,|\,A)=\frac{p(A|F_i)\,p(F_i)}{p(A)}\xrightarrow{p(A)=\sum_i p(A|F_i)p(F_i)}=\frac{p(A|F_i)\,p(F_i)}{\sum_i p(A|F_i)p(F_i)}$$

4. א. הוכח את הטענה הבאה: יהיו $X_1,X_2,X_3,\ldots X_n\sim N(0,\sigma^2)$ ב"ת כיצד מתפלג $\sum_{i=1}^{n}X_i$.

פתרון:

$$E\left(e^{t\sum_{i=1}^{n}X_i}\right)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right)=E\left(\prod_{i=1}^{n}e^{tx_i}\right)\xrightarrow{I.D}\prod_{i=1}^{n}E\left(e^{tx_i}\right)$$

$$\xrightarrow{X_1,X_2,X_3,\ldots X_n\sim N\left(0,\sigma^2\right)}\prod_{i=1}^{n}E\left(e^{tx}\right)\xrightarrow{M_X(t)=\exp\left[\frac{1}{2}\sigma^2t^2\right]}\prod_{i=1}^{n}e^{\frac{1}{2}\sigma^2t^2}=\left(e^{\frac{1}{2}\sigma^2t^2}\right)^n$$

$$=e^{\frac{1}{2}n\sigma^2t^2}$$

קבלנו פונקציה יוצרת מומנטים של התפלגות נורמאלית עם הפרמטרים המבוקשים ולכן

$$\sum_{i=1}^{n}X_i\sim N\left(0,n\sigma^2\right)$$

ב. הוכח כי טרנספורמציות ליניאריות על שני המשתנים משנות את ערך השונות המשותפת באופן הבא: $x'=b\cdot x_i+a \quad , \quad y'=c\cdot y_i+d \quad \rightarrow \operatorname{cov}(x',y')=b\cdot c\cdot \operatorname{cov}(x,y)$ כאשר b ו- c חיובים.

הוכחה: נתון כי ישנן טרנספורמציות ליניאריות על המשתנים: x ו- y באופן הבא: $x'=b\cdot x_i+a \quad , \quad y'=c\cdot y_i+d$ ולכן מתקיים:

$$\operatorname{cov}(x',y')=\frac{\sum_{i=1}^{n}\left(x'_i-\bar{X}'\right)\left(y'_i-\bar{Y}'\right)}{n}\xrightarrow{x'=b\cdot x_i+a \quad , \quad y'=c\cdot y_i+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\bar{X}'\right)\left[c\cdot y_i+d-\left(\bar{Y}'\right)\right]}{n}$$

$$\xrightarrow{\bar{X}'=b\cdot\bar{X}+a \quad , \quad \bar{Y}'=c\cdot\bar{Y}+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\left(b\cdot\bar{X}+a\right)\right)\left[c\cdot y_i+d-\left(c\cdot\bar{Y}+d\right)\right]}{n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-b\cdot\bar{X}-a\right)\left[c\cdot y_i+d-c\cdot\bar{Y}-d\right]}{n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i-b\cdot\bar{X}\right)\left[c\cdot y_i-c\cdot\bar{Y}\right]}{n}\xrightarrow{gorem \ meshutaf \ b\ \&\ c}\frac{bc\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[y_i-\bar{Y}\right]}{n}=b\cdot c\cdot\operatorname{cov}(x,y)$$

כאשר הקשר מושלם אך שלילי כל ההבדל בהוכחה הוא בעובדה כי צריך להציב $-b$ ובאותו אופן נקבל תשובה סופית של -1.

ג. הוכח כי תוחלת של מ.מ המתפלג גיאומטרית הינה $E\left(X\right)=\frac{1}{p}$

פתרון: **תוחלת:** $E(X)=\sum_{r=1}^{\infty}p\cdot r\cdot(1-p)^{r-1}\xrightarrow{1-p=q}p\sum_{r=1}^{\infty}r\cdot(q)^{r-1}$ נציב $1-p=a$

$$p\sum_{r=1}^{\infty}r\cdot(q)^{r-1}=p\sum_{r=0}^{\infty}r\cdot q^{r-1}$$

כעת, נעצור לרגע, ונתבונן בעובדה הידועה הבאה: ידוע לנו כי $\sum_{r=0}^{\infty}a^r=\frac{1}{1-a}$, אם נגזור את שני האגפים לפי a נקבל: $\sum_{r=0}^{\infty}ra^{r-1}=\frac{1}{(1-a)^2}$. במקרה זהו הסכום שקבלנו ולכן:

$$p\sum_{r=0}^{\infty}r\cdot q^{r-1}\xrightarrow{\sum_{r=0}^{\infty}ra^{r-1}=\frac{1}{(1-a)^2}}\frac{p}{(1-q)^2}\xrightarrow{1-a=p}\frac{p}{p^2}=\frac{1}{p}$$

מועד א' 11.2.2013 משך הבחינה 3.5 שעות

מועד א'

1. יהי X מ.מ בעל פונקצית ההצטברות (30 נקודות)

$$F(x)=\begin{cases}0 & x\le c\\ 1-\left(\frac{c}{x}\right)^a & x>c\end{cases}$$

כאשר $a>0$ ו- $c>0$.

א. מצא את פונקצית הצפיפות של X. (3 נקודות)

ב. מצאת את התוחלת בהנחה ש $a>1$ (5 נקודות)

ג. מצאת את התוחלת בהנחה ש $a<1$ (5 נקודות)

ד. מצא את $p\left(X^2<4c^2\right)$ (8 נקודות)

ה. (סעיף זה לא קשור לסעיפים הקודמים) נגדיר את X להיות מ.מ המתפלג אחידה רציפה עם הפרמטרים 1 ו- d ($X\sim U(1,d)$ כאשר $d>2$). נגדיר את Y להיות מ.מ באופן הבא $y=\frac{1}{x}$. מצא את פונקצית **הצפיפות וההצטברות** עבור Y. (9 נקודות)

פתרון השאלה בעמוד 119 !!

פתרון: א: נגזור, $(F(x))'=f(x)=\begin{cases}0 & x\le c\\ ac^ax^{-a-1} & x>c\end{cases}$

ב. $E(X)=\int_{-\infty}^{\infty}xf(x)=\int_{c}^{\infty}xac^ax^{-a-1}dx=\int_{c}^{\infty}ac^ax^{-a}dx=ac^a\left(\frac{x^{-a+1}}{-a+1}\right)_c^{\infty}$. כעת, נפריד מקרים:

$$\xrightarrow{a>1} ac^a\left[\left(\underbrace{\lim_{x\to\infty}\frac{x^{\overbrace{-a+1}^{<0}}}{-a+1}}_{0}\right)-\frac{c^{-a+1}}{-a+1}\right]=ac^a\frac{c^{-a+1}}{a-1}=\frac{ac}{a-1}$$

$$\xrightarrow{0<a<1} ac^a\left[\left(\underbrace{\lim_{x\to\infty}\frac{x^{\overbrace{-a+1}^{>0}}}{-a+1}}_{\infty}\right)-\frac{c^{-a+1}}{-a+1}\right]\to\infty$$

נשים לב כי עבור $0<a<1$ נקבל תוחלת אינסופית.

סעיף ג:

$$p\left(X^2<4c^2\right)=p\left(-2c<X<2c\right)=F(2c)-F(-2c)=F(2c)-0=$$

$$\xrightarrow{F(x)=\begin{cases}0 & x\le c\\ 1-\left(\frac{c}{x}\right)^a & x>c\end{cases}} 1-\left(\frac{c}{2c}\right)^a$$

סעיף ד:

X מתפלג אחיד רציף, פונקציית ההצטברות של X היא:

$$F_X(x)=\begin{cases}0 & x\le 1\\ \frac{x-1}{d-1} & 1<x<d\\ 1 & x\ge d\end{cases}$$

אנו מחפשים את פונ' ההצטברות של Y:

$$F_Y(a)=P(Y<a)=P\left(\frac{1}{X}<a\right)\overset{\substack{X>0\\a>0}}{=}P\left(\frac{1}{a}<X\right)=1-P\left(X\le\frac{1}{a}\right)=1-F_X\left(\frac{1}{a}\right)=$$

$$=1-\begin{cases}0 & \frac{1}{a}\le 1\\ \frac{\frac{1}{a}-1}{d-1} & 1<\frac{1}{a}<d\\ 1 & \frac{1}{a}\ge d\end{cases}=\begin{cases}1 & 1\le a\\ \frac{ad-1}{ad-a} & \frac{1}{d}<a<1\\ 0 & a\le\frac{1}{d}\end{cases}$$

נמצא את פונקצית הצפיפות של Y לפי: $f_Y(a)=F'_Y(a)$

$$\frac{\partial}{\partial a}\left(\frac{ad-1}{ad-a}\right)=\frac{\partial}{\partial a}\left(\frac{d}{d-1}-\frac{1}{a}\cdot\frac{1}{d-1}\right)=-\frac{\partial}{\partial a}\left(a^{-1}\cdot\frac{1}{d-1}\right)=\frac{1}{1-d}\frac{\partial}{\partial a}\left(a^{-1}\right)=\frac{1}{(d-1)a^2}$$

$$\Rightarrow f_Y(a)=\begin{cases}\frac{1}{(d-1)a^2} & \frac{1}{d}<a<1\\ 0 & otherwise\end{cases}$$

נשים לב שהאינטגרל של פונקצית הצפיפות הינו לכל d :

$$\int_{-\infty}^{\infty} f_Y(x)dx = \int_{\frac{1}{d}}^{1} \frac{1}{(d-1)a^2}dx = \frac{-1}{d-1}\left(\frac{1}{a}\right)_{\frac{1}{d}}^{1} = \frac{-1}{d-1}\cdot(1-d) = 1$$

2. 60% מהסטודנטים בקורס מופיעים להרצאות. מתוך הסטודנטים המופיעים 60% רושמים, 30% מקשיבים והשאר לא בעניינים. מתוך אלו שאינם מופיעים 70% משלימים את החומר, 10% יודעים את החומר ממקור קודם והשאר לא בעניינים. מתוך הרושמים בכיתה 90% מצליחים ומתוך המקשיבים בכיתה 80% מצליחים. מתוך הנעדרים המשלימים חומר 70% מצליחים ומתוך הנעדרים בעלי הידע הקודם 60% מצליחים. אלו שלא בעניינים נכשלים בין אם נכחו בכיתה ובין אם לאו. (26 נקודות)

א. מהו אחוז הנכשלים בקורס? (3 נקודות)

ב. מה אחוז הנכשלים מבין הנוכחים בהרצאות? (3 נקודות)

ג. ידוע כי הסטודנט נכשל, מה הסיכוי שנכח בהרצאות? (3 נקודות)

ד. לאחר המבחן, נבחרים מדגם של n סטודנטים באקראי. מהי התוחלת ומהי השונות של מספר הנכשלים כפונקציה של n? (8 נקודות)

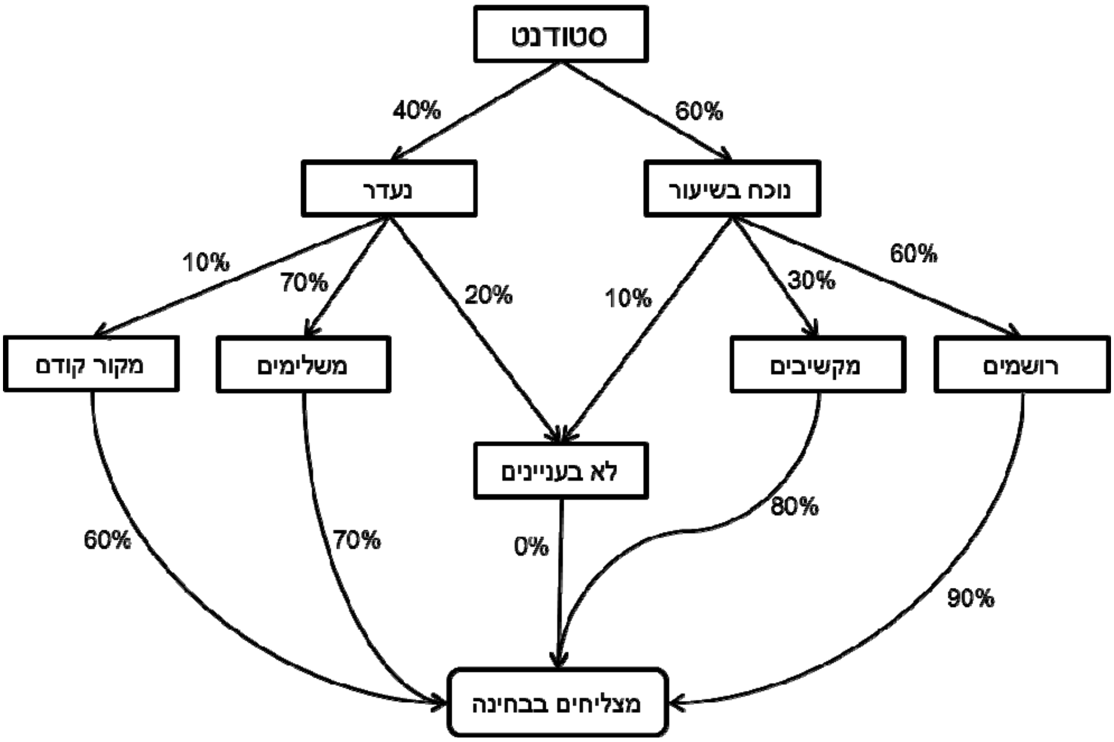
ה. נגדיר את המאורע A – {מספר נכשלים במדגם בגודל n קטן מ-1}. מהו n המינימלי עבורו ההסתברות למאורע A קטנה מ- 0.6? (9 נקודות)

פתרון רוב השאלה בעמוד 53

<u>פתרון שאלה 2:</u>

יסמן מאורעות לתלמיד אקראי : A – סטודנט מופיע בשיעור.

נצייר סכמת עץ, לסטודנט שנבחר באקראי :



א. ניתן לראות שאחוז הנכשלים יקיים:

$$P(fail)=0.6\cdot(0.6\cdot0.1+0.3\cdot0.2+0.1)+0.4\cdot(0.7\cdot0.3+0.1\cdot0.4+0.2)=0.312$$

ב. ואחוז הנכשלים מתוך הנוכחים בהרצאות:

$$P(fail\,|\,in\,class)=(0.6\cdot0.1+0.3\cdot0.2+0.1)=0.22$$

מי שהבין שזה "וגם" –

$$p(\,failed \cap \text{inclass}) = 0.6\left[0.1\cdot0.6+0.2\cdot0.3+0.1\cdot1\right]=0.132$$

ג. נשתמש בחוק בייס:

$$P(in\,class\,|\,fail)=\frac{P(fail\,|\,in\,class)\cdot P(in\,class)}{P(fail)}=\frac{0.22\cdot0.6}{0.312}=0.423$$

מכאן ההסתברות שסטודנט שנכשל נכח בהרצאות קטנה מהמשלים של מאורע זה. כלומר הסטודנט כנראה לא נכח בהרצאות.

ד. נבחר n סטודנטים באקראי, נגדיר מ"מ בדיד X, המתאר את מספר הנכשלים במדגם.

מכאן זהו משתנה בעל התפלגות בינומית: $X\sim Bin(n,p=0.312)$

מכן התוחלת והשונות של X:

$$E(X)=np=0.312n$$
$$Var(X)=np(1-p)=0.2147n$$

ה. עדיין מדובר ב- $X \sim Bin(n, p=0.312)$ ומכאן מבקשים :

$$p(A)<0.6 \rightarrow p(X=0)<0.6 \xrightarrow{natsiv:\ p_r=p(X=r)=\binom{n}{r}p^r\cdot q^{n-r}} \binom{n}{0}p^0\cdot 0.688^n<0.6$$

מתקיים החל מ n=2.

3. יהיו $Y \sim N(2,25)$ ו $X \sim N(4,16)$ מ.מ בלתי תלויים. (24 נקודות)

א. כיצד מתפלג $3X,\ 2Y$? (4 נקודות)

ב. יהיו $K \sim N(\mu_K,\sigma_K^2)$ ו $W \sim N(\mu_W,\sigma_W^2)$ מ.מ בלתי תלויים. הוכח כי $W-K$ גם מתפלג נורמאלית ומצא את הפרמטרים המתאימים. (5 נקודות)

ג. כיצד מתפלג $Y-X$? (2 נקודות)

ד. מצאו $p(2Y+3X+40<10)$. (5 נקודות)

ה. יהיו משתנים מקריים Z_i כך שלכל $i=1\ldots n$ מתקיים $Z_i \sim N(0,1)$ מצא כיצד מתפלג X אם מוגדר באופן הבא : $X=\dfrac{Z_3\sqrt{20}}{\sqrt{\sum_{i=1}^{20}(Z_i)^2}}$ (8 נקודות)

פתרון השאלה מורכב משאלה 3 של המבחן לדוגמא (כמו סעיפים א,ג ו-ד אצלנו במבחן), סעיף ב' שאלה 3 במבחן לדוגמא(כמו סעיף ה' אצלנו במבחן) וסעיף שאלה בספר עמוד 108 למטה

א.

$$X \sim N(4,16)\xrightarrow{3X}3X \sim N(3\cdot 4,9\cdot 16)\rightarrow 3X \sim N(12,144)$$
$$Y \sim N(2,25)\xrightarrow{2Y}2Y \sim N(2\cdot 2,4\cdot 25)\rightarrow 2Y \sim N(4,100)$$

ב. פתרון: נבדוק האם הפונקציה יוצרת מומנטים של $X-Y$ היא פונקציה יוצרת של ההתפלגות הנורמאלית: $E\left(e^{t(W-K)}\right)=E\left(e^{tW}e^{-tK}\right)$ בגלל שנתון כי המ.מ-ים בלתי תלויים ניתן להפריד $E\left(e^{t(W-K)}\right)=E\left(e^{tW}\right)E\left(e^{-tK}\right)$. כעת, כל מ.מ מתפלג נורמאלית ולכן ניתן להציב את הפונקציה היוצרת המתאימה לו :

$$E\left(e^{t(W-K)}\right)=E\left(e^{tW}\right)E\left(e^{-tK}\right)\xrightarrow{M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2t^2\right]}\exp\left[\mu_W t+\frac{1}{2}\sigma_W^2t^2\right]\exp\left[-\mu_K t+\frac{1}{2}\sigma_K^2t^2\right]$$
$$\exp\left[(\mu_W-\mu_K)t+\frac{1}{2}(\sigma_W^2+\sigma_K^2)t^2\right]$$

קבלנו התפלגות נורמאלית עם תוחלת $\mu_W-\mu_K$ ושונות $\sigma_W^2+\sigma_K^2$. ז״א בסה״כ $X-Y \sim N(\mu_W-\mu_K,\sigma_W^2+\sigma_K^2)$.

ג. $Y \sim N(2,25),\quad X \sim N(4,16) \to Y-X \sim N(-2,\ 41)$

$$p(2Y+3X+40<10) \longrightarrow p(2Y+3X<-30)$$

ד. $3X \sim N(12,144),\quad 2Y \sim N(4,100) \to 2Y+3X \sim N(16,244)$

$$\xrightarrow{tiknun} p\left(Z<\frac{-30-16}{\sqrt{244}}\right)=p(Z<-2.9)=1-0.9981=0.0019$$

הערה, ניתן לפתור גם כך, ברור שזה נכון ומתקבל ☺

$$p(2Y+3X+40<10)$$

$$3X \sim N(12,144),\quad 2Y \sim N(4,100) \to 2Y+3X+40 \sim N(56,244)$$

$$\xrightarrow{tiknun} p\left(Z<\frac{10-56}{\sqrt{244}}\right)=p(Z<-2.9)=1-0.9981=0.0019$$

ה. מבחינה אלגברית הביטוי $X=\frac{Z_{34}\sqrt{20}}{\sqrt{\sum_{i=1}^{20}(Z_i)^2}}$ שווה בדיוק ל $X=\frac{Z_{34}}{\sqrt{\frac{\sum_{i=1}^{20}(Z_i)^2}{20}}}$ לפי

המשפט בעמוד 129 " משפט: יהיו Z ו U מ.מ-ים בלתי תלויים המתפלגים באופן הבא $Z \sim N(0,1)$ ו- $U \sim \chi^2_m$ אזי X המוגדר $X=\frac{Z}{\sqrt{U/m}}$ הינו מ.מ המתפלג T עם m דרגות חופש"

ההתפלגות היא T עם 20 דרגות חופש.

4. ענה על הסעיפים הבאים: (20 נקודות)

א. הוכח את הטענה הבאה: יהיו $X_1 \sim N(0,\sigma_1^2),\ X_2 \sim N(0,\sigma_2^2),\dots,X_n \sim N(0,\sigma_n^2)$

משתנים מקריים ב"ת. כיצד יתפלג $\sum_{i=1}^{n}X_i$ ומהם הפרמטרים המתאימים? (6 נקודות) לפי

שאלה 4 במבחן לדוגמא 3 וגם לפי תרגיל משיעורי הבית

$$E\left(e^{t\sum_{i=1}^{n}X_i}\right)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right)=E\left(\prod_{i=1}^{n}e^{tx_i}\right)\xrightarrow{I.D}\prod_{i=1}^{n}E\left(e^{tx_i}\right)$$

$$\xrightarrow{X_1\sim N(0,\sigma_1^2),\ X_2\sim N(0,\sigma_2^2),\ldots,X_n\sim N(0,\sigma_n^2)}\prod_{i=1}^{n}E\left(e^{tx_i}\right)$$

$$\xrightarrow{M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2t^2\right]\xrightarrow{\mu=0}M_X(t)=\exp\left[\frac{1}{2}\sigma^2t^2\right]} e^{\left[\frac{1}{2}\sigma_1^2t^2\right]}\cdot e^{\left[\frac{1}{2}\sigma_2^2t^2\right]}\cdots e^{\left[\frac{1}{2}\sigma_n^2t^2\right]}=e^{\left[\frac{1}{2}t^2\right]\sum_{i=1}^{n}\sigma_i^2}$$

קבלנו התפלגות נורמאלית $\sum_{i=1}^{n} X_i \sim X_n \sim N\left(0, \sum_{i=1}^{n} \sigma_i^2\right)$

ב. יהיו מ.מ X ו- Y. הוכח כי טרנספורמציות ליניאריות על שני המשתנים באופן הבא: $y' = c \cdot y_i + d \quad , \quad x' = b \cdot x_i + a$, משנות את ערך השונות המשותפת כך שמתקיים: $\operatorname{cov}(x', y') = b \cdot c \cdot \operatorname{cov}(x, y)$ כאשר b ו- c חיובים. (6 נקודות)

עמוד 147-148

ג. מצא את המומנט הראשון והשני עבור מ.מ המתפלג מעריכית, והסבר את משמעותו של המומנט הראשון. (לפי הגדרה – לא לפי פונקציה יוצרת מומנטים) (8 נקודות) עמוד 98

$$E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f\left(x\right) dx = \int_{0}^{\infty} \lambda x^2 e^{-\lambda x} \xrightarrow{\text{Integration by parts}} \begin{cases} u = x^2 & u' = 2x \\ v' = \lambda e^{-\lambda x} & v = -e^{-\lambda x} \end{cases}$$

$$-x^2 e^{-\lambda x} - \int -2x e^{-\lambda x} dx = -x^2 e^{-\lambda x} + 2 \underbrace{\int x e^{-\lambda x} dx}_{A} =$$

$$A \xrightarrow{\text{Integration by parts}} \begin{cases} u = x & u' = 1 \\ v' = e^{-\lambda x} & v = \dfrac{-e^{-\lambda x}}{\lambda} \end{cases}$$

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

נחזור לאינטגרל הבסיסי ונציב

$$\left(-x^2 e^{-\lambda x} + 2\left(-x \frac{e^{-\lambda x}}{\lambda} - \frac{e^{-\lambda x}}{\lambda^2}\right)\right)_0^{\infty} = \frac{2}{\lambda^2}$$

בהצלחה!!!

בס"ד

הסתברות וסטטיסטיקה – מועד א סמסטר ב 2013

1. (24 נקודות) יהי X מ.מ בעל פונקצית צפיפות $|x|<a$, $f(x)=\frac{c}{\sqrt{a^2-x^2}}$ כאשר

a הינו פרמטר חיובי.

א. מצא את הקבוע c (5 נקודות)

ב. מצא את פונקצית ההצטברות של X (6 נקודות)

ג. חשב את ההסתברויות $p(0<X<a)$ ו- $p\left(|X|\le\frac{a}{2}\right)$ (5 נקודות)

ד. ללא קשר לסעיפים הקודמים: יהי משתנה מקרי W המתפלג מעריכית עם פרמטר λ

המקיים $Y=\arcsin W$, מצא את פונקצית **הצפיפות וההצטברות** של Y. (8 נקודות)

פתרון: א. בגלל שמדובר בפונקצית צפיפות אזי האינטגרל על כל התחום שווה ל-1:

$$\int_{-a}^{a}\frac{c}{\sqrt{a^2-x^2}}dx=1\xrightarrow{\int\frac{1}{\sqrt{a^2-x^2}}dx=\arcsin\frac{x}{a}+c}c\left(\arcsin\frac{x}{a}\right)_{-a}^{a}=c\left(\arcsin1-\arcsin(-1)\right)$$

$$\xrightarrow{\arcsin(-1)=-\arcsin1}c\left(2\arcsin1\right)\xrightarrow{\arcsin1=\frac{\pi}{2}}=c\pi=1\to c=\frac{1}{\pi}$$

ב. פונקצית ההצטברות:

$$F(x)=\frac{1}{\pi}\int_{-a}^{x}\frac{1}{\sqrt{a^2-x^2}}dx=\frac{1}{\pi}\left(\arcsin\frac{x}{a}\right)_{-a}^{x}=\frac{1}{\pi}\left(\arcsin\frac{x}{a}-\arcsin(-1)\right)$$

$$\xrightarrow{\arcsin(-1)=-\arcsin1\ ,\ \arcsin1=\frac{\pi}{2}}\frac{1}{\pi}\left(\arcsin\frac{x}{a}+\frac{\pi}{2}\right)=\frac{1}{\pi}\arcsin\frac{x}{a}+\frac{1}{2}$$

ובסה"כ נקבל:
$$F(x)=\begin{cases}0 & x<-a\\ \frac{1}{\pi}\arcsin\frac{x}{a}+\frac{1}{2} & -a\le x\le a\\ 1 & x>a\end{cases}$$

ג.
$$p(0<X<a)=F(a)-F(0)=\frac{1}{\pi}\underbrace{\arcsin\frac{a}{a}}_{\frac{\pi}{2}}+\frac{1}{2}-\left(\frac{1}{\pi}\underbrace{\arcsin\frac{0}{a}}_{0}+\frac{1}{2}\right)=\frac{1}{2}$$

$$p\left(|X|\le\frac{a}{2}\right)=p\left(-\frac{a}{2}\le X\le\frac{a}{2}\right)=F\left(-\frac{a}{2}\right)-F\left(\frac{a}{2}\right)=\frac{1}{\pi}\underbrace{\arcsin\frac{-1}{2}}_{\frac{\pi}{6}}+\frac{1}{2}-\left(\frac{1}{\pi}\underbrace{\arcsin\frac{1}{2}}_{\frac{\pi}{6}}+\frac{1}{2}\right)=\frac{1}{3}$$

ד. ידוע כי אם לW ישנה התפלגות מעריכית אזי פונקצית הצפיפות הינה

$f(w)=\begin{cases}\lambda e^{-\lambda w} & w>0\\ 0 & w\le 0\end{cases}$ ופונקצית ההצטברות הינה $F(w)=\begin{cases}1-e^{-\lambda w} & w>0\\ 0 & w\le 0\end{cases}$.

כעת נרצה למצוא את התפלגותו של $Y=\arcsin W$

$$F_Y(a)=p(Y\le a)=p(\arcsin W\le a)=p(W\le \sin a)=1-e^{-\lambda\cdot\sin a}$$

כעת, נגזור לפי המשתנה a על מנת לקבל את פונקצית הצפיפות

$$f_Y(a)=\left(1-e^{-\lambda\cdot\sin a}\right)'=\lambda\cos a\cdot e^{-\lambda\cdot\sin a}$$

2. (23 נקודות) משורר כותב שיר. הוא מסובב 2 סביבונים. אם יוצאת לו תוצאה זהה הוא כותב 10 שורות. אם יוצא לו נ' אחת בדיוק, לא משנה אם ראשונה או שנייה הוא כותב 6 שורות. אם יוצא לו ג' ו ה', לא חשוב באיזה סדר הוא כותב 3 שורות. בשאר המקרים הוא כותב שורה אחת.

תנוחות הסביבונים	x -מספר שורות	$x+2$	$p(x)$
נ' ו-נ' ה' ו- ה' ג' ו- ג' פ' ו- פ'	10	12	$\underbrace{\frac{1}{4}\cdot\frac{1}{4}}_{n\ \&\ n}+\underbrace{\frac{1}{4}\cdot\frac{1}{4}}_{g\ \&\ g}+\underbrace{\frac{1}{4}\cdot\frac{1}{4}}_{h\ \&\ h}+\underbrace{\frac{1}{4}\cdot\frac{1}{4}}_{h\ \&\ h}=\frac{1}{4}$
נ' ו- פ' ה' ו- נ' נ' ו- ה' פ' ו- נ' נ' ו- ג' ג' ו- נ'	6	8	$\frac{3}{8}$
ג' ו- ה' ה' ו- ג'	3	5	$\frac{1}{8}$
כל שאר האפשרויות	1	3	$1-\left[\frac{1}{4}+\frac{3}{8}+\frac{1}{8}\right]=\frac{1}{4}$

א. מה ההסתברות שיכתוב יותר מ-3 שורות? $\frac{1}{4}+\frac{3}{8}=0.625$ (5 נקודות)

ב. המשורר החליט לשנות את הכללים, ובכל תוצאה של הטלת הסביבונים הוא כותב 2 שורות יותר מאשר בתכנון המקורי. מהי תוחלת מספר השורות שיכתוב? (4 נקודות)

$$12\cdot\frac{1}{4}+8\cdot\frac{3}{8}+5\cdot\frac{1}{8}+3\cdot\frac{1}{4}=7.375$$

ג. (ללא קשר לנתוני השאלה המקוריים) 20% מהמשוררים מעדיפים את שעות היום, 50% מעדיפים את שעות הערב והשאר את שעות הלילה. מתוך אוהבי היום 40% עירונים והשאר לא. מתוך אוהבי הערב 70% עירונים. אחוז העירונים מכלל המשוררים הוא 60% . פגשתם משורר אוהב לילה, מה הסיכוי שהוא עירוני? (0.566)) (5 נקודות)

ד. (ביחס לנתונים בסעיף ג) מה אחוז המשוררים שגם אוהבים לילה וגם אינם עירוניים? (0.13)) (3 נקודות)

ה. מהו אחוז אוהבי היום מבין הלא עירוניים? (0.3).) (3 נקודות)

ו. נתון כי העירייה בדקה והחליטה כי משורר שמצליח הינו משורר שאינו עירוני ושמעדיף את שעות הלילה. אם נתון כי ישנם n משוררים. מהי התוחלת והשונות של מספר המשוררים המצליחים? (4 נקודות)

מאחר ומדובר בהתפלגות בינומית אזי תוחלת שווה ל והשונות שווה ל - $n\cdot 0.13$ שווה ל $n\cdot 0.13\cdot 0.87$

3. (32 נקודות)

א. הוכח את הטענה הבאה: יהי $Z\sim N(0,1)$ אזי $Z^2\sim\chi_1^2$ (12 נקודות)

הוכחה: נפתח את הפונקציה יוצרת מומנטים של Z^2 ונבדוק אם נגיע לפונקציה יוצרת של מ.מ המתפלג χ_1^2, ז״א נרצה להגיע בסוף הפיתוח שמתקיים $M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$.

$$M_{Z^2}(t)=\int_{-\infty}^{\infty}e^{tz^2}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\}dz=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2-2tz^2}{2}\right\}dz=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2(1-2t)}{2}\right\}dz$$

$$\xrightarrow{\frac{1}{1-2t}=y^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2y^2}\right\}dz\xrightarrow{\frac{z}{y}=w\ ,\ dz=ydw}y\underbrace{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{w^2}{2}\right\}dw}_{1}=y$$

$$M_{Z^2}(t)=\sqrt{\frac{1}{1-2t}}\quad\Rightarrow\quad Z^2\sim\chi_1^2$$

ב. נתון כי רווח בר סמך של התוחלת הינו $-3.1\le\mu\le 2.4$ כאשר $\alpha=0.05$. הוחלט לבצע את הטרנספורמציה הבאה $y=3x+4$. כיצד ישתנו ערכי הרווח ואורך הרווח בהנתן שגודל המדגם נשאר קבוע?. (אין צורך בחישוב רק הסבר) (4 נקודות)

פתרון: הממוצע ישתנה בהתאם לטרנספורמציה $\bar{X}_{new}=3\bar{X}_{old}+4$ ולכן הממוצע יגדל. השונות תוכפל ב9. מאחר והאורך תלוי רק בסטיית התקן אזי הוא יגדל

ג. עבור נתוני סעיף ב', כיצד ישתנה אורך הרווח הנתון ($-3.1\le\mu\le2.4$) אם נקטין את α? (אין צורך בחישוב רק הסבר) (2 נקודות).

פתרון: אם נקטין את אלפא אזי הרווח יגדל

ד. עבור נתוני סעיף ב', ועבור הרווח סמך הנתון ($-3.1\le\mu\le2.4$) האם ישנה אופציה כי התוחלת שווה ל-0? הסבר מדוע. (3 נקודות)

פתרון: בהסתברות של 95% ערך התוחלת יכול להימצא ברווח הנקוב. בתוך רווח זה נמצא הערך אפס ולכן מצב זה אפשרי בהחלט

ד. מצא את המומנט הראשון והשני עבור מ.מ המתפלג מעריכית, והסבר את משמעותו של המומנט הראשון ובאיזה מדד בא לידי ביטוי המומנט השני. (6 נקודות)

$$E\left(X^2\right)=\int_{-\infty}^{\infty}x^2f\left(x\right)dx=\int_{0}^{\infty}\lambda x^2e^{-\lambda x}\xrightarrow{\text{Integration by parts}}\begin{cases}u=x^2 & u'=2x\\ v'=\lambda e^{-\lambda x} & v=-e^{-\lambda x}\end{cases}$$

$$-x^2e^{-\lambda x}-\int-2xe^{-\lambda x}dx=-x^2e^{-\lambda x}+2\underbrace{\int xe^{-\lambda x}dx}_{A}=$$

$$A\xrightarrow{\text{Integration by parts}}\begin{cases}u=x & u'=1\\ v'=e^{-\lambda x} & v=\dfrac{-e^{-\lambda x}}{\lambda}\end{cases}$$

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

נחזור לאינטגרל הבסיסי ונציב

$$\left(-x^2e^{-\lambda x}+2\left(-x\frac{e^{-\lambda x}}{\lambda}-\frac{e^{-\lambda x}}{\lambda^2}\right)\right)\Bigg|_0^{\infty}=\frac{2}{\lambda^2}$$

ה. הוכח את הטענה הבאה: יהיו $X_1\sim\sim Poisson\left(\lambda_1\right),\ X_2\sim Poisson\left(\lambda_2\right),\ldots,X_n\sim Poisson\left(\lambda_n\right)$ ב"ת כיצד יתפלג $\sum_{i=1}^{n}X_i$ ומהם הפרמטרים המתאימים? (5 נקודות)

פתרון: נשים לב כי לא מדובר במשתנים שווי התפלגות!!!!

$$E\left(e^{t\sum_{i=1}^{n}X_i}\right)=E\left(e^{t\left(X_1+X_2+X_3+\ldots+X_n\right)}\right)=E\left(\prod_{i=1}^{n}e^{tx_i}\right)\xrightarrow{I.D}\prod_{i=1}^{n}E\left(e^{tx_i}\right)$$

$$\xrightarrow{X_1\sim\sim Poisson\left(\lambda_1\right),\ X_2\sim Poisson\left(\lambda_2\right),\ldots,X_n\sim Poisson\left(\lambda_n\right)}\prod_{i=1}^{n}E\left(e^{tx_i}\right)$$

$$\xrightarrow{M_X(t)=E\left(e^{tx}\right)=e^{\lambda\left(e^t-1\right)}}e^{\lambda_1\left(e^t-1\right)}\cdot e^{\lambda_2\left(e^t-1\right)}\ldots\cdot e^{\lambda_n\left(e^t-1\right)}=e^{\left(e^t-1\right)\sum_{i=1}^{n}\lambda_i}$$

קבלנו התפלגות פואסון עם פרמטר : $\sum_{i=1}^{n}\lambda_i$ ז"א $\sum_{i=1}^{n}X_i \sim Poisson\left(\sum_{i=1}^{n}\lambda_i\right)$

4. (21 נקודות)

א. הוכח כי כאשר מקדם המתאם של פירסון הינו חיובי ומושלם אזי מקדם זה שווה ל 1 (5 נקודות)

הוכחה: כאשר הקשר בין X ל- Y חיובי ומושלם, הרי שמתקיים $y_i = b\cdot x_i + a$ לכל x_i ואנו כבר יודעים כי כאשר ישנה טרנספורמציה ליניארית אזי מתקיים : $s_y = |b|\cdot s_x$ וגם $s_y{}^2 = b^2\cdot s_x{}^2$. כמו כן אנו יודעים כי $\bar{Y} = b\bar{X} + a$

ואזי נוכל לרשום את $r = \frac{\operatorname{cov}(x,y)}{s_x\cdot s_y} = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left(y_i-\bar{Y}\right)}{n\cdot s_x\cdot s_y}$ באופן הבא :

$$r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left(y_i-\bar{Y}\right)}{n\cdot s_x\cdot s_y} \xrightarrow{y_i=b\cdot x_i+a\ ,\ \bar{Y}=b\bar{X}+a} r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[b\cdot x_i+a-\left(b\bar{X}+a\right)\right]}{n\cdot s_x\cdot s_y}$$

$$r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[b\cdot x_i+a-b\bar{X}-a\right]}{n\cdot s_x\cdot s_y} \rightarrow r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[b\cdot x_i-b\bar{X}\right]}{n\cdot s_x\cdot s_y}$$

$$\xrightarrow{gorem\ \ meshutaf\ \ b} r = \frac{b\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[x_i-\bar{X}\right]}{n\cdot s_x\cdot s_y} \xrightarrow{s_y=|b|\cdot s_x\ ,\ \ s_y{}^2=b^2\cdot s_x{}^2} r = \frac{b\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n\cdot s_x\cdot|b|\cdot s_x}$$

$$r = \frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)^2}{n\cdot s_x{}^2} = \frac{s_x{}^2}{s_x{}^2} = 1$$

ב. יהיו מ.מ X ו- Y. הוכח כי טרנספורמציות ליניאריות על שני המשתנים באופן הבא:

$y' = c\cdot y_i + d \quad , \quad x' = b\cdot x_i + a$, משנות את ערך השונות המשותפת כך שמתקיים:

$\operatorname{cov}(x',y') = b\cdot c\cdot\operatorname{cov}(x,y)$ כאשר b ו- c חיוביים. (5 נקודות)

הוכחה: נתון כי ישנן טרנספורמציות ליניאריות על המשתנים: x ו- y באופן הבא:

$x'=b\cdot x_i+a \quad , \quad y'=c\cdot y_i+d$ ולכן מתקיים:

$$\operatorname{cov}(x',y')=\frac{\sum_{i=1}^{n}\left(x'_i-\bar{X}'\right)\left(y'_i-\bar{Y}'\right)}{n}\xrightarrow{x'=b\cdot x_i+a \quad , \quad y'=c\cdot y_i+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\bar{X}'\right)\left[c\cdot y_i+d-\left(\bar{Y}'\right)\right]}{n}$$

$$\xrightarrow{\bar{X}'=b\cdot\bar{X}+a \quad , \quad \bar{Y}'=c\cdot\bar{Y}+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\left(b\cdot\bar{X}+a\right)\right)\left[c\cdot y_i+d-\left(c\cdot\bar{Y}+d\right)\right]}{n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-b\cdot\bar{X}-a\right)\left[c\cdot y_i+d-c\cdot\bar{Y}-d\right]}{n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i-b\cdot\bar{X}\right)\left[c\cdot y_i-c\cdot\bar{Y}\right]}{n}\xrightarrow{gorem\ \ meshutaf\ \ b\ \&\ c}\frac{bc\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[y_i-\bar{Y}\right]}{n}=b\cdot c\cdot\operatorname{cov}(x,y)$$

ג. יהיו $Y\sim N\left(3,\frac{1}{9}\right)$ ו $X\sim N\left(5,1\right)$ מ.מ בלתי תלויים, ויהי W משתנה מקרי המוגדר באופן הבא: $W=2X-3Y$. מצאו כיצד מתפלג W ומצאו את $p\left(W<2.5\right)$. (5 נקודות)

$W\sim N\left(1,5\right)$.

פתרון: נבצע תקנון ונקבל כי $p\left(W<2.5\right)=p\left(Z<\frac{2.5-1}{\sqrt{5}}\right)=p\left(Z<0.67\right)\approx 0.74$

ד. כיצד מתפלג $\sum_{i=3}^{13}S_i^{\,2}$ בהנתן כי $\forall i \quad i=1\ldots n: S_i\sim N\left(0,1\right)$? (3 נקודות)

פתרון : לפי המשפט סכום הריבועים של מ.מ המתפלגים נורמאלית סטנדרטית מתפלגים חי בריבוע עם דרגות חופש השווה לכמות האיברים, במקרה שלנו חי בריבוע עם 11 דרגות חופש

ה. כיצד יתפלג $Y=\frac{S_1\sqrt{9}}{\sqrt{\sum_{i=3}^{11}\left(S_i\right)^2}}$ ומדוע? (3 נקודות)

פתרון: לפי המשפט" יהיו Z ו U מ.מ-ים בלתי תלויים המתפלגים באופן הבא

$Z \sim N(0,1)$ ו- $U \sim \chi_m^2$ אזי X המוגדר $X = \frac{Z}{\sqrt{U/m}}$ הינו מ.מ המתפלג T עם m

דרגות חופש", Y מתפלג t עם 9 דרגות חופש.

פתרון מועד ב' ו- ג' 2014

1. (27 נקודות) נתונה פונקצית הצפיפות הבאה $f(x) = 0.5cx$ כך ש $0 < x < 1$ ו- $c > 0$

א. מצא את c על מנת שפונקציה זו תקיים את כל תכונות ההתפלגות (2 נקודות)

ב. מצא את פונקצית ההצטברות - $F(x)$ (3 נקודות)

ג. מצא רבעון עליון ורבעון תחתון (4 נקודות)

ד. מצא את התוחלת ואת השונות (4 נקודות)

ה. מצא את ההסתברות הבאה $P(X < 0.5)$ (2 נקודות)

ו. העזר בסעיף ה' וקבע ללא חישוב, האם ערך החציון מעל או מתחת לערך 0.5? (3 נקודות)

ז. דגמו את X 64 פעמים מתוך ההתפלגות הנתונה, **מצא** את התפלגות ממוצע הדגימות והסבר כיצד הסקת עבור התפלגות זו. **מהו** בקירוב העשירון העליון של ממוצע הדגימות? (4 נקודות)

ח. (סעיף זה לא קשור לסעיפים הקודמים) נגדיר את X להיות מ.מ המתפלג אחידה רציפה עם הפרמטרים הבאים $X \sim U(0,d)$ כאשר $d > 2$. נגדיר את Y להיות מ.מ באופן הבא $Y = X^2$. מצא את פונקצית **הצפיפות וההצטברות** עבור Y. (5 נקודות)

2. (21 נקודות) חברת אינטרנט מספקת ללקוחותיה גישה למספר מאגרי מידע. נתון כי N_1 הוא מספר המילים בשאילתא של הלקוח, N_2 הוא גודל המאגר בו הוא מחפש (גודל זה מתעדכן כל שניה ולכן הוא משתנה רציף), ו S מודד את מהירות המחשב בו נערך החיפוש.

להלן הנתונים הבאים:

לN_1 יש התפלגות פואסונית עם פרמטר 0.9.

ל S ישנה התפלגות מעריכית עם פרמטר 0.6.

עבור המשתנה N_2 ידוע כי ההסתברות שגודל המאגר קטן מ30 הינו 0.3, ההסתברות כי גודל המאגר בין 30 ל100 היא 0.6 וההסתברות לגודל מאגר גדול מ100 היא 0.1. עבור כי חיפוש מאגר מידע יש הסתברות לטעות בהעברת המידע ללקוחות. ההסתברות לטעות עבור מאגר קטן מ30 היא 0.2 . ההסתברות לטעות עבור מאגר בגודל בין 30 ל100 היא 0.5. בהנתן כי ישנה טעות בהעברת המידע, אזי ההסתברות שהיא ממאגר קטן מ-30 היא 0.136.

א. מה ההסתברות לטעות כאשר גודל המאגר גדול מ-100? (5 נקודות)

ב. בהנתן כי ישנה טעות בהעברת המידע, מה ההסתברות לגודל המאגר בין 30 ל100 ? (4 נקודות)

ג. הוכח כי תוחלת של משתנה מקרי המתפלג פואסונית הינו $E(X)=\lambda$ (ניתן להוכיח ע"י פונקציה יוצרת מומנטים) ומצא את התוחלת עבור משתנה N_1 (5 נקודות)

ד. מצא את ההסתברות כי מהירות המחשב בו נערך החיפוש קטן מ1.5 (3 נקודות)

ה. בהנתן כי מהירות המחשב קטנה מ3, מה ההסתברות כי גדולה מ1.5 (4 נקודות)

<u>פתרון שאלה 2</u>

ע"מ לענות על סעיפים **א ו-ב**, יש להשתמש רק בנתונים שבחצי השני של השאלה.

נסמן את המאורעות הבאות עבור N_2 :

A: $N_2 \le 30$

B: $30 < N_2 \le 100$

C: $100 < N_2$

נגדיר גם את המאורעות הבאות:

שגיאה: E

אין שגיאה: R

נתון כי N_2 מתפלג באופן הבא:

מאורע	$P(N_2)$	$P(Error \mid N_2)$
A	0.3	0.2
B	0.6	0.5
C	0.1	?

כמו כן נתון:

$$P(A|E)=0.136$$

מתוך נוסחת בייס:

$$\Rightarrow 0.136=P(A|E)=\frac{P(E|A)P(A)}{P(E)}=\frac{0.2\cdot 0.3}{P(E)}\Rightarrow P(E)=\frac{0.06}{0.136}=0.4411$$

א. מתוך נוסחת ההסתברות השלמה:

$$P(E)=P(A)\cdot P(E|\mathrm{A})+P(B)\cdot P(E|\mathrm{B})+P(C)\cdot P(E|\mathrm{C})$$
$$\Rightarrow 0.4411=0.3\cdot 0.2+0.6\cdot 0.5+0.1\cdot P(E|\mathrm{C})$$
$$\Rightarrow P(E|\mathrm{C})=0.811$$

ב. $$P(\mathrm{B}|E)=\frac{P(E|\mathrm{B})P(B)}{P(E)}=\frac{0.5\cdot 0.6}{0.4411}=\frac{0.3}{0.4411}=0.68$$

ג. נשתמש בפונקציה יוצרת מומנטים של משתנה פואסוני:

$$M_{N_1}(t)=e^{\lambda(e^t-1)} \Rightarrow \frac{\partial}{\partial t}M_{N_1}(t)=e^{\lambda(e^t-1)}\lambda e^t$$
$$\Rightarrow E[N_1]=\frac{\partial}{\partial t}M_{N_1}(0)=e^{\lambda(e^0-1)}\lambda e^0=e^{\lambda(1-1)}\lambda=\lambda$$
$$\Rightarrow E[N_1]=0.9$$

ד. S מתפלג מעריכית ולכן:

$$F(t)=P(S<t)=\int_{-\infty}^{t} f_S(x)dx=\int_{0}^{t}\lambda e^{-\lambda x}dx=\lambda\left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^t=1-e^{-\lambda t}$$
$$\Rightarrow P(S<1.5)=1-e^{-0.6\cdot 1.5}=1-0.406=0.593$$

ה. S מתפלג מעריכית:

$$P(S>1.5\,|\,S<3)=\frac{P(S>1.5\cap S<3)}{P(S<3)}=\frac{P(1.5<S<3)}{P(S<3)}=\frac{F(3)-F(1.5)}{F(3)}$$

כאשר במעבר הראשון השתמשנו בנוסחת הסתברות מותנית.

$$P(S<1.5)=1-e^{-0.6\cdot 1.5}=0.593$$
$$P(S<3)=\left(1-e^{-0.6\cdot 3}\right)=0.834$$
$$\Rightarrow P(S>1.5\,|\,S<3)=\frac{0.834-0.593}{0.834}=0.289$$

3. (28 נקודות) יהיו $Y\sim N(-2,36)$ ו $X\sim N(2,4)$ מ.מ בלתי תלויים.

א. כיצד מתפלג $2Y,\ 0.5X$? (4 נקודות)

$$0.5X\sim N(0.5\cdot 2,0.25\cdot 4)\Rightarrow 0.5X\sim N(1,1)$$
$$2Y\sim N(2\cdot(-2),4\cdot 36)\Rightarrow 0.5X\sim N(-4,144)$$

ג. כיצד מתפלג $Y-X$? (4 נקודות)

$$Y-X\sim N(0.5-(-2),4+36)\Rightarrow Y-X\sim N(2.5,40)$$

ד. מצאו $p(0.5Y+0.5X-1<2)$. (4 נקודות)

$$P(0.5Y+0.5X-1<2)=P(0.5Y+0.5X-3<0)$$
$$0.5Y+0.5X-3\sim N(-3,10)$$

מכאן:

$$P(0.5Y+0.5X-3<0)=\Phi\left(\frac{3}{\sqrt{10}}\right)\cong\Phi(0.95)=0.8289$$

ו. יהיו משתנים מקריים Z_i כך שלכל

$i=1\dots n$ מתקיים $Z_i \sim N(0,1)$ מצא כיצד מתפלג X אם מוגדר באופן הבא:

$$X=\frac{7\cdot Z_{75}}{\sqrt{\sum_{i=1}^{49}(Z_i)^2}}$$ (4 נקודות)

$$\sum_{i=1}^{n} Z_i^{\,2} \sim \chi_n^2 \Rightarrow \sum_{i=1}^{49} Z_i^{\,2} \sim \chi_{49}^2$$

$$\frac{Z}{\sqrt{\frac{1}{\upsilon}\sum_{i=1}^{\upsilon} Z_i^{\,2}}} \sim t_\upsilon \Rightarrow \frac{7\cdot Z}{\sqrt{\sum_{i=1}^{49} Z_i^{\,2}}} = \frac{Z}{\sqrt{\frac{1}{49}\sum_{i=1}^{49} Z_i^{\,2}}} \sim t_{49}$$

ו. ידוע כי סטיית התקן בהשקעה של מניה היא 0.25. לשם בניית רווח סמך לתוחלת התשואה, חוקר דגם מדגם של 100 משקיעים מתוך אוכלוסיה זו. ממוצע התשואה שהתקבל היה 0.2.

א. מצא את רווח הסמך עם רמת ביטחון של 95% (3 נקודות).

ב. מצא את רווח הסמך עם רמת ביטחון של 99% (3 נקודות).

$$For\ 95\%\quad the\ CI\ \ is\quad 0.2-1.96\frac{0.25}{\sqrt{100}}<\mu<0.2+1.96\frac{0.25}{\sqrt{100}}$$

$$0.151<\mu<0.249 \quad\rightarrow\ L=0.098$$

$$For\ 99\%\quad the\ CI\ \ is\quad 0.2-2.57\frac{0.25}{\sqrt{100}}<\mu<0.2+2.57\frac{0.25}{\sqrt{100}}$$

$$0.13575<\mu<0.26425 \quad\rightarrow\ L=0.1285$$

ג. איזה רווח רחב יותר? הסבר (3 נקודות).

פתרון: הרווח השני רחב יותר. הסבר: מאחר ש נלקחה רמת בטחון גדול יותר – כך שב99% אנחנו רוצים להיות בטוחים כי התוחלת האמיתית נמצאת ברווח הנקוב – אזי נצטרך להגדיר רווח רחב יותר, אם "מספיק" לנו להיות בטוחים ב 95% אזי נוכל להגדיר רווח פחות רחב.

ד. אם החוקר היה דוגם גודל מדגם גדול יותר אורך הרווח שמצאת בסעיף א' היה גדול יותר או קטן יותר משמצאת? הסבר מדוע. (תשובה ללא הסבר סטטיסטי לא תתקבל – שיקול אלגברי לא מספיק) (3 נקודות)

פתרון: הרווח היה פחות רחב כיון שהממוצע שהתקבל, התקבל על סמך מספר רב יותר של תצפיות ועל כן הוא מדויק יותר ולכן מספיק לקחת סביבו רווח פחות רחב.

4. ענה על הסעיפים הבאים: (24 נקודות)

א. יהיו $X_1,X_2,X_3,\ldots X_r\sim G(1,\lambda)$ משתנים מקריים ב"ת. כיצד יתפלג $\sum_{i=1}^{n}X_i$ ומהם הפרמטרים המתאימים? (6 נקודות)

פתרון: הוכחה: ע"י פונקציה יוצרת מומנטים $M_{\sum_i^r X_i^2}(t)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_r)}\right)$ בגלל שהם

ב"ת אזי $E\left(e^{t(X_1+X_2+X_3+\ldots+X_r)}\right)=\prod_{i=1}^{r}E\left(e^{tX_i}\right)$ ובגלל שנתון שהם שווי התפלגות אזי

$\prod_{i=1}^{r}E\left(e^{tX_i}\right)=\prod_{i=1}^{r}\frac{\lambda}{\lambda-t}=\left(\frac{\lambda}{\lambda-t}\right)^r$. בסה"כ קבלנו פונקציה יוצרת מומנטים של התפלגות

גמא, משמע, שאם סוכמים מ.מ-ים המתפלגים מעריכית עם פרמטר λ אזי התפלגותם היא

גמא עם פרמטרים r ו- λ.

ב. טרנספורמציות ליניארית חיוביות על המשתנים אינן משנות את ערכו של מקדם המתאם של פירסון. (6 נקודות)

הוכחה:

$$r=\frac{\mathrm{cov}(x',y')}{s'_x\cdot s'_y}=\frac{\sum_{i=1}^{n}(x'_i-\bar{X}')(y'_i-\bar{Y}')}{s'_x\cdot s'_y\,n}\xrightarrow{x'=b\cdot x_i+a\quad,\quad y'=c\cdot y_i+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\bar{X}'\right)\left[c\cdot y_i+d-\left(\bar{Y}'\right)\right]}{s'_x\cdot s'_y\,n}$$

$$\xrightarrow{\bar{X}'=b\cdot\bar{X}+a\ ,\ \bar{Y}'=c\cdot\bar{Y}+d}\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-\left(b\cdot\bar{X}+a\right)\right)\left[c\cdot y_i+d-\left(c\cdot\bar{Y}+d\right)\right]}{s'_x\cdot s'_y\,n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i+a-b\cdot\bar{X}-a\right)\left[c\cdot y_i+d-c\cdot\bar{Y}-d\right]}{s'_x\cdot s'_y\,n}$$

$$\frac{\sum_{i=1}^{n}\left(b\cdot x_i-b\cdot\bar{X}\right)\left[c\cdot y_i-c\cdot\bar{Y}\right]}{s'_x\cdot s'_y\,n}\xrightarrow{gorem\ \ meshutaf\ \ b\ \&\ c}\frac{bc\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left[y_i-\bar{Y}\right]}{bcs_x\cdot s_y n}=\frac{\mathrm{cov}(x,y)}{s_x\cdot s_y}=r$$

ג. הסבר מה משמעותה של פונקצית ההצטברות ורשום את הגדרתה (4 נקודות)

פתרון: $F(x)$ מוגדרת כפונקצית ההתפלגות ודרכה נוכל לחשב התפלגויות וגם להגדיר

את פונקצית הצפיפות: $F(a)$ המסומנת גם $F_x(a)$ מקיימת $F(a)=F_x(a)=p(X\le a)$

(השטח משמאל). באופן דומה: $F(a\le X\le b)=p(X\le b)-p(X\le a)=F(b)-F(a)$.

הפונקציה $F(x)$ מחזירה את השטח משמאל לx = הסתברות לקבל ערך הקטן מ x.

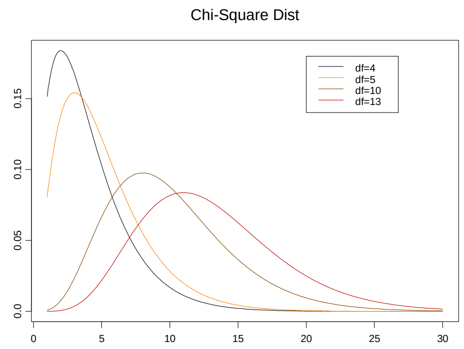
משמע, אנו מכניסים לפונקציה את x והיא מחזירה לנו את ההסתברות לקבל ערך הנמוך

מ x. זאת עבור כל ערך בהתפלגות הנתונה.

ד. מהו הקשר בין ההתפלגות הנורמאלית להתפלגות חי בריבוע הסבר (4 נקודות)

פתרון: כאשר ניקח מ"מ המתפלג נורמאלית סטנדרטית, משמע $Z \sim N(0,1)$ אזי ההתפלגות של אותו משתנה בריבוע הינה חי בריבוע, משמע $Z^2 \sim \chi_1^2$

הוכחנו זאת בכיתה ע"י פונקציה יוצרת מומנטים. ההיגיון שעומד מאחורי טענה זו הוא שההתפלגות הופכת חיובית



ז. הסבר את ההבדל בין "ממוצע מדגם" ל – "תוחלת אוכלוסייה" (4 נקודות)

מבחינת חישוב מתמטי, שני המושגים זהים, בשניהם סוכמים את כל האברים ומחלקים בכמות האיברים. ההבדל נמצא באיברים עצמם. באוכלוסיה, אמורים לבצע את חישוב הממוצע על כל איברי האוכלוסייה, ובמדגם מבצעים את חישוב הממוצע על כמות קטנה יותר, הלא היא כמות תצפיות המדגם. עוד נציין כי ממוצע האוכלוסייה הינו פרמטר- מספר קבוע - לא ידוע ששמו "תוחלת" וממוצע המדגם הינו מספר (המהווה משתנה מקרי – שהרי יכול להשתנות ממדגם למדגם) שאנו מחשבים ואמור לשערך את התוחלת האמיתית הלא ידועה של האוכלוסייה

מבחן בהסתברות- מועד ב' סמסטר א' 2014

1. (26 נקודות) יהי X מ"מ (משתנה מקרי רציף) לשינוי הטמפרטורה בכל יום אביבי באנטרקטיקה.
פונקצית הצפיפות של X היא:

$$f_X(x)=\begin{cases} A(2+x) & -2\le x\le 2 \\ 0 & otherwise \end{cases}$$

א. מצא את A (4 נקודות)

ב. מהי ההסתברות כי ביום אביבי שנבחר באקראי יהיה שינוי הטמפרטורה חיובי. (4 נקודות)

ג. אם ידוע שהשינוי טמפרטורה הוא חיובי, מהי ההסתברות ששינוי הטמפרטורה יהיה מעל מעלה 1. (6 נקודות)

ד. חשבו תוחלת ושונות של שינויי הטמפרטורה היומיים . (6 נקודות)

ה. (סעיף זה לא קשור לסעיפים הקודמים) נגדיר את X להיות משתנה מיקרי המתפלג התפלגות אחידה רציפה $X\sim U(0,1)$. נגדיר את Y להיות מ.מ באופן הבא $Y=2x-4$.
מצא את פונקצית הצפיפות עבור Y . (6 נקודות)

2. (20 נקודות) עם רדת הערב המשורר כותב שיר. הוא מטיל שתי קוביות וכותב שיר שבו מספר השורות שווה סכום התוצאות בשתי ההטלות.

א. מה ההסתברות שיכתוב שיר שבו יותר מ - 10 שורות? (4 נקודות)

ב. בסיומה של כתיבת שורה המשורר שותה ½ כוס יין. מהי תוחלת והשונות של מספר הכוסות שישתה המשורר? (4 נקודות)

ג. (ללא קשר לנתוני השאלה המקוריים) יש לכותב 3 דרכים אפשריות. בכל דרך אפשר בצומת הבאה לפנות ימינה, או שמאלה, או להמשיך ישר. הכותב בוחר באופן מקרי את אחת הדרכים. אם בחר בראשונה הסיכוי שבהמשך יפנה ימינה הוא 0.2 והסיכוי שיפנה שמאלה הוא 0.3. אם בחר בשנייה הסיכוי שבהמשך יפנה ימינה הוא 0.4 והסיכוי שיפנה שמאלה הוא 0.3. אם בחר בשלישית הסיכוי שיפנה ימינה הוא 0.6 והסיכוי שיפנה שמאלה הוא 0.4. מה ההסתברות שימשיך ישר? (4 נקודות)

ד. (ביחס לנתונים בסעיף ג) מסתבר שהכותב המשיך ישר. מה ההסתברות שבחר בדרך השלישית? (3 נקודות)

ה. מה ההסתברות לבחור בדרך השנייה ולא לפנות ימינה? (3 נקודות)

ו. ללא קשר לנתוני לסעיפים הקודמים: תוחלת זמן הבערה של פנס רחוב היא 3 שעות. ידוע כי משך הבערה מתפלג מעריכית. מה ההסתברות שפנס הרחוב יבער בין שעה לשלוש וחצי שעות? (2 נקודות)

3. (27 נקודות) נתון כי סטיית התקן שווה ל- $\sigma = 0.45$. במדגם בגודל 50 התקבל ממוצע של 1.2 חשב רווח סמך לתוחלת:

א. ברמת ביטחון של 95% (5 נקודות)

ב. ברמת ביטחון של 99% (2 נקודות)

ג. איזה רווח סמך רחב יותר ומדוע? (2 נקודות)

ד. מהו גודל המדגם שיש לדגום על מנת להבטיח ששגיאת האמידה לא תעלה על 0.001 כאשר נתון כי $\sigma = 0.5$. (6 נקודות)

ה. נתון כי קבוצת תצפיות מתפלגת נורמאלית על ממוצע m וסטית תקן s, מצא קבוע a כך שאחוז התצפיות,

א. בתוך הטווח $m \pm a \cdot s$ היא 75% (6 נקודות)

ב. הקטנה מ $m - a \cdot s$ היא 25% (6 נקודות)

פתרון: גודל המדגם גדול מספיק כדי שנוכל לומר כי מדובר בהתפלגות דגימה עבור נורמאלית עבור הממוצע. התפלגות האוכלוסיה: $X \sim N(?, 0.2025)$ ולכן

$$\bar{X} \sim N\left(?, \frac{0.2025}{50}\right)$$

א. נשתמש בנוסחא לרווח סמך:

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \xrightarrow{n=50,\ \sigma=0.45,\ \alpha=0.05,\ \bar{X}=1.2} 1.2 - z_{\frac{0.05}{2}} \frac{0.45}{\sqrt{50}} < \mu < 1.2 + z_{\frac{0.05}{2}} \frac{0.45}{\sqrt{50}}$$

$$1.2 - z_{0.025} \frac{0.45}{\sqrt{50}} < \mu < 1.2 + z_{0.025} \frac{0.45}{\sqrt{50}} \xrightarrow{z_{0.025}=1.96} 1.2 - 1.96 \frac{0.45}{\sqrt{50}} < \mu < 1.2 + 1.96 \frac{0.45}{\sqrt{50}}$$

$$\xrightarrow{1.96\frac{0.45}{\sqrt{50}}=0.1247} 1.2 - 0.1247 < \mu < 1.2 + 0.1247 \rightarrow 1.075 < \mu < 1.324$$

ז"א שבמדגם בעל גודל של 50 מתוך משתנה מקרי זה התפלגות הדגימה היא נורמאלית שסטיית התקן שלו שווה 0.45 ותוחלתו אינה ידועה, ההסתברות שהתוחלת תהיה שונה ממוצע המדגם בלא יותר מ 0.1247 יחידות היא 0.95.

ב.

$$\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \xrightarrow{n=50,\ \sigma=0.45,\ \alpha=0.01,\ \bar{X}=1.2} 1.2 - z_{\frac{0.01}{2}} \frac{0.45}{\sqrt{50}} < \mu < 1.2 + z_{\frac{0.01}{2}} \frac{0.45}{\sqrt{50}}$$

$$1.2 - z_{0.005} \frac{0.45}{\sqrt{50}} < \mu < 1.2 + z_{0.005} \frac{0.45}{\sqrt{50}} \xrightarrow{z_{0.005}=2.58} 1.2 - 2.58 \frac{0.45}{\sqrt{50}} < \mu < 1.2 + 2.58 \frac{0.45}{\sqrt{50}}$$

$$\xrightarrow{2.58\frac{0.45}{\sqrt{50}}=0.164} 1.2 - 0.164 < \mu < 1.2 + 0.164 \rightarrow 1.035 < \mu < 1.364$$

ג. נוכל לראות כי ככל שרמת הסמך עולה כך מתרחב רווח הסמך. ככל שנרצה להגדיל את הביטחון ($1-\alpha$) בהימצאות הערך האמיתי של התוחלת, μ, בתוך הרווח, נצטרך להגדיל את הרווח. באותו אופן נוכל לומר כי אם הרווח צר יותר כן גדלה ההסתברות שהוא לא יכיל את התוחלת, μ, כלומר רמת הסמך פוחתת.

ד. $$n \ge \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon} \right)^2 \xrightarrow{\varepsilon=0.001,\ z_{0.025}=1.96\ ,\ \sigma=0.5} n \ge \left(\frac{1.96 \cdot 0.5}{0.001} \right)^2 \rightarrow n \ge 960400$$

ה. פתרון: נרצה לחשב, $p\left(m-as \le X \le m+as\right)=0.75$ אך תחילה נעשה תקנון $Z=\frac{X-m}{s}$ אנו עושים זאת כי אנו יודעים לעבוד רק עם התפלגות נורמאלית סטנדרטית.

$$p\left(\frac{m-as-m}{s} \le \frac{X-m}{s} \le \frac{m+as-m}{s} \right) = p\left(-a \le Z \le a\right) \rightarrow \Phi(a)-\Phi(-a)=$$

$$\Phi(a)-\left(\underbrace{1-\Phi(a)}_{\Phi(-a)} \right)=2\Phi(a)-1 \rightarrow 2\Phi(a)-1=0.75 \rightarrow \Phi(a)=0.875 \rightarrow a=1.15$$

ב.

$$p\left(X \le m-as\right)=0.22 \xrightarrow{Z=\frac{X-m}{s}} p\left(\frac{X-m}{s} \le \frac{m-as-m}{s} \right)=0.22$$

$$p\left(Z \le -a\right)=\Phi(-a)=1-\Phi(a) \rightarrow \Phi(-a)=0.78 \rightarrow a=0.77$$

4.א. (27 נקודות) יהיו $Y \sim N(2,1)$ ו $X \sim N(4,25)$ מ.מ בלתי תלויים, ויהי W משתנה מקרי המוגדר באופן הבא: $W = X - 4Y$. מצאו כיצד מתפלג W (3 נקודות)

$$Y \sim N(2,1), \quad X \sim N(4,25) \xrightarrow{W=X-4Y} W \sim N(-4,29)$$

.

ב. מצאו את $p(W<2.5)$. (4 נקודות)

$$W \sim N(-4,29)$$

$$p(W<2.5) \xrightarrow{tiknun} p\left(Z < \frac{2.5-(-4)}{\sqrt{29}} \right) \rightarrow p(Z<1.2)=0.8849$$

ג. יהיו $K \sim N\left(\mu_K, \sigma_K^2\right)$ ו $D \sim N\left(\mu_D, \sigma_D^2\right)$ מ.מ בלתי תלויים. הוכח כי $D-K$ גם מתפלג נורמאלית ומצא את הפרמטרים המתאימים. (4 נקודות)

פתרון ב. פתרון: נבדוק האם הפונקציה יוצרת מומנטים של $X-Y$ היא פונקציה יוצרת של ההתפלגות הנורמאלית: $E\left(e^{t(D-K)}\right)=E\left(e^{tD}e^{-tK}\right)$ בגלל שנתון כי המ.מ-ים בלתי תלויים ניתן להפריד $E\left(e^{t(D-K)}\right)=E\left(e^{tD}\right)E\left(e^{-tK}\right)$. כעת, כל מ.מ מתפלג נורמאלית ולכן ניתן להציב את הפונקציה היוצרת המתאימה לו:

$$E\left(e^{t(D-K)}\right)=E\left(e^{tD}\right)E\left(e^{-tK}\right)\xrightarrow{M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2t^2\right]}\exp\left[\mu_D t+\frac{1}{2}\sigma_D^2t^2\right]\exp\left[-\mu_K t+\frac{1}{2}\sigma_K^2t^2\right]$$

$$\exp\left[\left(\mu_D-\mu_K\right)t+\frac{1}{2}\left(\sigma_D^2+\sigma_K^2\right)t^2\right]$$

קבלנו התפלגות נורמאלית עם תוחלת $\mu_D-\mu_K$ ושונות $\sigma_D^2+\sigma_K^2$. ז"א בסה"כ $X-Y\sim N\left(\mu_D-\mu_K,\sigma_D^2+\sigma_K^2\right)$.

ד. כיצד מתפלג $Y-X$? (2 נקודות)

$$Y\sim N(2,1),\quad X\sim N(4,25)\xrightarrow{A=Y-X}A\sim N(-2,26)$$

ה. מצאו $p\left(2Y+X+4<6\right)$. (5 נקודות)

$$p\left(2Y+X+4<6\right)\xrightarrow{Y\sim N(2,1),\quad X\sim N(4,25)\xrightarrow{B=2Y+X+4}B\sim N(16,27)}p\left(B<6\right)$$

$$\xrightarrow{tiknun}p\left(Z<\frac{6-16}{\sqrt{27}}\right)=p\left(Z<-1.9\right)=1-0.9713=0.0287$$

ו. הוכח כי אם $Z_1,Z_2,Z_3,\ldots Z_k$ מ.מ-ים ב"ת כך ש $\forall i=1\ldots k\quad Z_i\sim N(0,1)$ אזי $\sum_i^k Z_i^2\sim\chi_k^2$. (5 נקודות)

פתרון: נתחיל מפונקציה יוצרת מומנטים, $M_{\sum_i^k Z_i^2}(t)=E\left(e^{t\sum_i^k Z_i^2}\right)$ בגלל שנתון שהמ.מ-ים ב"ת אזי מתקיים $E\left(e^{t\sum_i^k Z_i^2}\right)=E\left(e^{t\left(Z^2{}_1+Z^2{}_2+\ldots+Z^2{}_k\right)}\right)=E\left(e^{tZ^2{}_1}\right)\cdot E\left(e^{tZ^2{}_2}\right)\cdot\ldots\cdot E\left(e^{tZ^2{}_k}\right)$ ובגלל שהמ.מ-ים שוויי התפלגות (בעלי התפלגות זהה לחלוטין) לפי הנתון $\forall i=1\ldots k\quad Z_i\sim N(0,1)$ אזי $E\left(e^{tZ^2{}_1}\right)\cdot E\left(e^{tZ^2{}_2}\right)\cdot\ldots\cdot E\left(e^{tZ^2{}_k}\right)=\left[M_{Z^2}(t)\right]^k$. כעת מההוכחה הקודמת ראינו כי $M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$ ולכן $\left[M_{Z^2}(t)\right]^k\xrightarrow{M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}}\left(\frac{1}{1-2t}\right)^{\frac{k}{2}}$. *משמעות הדבר היא כי סכום של מ.מ-ים ב"ת המתפלגים נורמאלית סטנדרטית מתפלג חי בריבוע עם מספר הדרגות החופש השווה למספר המחוברים.*

ז. יהיו משתנים מקריים Z_i ב"ת כך שלכל $i=1\ldots n$ מתקיים $Z_i \sim N(0,1)$ ויהיו משתנים מקריים A_i ב"ת כך שלכל $i=1\ldots n$ מתקיים $A_i \sim N(0,1)$, עוד נתון כי כל ה-מ"מ A_i וה-מ"מ Z_i ב"ת. מצא כיצד מתפלג X אם מוגדר באופן הבא : $X=\dfrac{42\sum_{i=1}^{52}(Z_i)^2}{52\sum_{i=1}^{42}(A_i)^2}$ (4 נקודות)

פתרון :נשתמש בשני משפטים :

5. אם $Z_1,Z_2,Z_3,\ldots Z_k$ מ.מ-ים ב"ת כך ש $\forall i=1\ldots k$ $Z_i \sim N(0,1)$ אזי $\sum_i^k Z_i^2 \sim \chi_k^2$. לכן כל סכום המוצג במשתנה המקרי מתפלג חי בריבוע עם מספר דרגות חופש כמספר המחוברים : $\sum_{i=1}^{52}(Z_i)^2 \sim \chi_{52}^2$ ו $\sum_{i=1}^{42}(Z_i)^2 \sim \chi_{42}^2$

6. יהי מ.מ U המתפלג χ_m^2 ויהי מ.מ V המתפלג χ_n^2 . מ.מ אלו הם ב"ת. אזי למ.מ X המוגדר $\dfrac{U/m}{V/n}$ יש התפלגות F עם m,n דרגות חופש. נסמן $X \sim F_{m,n}$

עם קצת אלגברה בסיסית... נסיק כי הביטוי המוצג בתרגיל מתפלג $X \sim F_{52,42}$ נסו!!

מבחן בהסתברות- מועד א' סמסטר ב' 2014

<u>שאלה 1 : 25 נקודות</u>

יהי X מ.מ בעל פונקצית צפיפות $f\left(x\right)=c\left(1-x^2\right),\ \ 0<x<1$

ה. מצא את c (2 נקודות)

ו. מצא את פונקצית ההצטברות של X (3 נקודות)

ז. מצא את החציון (2 נקודות)

ח. מצא את התוחלת ואת השונות של X (4 נקודות)

ט. מצאו את ההסתברות ש $p\left(x^2\le 0.25\right)$ (2 נקודות)

י. מצא את הטווח הבינרבעוני של X (2 נקודות)

יא. דוגמים את X 100 פעמים מהו בקירוב האחוזון ה 12.5 העליון של ממוצע הדגימות? (4 נקודות).

יב. נגדיר את W להיות מ.מ המתפלג אקספוננציאלית. $W\sim \exp\left(\lambda\right)$. נגדיר את Y להיות מ.מ באופן הבא $Y=e^w$. מצא את פונקצית הצפיפות וההצטברות עבור Y **ללא מציאת פונקצית ההצטברות של** W . (4 נקודות)

יג. הוכח כי תוחלת של משתנה אקספוננציאלי הינו $\frac{1}{\lambda}$, ניתן על פי הגדרה או על פי פונקציה יוצרת מומנטים (2 נקודות)

פתרון שאלה 1:

א. בגלל שמדובר בפונקצית צפיפות אזי האינטגרל על כל התחום שווה ל-1:

$$\int_0^1 c\left(1-x^2\right)dx=1\ \rightarrow c\left(x-\frac{x^3}{3}\right)_0^1=\frac{2c}{3}\ \rightarrow\ \frac{2c}{3}=1\ \ \rightarrow c=\frac{3}{2}$$

ב. $F\left(x\right)=\int_0^x \frac{3}{2}\left(1-x^2\right)dx\ \rightarrow \frac{3}{2}\left(x-\frac{x^3}{3}\right)_0^x=\frac{3x}{2}-\frac{x^3}{2}$ ולכן נגדירה באופן מצא:

$$F\left(x\right)=\begin{cases}0 & x<0\\ \frac{3x}{2}-\frac{x^3}{2} & 0\le x\le 1\\ 1 & x>1\end{cases}$$

ג. על מנת למצוא את החציון נחפש את ה- a הראשון המקיים $F\left(a\right)=\frac{3a}{2}-\frac{a^3}{2}\ge 0.5$

מתקבל: $a\approx 0.347$

ד.התוחלת של X : $E(X)=\int_0^1 \frac{3}{2}x(1-x^2)dx \;\to\; \frac{3}{2}\left(\frac{x^2}{2}-\frac{x^4}{4}\right)_0^1=\frac{3}{8}$. עבור השונות נחשב

תחילה את המומנט השני : $E(X^2)=\int_0^1 \frac{3}{2}x^2(1-x^2)dx \;\to\; \frac{3}{2}\left(\frac{x^3}{3}-\frac{x^5}{5}\right)_0^1=\frac{1}{5}$. נציב בנוסחת

השונות $V(X)=E(X^2)-(E(X))^2=\frac{1}{5}-\left(\frac{3}{8}\right)^2\approx 0.06$

ה. נמצא $p(X^2\le 0.25)\to p(-0.5\le X\le 0.5)$ ומאחר ש- $0<x<1$ אזי נחפש את

ההסתברות הבאה $p(X\le 0.5)$. להלן החישוב

$$p(X\le 0.5)\;\to\; F(0.5)=\frac{3\cdot 0.5}{2}-\frac{(0.5)^3}{2}=\frac{11}{16}=0.6875$$

ו. הטווח הבינרבעוני : על מנת למצוא את הרבעון העליון והתחתון נחפש את ה- a וה b הראשונים המקיימים

$$\left.\begin{aligned}F(a)&=\frac{3a}{2}-\frac{a^3}{2}<\frac{3}{4}\\ F(b)&=\frac{3b}{2}-\frac{b^3}{2}<\frac{1}{4}\end{aligned}\right\}\to a-b$$

ז. אם דוגמים 100 פעמים את X אזי התפלגות הדגימה של הממוצע הינה נורמאלית עם תוחלת

השווה ל $\frac{3}{8}$ ושונות השווה $\frac{0.06}{100}$. עבור התפלגות נורמאלית סטנדרטית,מתקיים כי

$P(X>a)=0.125\to a=1.15$ כעת, נמצא את $\dfrac{x_i-\frac{3}{8}}{\sqrt{\frac{0.06}{100}}}=a=1.15$

כלומר: $x_i\approx 0.4$

ח. ידוע כי אם ל- W ישנה התפלגות מעריכית אזי פונקצית הצפיפות הינה

$f(w)=\begin{cases}\lambda e^{-\lambda w} & w>0\\ 0 & w\le 0\end{cases}$ ופונקצית ההצטברות הינה $F(w)=\begin{cases}1-e^{-\lambda w} & w>0\\ 0 & w\le 0\end{cases}$.

כעת נרצה למצוא את התפלגותו של $Y=e^W$

$$F_Y(a)=p(Y\le a)=p(e^W\le a)=p(W\le\ln a)=F_W(\ln a)$$

$$f_Y(a)=(F_Y(a))'=(F_W(\ln a))'\cdot\frac{1}{a}=f_W(\ln a)\cdot\frac{1}{a}\xrightarrow{f(w)=\lambda e^{-\lambda w}}\lambda e^{-\lambda\ln a}\cdot\frac{1}{a}=\lambda a^{-\lambda-1}$$

$$\Rightarrow F_Y(a)=\int_{-\infty}^{t}\lambda a^{-\lambda-1}=\ldots=1-a^{-\lambda}$$

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_{0}^{\infty}\lambda xe^{-\lambda x}\xrightarrow{\text{Integration by parts}}\begin{cases}u=x & u'=1\\ v'=\lambda e^{-\lambda x} & v=-e^{-\lambda x}\end{cases}$$

ט. $$-xe^{-\lambda x}-\int -e^{-\lambda x}dx=\left(-xe^{-\lambda x}-\frac{1}{\lambda}e^{-\lambda x}\right)_0^{\infty}\rightarrow\lim_{x\to\infty}\frac{x}{e^{\lambda x}}\xrightarrow{lopital}\lim_{x\to\infty}\frac{1}{\lambda e^{\lambda x}}\rightarrow 0$$

$$0-0-\left(0-\frac{1}{\lambda}\right)=\frac{1}{\lambda}$$

שאלה 2 : 23 נקודות

בחברת "העברת מסרים בע"מ", ידוע כי ההסתברות כי מסר עובר ברשת עמוסה בפחות מ -3 שניות היא 0.3, בין 3 ל- 7 שניות היא 0.6 ומעל 7 שניות היא 0.1. ידוע כי כל מסר עובר באופן מדויק או שגוי. אם המסר עבר בפחות מ 3 שניות, אזי ההסתברות כי עבר באופן מדויק היא 0.6. אם המסר עבר בין 3 ל- 7 שניות, אזי ההסתברות כי עבר באופן מדויק היא 0.5. ההסתברות כי המסר עבר מעבר ל- 7 שניות בהנתן כי עבר באופן מדויק, היא 0.04.

א. מה ההסתברות כי מסר עבר במדויק וגם עבר ביותר מ 7 שניות? (3 נקודות)

ב. מה ההסתברות כי מסר עבר במדויק (במספר דקות כלשהו)? (3 נקודות)

ג. מסר הגיע באופן שגוי, מה הסיכוי שעבר בפחות מ 3 שניות? (3 נקודות)

חברת "העברת מסרים בע"מ" מפצה על מסרים שנשלחו באופן שגויי. אם המסר נשלח בפחות מ 3 שניות והגיע שגויי היא מפצה ב 15 ₪. אם המסר נשלח בין 3 ל- 7 שניות והגיע שגויי היא מפצה ב 25 ₪. אם המסר נשלח מעבר ל 7 שניות והגיע שגויי היא מפצה ב 40 ₪.

ד. בנה טבלת התפלגות מתאימה כאשר המשתנה המקרי הינו הפיצוי הכספי. ומצא את התוחלת. (4 נקודות)

ה. מנהל החברה החליט כי ישנה את מדיניות החברה ויכפיל כל פיצוי ב 0.6 ולאחר זאת יוסיף 6 ₪. מצא את תוחלת סכום כסף הפיצוי. (3 נקודות)

ו. מצא את השונות לפני ואחרי שינוי מדיניות החברה. (2 נקודות)

ז. ידוע כי מספר המסרים מתפלג פואסונית, עוד ידוע כי ממוצע מספר קבלת מסרים שגויים בשעה היא 2. מצא את ההסתברות לימה ללא הודעות שגיאה (3 נקודות)

ח. מה ההסתברות שיתקבל יותר ממסר שגויי אחד במשך 6 שעות? (2 נקודות)

פתרון שאלה 2:

נסמן את המאורעות:

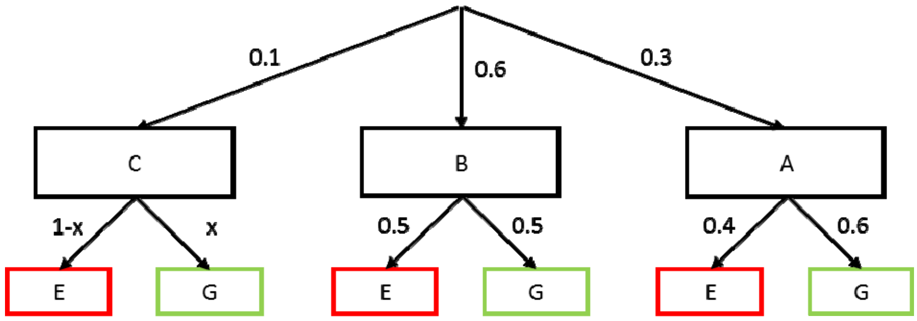
A – מסר עבר בפחות מ- 3 שניות.

B – מסר עבר ביותר מ-3 ופחות מ-7 שניות.

C – מסר עבר ביותר מ- 7 שניות.

G – המסר עבר במדויק. E – המסר עבר עם שגיאה.

נצייר את הנתונים בסכמת עץ:



כאשר סימנו את $P(\mathrm{G}\mid\mathrm{C})=x$

כמו כן נתון: $P(\mathrm{C}\mid\mathrm{G})=0.04$

דבר ראשון נמצא את x:

מנוסחת הסתברות שלמה: $P(G)=0.1\cdot x+0.6\cdot 0.5+0.3\cdot 0.6$

לפי נוסחת בייס: $P(\mathrm{C}\mid\mathrm{G})=\dfrac{x\cdot P(\mathrm{C})}{P(G)}=0.04\Rightarrow P(G)=\dfrac{x}{0.4}$

מכאן ניתן לחלץ את x: $\dfrac{x}{0.4}=0.1\cdot x+0.48\Rightarrow 10x=0.4x+1.92\Rightarrow x=0.2$

א) $P(\mathrm{G}\cap\mathrm{C})=0.1\cdot 0.2=0.02$

ב) $P(G)=\dfrac{x}{0.4}=0.5$

ג) $P(\mathrm{A}\mid E)=\dfrac{P(E\mid A)P(A)}{P(E)}=\dfrac{0.4\cdot 0.3}{1-0.5}=0.24$

ד) נסמן את הפיצוי באות Y, להלן טבלת ההתפלגות של Y:

Y	0	15	25	40
P(Y)	0.5	0.12	0.3	0.08

והתוחלת: $E(Y)=0+15\cdot 0.12+25\cdot 0.3+40\cdot 0.08=12.5$

ה) ניתן להשתמש בתכונות התוחלת ולקבל: $E(0.6Y+6)=0.6E(Y)+6=13.5$

ו) את השונות נמצא מתוך: $V(Y)=E(Y^2)+E(Y)^2$

$$E(Y^2)=15^2\cdot 0.12+25^2\cdot 0.3+40^2\cdot 0.08=342.5\Rightarrow V(Y)=186.25$$

$$V(0.6Y+6)=0.36V(Y)\Rightarrow V(0.6Y+6)=67.05$$

ז) נסמן באות X את מספר ההודעות השגויות במשך יממה. נתון כי X מתפלג פואסונית,

כלומר: $X\sim Pois\left(\lambda=2\frac{Error}{hour}\right)\Rightarrow X\sim Pois\left(\lambda=48\frac{Error}{day}\right)$

$$\Rightarrow P(X=0)=\frac{48^0}{0!}e^{-48}=1.425\cdot10^{-21}$$

ח) $X\sim Pois\left(\lambda=2\frac{Error}{hour}\right)\Rightarrow X\sim Pois\left(\lambda=12\frac{Error}{6hours}\right)$

$$\Rightarrow P(X>1)=1-P(X=1)-P(X=0)=1-\frac{12^1}{1!}e^{-12}-\frac{12^0}{0!}e^{-12}=0.99992$$

שאלה 3 : 23 נקודות

ידוע כי בקרב אוכלוסיית הסטודנטים בארץ ישראל, סטיית התקן של הציונים בקורס "מבוא להסתברות וסטטיסטיקה" הינה 1.2. לשם בניית רווח סמך לתוחלת ברמת ביטחון של 95%, חוקר דגם מדגם של 100 סטודנטים מתוך אוכלוסיה זו. ממוצע המדגם שהתקבל היה 80.

א. מצא את רווח סמך זה עם רמת ביטחון של 95% (3 נקודות). רווח הסמך הוא: $79.7648<\mu<80.2352$

ב. הסבר מה משמעותו של רווח זה שקבלת בסעיף א' (3 נקודות). ברמת ביטחון של 95% ניתן לקבוע כי התוחלת האמיתית של האוכלוסייה תמצא ברווח שמצאנו בסעיף א

ג. מצא את ההסתברות כי ציונו של אחד הסטודנטים הינו נמוך מ-85 (2 נקודות). לא ניתן לחשב כיון שלא נתונה התפלגות האוכלוסייה

ד. מצא את ההסתברות כי ציונו של אחד הסטודנטים הינו בין 50 ל 70 (2 נקודות). כמו שרשמתי במייל לא נחשב סעיף זה

ה. החוקר החליט לשנות את תוצאות **המדגם** והכפיל כל ציון ב 0.8 ולאחר מכן הוסיף 18 נקודות, האם רווח הסמך ישתנה? אם כן, כיצד, ומצא אותו. תן נימוק על כל אחת מהחלטותיך (4 נקודות).

רווח הסמך ישתנה – כך שממוצע המדגם החדש יוכפל ב0.8 ויתווסף לו 18. אך אורך הרווח לא ישתנה. נקבל: $81.7648<\mu<82.2352$

ו. החוקר פנה למועצה להשכלה גבוהה וביקש לשנות כל ציון לכלל אוכלוסיית הסטודנטים באופן הבא: הכפלת כל ציון ב 0.8 ולאחר מכן הוספה של 18 נקודות. האם רווח הסמך ישתנה? אם כן, כיצד, ומצא אותו. תן נימוק על כל אחת מהחלטותיך (4 נקודות).

רווח הסמך ישתנה – כך שממוצע המדגם החדש יוכפל ב 0.8 ויתווסף לו 18 ושונות האוכלוסייה גם תשתנה ותוכפל ב0.64. אורך הרווח במקרה הזה ישתנה.

חשוב להדגיש כי השוני בין הסעיפים הוא כי בסעיף ה' מדובר רק על טרנספורמציה בתצפיות המדגם ולכן השונות אינה משתנה (היא שונות האוכלוסייה) אך בסעיף ו' מדובר על טרנספורמציה של האוכלוסייה כולה, משמע גם המדגם משתנה וגם שונות האוכלוסייה כולה

נקבל: $81.8118 < \mu < 82.1882$

ז. ללא חישוב מפורש, הסבר מה יקרה לאורך רווח הסמך במידה והחוקר ידגום 500 סטודנטים, הסבר מדוע - תן הסבר סטטיסטי **ולא אלגברי** (3 נקודות). גודל המדגם גדל ועל כן התוצאות מדויקות יותר ולכן אורך הרווח יקטן

ח. הסבר מדוע כאשר מחשבים רווח סמך לתוחלת משתמשים בהתפלגות נורמאלית? (2 נקודות). לפי משפט הגבול המרכזי, אנו עורכים התפלגות דגימה של הממוצע והוא מתפלג נורמאלי

<u>שאלה 4.1</u> : בסעיפים א'-ו' נתונים מושגים, הסבר את משמעותם. (29 נקודות)

א. פונקצית הצטברות $F(x)$ ורשום את הגדרתה (2 נקודות)

ב. רמת ביטחון (בהקשר של רווח סמך) (2 נקודות)

ג. פונקצית צפיפות $f(x)$ (2 נקודות)

ד. התפלגות דגימה של הממוצע (2 נקודות)

ה. משתנה מקרי בדיד (2 נקודות)

ו. שונות המדגם (2 נקודות)

ז. מקדם המתאם של פירסון (2 נקודות)

פתרון שאלה 4.1:

א. $F(x)$ מוגדרת כפונקצית ההתפלגות ודרכה נוכל לחשב התפלגויות וגם להגיד לפונקצית הצפיפות: $F(a)$ המסומנת גם $F_x(a)$ מקיימת $F(a)=F_x(a)=p(X\le a)$ (השטח משמאל).

באופן דומה: $F(a\le X\le b)=p(X\le b)-p(X\le a)=F(b)-F(a)$.

הגדרה פורמאלית: תהי $f(x)$ פונקציה רציפה כך שעבור כל מתקיים $\int_a^b f(x)dx=F(b)-F(a)$ אזי קוראים ל $f(x)$ פונקצית צפיפות של שמ.מ רציף X

ו- $F(x)$ נקראת פונקצית ההצטברות או ההתפלגות.

ב. בהינתן שקבלנו $\bar{X}$ מסוים מתוך מדגם שדגמנו אזי μ שהיא התוחלת האמיתית תמצא ברווח בר הסמך בביטחון\רמת ביטחון של $1-\alpha$.

ג. בתורת ההסתברות, פונקצית צפיפות של משתנה מקרי היא פונקציה המתארת את צפיפות המשתנה בכל נקודה במרחב המדגם. ההסתברות שמשתנה מקרי יימצא בקטע מסוים היא האינטגרל של הצפיפות בקטע, ולכן המשתנה נוטה יותר לקבל ערכים שבהם הצפיפות גבוהה.

ד. התפלגות דגימה של סטטיסטי מסוים היא פונקצית התפלגות שלו. התפלגות זו תלויה בצורתו המתמטית של הסטטיסטי, בגודל המדגם ובתכונות האוכלוסייה ממנה נדגם. **עבור הממוצע** נקבל כי תוחלת ממוצע המדגם שווה לתוחלת המשתנה המקרי שממנו דוגמים.

במילים פשוטות יותר נוכל לומר כי ממוצע של הממוצעים של כל המדגמים האפשריים הינו בדיוק ממוצע האוכלוסייה זאת ללא תלות בגודל המדגם.

ה. אנו עורכים ניסוי כאשר התוצאות שלו מקריות, נסמן ב-X את המספר המצורף לתוצאה, או את התוצאה עצמה- אם היא מספרית.

הערך שקיבל- X יהיה תלוי בתוצאת הניסוי- שהוא **מקרי,** וכתוצאה מזה אנו מכנים את **X** – **משתנה מקרי, מ.מ**. משתנה מקרי בדיד הינו משתנה מקרי כמו שהגדרנו אשר מקבל רק ערכים בדידים.

ו. שונות הינה ממוצע המרחקים בריבוע של כל תצפית מממוצע המדגם. אנו נשאוף לשונות קטנה ככל שניתן

ז. מדד זה הינו מדד קשר והוא בודק קשר ליניארי בלבד נסמנו, r

r - מוגדר להיות כיחס בין השונות המשותפת למכפלת סטיות התקן של X ו- Y. נכתוב זאת כנוסחא :

$$r=\frac{\operatorname{cov}(x,y)}{s_x\cdot s_y}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{X}\right)\left(y_i-\bar{Y}\right)}{n\cdot s_x\cdot s_y}$$

ידוע כי מקדם המתאם של פירסון נע בין 1 ל- 1-. כאשר ככל שקרוב לערך 1 ישנו קשר חיובי וככל שקרוב ל1- ישנו קשר שלילי. אם שווה ל1 ו ל1- אזי נאמר כי הקשר מושלם.

__**שאלה 4.2.**__ ענה על הסעיפים הבאים (15 נקודות) :

א. נתון כי $Z_1,Z_2,Z_3,\ldots Z_k$ מ.מ-ים ב"ת כך ש $\forall i=1\ldots k \quad Z_i\sim N(0,1)$. הוכח ומצא את שם ההתפלגות ואת הפרמטרים המתאימים עבור סכום הריבועים של אלו, משמע $\sum_i^k Z_i^2$ (5 נקודות).

ב. יהיו $X_1,X_2,X_3,\ldots X_r\sim Gamma(1,\lambda)$ משתנים מקריים ב"ת. מצא ע"י הוכחה כיצד יתפלג $\sum_{i=1}^{n}X_i$ ומהם הפרמטרים המתאימים? (5 נקודות)

ג. הוכח כי טרנספורמציות ליניארית חיוביות על משתנים, אינן משנות את ערכו של מקדם המתאם של פירסון (5 נקודות).

פתרון שאלה 4.2:

א. **הוכחה**: שוב נתחיל מפונקציה יוצרת מומנטים, $M_{\sum_i^k Z_i^2}(t)=E\left(e^{t\sum_i^k Z_i^2}\right)$ בגלל שנתון שהמ.מ-ים ב"ת אזי מתקיים:

$E\left(e^{t\sum_i^k Z_i^2}\right)=E\left(e^{t\left(Z^2{}_1+Z^2{}_2+\ldots+Z^2{}_k\right)}\right)=E\left(e^{tZ^2{}_1}\right)\cdot E\left(e^{tZ^2{}_2}\right)\cdot\ldots\cdot E\left(e^{tZ^2{}_k}\right)$ ובגלל שהמ.מ-ים שוויי התפלגות (בעלי התפלגות זהה לחלוטין) לפי הנתון $\forall i=1\ldots k$ $Z_i\sim N(0,1)$ אזי $E\left(e^{tZ^2{}_1}\right)\cdot E\left(e^{tZ^2{}_2}\right)\cdot\ldots\cdot E\left(e^{tZ^2{}_k}\right)=\left[M_{Z^2}(t)\right]^k$. כעת מההוכחה הקודמת ראינו כי $M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$ ולכן $\left[M_{Z^2}(t)\right]^k\xrightarrow{M_{Z^2}(t)=\left(\frac{1}{1-2t}\right)^{\frac{1}{2}}}\left(\frac{1}{1-2t}\right)^{\frac{k}{2}}$. *משמעות הדבר היא כי סכום של מ.מ-ים ב"ת המתפלגים נורמאלית סטנדרטית מתפלג חי בריבוע עם מספר הדרגות החופש השווה למספר המחוברים.*

ב.

$$E\left(e^{t\sum_{i=1}^n X_i}\right)=E\left(e^{t(X_1+X_2+X_3+\ldots+X_n)}\right)=E\left(\prod_{i=1}^n e^{tx_i}\right)\xrightarrow{I.D}\prod_{i=1}^n E\left(e^{tx_i}\right)$$

$$\xrightarrow{X_1,X_2,X_3,\ldots X_r\sim Gamma(1,\lambda)}\prod_{i=1}^n E\left(e^{tx_i}\right)=\prod_{i=1}^n E\left(e^{tx}\right)\xrightarrow{M_X(t)=E\left(e^{tx}\right)=\left(\frac{\lambda}{\lambda-t}\right)^1}\prod_{i=1}^n\left(\frac{\lambda}{\lambda-t}\right)^1=\left(\frac{\lambda}{\lambda-t}\right)^n$$

קבלנו $\sum_{i=1}^n X_i\sim Gamma(r,\lambda)$

ג.

$$r=\frac{\operatorname{cov}(x',y')}{s'_x\cdot s'_y}=\frac{\sum_{i=1}^n\left(x'_i-\bar{X}'\right)\left(y'_i-\bar{Y}'\right)}{s'_x\cdot s'_y\,n}\xrightarrow{x'=b\cdot x_i+a\quad,\quad y'=c\cdot y_i+d}\frac{\sum_{i=1}^n\left(b\cdot x_i+a-\bar{X}'\right)\left[c\cdot y_i+d-\left(\bar{Y}'\right)\right]}{s'_x\cdot s'_y\,n}$$

$$\xrightarrow{\bar{X}'=b\cdot\bar{X}+a\quad,\quad\bar{Y}'=c\cdot\bar{Y}+d}\frac{\sum_{i=1}^n\left(b\cdot x_i+a-\left(b\cdot\bar{X}+a\right)\right)\left[c\cdot y_i+d-\left(c\cdot\bar{Y}+d\right)\right]}{s'_x\cdot s'_y\,n}$$

$$\frac{\sum_{i=1}^n\left(b\cdot x_i+a-b\cdot\bar{X}-a\right)\left[c\cdot y_i+d-c\cdot\bar{Y}-d\right]}{s'_x\cdot s'_y\,n}$$

$$\frac{\sum_{i=1}^n\left(b\cdot x_i-b\cdot\bar{X}\right)\left[c\cdot y_i-c\cdot\bar{Y}\right]}{s'_x\cdot s'_y\,n}\xrightarrow{gorem\ \ meshutaf\ \ b\ \&\ c}\frac{bc\sum_{i=1}^n\left(x_i-\bar{X}\right)\left[y_i-\bar{Y}\right]}{bcs_x\cdot s_y n}=\frac{\operatorname{cov}(x,y)}{s_x\cdot s_y}=r$$

דפי נוסחאות לבחינה – זה הדף שמופיע בבחינה!

בכד שבו n כדורים ממוספרים מ1 עד n בוחרים k כדורים בתנאים הבאים:

	עם החזרה	בלי החזרה
עם חשיבות לסדר	n^r	$n_r=\dfrac{n!}{(n-r)!}$
ללא חשיבות לסדר	$\binom{n+r-1}{r}$	$\binom{n}{r}=\dfrac{n!}{r!(n-r)!}$

1) $p(A^C)=1-p(A)\ ,\quad p(A)=1-p(A^C)$

2) $p(A\cup A^C)=p(\Omega)=1$

3) $p(A\cup B)=p(A)+p(B)-p(A\cap B)$

4) $p(A\cap B)+p(A\cap B^C)=p(A)$

5) $p(A^C\cap B^C)=p(A\cup B)^C$

בייס:

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

$$p(A)=p(A\mid B)\,p(B)+p(A\mid B^c)\,p(B^c)$$

$$P(B\mid A)=\frac{P(A\mid B)P(B)}{P(A\mid B)P(B)+P(A\mid B^c)\big(1-P(B)\big)}$$

חוק ההסתברות השלמה או חוק ההסתברות הכוללת:

יהיו $F_1,F_2,F_3\ldots$ מאורעות זרים, $\forall\, i\neq j$ $\ F_i\cap F_j=\varnothing$ ומתקיים $\bigcup_i F_i=\Omega$. נתון

מאורע מסוים A אזי בהכרח מתקיים: $A=(F_1\cap A)\cup(F_2\cap A)\cup(F_3\cap A)\ldots$.

אם Ω ניתן לפירוק של איחוד זר של F_i -ים (סופי או אינסופי) אזי

$$p(F_i\mid A)=\frac{p(A\mid F_i)\,p(F_i)}{\sum_i p(A\mid F_i)p(F_i)}$$

תקנון מוגדר: אם ל-X יש תוחלת μ ושונות σ^2 אזי Y המתוקנן מוגדר: $Y=\dfrac{X-\mu}{\sigma}$

התפלגויות בדידות:

הגדרתה של פונקציה יוצרת מומנטים עבור מ.מ בדיד: $M_X(t)=E\left(e^{tX}\right)=\sum_r e^{tr}\cdot p(X=r)$

הגדרת תוחלת ושונות עבור מ.מ בדיד : $\mu = E(X) = \sum_r r \cdot p(X=r)$

$$\sigma^2 = V(X) = E(X^2) - (E(X))^2$$

התפלגויות רציפות :

הגדרת פונקצית הצטברות : $F(a) = F_x(a) = p(X \le a)$

תוחלת ושונות של מ.מ רציף : $E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx \quad V(X) = E(X^2) - (E(X))^2$

הגדרתה של פונקציה יוצרת מומנטים עבור מ.מ רציף $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tX} f_X(x)dx$

טבלת התפלגות נורמאלית סטנדרטית המצטברת :

	0.00	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6103	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

טענה יהי מ.מ U המתפלג χ_m^2 ויהי מ.מ V המתפלג χ_n^2 . מ.מ אלו הם ב"ת. אזי

למ.מ X המוגדר $\frac{U/m}{V/n}$ יש התפלגות F עם m,n דרגות חופש. נסמן $X \sim F_{m,n}$

משפט: יהיו Z ו U מ.מ-ים בלתי תלויים המתפלגים באופן הבא $Z \sim N(0,1)$ ו-

$U \sim \chi_m^2$ אזי X המוגדר $X = \frac{Z}{\sqrt{U/m}}$ הינו מ.מ המתפלג T עם m דרגות חופש.

רווח סמך: $\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

נוסחא לגודל מדגם: $n \geq \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{\varepsilon} \right)^2$

שונות משותפת מסומנת: $\text{cov}(x,y)$ או S_{xy} מוגדרת באופן הבא: $\text{cov}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n}$

r - מקדם המתאם של פירסון $r = \frac{\text{cov}(x,y)}{s_x \cdot s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{n \cdot s_x \cdot s_y}$

רגרסיה ליניארית פשוטה: משוואת הרגרסיה היא $Y = \hat{\beta} X + \hat{\alpha}$ אם $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ ו-

$$\hat{\beta} = \frac{r \cdot s_y}{s_x}$$

אינטגרלים

1. $\int x^n dx = \frac{x^{n+1}}{n+1} + c \quad , \quad n \neq -1$

2. $\int (ax+b)^n dx = \frac{(ax+b)^{n+1}}{a(n+1)} + c \quad , \quad n \neq -1$

3. $\int \frac{1}{ax+b} dx = \frac{\ln|ax+b|}{a} + c$

4. $\int a^{Ex+b} dx = \frac{a^{Ex+b}}{E \ln a} + c$

5. $\int e^{Ex+b} dx = \frac{e^{Ex+b}}{E} + c$

$$\int \sin(ax+b)dx = \frac{-\cos(ax+b)}{a}+c$$ .6

$$\int \cos(ax+b)dx = \frac{\sin(ax+b)}{a}+c$$ .7

$$\int \frac{1}{x^2+a^2}dx = \frac{1}{a}\arctan\frac{x}{a}+c$$ .8

$$\int \frac{1}{x^2+1}dx = \arctan x+c$$ .9

$$\int \frac{1}{\sqrt{1-x^2}}dx = \arcsin x+c$$ .10

$$\int \frac{1}{x^2-a^2}dx = \frac{1}{2a}\ln\left|\frac{x-a}{x+a}\right|+c$$ .11

$$\int \frac{1}{\sqrt{a^2-x^2}}dx = \arcsin\frac{x}{a}+c$$ .12

$$\int \frac{1}{\sin x}dx = \ln\left|\tan\frac{x}{2}\right|+c$$ .13

$$\int \frac{1}{\cos x}dx = \ln\left|\tan\left(\frac{x}{2}+\frac{\Pi}{4}\right)\right|+c$$ .14

$$\int \tan x dx = -\ln|\cos x|+c$$ .15

$$\int \cot x dx = \ln|\sin x|+c$$ .16

$$\int \frac{1}{\sqrt{a+x^2}}dx = \ln\left|x+\sqrt{a+x^2}\right|+c$$ .17

$$\int \frac{1}{\sqrt{a^2-x^2}}dx = \arcsin\frac{x}{a}+c$$ .18

התפלגויות בדידות: זה דף ההתפלגויות שמופיע בבחינה

שם	סימון	צפיפות	תוחלת	שונות	פונקציה יוצרת מומנטים
אחידה	$X\sim U(a,b)$	$\frac{1}{b-a+1}$	$E(X)=\frac{b+a}{2}$	$V(X)=\frac{(b-a+1)^2-1}{12}$	$M_X(t)=\sum_{i=1}^{b-a+1}\frac{1}{b-a+1}e^{it}$
בינומית	$X\sim B(n,p)$	$p(X=r)=\binom{n}{r}p^r\cdot q^{n-r}$	$E(X)=np$	$V(X)=npq$	$M_X(t)=(q+p\cdot e^t)^n$
גיאומטרית	$X\sim G(p)$	$p(x=r)=p\cdot(1-p)^{r-1}$	$E(X)=\frac{1}{p}$	$V(X)=\frac{q}{p^2}$	$M_X(t)=\frac{pe^t}{1-qe^t}$
פואסון	$X\sim P(\lambda)$	$p(x=r)=\frac{\lambda^r}{r!}e^{-\lambda}$	$E(X)=\lambda$	$V(X)=\lambda$	$M_X(t)=e^{\lambda(e^t-1)}$
ברנולי	$X\sim Ber(p)$	$p(X=1)=p$	$E(X)=p$	$V(X)=pq$	$M_X(t)=q+pe^t$
היפרגיאומטרית	$X\sim HG(N,m,n)$	$p(X=k)=\frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}$	$E(X)=\frac{nm}{N}$	$V(X)=\frac{n(m/N)(1-m/n)(N-n)}{N-1}$	
בינומית שלילית	$X\sim NB(r,p)$	$P_X(k)=\binom{r+k-1}{k}(1-p)^r p^k$	$E(X)=\frac{rq}{p}$	$V(X)=\frac{rq}{p^2}$	$M_X(t)=\left(\frac{p}{1-pe^t}\right)^r$

התפלגויות רציפות:

שם	סימון	צפיפות	תוחלת	שונות	פונקציה יוצרת מומנטים
אחידה רציפה	$X\sim U(a,b)$	$f(x)=\begin{cases}0 & x<a \;\; or \;\; x>b\\ \frac{1}{b-a} & a\le x\le b\end{cases}$	$E(X)=\frac{b+a}{2}$	$V(X)=\frac{(b-a)^2}{12}$	$M_X(t)=\frac{e^{bt}-e^{at}}{t(b-a)}$
מעריכית	$X\sim \exp(\lambda)$	$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$	$E(X)=\frac{1}{\lambda}$	$V(X)=\frac{1}{\lambda^2}$	$M_X(t)=\frac{\lambda}{\lambda-t}$
נורמאלית	$X\sim N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$E(X)=\mu$	$V(X)=\sigma^2$	$M_X(t)=\exp\left[\mu t+\frac{1}{2}\sigma^2t^2\right]$
וייבול	$X\sim \mathrm{Weibull}(a,b)$	$f(x)=abx^{b-1}e^{-ax^b}$	$E(X)=a^{\frac{-1}{b}}\Gamma(1+b^{-1})$	$V(X)=a^{\frac{-2}{b}}\left[\Gamma(1+2b^{-1})-\Gamma^2(1+b^{-1})\right]$	$M_X(t)=a^{\frac{-t}{b}}\Gamma\left(1+\frac{t}{b}\right)$
גמא	$X\sim Gamma(r,\lambda)$	$f_X(x)=\frac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}$	$E(X)=\frac{r}{\lambda}$	$V(X)=\frac{1}{\lambda^2}$	$M_X(t)=\left(\frac{\lambda}{\lambda-t}\right)^r$
ביתא	$X\sim Beta(a,b)$	$f_X(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1}$	$E(X)=\frac{a}{a+b}$	$V(X)=\frac{ab}{(a+b+1)(a+b)^2}$	
קושי	$X\sim ch(x_o,\gamma)$	$f(x)=\frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$	לא קיים		

Chi-square	$X \sim \chi_k{}^2$	$f_X(x)=\frac{1}{\Gamma(k/2)}\left(\frac{1}{2}\right)^{(k/2)} x^{(k/2)-1}e^{-x/2}$	$E(x)=k$	$V(x)=$ $2k$	$M_X(t)=\left(\frac{1}{1-2t}\right)^{\frac{k}{2}}$
F	$X \sim F_{m,n}$	$f_X(x)=\frac{\Gamma\left[(m+n)/2\right]}{\Gamma[m/2]\Gamma[n/2]}\left(\frac{m}{n}\right)^{m/2}\frac{x^{(m-2)/2}}{\left[1+(m/n)x\right]^{(m+n)/2}}$	$E(X)=\frac{n}{n-2}$	$V(X)=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$	

T	$X \sim t_m$	$f_X(x)=\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma(m/2)}\frac{1}{\sqrt{m\pi}}\frac{1}{\left(1+\frac{x^2}{m}\right)^{\frac{m+1}{2}}}$	$E(x)=0$	$V(x)=\frac{m}{m-2}$	לא קיימת
לוג נורמאלית	$\ln X \sim N\left(\mu,\sigma^2\right)$	$f_X(x)=\frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\right\}$	$E(X)=\exp\left(\mu+\frac{1}{2}\sigma^2\right)$	$V(X)=$ $(e^{\sigma^2}-1)e^{2\mu+\sigma^2}$	לא קיימת

הוכחות למבחן

א. חוק ההסתברות השלמה או חוק ההסתברות הכוללת - עמוד 48

ב. טענה בעמוד 124

*** מה צריך לדעת לפתח עבור ההתפלגויות?

פיתוחים נוספים לבחינה - אלו פיתוחים ולא משפטי הוכחה ולכן הם לא נכללו ב" משפטים להוכחה"

להלן התוחלות והשונויות לבחינה

התפלגות בינומית: תוחלת לפי הגדרה ושונות ע"י פונקציה יוצרת מומנטים (עמוד 73)

התפלגות גיאומטרית: תוחלת לפי הגדרה ושונות ע"י פונקציה יוצרת מומנטים ולהראות שסכום הסתברויות שווה ל- 1 (עמוד 75)

התפלגות פואסון: תוחלת לפי הגדרה ושונות ע"י פונקציה יוצרת מומנטים ולהראות שסכום הסתברויות שווה ל-1 (עמוד 79)

התפלגות אחידה רציפה: תוחלת לפי הגדרה ושונות על פי הגדרה ולהראות שאינטגרל על כל תחום ההתפלגות שווה ל- 1 (עמוד 97)

התפלגות מעריכית: תוחלת לפי הגדרה ושונות על פי הגדרה ולהראות שאינטגרל על כל תחום ההתפלגות שווה ל-1 (עמוד 109)

התפלגות נורמאלית: מציאת תוחלת ושונות ע"י פונקציה יוצרת מומנטים (עמוד 103)

טענה: מקדם המתאם של פירסון נע בין 1 ל- 1-.

משפט 1: טרנספורמציות ליניארית חיוביות על המשתנים אינן משנות את ערכו של מקדם המתאם של פירסון (הוכחה בשיעורי בית).

משפט 2: טרנספורמציות ליניאריות על שני המשתנים משנות את ערך השונות המשותפת באופן הבא: כאשר ו- חיובים.

משפט 3: שונות משותפת של משתנה עם עצמו שווה לשונות שלו.

משפט 4: כאשר נרצה למצוא מקדם מתאם של פירסון של משתנה עם עצמו נמצא כי r שווה ל- 1.

אין צורך בהוכחת משפט הגבול המרכזי - רק הבנתו ויישומו.

אין צורך בהוכחת אי- שוויון צבישב - רק הבנתו ויישומו.

- לגבי הוכחות של תוחלות ושונויות מתוך פונקציות יוצרות מומנטים - מדובר בגזירה בסיסית! והפונקציות מופיעות בדף נוסחאות.

צריך לזכות נוסחאות של סטיית תקן, שונות וחציון (לא מופיע בדף נוסחאות)

אין נוסחאות גזירה כי יש נוסחאות אינטגרלים.