

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

סיכומי הרצאות

מבוא לסטטיסטיקה
והסתברות

הוקלד ע"י דביר חדד

מרצה: ד"ר עינת אביאל
סמסטר קיץ תשע"ב

26.7.12-הוצאה 1

avielei@cs.biu.ac.il

סטטיסטיקה:

מדע העוסק בנתונים כמותיים. את הנתונים אנו אוספים, מעבדים, מציגים ומסיקים מהם מסקנות.

הגדרות:

משתנה: תכונה המקבלת ערכים שונים. כגון טמפר', גובה, משקל, וכו'..
קבוע: גודל שערכו יחיד.



מיון משתנים:

1. מיון לפי מהות משתנה: כמותי ואיכותי.
כמותי: ערכי המשתנה מציינים כמות- גובה, משקל, וכו'..
איכותי: ערכי המשתנים נבחנים לפי סוג איכות ולא ע"י ביטוי מספרי- מין, מצב משפחתי, וכו'..
המשתנה הכמותי מתחלק לרציף ובדיד. רציף הכוונה שבין כל שני ערכים יש ∞ ערכים. (גיל, גובה). בדיד הכוונה שבין כל שני ערכים מספר סופי של ערכים. (מס' ילדים, מס' חדרים בבית)
2. מיון לפי רמת המדידה:
 - א. משתנה נורמלי (שמי) – ערכי המשתנה נבחנים לפי שמות, באשר אין משמעות לסדר בין הערכים. (מצב משפחתי, מוצא, דת...)
 - ב. משתנה אורדינאלי (סדורי) – ניתן לדרג את הערכים מהגבוה לנמוך. (דרגות בצבא, שביעות רצון...)
 - ג. משתנה אינטרוואלי (מרווחי) – ערכי המשתנים הם מספרים ויש משמעות למרווחים ביניהם. (IQ, טמפרטורה...)
 - ד. משתנה יחס- ערכי המשתנה הם במספרים ויש משמעות בחלוקת ערך אחד בשני וציון היחס ביניהם (גיל, גובה, משקל)...
3. מיון לפי כיוון התיאור:
 - משתנה תלוי: (מוסבר) משתנה המתואר על ידי המשנה הבלתי תלוי
 - משתנה בלתי תלוי: (מסביר) נשנה המסביר את המשנה התלוי.

סטטיסטיקה תיאורטית:

שכיחות: מספר המקרים במדגם עבור קבוצה מסוימת. התחלקות השכיחויות לפי הערכים של המשתנה.
תסומן f/F .

דוגמה:

שכיחות מצטברת $\frac{F(x)}{n}$ יחסית	$F(x)$ שכיחות מצטברת	שכיחות יחסית $\frac{f(x)}{n}$	שכיחות $f(x)$	X משתנה
$\frac{3}{22}$	3	$\frac{3}{22}$	3	5
$\frac{11}{22}$	11	$\frac{8}{22}$	8	$\mu_0=6$
$\frac{15}{22}$	15	$\frac{4}{22}$	4	7
$\frac{18}{22}$	18	$\frac{3}{22}$	3	8
$\frac{20}{22}$	20	$\frac{2}{22}$	2	9
$\frac{22}{22}$	22	$\frac{2}{22}$	2	10
		1	n=22 (גודל המדגם)	

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

מדדים:

1. שכית - M_0 . הערך / נקרה ששכיחותו היא הגבוהה ביותר. דיאגרמת מקלות- הצגה גרפית למשתנה בדיד.

מדדי מרכז:

1. חציון - Md . חציון הוא הערך המשתנה הנחקר שמחצית הנחקרים גדולים ממנו וחצי גדולים ממנו. החציון הוא ערך אמצעי בהתפלגות.
כלל: עבור קבוצת ערכים $\{X_1, \dots, X_n\}$ שמסודרים בסדר עולה. קיימות שתי אופציות: n זוגי או אי זוגי. אם n אי זוגי, אז החציון נמצא במקום ה- $\frac{n+1}{2}$ ז"א: $Md = X_{\frac{n+1}{2}}$
אם n זוגי, נקבל ש: $Md = \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}$
2. ממוצע: הממוצע הוא סכום כל ערכי המשתנה לכל הנחקרים חלקי מס' הנחקרים.

ובטבלת שכיחויות :

3. אמצע טווח: $MR = \frac{X_{min} + X_{max}}{2}$.

d צפיפות = $\frac{\text{שכיחות}}{\text{קבוצה רחב}}$	אמצע קטע- x חדש	I רוחב קבוצה	$f(x)$	X משתנה
$\frac{5}{10}$	45	10	5	40-50
$\frac{3}{10}$	55	10	3	50-60
$\frac{8}{10}$	65	10	8	60-70
$\frac{2}{10}$	75	10	2	70-80
$\frac{1}{10}$	85	10	1	80-90
			$n=19$	

דוגמה:

בהתחלה הנתונים היו רציפים ויצרנו x חדש כך שהמקרה בדיד ואנו יודעים לטפל בו. נמצא את השכיח, הממוצע, אמצע הטווח והחציון.

$$M_0 = 65 = \text{שכיח}$$

$$MR = \frac{40+90}{2} = 65 = \text{טווח אמצע}$$

$$\text{ממוצע} = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} = 60.25$$

$$Md = L_0 + \frac{\frac{n}{2} - F(x)}{f(x)} (L_1 - L_0) \quad \text{כדי לחשב את החציון ברציף, נשתמש בנוסחה הבאה}$$

נבחר מה כל אחד מהסימונים:

L_0 - גבול תחתון של קב' החציון

L_1 - גבול עליון של קב' החציון

$F(x)$ - שכיחות מצטברת עד קב' החציון (לא כולל)

$f(x)$ שכיחות של קבוצת החציון

n - גודל המדגם.

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

מדדי פיזור:

מהו פיזור? מידת השוני בקבוצה או גודל ההבדלים בין הנתונים לבין ערך מרכזי מסויים.

1. טווח: $R = X_{max} - X_{min}$

2. טווח ברבעוני: $IQR = Q_3 - Q_1$

3. שונות- ממוצע קיבועי הסטיות בין ערכי הסדרה לבין הממוצע. $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{x}^2$.
תמיד חיובית.

4. סטית התקן, s , שורש ריבועי של השונות. משמעות: רוב התצפיות נמצאות $\pm$ סטיית התקן מהממוצע.

בטבלת שכיחות:

בבדיד: $\bar{x} = \frac{\sum_{i=1}^n x_i f(x_i)}{n}$, $s^2 = \frac{\sum_{i=1}^n f(x_i)(x_i - \bar{x})^2}{n}$ ברציף אותו דבר בדיוק פשוט עם האינסוף החדש (אמצע קטע)

שאלה: נתונה קבוצה של 20 תלמידי. הממוצע שלהם הים 90, וסטיית תקן 5. מאוחר יותר נבחנו עוד 2 תלמידים וקיבלו את הציונים הבאים:

א'	ב'	ג'	ד'	ה'
87,93	83,97	97,98	81,82	85,95
לא השתנה	לא השתנה	עלה	ירד	לא השתנה
ירד	עלה	דרוש חישוב נוסף	לא השתנה	לא השתנה

ציון תקן- מתאר מיקום יחסי של תצפית ביחס לסדרה הסטטי אליה משתייך.

דוגמה: סטודנט נבחן בסטטיסטיקה ובביולוגיה. בסטטי קיבל 85 ובביו 80. בסטטי $\bar{x} = 82$ וגם $S_x = 5$

ובביולוגיה $\bar{x} = 70$ וגם $S_x = 3$.

מתקיים שציון התקן של התלמיד בסטטי קטן מציון התקן בביולוגיה. לכן מצטיין יותר בביולוגיה.

מקדם השתנות: מסמנים ב.c.v. $c.v. = \frac{S_x}{\bar{x}}$. מדד פיזור יחסי שנועד להשוות בין קבוצות מבחינת הומוגניות. ככל c.v. קטן יותר, הקבוצה יותר הומוגנית.

דוגמה:

באוכלוסיה מסוימת הגובה הממוצע הוא 174 ס"מ עם סטית תקן 15 ס"מ. המשקל הממוצע 60 קילו עם סטיית תקן של 10 קילו. היכן הקבוצה יותר הומוגנית? ברור שהגובה יותר הומוגני מחישוב פשוט.

טרנספורמציה ליניארית:

טרנספורמציה ליניארית זה שינוי קוויץ אם יש לנו משתנה x ומבצעים $y = ax + b$ אזי:

$Mdy = aMdx + b$ (החציון)

$\bar{y} = a\bar{x} + b$ (הממוצע)

$M_{oy} = aM_{0x} + b$ (השכיח)

$R_y = |a|R_x$ (הטווח)

$MR_y = aMR_x + b$ (אמצע טווח)

$S^2y = a^2S^2x$ (שונות)

$S_y = |a|S_x$ (סטית תקן)

$IQR_y = |a|IQR_x$ (טווח בין רבעוני)

2-31.7.12-הוצאה

התפלגות נורמאלית:

צורת ההתפלגות בעלת חשיבות רבה בתיאוריה הסטטי של הסקה ממדגם לאוכלוסיה, נקראת גם התפלגות פעמון/גאוס. התפלגות נורמאלית היא התפלגות סימטרית, חד שיאית, דמוית פעמון, למשתנה רציף בגבולות $-\infty < x < \infty$. המשתנה הנורמאלי מאופיין ע"י שני פרמטרים:

(1) תוחלת (ממוצע) של המשתנה המסומן

ב.מ.

(2) השונות-פיזור המסומן ע"י σ^2 .

סימון ההתפלגות: $x \sim N(\mu, \sigma^2)$.

השטח מתחת לעקומה שווה 100%. הקו המרכזי בתמונה הוא הממוצע, שכיח והחציון. זהו מרכז ההתפלגות.

ציון תקן: $Z_x = \frac{x-\mu}{\sigma}$, כאשר Z_x מתפלג

נורמאלית סטנדרטית, ז"א $Z_x \sim N(0,1)$, אפשר

להיעזר בטבלת ההתפלגות הנורמאלית. השטח שמצוין בטבלה הוא תמיד השטח משמאל. (את הטבלה ניתן למצוא בנספחים)

כללים שימושיים:

$$p(Z > Z_1) = 1 - \Phi(Z_1) \quad \bullet$$

$$p(-Z_1) = 1 - \Phi(Z_1) \quad \bullet$$

דוגמה:

גובה של חיילים מתפלג נורמאלי עם ממוצע 175 ס"מ. וסטיית תקן של 10 ס"מ. נחבר חייל באופן מקרי.

(א) מה ההסתברות שגובהו מעל 200 ס"מ?

(ב) מה ההסתברות שגובהו מתחת ל165 ס"מ?

(ג) מה גובה החיל ש90% מהחיילים נמוכים ממנו?

(ד) מה גובה החייל ש10% נמוכים ממנו?

פתרון:

$$x \sim N(175, 100) \quad (א)$$

$$p(x > 200) = p\left(z > \frac{200-175}{10}\right) = p(z > 2.5) = 1 - p(z \leq 2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062$$

$$p(x < 165) = p\left(z < \frac{165-175}{10}\right) = p(z < -1) = \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587 \quad (ב)$$

$$x=187.85, Z = \frac{x-175}{10} = 1.285 \quad (ג)$$

$$x=162.15, Z = \frac{x-175}{10} = -1.285 \quad (ד)$$

התפלגות הממוצע:

אם $x \sim N(\mu, \sigma^2)$ ז"א $Z_x = \frac{x-\mu}{\sigma}$ ואז $\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ וע"פ הגדרת ציון התקן $Z_{\bar{x}} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ ומתקבל כי:

$$Z_{\Sigma} = \frac{\Sigma x - n \cdot \mu}{\sigma \cdot \sqrt{n}} \text{ ואז } \Sigma x \sim N(n \cdot \mu, n \cdot \sigma^2)$$

דוגמה:

התפלגות משקל של ילדים הים נורמאלית עם ממוצע 40 ושונות של 49 ק"ג.

(א) אם נבחר ילד באופן מקרי מה ההסתברות שמשקלו מעל 50 ק"ג?

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

- (ב) מה ההסתברות שממוצע משקלם של עשרה ילדים לא יעלה על 45 ק"ג?
(ג) מה ההסתברות שסכום משקלם של שבעה ילדים יהיה לא פחות מ-270 ק"ג?
פתרון:

$$p(x > 50) = p\left(z > \frac{50-40}{7}\right) = 1 - p(z < 1.41) = 1 - 0.9207 = 0.0793 \quad \text{א)}$$

$$p(\bar{X}_{10} \leq 45) = p\left(Z_{\bar{x}} < \frac{45-40}{7/\sqrt{10}}\right) = p(Z < 2.26) = \Phi(2.226) = 0.9881 \quad \text{ב)}$$

$$p(\Sigma 7 \geq 270) = p\left(Z_{\Sigma} \geq \frac{270-7 \cdot 40}{7 \cdot \sqrt{7}}\right) = p(Z_{\Sigma} \geq -0.53) = 1 - p(Z_{\Sigma} < -0.53) = \Phi(0.53) = 0.7019 \quad \text{ג)}$$

משפט הגבול המרכזי:

אם לוקחים מדגם מקרי בגודל n ($n > 30$) מתוך אוכלוסייה שלא מתפלגת נורמאלית או שהתפלגותה לא ידועה לא ידוע. אם התוחלת μ והשונות σ^2 ידועות באותה אוכלוסייה אז $\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

הסקה סטטיסטית:

הסקה סטטיסטית עוסקת בשיטות להסקת מסקנות על כלל האוכלוסייה על סמך תוצאות של מדגם מייצג. רוב המחקרים מתבססים על תוצאות מדגם ולא על כל האוכלוסייה.

סימונים:	מדגם	אוכלוסייה
ממוצע	$\bar{x}$	μ
שונות	$\hat{\sigma}^2$	σ^2
פרופורציה	$\hat{p}$	p

שני מושגים הקשורים להסקה סטטיסטית:

1. רווח בר סמך
2. בדיקת השערות

רווח בר סמך:

הממוצע הכללי של אוכלוסייה נמצא: $\mu \leq \bar{x} + \text{סטייה}$ ו- $\mu \geq \bar{x} - \text{סטייה}$

כאשר אנו רוצים למצוא את הרווח אנחנו לוקחים בחשבון טעות שנקראת גם רמת מובהקות ומסומנת ב- α , או רמת בטחון/סמך שמסומנת ב- $1 - \alpha$. רווח הסמך הוא תחום סימטרי שבו נמצא הממוצע האמיתי של האוכלוסייה כאשר המטרה שלנו היא למצוא את גבולות הרווח, כאשר: $p(a \leq \mu \leq b) = 1 - \alpha$. לצורך בניית הגבולות נשתמש במשפט הגבול המרכזי.

מקרה I: רווח סמך לתוחלת כאשר השונות באוכלוסייה ידועה: (σ^2)

$$\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{אורך רווח הסמך: } L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \text{ והסטייה בין ממוצע}$$

המדגם לאוכלוסייה $\frac{L}{2} = \epsilon = Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$. כאשר σ -סטיית התקן. n -גודל המדגם. z -ערך מטבלת ההתפלגות הנורמאלית.

תרגיל:

רופא רוצה לאמוד את ממוצע לחץ השם של אוכלוסיית מבוגרים. לשם כך לשח מדגם מקרי של 225 מבוגרים ונמצא ממוצע של 148. ידוע כי סטטיית התקן באוכלוסייה היא 20.

- א) חשבו רווח סמך ברמת בטחון של 95%
ב) מהו אורך רווח הסמך?
ג) מה הסטייה בין ממוצע המדגם לתוחלת האוכלוסייה?

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

(ד) אם ניקח מדגם גדול פי ארבע, מה יקרה לאורך הרווח סמך?

(ה) חשבו את סעיף א עבור רמת בטחון של 99%

פתרון:

(א) $\alpha = 0.05$, נציב בנוסחא לרווח סמך כאשר $\sigma = 20$, $n = 148,225$, $\bar{x} = 148$

$$\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$148 - Z_{1-\frac{0.05}{2}} \cdot \frac{20}{\sqrt{225}} \leq \mu \leq 148 + Z_{1-\frac{0.05}{2}} \cdot \frac{20}{\sqrt{225}}$$

הנורמאלית כדי למצוא את Z . $Z_{1-\frac{0.05}{2}} = Z_{0.975} = 1.96$. שוב נציב. ונקבל כי:

$$145.387 \leq \mu \leq 150.613 \text{ ברמת בטחון של } 95\%$$

$$L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} = 2 \cdot 1.96 \cdot \frac{20}{\sqrt{225}} = 5.226 \quad (ב)$$

$$\frac{L}{2} = \epsilon = \frac{5.226}{2} = 2.613 \quad (ג)$$

$$L = 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \text{ ע"פ הנוסחא נקבל כי } L \text{ יקטן פי } 2. \quad (ד)$$

(ה) $\alpha = 0.01$, נציב בנוסחא לרווח סמך כאשר $\sigma = 20$, $n = 148,225$, $\bar{x} = 148$

$$\bar{x} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

$$148 - Z_{1-\frac{0.01}{2}} \cdot \frac{20}{\sqrt{225}} \leq \mu \leq 148 + Z_{1-\frac{0.01}{2}} \cdot \frac{20}{\sqrt{225}}$$

הנורמאלית כדי למצוא את Z . $Z_{1-\frac{0.01}{2}} = Z_{0.995} = 2.575$. שוב נציב. ונקבל כי:

$$144.57 \leq \mu \leq 151.43 \text{ ברמת בטחון של } 99\%$$

הערה:

$$L \geq 2Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \text{ נבצע כמה חישובים ונקבל כי } n \geq \left(\frac{2Z_{1-\frac{\alpha}{2}} \cdot \sigma}{L} \right)^2$$

להגבלת אורך רווח הסמך.

דוגמה:

חוקר רוצה להעריך ברמת בטחון של 95% את ממוצע השכר בישראל. ידוע שסטיית התקן באוכלוסייה היא 2000 ₪. מהו גודל המדגם הנדרש כך ש (א) אורך רווח השמך לא יעלה על 400 שח? (ב) הסטייה בין המדגם לאוכלוסייה לא תעלה על 400 שח?

פתרון:

(א) $\alpha = 0.05$, נציב בנוסחא שפיתחנו כאשר $\sigma = 2000$, $Z_{1-\frac{0.05}{2}} = 1.96$. $n \geq \left(\frac{2 \cdot 1.96 \cdot 2000}{400} \right)^2$ מכך

$$n \geq 384.16 \approx 385$$

$$n \geq 96.04 \approx 97 \text{ מכך } n \geq \left(\frac{2 \cdot 1.96 \cdot 2000}{800} \right)^2 \quad (ב)$$

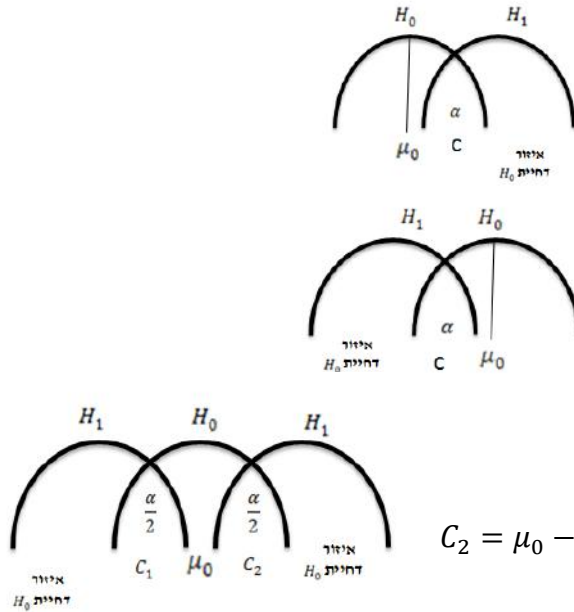
2.8.12-הוצאה 3

מקרה 1: בדיקת השערות לתוחלת האוכלוסייה כאשר השונות ידועה:

בבדידה השערות קיימות שתי השערות:

- H_0 , השערת בסיס, השערת האפס, מה שהיה ידוע ומקבול עד כה
 - H_1 , השערת אלטרנטיבית, השערת חוקר/השערה חדשה
- עלינו להחליט מי מבין שני ההשערות מתאימה יותר.

3 סוגי מבחנים:



(1) מבחן חד כיווני ימני:

$H_0: \mu = \mu_0$ השערת בסיס.

$H_1: \mu > \mu_0$ השערת חוקר.

$$C = \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$$

(2) מבחן חד כיווני שמאלי:

$H_0: \mu = \mu_0$ השערת בסיס.

$H_1: \mu < \mu_0$ השערת חוקר.

$$C = \mu_0 - Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}$$

(3) מבחן דו צדדי:

$H_0: \mu = \mu_0$ השערת בסיס.

$H_1: \mu \neq \mu_0$ השערת חוקר.

$$C_2 = \mu_0 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, C_1 = \mu_0 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

הטעויות הסטטיסטיות בבדיקה השערות:

מצב בפועל

	H_0 נכונה	H_1 נכונה
דוחים H_0	α טעות מסוג ראשון	$1 - \beta$ עצמת המבחן
דוחים H_1	$1 - \alpha$	β טעות מסוג שני

השאפה היא שהטעויות α, β יהיו כמה שיותר קטנות, במידת האפשר.

בין α, β קיים קשר הפוך.

- ככל α גדל, β קטן ו- $1 - \beta$ גדל

- ככל שקטן β , גדל $1 - \beta$ קטן

דוגמה:

ידוע שממוצע הכנסה חודשי הוא 6000 שח עם סטיית תקן של 500 שח. חוקר טוען כי ממוצע ההכנסה בשנה האחרונה עלה. לצורך בדיקת הטענה נלקח מדגם מקרי של 100 אנשים, ונמצא כי הממוצע 6125.

(א) כתבו השערות

(ב) בדקו את ההשערות ברמת מובהקות של 5%.

(ג) מה מסקנתכם עבור רמת מובהקות של 10%? האם יש צורך בחישוב נוסף?

פתרון:

(א) $H_0: \mu = 6000$. $H_1: \mu > 6000$

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

- ב) $C = 6000 + 1.645 \cdot \frac{500}{\sqrt{100}}$ מכאן $C = \mu_0 + Z_{1-\alpha} \cdot \frac{\sigma}{\sqrt{n}}, \alpha = 5\%, Z_{0.95} = 1.645$
מתקבל כי $C = 6082.25$ ולכן דחינו את H_0 ומסיקים את H_1 .
ג) אין צורך בחישוב נוסף כי אם דחינו ב-5% ברור שנדחה גם ב-10%.

הגדרה: p. Value: (α) : זוהי ה- α המינימאלית עבורה דוחים את H_0 . לכל α שגדולה מ-p.v. דוחים את H_0 .

$$\text{במבחן ימני: } p.v = p\left(z > \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

$$\text{במבחן שמאלי: } p.v = p\left(z < \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right)$$

במבחן דו צדדי: $p.v = 2p.v$ לצד המתאים.

בהמשך לדוגמה, חשבו את p.v.:

$$p.v = p\left(z > \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}\right) = p\left(z > \frac{6125 - 6000}{500/\sqrt{100}}\right) = p(z > 2.5) = 1 - p(z \leq 2.5) = 1 - \Phi(2.5) = 0.0062$$

עבור $\alpha > 0.62\%$ דוחים את H_0 .

מקרה 2: רווח סמך ובדיקת השערות לתוחלת μ כאשר השונות באוכלוסייה לא ידועה.

במקרה זה בו השונות באוכלוסייה לא ידועה נצטרך להשתמש במקום בשונות של המדגם:

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n-1}$$

בהתפלגות t , כאשר התפלגות t היא התפלגות סימטרית שטוחה יותר בהשוואה להתפלגות נורמאלית. התפלגות t תלויה בגודל המדגם.

$$\text{רווח סמך: } \bar{x} - t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}$$

את $Z_{\bar{x}}$ מחליף $t_{\bar{x}} = \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}$.

דוגמה:

משקל של 7 קופסאות מסוימות נתון: 100, 98, 104, 104, 98, 102, 96. בהנחה שמספר הקופסאות מתפלג נורמאלית, בנו רווח סמך לתוחלת ברמת בטחון של 99%.

פתרון:

$$\hat{s}^2 = \frac{104^2 + \dots + 7 \cdot 100^2}{7-1} = \bar{x} = \frac{100 + \dots + 96}{7} = 100$$

מאחר והשונות לא ידועה, נחשב את הערכים הבאים: $\hat{s} = \sqrt{8} = 2.83$. רוצים רמת בטחון של 99%, לכן נחשב את $t_{7-1, 1-\frac{0.01}{2}} = t_{6, 0.995} = 3.707$ ומכאן נקבל כי $\hat{s} = \sqrt{8} = 2.83$.

$$96.034 \leq \mu \leq 100 + 3.707 \cdot \frac{2.83}{\sqrt{7}} \leq 103.965$$

נציב הכל ונקבל כי: $100 - 3.707 \cdot \frac{2.83}{\sqrt{7}} \leq \mu \leq 100 + 3.707 \cdot \frac{2.83}{\sqrt{7}}$ ולכן $96.034 \leq \mu \leq 103.965$ ברמת בטחון של 99%.

בבדיקת השערות:

$$\text{במבחן ימני: } p.v = p\left(t_{n-1} > \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}\right)$$

$$\text{במבחן שמאלי: } p.v = p\left(t_{n-1} < \frac{\bar{x} - \mu_0}{\hat{s}/\sqrt{n}}\right)$$

במבחן דו צדדי: $p.v = 2p.v$ לצד המתאים.

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

דוגמה :

בעיר מסוימת ידוע שהציון הממוצע במתמטיקה מתפלג נורמאלית עם $\mu = 70$. מנהל בית הספר בעיר טוען שהממוצע גבוה יותר. לצורך כך נלקח מדגם מקרי של 8 תלמידים, להלן ציוניהם:
68,71,79,76,81,69,72,66

(א) כתבו השערות

(ב) מצאו את p.v.

פתרון :

(א) $H_1: \mu > 70$. $H_0: \mu = 70$

(ב) $p.v = p\left(t_7 > \frac{72.75-70}{5.391/\sqrt{8}}\right) = p(t_7 > 1.44)$, $\hat{s} = 5.391$, $\bar{x} = 72.75$

$0.05 \leq p.v. \leq 0.1$, $0.9 \leq 1 - p.v. \leq 0.95$ מתחת לגבול תחתון, לא דוחים את H_0 . בין לגבולות, לא ידוע, ומעל לגבול העליון בטוח דוחים את השערת הבסיס. מסקנה: לכל α שגדולה מ-10% דוחים את H_0 .

הקשר בין רווח סמך לבדיקת השערות:

לפעמים ניתן לדעת אם דוחים את H_0 או לא לפי רווח סמך כאשר יש התאמה ברמת המובהקות α . אם μ_0 נמצא בגבולות בתוך גבולות רווח הסמך אזי לא דוחים את H_0 , ואם הוא נמצא מחוץ לגבולות, ניתן לדחות את השערת הבסיס.

דוגמה :

ברמת בטחון של 96% התקבל רו"ס: $76 \leq \mu \leq 84$. לפי רו"ס זה בדקו האם נדחה H_0 במקרים הבאים:

(א) $\alpha = 4\%$, $H_1: \mu \neq 82$ $H_0: \mu = 82$

(ב) $\alpha = 6\%$, $H_1: \mu \neq 86$ $H_0: \mu = 86$

פתרון :

(א) השערת הבסיס נמצאת בתוך גבולות הרו"ס ולכן לא דוחים אותה. (רווח הסמך הוא ברמת בטחון 96%)

(ב) אם ב-4% מחוץ לגבולות רווח הסמך, ברור שב-6% זה גם כן יקרה, ולכן H_0 נדחה.

נוסחה חשובה: (כאשר השונות באוכלוסייה ידועה)

$$n \geq \left(\frac{\sigma(Z_{1-\alpha} + Z_{1-\beta})}{\mu_1 - \mu_0} \right)^2$$

הטעויות α, β .

דוגמה :

מהו גודל המדגם המינימאלי שחוקר צריך לקחת כאשר $\sigma = 200$ ומעוניין ש $\alpha < 5\%$, $\beta < 1\%$. עבור ההשערות $H_1: \mu = 3600$ $H_0: \mu = 3500$

פתרון :

$$n \geq 63.07 \approx 64, n \geq \left(\frac{200(Z_{0.95} + Z_{0.99})}{3600 - 3500} \right)^2, n \geq \left(\frac{\sigma(Z_{1-\alpha} + Z_{1-\beta})}{\mu_1 - \mu_0} \right)^2$$

נציב בנוסחה

מקרה 3: בדיקת השערות ורו"ס למדגמים תלויים=מזווגים:

במדגמים אלו לוקחים 2 קבוצות התלויות אחת בשנייה. נגדיר 2 סדרות, x,y, כאשר סדרת ההפרשים = d.

$$\left(\bar{d} = \frac{\sum_{i=1}^n d_i}{n} \right) . \mu_d$$

אני נעסוק בתוחלת ההפרשים, μ_d

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

רו"ס ל μ_d מוגדר כך: $\bar{d} - t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{\hat{s}_d}{\sqrt{n}} \leq \mu_d \leq \bar{d} + t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{\hat{s}_d}{\sqrt{n}}$
ועבור בדיקת השערות: $H_0: \mu_d = 0$, $H_1: \mu_d > 0$, $H_1: \mu_d < 0$, $H_1: \mu_d \neq 0$ בהתאם לסוג המבחן.

מניחים שסדרת ההפרשים $d_i \sim N(\mu_d, \sigma_d^2)$, כאשר $\hat{\sigma}_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}$ ונסמן גם $t_{\bar{d}} = \frac{\bar{d} - d_0}{\hat{s}_d / \sqrt{n}} \sim t_{n-1}$

ימני: $p.v. = p(t_{n-1} > \frac{\bar{d} - d_0}{\hat{s}_d / \sqrt{n}})$, שמאלי: $p.v. = p(t_{n-1} < \frac{\bar{d} - d_0}{\hat{s}_d / \sqrt{n}})$. דו"צ: $p.v. = 2p.v.$ לצד המתאים.

דוגמה:

רופא סיני טוען כי בידו טיפול המפחית עישון. להלן נתוני מטופליו על מס' הסיגריות היומיות לפני ואחרי הטיפול.

6	5	4	3	2	1	מטופל
40	30	20	15	20	10	X לפני
10	15	10	10	10	5	Y אחרי
30	15	10	5	10	5	d=x-y

(א) נסחו את ההשערות ומצאו את $p.v.$:

(ב) בנו רו"ס לממוצע השינוי בכמות הסיגריות ברמת בטחון של 98%:

פתרון:

(א) $H_0: \mu_d = 0$, $H_1: \mu_d > 0$. $\bar{d} = \frac{30+15+\dots+5}{6} = 12.5$. $\hat{s}_d = 9.35$. $p.v. = p(t_5 > \frac{12.5-0}{9.35/\sqrt{6}})$

$p.v. = p(t_5 > 3.27)$ מכאן ש $99\% \leq 1 - p.v. \leq 97.5\%$, ולכן $2.5\% \leq p.v. \leq 1\%$. לכל α שגדול מ-2.5% נקבל שהטיפול אכן יעיל.

(ב) $1 - \frac{\alpha}{2} = 0.99$. $t_{5, 99\%} = 3.365$. $\bar{x} - t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1, 1-\frac{\alpha}{2}} \cdot \frac{\hat{s}}{\sqrt{n}}$. נציב ונקבל כי: $-0.34 \leq \mu \leq 25.34$, ברמת בטחון של 98%.

מקרה 3: בדיקת השערות ורו"ס לפרופורציה (p) באוכלוסייה

נגדיר: $\hat{p} = \frac{x}{n}$, ונסמן $q = 1 - p$.

נניח כי $\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$, וגם כי $Z_{\hat{p}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$.

רו"ס: $\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

בדיקת השערות: $H_0: P = p_0$, $H_1: P > p_0$, $H_1: P < p_0$, $H_1: P \neq p_0$ בהתאם לסוג המבחן.

ימני: $p.v. = p(Z > \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}})$, שמאלי: $p.v. = p(Z < \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}})$. דו"צ: $p.v. = 2p.v.$ לצד המתאים.

מציאת גודל מדגם מינימאלי:

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

$$(1) \text{ להגבלת אורך רווח הסמך (L): } n \geq \left(\frac{2Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\hat{p}(1-\hat{p})}}{L} \right)^2 \text{ כשהפרופורציה במדגם לא ידועה}$$

לוקחים $p=0.5$.

$$(2) \text{ להגבלת הטעויות } \alpha, \beta, \text{ כאשר } H_1: P = P_1 \text{ } H_0: P = P_0$$

$$n \geq \left(\frac{\sqrt{p_0(1-p_0)} \cdot Z_{1-\alpha} + \sqrt{p_1(1-p_1)} \cdot Z_{1-\beta}}{P_1 - P_0} \right)^2 \text{ ולהציב } \odot$$

דוגמה:

רוצים לבדוק האם מוטיבציה של מתגייסים לרבי ירדה. מנתוני העבר ידוע כי 40% משרתים בקרבי. נלקח מדגם של 100 מתגייסים חדשים ונמצא כי 30 מתוכם הלכו לקרבי.

(א) נסחו השערות ומצאו את רמת המובהקות המינימאלית לדחיית H_0 .

(ב) בנו רי"ס לפרופורציית המתגייסים לקרבי ברמת בטחון של 95%.

פתרון:

(א) השערות: $H_1: P < 0.4$, $H_0: P = 0.4$. ורמת מובהקות מינימלית:

$$p.v. = p \left(Z < \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right) = p \left(Z < \frac{0.3 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{100}}} \right) = \phi(-2.04) = 1 - \phi(2.04) =$$

0.0207. עבור כל α שגדולה מ-2.07% דוחים את השערת הבסיס.

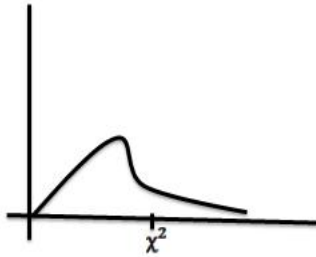
$$(ב) Z_{1-\frac{\alpha}{2}} = 1.96, 1 - \frac{\alpha}{2} = 0.975, 1 - \alpha = 0.95$$

$$\hat{p} - Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + Z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.21 \leq p \leq 0.39, 0.3 - 1.96 \cdot \sqrt{\frac{0.3 \cdot 0.7}{100}} \leq p \leq 0.3 + 1.96 \cdot \sqrt{\frac{0.3 \cdot 0.7}{100}}$$

95%

4-5.8.12-הוצאה



התפלגות χ^2 , (חי בריבוע) היא התפלגות אי סימטרית ימנית (חיובית) בעלת ערכים חיוביים בלבד. התפלגות זו תלויה בגודל המדגם. השימוש בטבלת χ^2 היא כמו בטבלת t .

χ^2 לטיב התאמה:

במבחן זה נרצה לבדוק האם תופעה מסוימת מתפלגת כצפוי. יש טיב התאמה H_0 , או אין טיב התאמה H_1 .

כלל ההכרעה במבחן זה: $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} > \chi_{k-1, 1-\alpha}^2$ אזי

דוחים את H_0 . [K -מספר הקטגוריה], [O -מה שנצפה בפועל ($observe$) E זה המצופה ($expected$)].

לדוגמה:

על ידי הטלת קובייה 120 פעמים התקבלו התוצאות הבאות:

סה"כ	6	5	4	3	2	1	תוצאה
O_i	23	17	25	22	15	18	
E_i	20	20	20	20	20	20	

(א) בדקו ברמת מובהקות של 5% האם הקובייה הוגנת, סימטרית, וכן נסחו השערות.

(ב) מצא את α מינימאלית עבורה נוכל לדחות את H_0 .

פתרון:

(א) השערת הבסיס, H_0 היא שהקובייה סימטרית, ו H_1 הוא שאין טיב התאמה (קובייה לא סימטרית). $\chi_{5, 0.95}^2 = 11.07$, זה של הטבלה, ועכשיו למחשוב: (ע"פ הנוסחה) $\chi^2 = 3.8$. קיבלנו

כי המחשוב קטן מהערך של הטבלה, ולכן לא דוחים את H_0 .

(ב) נחפש ערך χ^2 של טבלה שיהיה קטן מהערך המחושב: 3.8. ז"א, $\chi^2 < 3.8$ טבלה. נביט בטבלה

ונקבל כי $0.5 \leq 1 - p.v. \leq 0.75$, ז"א ש $0.25 \leq p.v. \leq 0.75$, ז"א שלכל α גדולה מ-75%

דוחים את H_0 .

מבחן χ^2 לאי תלות (אי קשר):

מבחן זה בודק האם קשר (תלות) בין שני משתנים איכותיים.

השערת הבסיס היא שאין קשר (H_0), ו H_1 היא שיש קשר.

כלל ההכרעה הוא: אם $\chi^2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} > \chi_{(R-1)(C-1), 1-\alpha}^2$ כאשר R מספר השורות C מספר

העמודות, אזי דוחים את השערת הבסיס H_0 .

דוגמה: רוצים לבדוק אם קיים קשר/תלות בין מגדר (זכר/נקבה) לבין צבע שיער (שחור/חום/בלונד) לצורך כך נלקח מדגם של 200 והתקבלו התוצאות הבאות:

E מגדר / O צבע	שחור	חום	בלונד	סה"כ
זכר	50	40	30	120
נקבה	30	30	20	80
סה"כ	80	70	50	200

(א) כתבו השערות ובדקו ברמה של 5%.

(ב) מצאו את $P.V.$

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

פתרון :

א) H_0 : אין תלות, H_1 : יש תלות בין המגדר לצבע השיער. $\chi^2_{(2-1)(3-1), 95\%} = 5.991$, ועכשיו נמצא

$$\chi^2_{\text{מחושב}} = \chi^2_{\text{טבלה}} = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{(50-48)^2}{48} + \dots + \frac{(30-32)^2}{32} = 0.466$$

ולכן לא דוחים את H_0 .

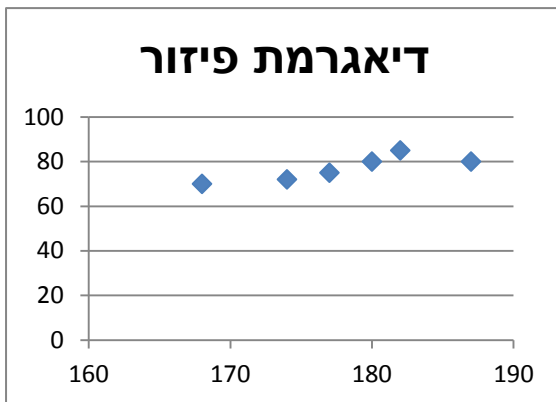
ב) $\chi^2_{\text{מחושב}} = 0.466 < \chi^2_{\text{טבלה}}$, לכן $0.1 \leq 1 - p.v. \leq 0.25$, ומכאן ש $0.75 \leq p.v. \leq 0.9$, לכן

עבור כל α שגדולה מ-90% נדחה את השערת הבסיס ונסיק כי יש תלות בין המשתנים.

מתאם רגרסיה :

נרצה לבדוק האם קיים קשר/תלות ליניארית בין 2 משתנים כמותיים. ישנם שני סוגי קשרים : חיובי ושלילי.
בקשר חיובי : ככל שערך משתנה 1 עולה כך עולה גם השני.
בקשר שלילי : ככל שערך משתנה אחד עולה כך יורד השני.

דוגמה :



	6	5	4	3	2	1	
גובה	187	182	180	177	174	168	
משקל	80	85	80	75	72	70	

מקדם המתאם של פירסון : $-1 \leq r_{xy} \leq 1$.

כאשר $r=1$ קשר חיובי מושלם.

$r=-1$ קשר שלילי מושלם.

$0.7 \leq |r| < 1$ קשר חזק.

$0.4 \leq |r| < 0.7$ קשר בינוני.

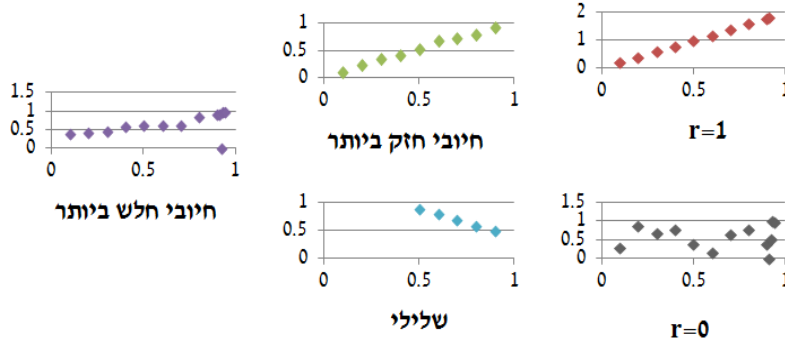
$0 < |r| < 0.4$ קשר חלש.

$r=0$ אין קשר.

(כאשר אין קשר יכול להתקבל

גם קו אופקי מקביל לציר

איקס, הציר התחתון)



חישוב $r_{xy} = \frac{\text{cov}(x,y)}{S_x \cdot S_y}$: כאשר $r_{xy} = \frac{\sum_{i=1}^n X_i \cdot Y_i}{n} - \bar{X} \cdot \bar{Y}$, זה יכול להיות גם שלילי וגם חיובי...

$$S_x = \frac{\sum_{i=1}^n X_i^2}{n} - \bar{X}^2, S_y = \frac{\sum_{i=1}^n Y_i^2}{n} - \bar{Y}^2$$

דוגמה :

הגובה (x) הממוצע של המדגם מהשאלה הקודמת הוא 178 ס"מ, המשקל (y) הממוצע הוא 77 קילוגרם, $S_y = 5.164$, $S_x = 6.027$ ו $\text{Cov}(X,Y)=26.166$ וע"פ נוסחא, $r_{xy} = 0.84$ ולכן קשר חיובי חזק.

רגרסיה :

אם קיים קשר בין שני משתנים נרצה למצוא את הקשר ע"י משוואת רגרסיה פשוטה ליניארית מהצורה

ד"ר עינת אביאלי הוקלד ע"י דביר חדד

$\hat{y} = bx + a$ ניבוי y ע"י x או $\hat{x} = b'y + a'$ ניבוי x ע"י y .
חישוב a, b : עבור $\hat{y} = bx + a$, $b = \frac{\text{cov}(x,y)}{S_x^2} = r \frac{S_y}{S_x}$, ואת a מחשבים ע"י הצבה של הממוצעים
 $\bar{y} - b\bar{x} = a$. ועבור $\hat{x} = b'y + a'$, נגדיר $b' = \frac{\text{cov}(x,y)}{S_y^2} = r \frac{S_y}{S_x}$ ושוב את a' ע"י הצבה של הממוצעים.

דוגמה:

המשך גובה ומשקל (דוגמה קודמת). על סמך הנתונים הקודמים:
(א) מצאו את משוואת הרגרסיה לניבוי משקל y לפי גובה.
(ב) מה תהיה הערכתם למשקל של אדם שגובהו 181 ס"מ?
פתרון:

(א) $b = \frac{26.166}{6.027^2} = 0.72$ נמצא את a ע"פ הנוסחא ונקבל $a = -51.16$, ולכן משוואת הרגרסיה לניבוי הגובה היא: $\hat{y} = 0.72x - 51.16$.
(ב) נציב במקום x את הגובה 181 ונקבל משקל 79.21 קילוגרם.

הסתברות:

ניתן להגדיר את המושג הסתברות של מאורע באמצעותו שכיחותו היחסית. נתאר לנו ניסוי, שמרחב המדגם שלו הוא S . נניח שמצעים ניסוי זה שוב ושוב באותם תנאים בדיוק. לכל מאורע E במרחב המדגם S נסמן את השכיחות של מאורע E ב $V(E)$ דהיינו מספר הפעמים שהמאורע E מתרחש בחזרות הראשונות על הניסוי. ואז נגדיר את ההסתברות של המאורע E ע"י $p(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$. ז"א ההסתברות $p(E)$ מוגדרת כגבול של המספר היחסי של הפעמים שבהן E מתרחש. $p(E)$ היא השכיחות היחסית הגבולית של E .
 $p(E)$ מקיים את 3 האקסיומות הבאות:

$$0 \leq p(E) \leq 1 \quad (\text{א})$$

$$p(S) = 1 \quad (\text{ב})$$

$$E_i, E_2, \dots \text{ כלומר מאורעות המקיימים } E_i \cap E_j = \emptyset \text{ לכל } i \neq j \quad (\text{ג})$$

מתקיים $p(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} p(E_i)$ (לכל סדרת מאורעות זרים, ההסתברות שלפחות אחד מהם יתרחש שווה לסכום ההסתברות של המאורעות המהווים את הסדרה)

הסתברות מותנית ואי תלות:

הגדרה: ההסתברות המותנית של מאורע E בתנאי/בהנחה F מסומנת $p(E/F)$ אם $p(F) > 0$ אזי ניתן לחשב זאת ע"י הנוסחה $p(E/F) = \frac{p(E \cap F)}{p(F)}$

דוגמה:

מטילים שני קוביות, אדומה וכחולה. נניח שלכל אחת מ-36 האפשרויות סיכויים שווים להתרחש, כלומר ההסתברות של כל תוצאה היא $1/36$. נניח שראינו שסכום מה שהתקבל בקוביות שווה לשמונה. מה הסיכוי שהקובייה האדומה הראתה שלוש?
פתרון:

$$E = \text{סכום של 8 ב-2 הקוביות}$$

$$F = \text{תוצאת הקובייה האדומה היא 3. } p(F) = \{(6,2), (5,3), (4,4), (3,5), (2,6), (4,4)\}$$

$$p(E/F) = \frac{p(E \cap F)}{p(F)} = \frac{1/36}{6/36} = \frac{1}{6}$$

$$\text{טענה: } p(E \cap F) = p(E/F) \cdot p(F)$$

דוגמה: קרן מהססת אם לבחור קורס בצרפתית או בכימיה. היא מעריכה שתקבל 10 בצרפתית

בהסתברות של 0.5 ותקבל 10 בכימיה בהסתברות של $\frac{2}{3}$. אם קרן מחליטה לבחור בין הקורסים ע"י הטלת מטבע, מה ההסתברות שתקבל 10 בכימיה?

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

פתרון: ע"פ הטענה נקבל כי $p(E \cap F) = p(E/F) \cdot p(F)$ כאשר E זה לקבל F 10 לבחור בכימיה. לכן ההסתברות היא מכפלה של $\frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$.

נרחיב את הטענה לדבר הבא: לחישוב ההסתברות של חיתוך מספר כלשהו של מאורעות:
נוסחת הכפל: $p(E_1 \cap \dots \cap E_n) = p(E_1) \cdot p(E_2/E_1) \cdots p(E_n / E_1 \cap \dots \cap E_{n-1})$. ונמחיש זאת בדוגמה:

דוגמה:

מחלקים באקראי חפיסה של 52 קלפים 48 ערמות, 13 בכל אחת. חשב את ההסתברות שבכל ערימה יש אס אחד בדיוק.

פתרון:

נגדיר E_1 = אס עלה נמצא באחד הערמות. E_2 = אב עלה ואס לב בערמות שונות. E_3 = אס עלה, לב ומעוין בערמות שונות. E_4 = ארבעת האסים בערמות שונות.

ע"פ הנוסחה: $p(E_1 \cap \dots \cap E_4) = p(E_1) \cdot p(E_2/E_1) \cdot p(E_3/E_1 \cap E_2) \cdot p(E_4 / E_1 \cap E_2 \cap E_3)$.
נציב ונקבל כי $p(E_1 \cap \dots \cap E_4) = 1 \cdot \frac{39}{51} \cdot \frac{26}{50} \cdot \frac{13}{49} = 0.1054$.

נוסחת בייס (Bayes):

יהיו E ו- F מאורעות. אנו יכולים לבטא את המאורע E ע"י $E = (E \cap F) \cup (E \cap F^C)$. המאורעות זרים ולכן מאקסיומה שלוש ומהטענה הקודמת נקבל כי: $p(E) = p(E \cap F) + p(E \cap F^C)$.
 $p(E) = p(E/F) \cdot p(F) + p(E/F^C) \cdot p(F^C)$ ולכן $p(E/F) \cdot p(F) + p(E/F^C) \cdot [1 - p(F)]$.

דוגמה:

חברת ביטוח מאמינה שניתן לחלק בני אדם לקבוצות: מועדים לתאונות ושאינם. מהסטטיסטיקה של חברת הביטוח עולה שבתוך תקופה של שנה עבור קבוצה 1 קורות תאונות בהסתברות 0.4 וקבוצה 2 בהסתברות 0.2. אם נניח ש-30% מהאוכלוסייה מועדים לתאונות, מה ההסתברות שלמבוטח חדש תקרה תאונה בשנה הראשונה?

פתרון:

נסמן, E למבוטח תקרה תאונה בשנה הראשונה. F מסמל את המאורע "לקוח מועד לתאונות". נציב בנוסחה ונקבל כי $p(E) = 0.2 \cdot 0.7 + 0.3 \cdot 0.4 = 0.26$ זוהי ההסתברות המבוקשת.

המשך:

מה ההסתברות שהוא היה מועד לתאונות אם הוא (המבוטח) עבר תאונה בשנה הראשונה?
פתרון:

ע"פ נוסחת ההסתברות המותנית נקבל כי $p(F/E) = \frac{p(E \cap F)}{p(E)} = \frac{0.3 \cdot 0.4}{0.26} = 0.461$

הגדרה: יחס סיכויים של מאורע A מוגדר ע"י $\frac{p(A)}{p(A^C)} = \frac{p(A)}{1-p(A)}$, בהינתן עובדה חדשה B נקבל את הנוסחה

$p(A/B) = \frac{p(B/A) \cdot p(A)}{p(B)}$ ואז $p(A^C/B) = \frac{p(B/A^C) \cdot p(A^C)}{p(B)}$ ואז מספר הסיכויים החדש הוא:

$\frac{p(A/B)}{p(A^C/B)} = \frac{p(A)}{p(A^C)} \cdot \frac{p(B/A)}{p(B/A^C)}$. ז"א מתקבל יחס סיכויים ישן כפול יחס סיכויים מותנית.

הכללה של הטענה הקודמת:

יהיו $F_1, \dots, F_n$ מאורעות זרים במרחב מדגם S המקיימים $S = \bigcup_{i=1}^n F_i$, נסמן ב- E מאורע כלשהו במרחב המדגם: $\bigcup_{i=1}^n E \cap F_i = E$, המאורעות $E \cap F_i$ זרות לכל i , נקבל:

$p(E) = \sum_{i=1}^n p(E \cap F_i) = \sum_{i=1}^n p(E/F_i) \cdot p(F_i)$. נניח כעת E התרחש ואנו מעוניינים לקבוע איזה

מה- F_i ים התרחש, אזי $p(F_j/E) = \frac{p(E \cap F_j)}{p(E)} = \frac{p(E/F_j) \cdot p(F_j)}{\sum_{i=1}^n p(E/F_i) \cdot p(F_i)}$

5-7.8.12-הרצאה

מאורעות בלתי תלויים:

הגדרה: שני מאורעות נקראים בלתי תלויים אם מתקיים $p(E \cap F) = p(E) \cdot p(F)$, אחרת המאורעות נקראים תלויים. דבר זה נבע מהגדרת ההסתברות המותנית $p(E/F) = \frac{p(E \cap F)}{p(F)}$.

אם נציב שם את ההגדרה נקבל $p(E/F) = p(E)$.

דוגמה:

בוחרים קלף באקראי מתוך חבילה רגילה של 52 קלפים. אם E הוא המאורע שהקלף הנבחר הוא אס ו- F הוא המאורע שנבחר הוא עלה. האם המאורעות תלויים?

הם בלתי תלויים. $p(E) = \frac{4}{52} = \frac{1}{13}$, $p(F) = \frac{1}{4}$, $p(E \cap F) = \frac{1}{52}$. ניתן לראות בבירור כי $p(E \cap F) = p(E) \cdot p(F)$, לכן הם בלתי תלויים.

טענה: אם E ו- F מאורעות בת"ל אזי גם E ו- F^C הם מאורעות בת"ל.

הוכחה: נניח E, F הם מאורעות בת"ל. נגדיר כי $E = (E \cap F) \cup (E \cap F^C)$. לכן $E \cap F, E \cap F^C$ הם מאורעות זרים, ומכך נקבל: $p(E) = p(E \cap F) + p(E \cap F^C) = p(E) \cdot p(F) + p(E \cap F^C)$. העברת אגפים פשוטה ונקבל $p(E \cap F^C) = p(E) \cdot [1 - p(F)] = p(E) \cdot p(F^C)$ משל.

משתנים מקריים:

לפונקציות הממשיות המוגדרות על מרחב המדגם של ניסוי מקרי קוראים משתנים מקריים.

דוגמה:

נניח שהניסוי הוא הטלת 3 מטבעות תקינים אם Y מסמן את מספר הפעמים שהתקבל H או T הוא משתנה מקרי המקבל את אחד הערכים 0, 1, 2, 3.

$p(Y = 0) = \{(TTT)\}$.

פעם אחת: $p(Y = 1) = \{(HTT), (TTH), (THT)\}$

פעמיים: $p(Y = 2) = \{(HHT), (THH), (HTH)\}$

שלוש פעמים: $p(Y = 3) = \{(HHH)\}$

ניתן לראות שהסכום הוא 1.

פונקציית ההתפלגות המצטברת:

פונקציית ההתפלגות המצטברת של F של משתנה מקרי C מוגדרת לכל מספר ממשי b על ידי נושחה כך שבעצם זה מוגדר להיות ההסתברות שמשתנה מקרי X יקבל ערך קטן או שווה ל- b . $F(b) = p(x \leq b)$. תכונות:

1. F היא פונקציה לא יורדת, כלומר אם $a < b$ מתקיים $F(a) \leq F(b)$.

2. $\lim_{b \rightarrow \infty} F(b) = 1$

3. $\lim_{b \rightarrow -\infty} F(b) = 0$

4. F רציפה מימין. כלומר לכל b ולכל סדרה יורדת b_n כאשר n גדול מאחד, המתכנסת ל- b מתקיים

$\lim_{n \rightarrow \infty} F(b_n) = F(b)$ ש

תכונות 2-4 נובעות מתכונת הרציפות של ההסתברות.

תכונה 1 נובעת מכך שאם $a < b$, אזי המאורעים הבאים מקיימים $\{x \leq a\} \subseteq \{x \leq b\}$, ולכן $F(a) \leq F(b)$ לא יכולה להיות גדולה מ- $F(b)$.

הערות:

1. לכל $a < b$ מתקיים $p(a \leq x \leq b) = F(b) - F(a)$.

2. אם נרצה לחשב את ההסתברות שא קטן ממש מ- b נוכל להשתמש בתכונת הרציפות של

ההסתברות: $p(x < b) = p(\bigcup_{n=1}^{\infty} \{x \leq b - \frac{1}{n}\}) = p(\lim_{n \rightarrow \infty} \{x \leq b - \frac{1}{n}\})$

$\lim_{n \rightarrow \infty} p(x \leq b - \frac{1}{n}) = \lim_{n \rightarrow \infty} F(b - \frac{1}{n})$ שימו לב ש $F(b) = p(x < b) + p(x = b)$

דוגמה :

$$F(x) = \begin{cases} 0; x < 0 \\ x/2; 0 \leq x < 1 \\ 2/3; 1 \leq x < 2 \\ 11/12; 2 \leq x < 3 \\ 1; 3 \leq x \end{cases}$$

פונקציה ההתפלגות המצטברת של המשתנה המקרי x מוגדרת ע"י :

$$p(x < 3) = \lim_{n \rightarrow \infty} p(x \leq 3 - \frac{1}{n}) = \lim_{n \rightarrow \infty} 3 - \frac{1}{n} = \frac{11}{12}$$

$$p(x = 1) = p(x \leq 1) - p(x < 1) = F(1) - \lim_{n \rightarrow \infty} (1 - \frac{1}{n}) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6}$$

$$p(x > \frac{1}{2}) = 1 - p(x \leq \frac{1}{2}) = 1 - F(\frac{1}{2}) = 1 - \frac{1}{2} = \frac{3}{4}$$

משתנים מקריים בדידים :

משתני מקרי היכול לקבל לכל היותר מספר בן מניה של ערכים אפשריים נקרא משתנה מקרי בדיד. מגדירים את הפונקציה הסתברות $p(a)$ ע"י $p(a) = p(x=a)$. פונקציה ההסתברות חיובית לכל היותר למספר בן מניה של ערכי a . כלומר, אם המשתנה המקרי x חייב לקבל אחד מקבוצת ערכים כלשהם אז לכל ערך אחד של x $p(x)=0$. מכיוון שא חייב לקבל אחד מערכי x_i אזי מתקיים $\sum_{i=1}^{\infty} x_i = 1$ את פונקציה ההתפלגות המצטברת ניתן לבטא באמצעות $p(x)$ ע"י $F(a) = \sum_{x=x \leq a} p(x)$.

דוגמה :

$$p(1) = \frac{1}{4}, p(2) = \frac{1}{2}, p(3) = \frac{1}{8}, p(4) = \frac{1}{8}$$

פונקציה ההסתברות/התפלגות של x נתונה ע"י

$$F(a) = \begin{cases} 0; a < 1 \\ \frac{1}{4}; 1 \leq a < 2 \\ \frac{3}{4}; 2 \leq a < 3 \\ \frac{7}{8}; 3 \leq a < 4 \\ 1; a \geq 4 \end{cases}$$

פונקציה ההתפלגות המצטברת? פתרון :

תוחלת :

אם x הוא משתנה מקרי בדיד בעל פונקציה הסתברות $f(x)$ אז התוחלת של אהמסומנת ב $E(x)$ מוגדרת ע"י הנוסחה $E(x) = \sum_{x:p(x)>0} x \cdot p(x)$.

תוחלת של פונקציה משתנה מקרי :

נניח שנתונה לנו פונקציה הסתברות של משתנה מקרי בדיד x ואנו רוצים לחשב את התוחלת של פונקציה כלשהי של x , נאמר $g(x)$. הפתרון הוא חישוב ע"פ הגדרת התוחלת.

טענה : אם $x_1, x_2, \dots$ בהסתברות $p(x_1), p(x_2), \dots$ אחת משתנה מקרי בדיד, המקבל את הערכים $x_1, x_2, \dots$ בהסתברות $p(x_1), p(x_2), \dots$ בהתאמה, אזי לכל פונרציה ממשית g מתקיים $E(g(x)) = \sum_i g(x_i) \cdot p(x_i)$.

דוגמה :

$$p(x = 1) = 0.3, p(x = 0) = 0.5, p(x = -1) = 0.2$$

נשמך בא משתנה מקרי בדיד המקבל את הערכים $-1, 0, 1$ בהסתברות $0.3, 0.5, 0.2$ חשב את $E(x^2)$.

פתרון :

$$\sum_i x_i^2 \cdot p(x_i) = 1^2 \cdot 0.3 + (-1)^2 \cdot 0.2 + 0^2 \cdot 0.5 = 0.5$$

נבצע זאת ע"י הצבה בנוסחא שיש בטענה: $E(x^2) = 0.5$ ומצאנו את הדרוש.

$$E(x^2) \neq E^2(x)$$

הערה :

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

מסקנה: אם a, b הם קבועים, אזי $E(ax+b) = aE(x) + b$.
הוכחה: $E(ax + b) = \sum_{x:p(x)>0} (ax + b) \cdot p(x) = \sum ax \cdot p(x) + \sum b \cdot p(x) = a \cdot \sum x \cdot p(x) + b \cdot \sum p(x) = aE(x) + b$.
הגדרת התוחלת קיבלנו את הדרוש.

התוחלת של משתנה מקרי x , דהיינו $E(x)$ מכונה גם בשמות כגון הממוצע של x או המומנט הראשון של x .
כאשר $n \geq 1$ הגודל $E(x^n)$ נקרא המומנט החי של x ומהטענה נקבל כי $E(x^n) = \sum_{x:p(x)>0} x^n \cdot p(x)$.

שונות:

בהינתן משתנה מקרי x בעל פונקציית התפלגות מצטברת F . התוחלת $E(x)$ נותנת את הממוצע המשוקלל איך אינה אומרת דבר על תחום ההשתנות או על האופן בו מפוזרים הערכים בתחום.

$W=0$	$P(w=0)=1$	$E(w)=0$
$Y=-1,+1$	$P(y=-1)=0.5, p(y=1)=0.5$	$E(y)=0$
$Z=-100,+100$	$P(z=-100)=0.5, p(z=100)=0.5$	$E(z)=0$

דוגמה: 3 משתנים מקריים:

צריך להעריך את ההשתנות האפשרית של ערכי אבאמצעות גודל שהגדרתו מבוססת על הפרשים שבין ערכי x לתוחלתו. גודל זה מוגדר כתוחלת ריבועי ההפרש שבין x לתוחלתו.

הגדרה:

יהי x משתנה מקרי שתוחלתו μ . השונות של x מסומנת ב $V(x)$ ומוגדרת ע"י $V(x) = E(x - \mu)^2$. (יש נוהגים לסמנה ב $Var(X)$. ע"י פיתוח $V(x)$ מתקבל: $V(x) = \sum_x (x - \mu)^2 \cdot p(x) = \sum x^2 \cdot p(x) - 2\mu \sum x \cdot p(x) + \mu^2 = E(x^2) - 2\mu E(x) + \mu^2 = E(x^2) - 2\mu^2 + \mu^2 = E(x^2) - \mu^2$.
 $V(x) = E(x^2) - E^2(x)$ וקיבלנו נוסחה חדשה: $E(x^2) - 2\mu^2 + \mu^2 = E(x^2) - \mu^2$

טענה: עבור a, b קבועים מתקיים $v(ax+b)=a^2V(x)$.

הוכחה: נסמן $E(x)=\mu$. כמו כן ידוע כי $E(ax+b)=aE(x)+b$.
 $V(ax+b)=E[(ax+b)-(a\mu+b)]^2=E(a^2(x-\mu)^2)=a^2V(x)$. מ.ש.ל.

הערה: השורש הריבועי של השונות $V(x)$ מוגדר ע"י סטיית התקן ומסומן ב $SD(x) = \sqrt{V(x)}$.

6-9.8.12-הוצאה

התפלגויות בדידות מיוחדות:

1. משתנה מקרי ברנולי ובינומי:

ניסוי מקרי שלו 2 תוצאות אפשריות. "הצלחה" בהסתברות p או "כישלון" בהסתברות $1-p$.
כאשר p נמצא בין 0 ל-1, נקרא ניסוי ברנולי.
אם נגדיר $x=1$ עבור הצלחה, ו- $x=0$ עבור כישלון, אז פונקציית ההתפלגות של x נתונה ע"י $p(0)=1-p$ ו- $p(1)=p(x=1)=p$.
על משתנה מקרי זה, x , אומרים שהוא משתנה מקרי ברנולי. עכשיו נניח שעורכים n ניסויי ברנולי עם הסתברות p להצלחה והסתברות של $1-p$ לכישלון. אם x הוא מספר ההצלחות המתקבלות במניסויים אומרים על x שהוא משתנה מקרי בינומי בעל פרמטרים (n, p) . מכאן שמשתנה מקרי ברנולי אינו אלא משתנה מקרי בינומי עם הפרמטרים $(1, p)$.
פונקציית ההסתברות של מ"מ בינומי:

אשר מסומן ב- $x \sim \text{Bin}(n, p)$ היא $p(i) = p(x = i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$
לפי נוסחת הבינום נקבל כי:

$$\sum_{i=0}^n p(i) = \sum_{i=0}^n \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} = [p + (1-p)]^n = 1$$

דוגמה:

חברה מסוימת מייצרת ברגים ומוכרת אותם בחבילות המכילות 10 ברגים כל אחת. החברה מתחייבת להחליף כל חבילה שיש בה יותר מבורג פגום אחד. אם כל בורג שהחברה מייצרת פגום בהסתברות 0.01 באופן בת"ל בברגים האחרים, מה אחוז החבילות הנמכרות שהחברה צריכה להחליף?
פתרון:

$x \sim \text{Bin}(10, 0.01)$ מספר פגומים בחבילה. ההסתברות שהיה צריך להחליף חבילה הוא:
 $p(1 < x \leq 10) = p(x=2) + \dots + p(x=10)$ שזה בעצם: $1 - p(x=0) - p(x=1)$
נציב בפונקציית ההסתברות של משתנה מקרי בינומי ונקבל:

$$p(1 < x \leq 10) = 1 - \binom{10}{0} \cdot 0.01^1 \cdot 0.99^9 - \binom{10}{1} \cdot 1 \cdot 0.99^{10} = 4.266 \cdot 10^{-3}$$

ש-0.426% מהחבילות צריך להחליף.

תכונות:

כאשר $E(x^k) = \sum_{i=0}^n \binom{n}{i} i^k p^i (1-p)^{n-i} = np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np E((y+1)^{k-1})$
 $y \sim \text{Bin}(n-1, p)$ ומכאן שק- $E(x) = np$, $E(x^2) = (np)^2 - np^2 + np$, $V(x) = (np)^2 - [(np)^2 - np^2 + np]$
ולסיכום: $E(x) = np$, $V(x) = np(1-p)$. ניתן לסמן $q=1-p$.

פונקציית ההתפלגות המצטברת של משתנה מקרי בינומי היא:

$$p(x \leq i) = \sum_{k=0}^i \binom{n}{k} p^k (1-p)^{n-k}$$

$$p(x = k+1) = \frac{p}{1-p} \cdot \frac{n-k}{k+1} \cdot [p(x = k)]$$

. ההתפלגות המיוחדת הבאה היא הפאוסונית.

2. משתנה מקרי פואסוני:

על משתנה x שיכול לקבל ערך של אחד מהמספרים השלמים החיוביים $(0, 1, 2, \dots)$ אומרים שהוא משתנה מקרי פואסוני עם הפרמטר λ אם לכל λ גדול מאפס כלשהו מתקיים $p(i)$

$$p(x = i) = \frac{e^{-\lambda} \lambda^i}{i!}$$

כאשר $i=0, 1, 2, \dots$ וזו פונקציית הסתברות מאחר ואם נחשב את הבא נקבל כי

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

הסכום מתכנס ל-1: $\sum_{i=0}^{\infty} p(i) = \sum_{i=0}^{\infty} \frac{e^{-\lambda} \lambda^i}{i!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$. כדורש.

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

למשתנה מקרי פואסוני יש יישומים רבים בתחומים מרובים. ניתן להשתמש בו כקירוב למשתנה הבינומי (n, p) , כאשר n גדול ו- p קטן אזי $x \sim \text{Bin}(n, p) \approx \text{Poi}(\lambda)$ כאשר $\lambda = np$.

דוגמה:

בעמוד של סבפר בבית דפוס מסויים ידוע שמספר הטעויות בו מתפלג פאוסון עם $\lambda = \frac{1}{2}$. פתחתם את הספר בבית, מה ההסתברות שיש טעות אחת לפחות בעמוד שלפניכם?
פתרון:

$$p(x \geq 1) = 1 - p(x = 0) = 1 - (e^{-0.5} \cdot \frac{0.5^0}{0!}) \approx 0.393$$

דוגמה נוספת:

לפריט המיוצר ע"י מכונה מסוימת יש הסתברות 0.1 להיות פגום. חשבו את ההסתברות שבמדגם של 10 פריטים יימצאו לכל היותר פריט אחד פגום.

(א) בינומית: $x \sim \text{Bin}(10, 0.1)$ ולכן $p(x = 0) + p(x = 1) = 0.7361$

(ב) פאוסוני: $\lambda = 1, np = 1$, $p(x = 0) + p(x = 1) = 0.7358$.

תכונות:

$$E(x) = \sum_{i=0}^{\infty} i \frac{e^{-\lambda} \lambda^i}{i!} = \lambda \sum_{i=1}^{\infty} \frac{e^{-\lambda} \lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

$$E(x^2) = \sum_{i=0}^{\infty} i^2 \frac{e^{-\lambda} \lambda^i}{i!} = \lambda \sum_{j=0}^{\infty} \frac{(j+1)e^{-\lambda} \lambda^j}{j!} = \lambda \left(\sum_{j=0}^{\infty} \frac{j e^{-\lambda} \lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda} \lambda^j}{j!} \right) = \lambda(\lambda + 1)$$

וע"פ הנוסחה ל- $V(x)$ נקבל כי $V(x) = \lambda$.

3. משתנה מקרי גיאומטרי:

נניח שמבצעים ניסויים חוזרים בלתי תלויים שתוצאות כל אחד מהם היא "הצלחה" בהסתברות p כמובן בין 1 ל-0. מסמן בא את מספר הניסויים שיש לבצע עד שמתקבלת ההצלחה הראשונה.

פונקציית ההסתברות היא: $p(x = n) = (1 - p)^{n-1} \cdot p$.

כאשר אכן: $\sum_{n=1}^{\infty} p(x = n) = \sum_{n=1}^{\infty} p \cdot (1 - p)^{n-1} = p \sum_{n=1}^{\infty} (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1$.

הסימון: $x \sim \text{Geo}(p)$.

דוגמה:

כד מכיל N כדורים לבנים ו- M שחורים. מוציאים באקראי מהכד כדור אחד כדור עד שמתקבל כדור שחור. אם בחירת הכדורים נעשית עם החזרה, מה ההסתברות שיידרשו לכך:

(א) בדיוק n הוצאות.

(ב) לפחות k הוצאות.

פתרון:

נסמן בא את מספר ההוצאות של הכדורים עד שמתקבל כדור שחור ראשון. ההסתברות להוציא

כדור שחור היא $p = \frac{M}{N+M}$ (א) $p(x = n) = \left(\frac{N}{N+M}\right)^{n-1} \cdot \frac{M}{N+M}$.

(ב) $p(x \geq k) = \sum_{n=k}^{\infty} \left(\frac{N}{N+M}\right)^{n-1} \cdot \frac{M}{N+M} = \left(\frac{N}{N+M}\right)^{k-1}$.

תכונות:

נסמן $q = 1 - p$: $E(x) = \sum_{n=1}^{\infty} np(1 - p)^{n-1} = \sum_{n=0}^{\infty} npq^{n-1} = \sum_{n=0}^{\infty} (q^n)' = q = 1 - p$.

$E(x) = \frac{1}{p}$ ו"א ש $p(\sum_{n=0}^{\infty} q^n)' = p \left(\frac{1}{1-q}\right)' = \frac{p}{(1-q)^2} = \frac{1}{p}$.

נבצע חישוב דומה עבור $E(x^2)$.

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

$$E(x^2) = \sum_{n=1}^{\infty} n^2 q^{n-1} p = p \sum_{n=1}^{\infty} (nq^n)' = p \left(\frac{q}{p} \cdot \sum_{n=1}^{\infty} nq^{n-1} p \right)' = : \text{נחשב}$$

$$p \left(\frac{q}{p} E(x) \right)' = p \left(\frac{q}{(1-q)^2} \right)' = p \left(\frac{1}{(1-q)^2} + \frac{2q}{(1-q)^3} \right) = p \left(\frac{1}{p^2} + \frac{2(1-p)}{p^3} \right) = \frac{2}{p^2} - \frac{1}{p}$$

$$V(x) = \frac{1-p}{p^2} : \text{וע"פ הנוסחאות נקבל}$$

4. משתנה מקרי בינומי שלילי :

נניח שמבצעים ניסויים חוזרים בלתי תלויים, שתוצאות כל אחד מהם היא הצלחה בהסתברות p ושוב p בין 1 ל-0. נסמן בא את מספר הניסויים שיש לבצע עד שנשפרות בדיוק r הצלחות.

$$p(x = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r} : \text{כאשר } n=r, r+1, \dots$$

מתקיים $\sum_{n=r}^{\infty} p(x = n) = \sum_{n=r}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1$ משתנה מקרי שפונקציית ההסתברות שלו נתונה ע"י $P(x=n)$ מכונה משתנה מקרי בינומי שלילי ומסומן $x \sim NB(r, p)$. יש לשים לב שמשתנה מקרי גיאומטרי הוא מקרה פרטי של בינומי שלילי עם פרמטרים $(1, p)$.

תכונות :

$$E(x^k) = \sum_{n=r}^{\infty} n^k \binom{n-1}{r-1} p^r (1-p)^{n-r} = \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r} =$$

$$\frac{r}{p} \sum_{m=r+1}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)} = \frac{r}{p} E[(y-1)^{k-1}]$$

$$y \sim NB(r+1, p) : \text{עכשיו נחשב את } V(x)$$

$$E(x) = \frac{r}{p}, \text{ ועכשיו } E(x^2) = \frac{r}{p} \cdot \left(\frac{r+1}{p} - 1 \right) \text{ ולכן } V(x) = \frac{r(1-p)}{p^2}$$

5. משתנה מקרי היפר-גיאומטרי :

בוחרים באקראי (בלי החזרה) מדגם מגדול n (נ טבעי כלשהו) מתוך כד המכיל N כדורים. M לבנים ו- $N-M$ שחורים. נסמן בא את מספר הכדורים הלבנים שנבחרו :

$$p(x = i) = \frac{\binom{m}{i} \cdot \binom{N-m}{n-i}}{\binom{N}{n}} : \text{פונקציית ההסתברות}$$

$$E(x) = \frac{nm}{N}, V(x) = \frac{N-n}{N-1} \cdot n \cdot p \cdot (1-p) : \text{תכונות}$$

משתנה מקרי שההתפלגות שלו היא זו עבור שלמים n, m, N כך ש $0 \leq n, m \leq N$ מכונה

$$x \sim HG(m, N, n) \text{ התפלגות היפר גיאומטרית ומסומנת}$$

דוגמה :

קמעוני רוכש רכיבים חשמליים בחבילות של 10 יחידות. הרכישה של כל חבילה מתבצעת רק לאחר שבדק באקראי 3 רכיבים מתוכה ומצא שהם תקינים. אם ב-30% מהחבילות יש 4 רכיבים פגומים וב-70% יש רכיב אחד פגום. איזה אחוז מהחבילות שבדק הקמעוני אינו רוכש?
פתרון :

$$= \left(\text{יקנה חבילות עם 1 פגום} \right) \cdot p + \left(\text{יקנה חבילות של 4 פגומים} \right) \cdot p = 0.7 \cdot p + 0.3 \cdot p \left(\text{יקנה חבילה} \right) \text{ ובפני ההסתברות :}$$

$$= \frac{3}{10} \cdot \frac{\binom{4}{0} \cdot \binom{6}{3}}{\binom{10}{3}} + \frac{7}{10} \cdot \frac{\binom{1}{0} \cdot \binom{9}{3}}{\binom{10}{3}} = \frac{54}{100} = 54\%$$

ע"פ הנתונים. ז"א שהוא לא קונה 46%.

12.8.12-הרצאה 7

משתנים מקריים רציפים:

כעת נדון במשתנים מקריים שקבוצת הערכים האפשריים שלהם אינה בת מנייה, כגון: זמן חיים של מקרר, הרגע בו עוצרת רכבת בתחנה, וכו'. נאמר ש x הוא משתנה מקרי רציף אם קיימת פונקציה אי שלילית f המוגדרת לכל x ממשי, כלומר $f(x) \geq 0$. כאשר לכל קבוצה B של מספרים ממשיים מתקיים:

$$f(x \in B) = \int_B f(x) dx$$

המשתנה המקרי x .

f חייב לקיים:

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad 1.$$

$$p(a \leq x \leq b) = \int_a^b f(x) dx \quad 2.$$

$$p(x = a) = \int_a^a f(x) dx = 0 \quad 3. \text{ אם נציב } a=b \text{ אזי}$$

$$p(x < a) = p(x \leq a) = \int_{-\infty}^a f(x) dx \text{ ומשלוש נובע כי}$$

דוגמה:

$$f(x) = \begin{cases} c(4x - 2x^2); & 0 < x < 2 \\ 0; & \text{אחרת} \end{cases}$$

יהי x משתנה מקרי רציף בעל פונקציית צפיפות

(א) מצאו את c

(ב) חשבו את $p(x > 1)$.

פתרון:

$$\int_0^2 c(4x - 2x^2) dx = c \left[\frac{4x^2}{2} - \frac{2x^3}{3} \right]_{x=0}^{x=2} = 1 \quad \text{(א) ע"פ התכונות של פונקציית צפיפות נקבל:}$$

$$c = \frac{3}{8} \text{ וחישבו פשוט ייתן לנו ש}$$

$$p(x > 1) = \int_1^{\infty} f(x) dx = \int_1^2 \frac{3}{8} (4x - 2x^2) dx = \frac{3}{4} \quad \text{(ב) שוב, ע"פ הגדרה:}$$

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x) dx \quad \text{הגדרה:}$$

טענה: אם x משתנה מקרי רציף בעלי פונקציית צפיפות $f(x)$, אזי לכל פונקציה ממשית g מתקיים (בדומה

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx \text{ למשתנה מקרי בדיד)}$$

$$V(x) = E(x^2) - E^2(x) \quad \text{הגדרה:}$$

$$V(ax+b) = a^2 V(x), E(ax+b) = aE(x) + b \text{ הערות:}$$

$$F'(x) = f(x) \quad F(x) = \int_{-\infty}^x f(t) dt \text{ הגדרה: פונקציית ההתפלגות המצטברת מוגדרת להיות}$$

דוגמה:

$$X \text{ מ"מ בעל פונקציית צפיפות } f(x) = \begin{cases} 2x; & 0 \leq x \leq 1 \\ 0; & \text{אחרת} \end{cases} \text{ חשבו את (א) התוחלת ו(ב) השונות.}$$

פתרון:

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^1 x \cdot 2x dx = \left[\frac{2x^3}{3} \right]_0^1 = \frac{2}{3} \quad \text{(א)}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) dx = \int_0^1 2x^3 dx = \left[\frac{2x^4}{4} \right]_0^1 = \frac{1}{2} \quad \text{(ב) ראשית:}$$

$$V(x) = \frac{9}{18} - \frac{8}{18} = \frac{1}{18} \text{ כי}$$

דוגמה נוספת:

נתונה פונקציה צפיפות של x $f(x) = \begin{cases} 1; & 0 \leq x \leq 1 \\ 0; & \text{אחרת} \end{cases}$. חשבו את $E(e^x)$.
פתרון: $E(e^x) = \int_{-\infty}^{\infty} e^x \cdot f(x) dx = \int_0^1 e^x dx = e - 1$

התפלגויות רציפות מיוחדות:

1. משתנה מקרי אחיד

נאמר על משתנה מקרי שיש לו התפלגות אחידה בקטע $[0,1]$ אם פונקציה הצפיפות שלו נתונה ע"י

$$f(x) = \begin{cases} 1; & 0 \leq x \leq 1 \\ 0; & \text{אחרת} \end{cases} \text{ וגם } \int_{-\infty}^{\infty} f(x) dx = 1$$

מכיוון ש $f(x) > 0$ רק כאשר $x \in [0,1]$ הרי ש אלקבל בהכרח ערך כלשהו בקטע $[0,1]$. כמו כן, מכך ש $f(x)$ קבועה לכל $x \in [0,1]$ נובע שא יכול להימצא בהסתברות שווה בסביבות כל ערך בקטע $[0,1]$. על מנת לבדוק את הטענה, עבור כל a, b שבין 0 ל1 ונניח ש b גדול מ a , מתקיים שההסתברות להיות בקטע הספציפי הנ"ל שווה בערכה לאורך הקטע. $p(a \leq x \leq b) = b - a$. באופן כללי נאמר ש אהוא משתנה מקרי אחיד בקטע $[\alpha, \beta]$ אם פונקציה הצפיפות שלו נתונה על ידי,

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}; & \alpha \leq x \leq \beta \\ 0; & \text{אחרת} \end{cases} \text{ מכיוון ש } F(x) = \int_{-\infty}^x f(t) dt \text{ נקבל את פונקציה הצפיפות}$$

$$F(x) = \begin{cases} 0; & x < \alpha \\ \frac{x - \alpha}{\beta - \alpha}; & \alpha \leq x \leq \beta \\ 1; & x > \beta \end{cases} : [\alpha, \beta] \text{ איד קטע איד } [\alpha, \beta]$$

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} = \frac{\beta + \alpha}{2} \text{ (א) תכונות:}$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{\alpha}^{\beta} x^2 \cdot \frac{1}{\beta - \alpha} = \frac{\alpha^2 + \alpha\beta + \beta^2}{3} \text{ (ב)}$$

$$V(x) = E(x^2) - E^2(x) = \frac{(\beta - \alpha)^2}{12} \text{ (ג)}$$

דוגמה:

נניח שלא יש התפלגות אחידה ב $[0,10]$ $f(x) = \begin{cases} \frac{1}{10}; & 0 \leq x \leq 10 \\ 0; & \text{אחרת} \end{cases}$ חשבו את ההסתברויות

$$p(3 < x < 8) = \int_3^8 \frac{1}{10} dx = \frac{5}{10} = 0.5, p(x > 3) = \int_3^{\infty} \frac{1}{10} dx = \frac{3}{10}$$

2. משתנה מקרי נורמלי:

נאמר שא הוא משתנה מקרי נורמאלי או שלא יש התפלגות עם הפרמטרים (μ, σ^2) אם פונקציה הצפיפות שלו נתונה ע"י (עבור כל x ממשי) $f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ הגרף של פונקציה הצפיפות

הזו היא עקומת פעמון סימטרית לציר האנכי שעובר ב μ . סימון $x \sim N(\mu, \sigma^2)$.
וכן מתקיים $V(x) = \sigma^2, E(x) = \mu, \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1$

תכונה: $x \sim N(\mu, \sigma^2)$ אזי $y = ax + b$ מתפלג נורמאלי עם $y \sim N(a\mu + b, a^2\sigma^2)$ וכן מתקיים $z \sim N(0,1)$ כשאר z משתנה מקרי סטנדרטי.

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy : \text{פונקציה הצפיפות המצטברת}$$

הקירוב הנורמאלי להתפלגות הבינומית:

משפט חשוב בתורת ההסתברות, המכונה משפט הגבול של דה מואבר לפלס קובע שכשאר n גדול, ההתפלגות של משתנה מקרי נורמאלי השווה לו בתוחלתו ובשונותו.

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

המשפט הוא: $p\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \Phi(b) - \Phi(a)$ כאשר S_n זה מהספר ההצלחות עם הסתברות p להצלחה.

דהיינו, $x \sim \text{Bin}(n, p)$, אזי $E(x) = np$, $V(x) = np(1-p)$ ולכן $x \sim N(np, np(1-p))$ ונסמן $Z = \frac{x - np}{\sqrt{np(1-p)}}$ ונקבל ש $Z \sim N(0,1)$.

הערה: כשעוברים ממשתנה מקרי לרציף צריך לעשות תיקון רציפות.

דוגמה:

יהי x מספר הפעמים שמתקבל H ב-40 הטלות של מטבע תקין. מה ההסתברות ש $x=20$?
פתרון:

לפי בינומי: $x \sim (40, 0.5)$ ולכן $p(x=20) = \binom{40}{20} \left(\frac{1}{2}\right)^{20} \left(\frac{1}{2}\right)^{40-20} = 0.1254$

לפי הקירוב הנורמאלי: $p(x=20) = p(19.5 \leq x \leq 20.5) = p\left(\frac{19.5-20}{\sqrt{10}} \leq x \leq \frac{20.5-20}{\sqrt{10}}\right) = p(-0.158 \leq x \leq 0.158) \approx 0.1256$
הבינומי.

3. משתנה מקרי מעריכי:

על משתנה מקרי רציף שפונקציית הצפיפות שלו נתונה ע"י λ חיובי כלשהו נאמר שהוא משתנה

מקרי מעריכי עם הפרמטר λ . פונקציית הצפיפות הינה: $f(x) = \begin{cases} \lambda e^{-\lambda x} & ; x \geq 0 \\ 0 & ; x < 0 \end{cases}$. פונקציית

ההתפלגות המצטברת נתונה לכל x אי שלילי ע"י $F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$. כמו כן

מתקיים $1 = \int_0^\infty f(x) dx$, $V(x) = \frac{1}{\lambda^2}$, $E(x) = \frac{1}{\lambda}$. סימון: $x \sim \exp(\lambda)$

הערה: משתנה מקרי מעריכי הוא בעל תכונה של חוסר זיכרון. $p(x > s+t) = p(x > t)$.

דוגמה:

נניח שמשך הזמן בדקות של מספר טלפון היא מ"מ מעריכי עם פרמטר $\lambda=0.1$. אם מישהו נכנס לתא הטלפון הציבורי ברגע שהגעת לתא זה, מה ההסתברות שיהיה עליך לחכות:

(א) יותר מ-10 דקות

(ב) בין 10 ל-20 דקות

פתרון:

(א) $p(x > 10) = 1 - F(10) = 1 - 1 - e^{-1} = 0.368$ (וכמו כן $x \sim \exp(\frac{1}{10})$)

(ב) $p(10 < x < 20) = F(20) - F(10) = e^{-1} - e^{-2} = 0.233$

לסיכום: פונקציה מצטברת: $F(x) = 1 - e^{-\lambda x}$, וצפיפות $f(x) = \lambda e^{-\lambda x}$.

התפלגות משותפת של משתנים מקריים:

עד כה עסקנו בהתפלגויות של משתנים מקריים יחידים. אולם, לעתים קרובות אנו מעוניינים בהסתברויות של מאורעות המוגדרים באמצעות שני משתנים מקריים או יותר. כדי לדון בהסתברות כזו נגדיר, לכל שני משתנים מקריים x ו y את פונקציית ההתפלגות המצטברת המשותפת של x ו y ע"י:

$F(a, b) = p\{x \leq a, y \leq b\}$, a, b ממשיים. את פונקציית ההתפלגות המצטברת של x אפשר לקבל

מפונקציית ההתפלגות המצטברת המשותפת של x ו y . כדלקמן:

$f_x(a) = p\{x \leq a\} = p\{x \leq a, y < \infty\} = p(\lim_{b \rightarrow \infty} \{x \leq a, y \leq b\}) = \lim_{b \rightarrow \infty} F(a, b) =$

$F(a, \infty)$ וקיבלנו את המבוקש. ובאופן דומה $f_y(b) = F(\infty, b)$. קוראים לפונקציות אלה לעיתים

פונקציות ההתפלגות המצטברת השולית של x ו y בהתאמה.

ד"ר עינת אביאלי
הוקלד ע"י דביר חדד

כל תרגיל ניתן לפתירה ע"י פונקצית ההתפלגות המצטברת : $p\{x > a, y > b\} = 1 - p(\{x > a, y > b\}^c) = 1 - p(\{x \leq a\} \cup \{y \leq b\}) = 1 - f_x(a) - f_y(b) + F(a, b)$ וגם אפשר להרחיב ש : $p\{a_1 < x \leq a_2, b_1 < y \leq b_2\} = F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1)$ שמתקיים $b_1 < b_2, a_1 < a_2$.

אם x, y משתנים מקריים בדידים מגדירים :

$$p(x, y) = p\{X = x, Y = y\}$$

פונקצית ההסתברות המשותפת של x היא $p_x(x) = P\{X = x\} = \sum_{y: p(x, y) > 0} p(x, y)$

ובאופן דומה נקבל : $p_y(y) = P\{Y = y\} = \sum_{x: p(x, y) > 0} p(x, y)$

דוגמה :

בוחרים באקראי בלי החזרה 3 כדורים מתוך כד המכיל 3 אדומים, 4 לבנים ו-5 כחולים. X מספר הכדורים האדומים שנבחרו, ו- y מספר הלבנים. אזי, פונקצית ההסתברות המשותפת של x, y (נשתמש בטבלה).

מספר לבנים / מספר שחורים	0	1	2	3	סכום השורה
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
סכום העמודה	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	1

אומרים שלא ו- y יש התפלגות משותפת רציפה אם קיימת פונקציה אי שלילית $f(x, y)$ המוגדרת לכל x, y ממשיים ולה התכונה, שלכל קבוצה C של זוגות מספרים ממשיים (כלומר קבוצה במישור הדו מימדי) מתקיים $p(\{(X, Y) \in C\}) = \iint_{(x, y) \in C} f(x, y) dx dy$.

יהיו A, B קבוצות כלשהן של מספרים ממשיים. נגדיר $C = \{(x, y) : x \in A, y \in B\}$, ונקבל שמתקיים ע"פ המשוואה הקודמת $p\{x \in A, y \in B\} = \int_B \int_A f(x, y) dx dy$.

ומכיוון שמתקיים : $F(a, b) = p\{x \in (-\infty, a], y \in (-\infty, b]\} = \int_{-\infty}^b \int_{-\infty}^a f(x, y) dx dy$ נובע ע"י גזירה כי $f(a, b) = \frac{\partial^2}{\partial a \partial b} F(a, b)$.

ומכאן : $p\{a < x < a + da, b < y < b + db\} = \int_b^{b+db} \int_a^{a+da} f(x, y) dx dy \approx f(a, b) da db$ לפיכך, $f(a, b)$ היא מידה לסבירות שהערך של וקטור מקרי (x, y) יהיה קרוב ל- (a, b) .

לסיכום : פונקציות הצפיפות של x ו- y הם : $f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$ ובהתאם : $f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$