הסתברות - סיכום

$\boxed{I}$ - מושגים בסיסיים

$\mathcal{B}$ - מרחב המאורעות - כל המאורעות שיכולים להתקיים.

Ω - מרחב המדגם - כל המאורעות שיכולים להתקיים בניסוי מסוים.

A - (או כל אות אחרת) - מאורע מסוים מתוך מרחב המדגם. מאורע יכול להיות מורכב ממספר תת-מאורעות. היינו, מאורע מורכב הנו אוסף של מאורעות פשוטים.

- 3 האקסיומות הבסיסיות

1) $A \in \mathcal{B} \Rightarrow \bar{A} \in \mathcal{B}$

2) $A, B \in \mathcal{B} \Rightarrow A \cap B \in \mathcal{B}$

3) $\Omega \in \mathcal{B}$

- 3 אקסיומות למאורעות מש"ם

1) $\forall (A \in \mathcal{B}) \, [p(A) \geq 0]$ הסתברות איננה שלילית.

*2) $A \cap B = \phi \Rightarrow p(A \cup B) = p(A) + p(B)$

3) $p(\Omega) = 1$

- מסקנות מהאקסיומות

$\Rightarrow p(\bar{A}) = 1 - p(A)$

$\Rightarrow p(\phi) = 0$

$\Rightarrow p(A \setminus B) = p(A) - p(A \cap B)$

*$\Rightarrow p(A \cup B) = p(A) + p(B) - p(A \cap B)$

*) $p(A \cup B \cup C) = p(A) + p(B) + p(C) - p(A \cap B) - p(A \cap C) - p(B \cap C) + p(A \cap B \cap C)$

$\Rightarrow A \subset B \Rightarrow p(A) \leq p(B)$

- קבוצות זרות

1) הקבוצות A ו-B נקראות זרות אם אין להן איבר משותף: $A \cap B = \phi$.

2) אוסף הקבוצות $A_1, A_2, \dots, A_n$ נקראות זרות בזוגות אם $A_n \cap A_m = \phi$ לכל $n \neq m$.

- קבוצת החזקה

קבוצת החזקה של A הנה קבוצת כל תת-הקבוצות של A: $\Rightarrow 2^A = \{B : B \subseteq A\}$

אם בקבוצה יש n איברים אזי בקבוצת החזקה יהיו 2^n איברים.

$\boxed{\text{II}}$ קצת קומבינטוריקה

1) כמה סדרות מאורך n אפשר ליצור מ־m איברים? $m \cdot m \cdots m = \boxed{m^n}$

2) כמה קבוצות סדורות (ללא חזרה) מאורך n אפשר ליצור מ־m איברים?
בתנאי ש־ $m \geq n$ (אם $m < n$ התשובה היא אפס).

$$m(m-1)(m-2)\cdots[m-(n-1)] = m(m-1)(m-2)\cdots[m-(n-1)] \cdot \frac{(m-n)\cdots 1}{(m-n)\cdots 1} = \boxed{\frac{m!}{(m-n)!}}$$

3) בכמה דרכים אפשר לסדר קבוצה בגודל m? $m(m-1)(m-2)\cdots 1 = \boxed{m!}$

4) בכמה דרכים אפשר לבחור n איברים מתוך m קבוצות ($m \geq n$)?

1 נניח שיש x דרכים לבחור את האיברים. אם נסדר אותם ($n!$) נקבל קבוצות סדורות: $x \cdot n!$

2 לפי סעיף 2 יש $\frac{m!}{(m-n)!}$ דרכים לבחור קבוצות סדורות.

3 $x \cdot n! = \frac{m!}{(m-n)!} \Rightarrow \boxed{x = \frac{m!}{n!(m-n)!} = \binom{m}{n}}$

- תכונות של $\binom{m}{n}$

הבינום של ניוטון:
$\Rightarrow (a+b)^m = \sum_{n=0}^{m} \binom{m}{n} a^n b^{m-n}$

$\Rightarrow \binom{m}{n} = \binom{m}{m-n}$

$\Rightarrow \binom{m}{0} = \binom{m}{m} = 1$

$\Rightarrow \binom{m}{1} = \binom{m}{m-1} = m$

$\Rightarrow \binom{m}{n} = \binom{m-1}{n} + \binom{m-1}{n-1}$

III מרחב הסתברות סימטרי

1) מרחב המדגם סופי.
2) אוסף המאורעות הוא אוסף כל תתי-הקבוצות.
3) הסיכויים לכל שתי תוצאות פשוטות זהות.

- פשוט צריך לחשב כמה אירועים פשוטים מרכיב אירוע A ולחלק במספר האירועים סה"כ במרחב המדגם.

$$\Rightarrow \boxed{p(A) = \frac{\#A}{\#\Omega}}$$

IV הסתברות מותנית

- $p(B|A)$ - ההסתברות ש-B יתרחש, בהינתן A.
קודם צריך ש-A יתרחש, ואז החיתוך של A ו-B.

$$\Rightarrow \boxed{p(B|A) = \frac{p(A\cap B)}{p(A)}}$$

- הסתברות א-פריורית - לא תמיד יודעים מה בדיוק ההסתברות לאירוע הראשון, A. במקרה כזה משערים את ההסתברות - נותנים לה ערך א-פריורי ורואים איך הוא משתנה בהתאם לנתונים האחרים. מחשבים את $p(B)$ לפי $p(A)$ ואז:

$$\Rightarrow p(A) = p(B)p(A|B) + p(\bar{B})p(A|\bar{B})$$

$$\Rightarrow p(B|A) = \frac{p(A\cap B)}{p(A)} = \frac{p(B)p(A|B)}{p(A)}$$

ואם נכליל את שתי הנוסחאות: יהיו נתונים המאורעות $B_1,\dots,B_n$ כך שלכל $i\neq j$
$B_i\cap B_j=\phi$ וכן $\Omega=\bigcup_{i=1}^{n}B_i$ (כלומר, כל המאורעות זרים). $B_1,\dots,B_n$ הן הסתברויות א-פריוריות.

- נוסחת ההסתברות השלמה

$$\Rightarrow \boxed{p(A) = \sum_{i=1}^{n} p(B_i)p(A|B_i)}$$

נוסחת בייס להסתברויות אפוסטריוריות

$$\Rightarrow \boxed{p(B_i|A) = \frac{p(B_i\cap A)}{p(A)} = \frac{p(B_i)p(A|B_i)}{\sum_{j=1}^{n} p(B_j)p(A|B_j)}}$$

- אי-תלות סטטיסטית

הגדרה: שני אירועים A ו-B בלתי-תלויים אם: $p(A|B)=p(A)$ או אם $p(B)=0$ (ואז זה לא משנ
במקרה כזה הסתברות החיתוך שווה למכפלת ההסתברויות: $p(A\cap B) = p(A)p(B)$

הכללה: המאורעות $A_1,\dots,A_n$ יקראו בלתי-תלויים אם כל אחד מהם ב"ת בכל תת-קבוצה של האחרים

$$\Rightarrow \boxed{p(A_i \mid A_{j_1}\cap A_{j_2}\cap\dots\cap A_{j_k}) = p(A_i) \quad , \quad A_i\neq A_{j_k}}$$

אורגבאור ©

$$\Rightarrow \boxed{p\left(\bigcap_{i=1}^{n}A_i\right) = \prod_{i=1}^{n} p(A_i)}$$

ואז חיתוך ההסתברויות שווה למכפלתן

V התפלגויות - משתנים מקריים בדידים

1) התפלגות בינומית

הדוגמה המוכרת ביותר היא סדרת ניסויי ברנולי, בה הניסויים ב"ת סטטיסטית. p יציין את מס' ההצלחות ו-q את מס' הכשלונות.

$$\boxed{\begin{array}{l} X = 0,1,\dots,n \\ p\{X=k\} = \binom{n}{k} p^k q^{n-k} \\ X \sim B(n,p) \end{array}}$$

בהתפלגות הבינומית שואלים כמה ניסויים k מתוך n הסתיימו בהצלחה/כישלון.

X מוגדר כמשתנה מקרי בינומי. סכום ההסתברויות מתנהג כמו הביטוי של בינום: $(p+q)^n$.

2) התפלגות גיאומטרית

גם כאן מסתכלים על סדרת ניסויי ברנולי. הפעם שואלים: מה ההסתברות שנבצע k ניסויים עד להצלחה הראשונה.

$$\boxed{\begin{array}{l} Y = 1,2,3,\dots \\ p\{Y=k\} = p q^{k-1} \\ Y \sim G(p) \end{array}}$$

Y מוגדר כמשתנה מקרי גיאומטרי. סכום ההסתברויות מתנהג כמו טור גיאומטרי.

3) הקירוב הבינומי של פואסון –

מסתכלים על ההתפלגות הבינומית ל-n גדול ו-p קטן. מגדירים משתנה חדש: λ. ומסתכלים על הגבול של ההתפלגות הבינומית כש: $n \to \infty$.

$$\boxed{\begin{array}{l} X = 0,1,2,\dots,n \\ \lambda = np \\ p\{X=k\} = \binom{n}{k} p^k q^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda} \\ X \sim P(\lambda) \end{array}}$$

$$p\{X=k\} = \binom{n}{k} p^k q^{n-k} = [\lambda = np][n \to \infty] = \frac{\lambda^k}{k!} e^{-\lambda}$$

+ משתנה מקרי - הגדרה

יהי נתון מרחב מדגם $<\Omega, B, p>$. משתנה מקרי מעל מרחב זה הוא פונקציה מעל מרחב המדגם.

המקריות בשם המשתנה מצביעה על הבחירה המקרית של התוצאה מתוך מרחב המדגם.

+ משתנה מקרי בדיד - הגדרה

משתנה מקרי שמקבל סדרה (סופית או אינסופית) של ערכים.

אם אנחנו יודעים מהם הערכים וההסתברות של אירוע מקרי $p\{X=x_k\}$ ואומרים שאנחנו יודעים מהי התפלגות המשתנה המקרי.

+ תנאים בסיסיים

$$p\{X=x_k\} = p_k \geq 0 \;,\; \sum_k p_k = 1$$

4) <u>התפלגות פואסון</u> -

+ מסתכלים על תהליך כלשהו שמתרחש לאורך זמן או מרחב.

את הזמן/מרחב מחלקים ל-n חלקים.

שואלים: כמה נקודות x יש בכל תת-קטע n' $\Leftarrow$

מה ההסתברות שיש x נקודות בכל תת-קטע n'.

כש-$n \to \infty$ מקבלים את התפלגות פואסון.

+ <u>הגדרה</u>: תהליך שבפרק זמן קצר מאוד $[t,t+h]$ יש סיכוי $\lambda h + o(h)$ לאירוע אחד

וסיכוי $o(h)$ ליותר מאירוע אחד, כאשר: $o(h)=f(h): \lim_{h \downarrow 0} \frac{f(h)}{h} = 0$

($o(h)$ הנו הטעות).

- <u>דרישות בסיסיות</u>

I) אי-תלות באירוע עבר (תכונה מרקובית).

II) סטציונריות (כלומר, אי-תלות בזמן).

III) פשטות הזרם (כל אירוע מתרחש בנפרד. אין מקרה שבו שני אירועים מתרחשים בו-זמנית).

+ מתוך הדרישות וההגדרות הנ"ל, מסתכלים על קירוב פואסון:

$$\begin{cases} X = 0,1,2,\dots,n \\ \lambda = np \\ p\{X=k\} = \frac{\lambda^k}{k!} e^{-\lambda} \end{cases}$$

אבל עכשיו, במקום $X \sim P(\lambda)$ מסתכלים על $\boxed{X \sim P(\lambda t)}$ ומקבלים:

$$\boxed{p\{X(t)=k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}}$$

כלומר: עד זמן t התרחשו k אירועים.

מתוך נוסחה זו ניתן לקבל את שתי הנוסחאות הבאות:

הסיכוי שעד זמן t התרחשו 0 אירועים: 1) $\boxed{p_0(t) = p\{X(t)=0\} = e^{-\lambda t}}$

הסיכוי שעד זמן t התרחשו $k \neq 0$ אירועים: 2) $\boxed{p\{X(t)=k,\ k\neq 0\} = 1-e^{-\lambda t}}$

כלומר, שהתרחש מספר כלשהו, לא ידוע, של אירועים. לפי הנוסחה האחרונה אפשר להגדיר

להסתברות פונקציה $F(t)$ בעזרתה ניתן יהיה למצוא את הזמן T עד האירוע הראשון:

$$\boxed{T \sim e(\lambda)} := \boxed{\forall (t \geq 0)\left[F(t) = p(T \leq t) = 1-e^{-\lambda t}\right]}$$

אור סעיף 6 T מוגדר כמשתנה מקרי <u>מעריכי</u>.

5) <u>התפלגות מולטינומית</u>

הכללה של ההתפלגות הבינומית למקרה של n ניסויים ב"ת, כאשר לכל ניסוי יכולים להיות מספר תוצאות אפשריות (r). במקרה כזה מכלילים את הנוסחה הבינומית ומקבלים:

$$\boxed{p\{X=k_r\} = \frac{n!}{k_1!k_2!\cdots k_r!}\, p_1^{k_1}p_2^{k_2}p_3^{k_3}\cdots p_r^{k_r}}$$

כאשר k_r הם שכיחויות ומתקיים לפי: $k_1+k_2+k_3+\dots+k_r=n$

וההסתברויות p_r מתקיימות לפי: $p_1+p_2+\dots+p_r=1$

השימוש העיקרי בפיזור זה הוא עבור מקרים של <u>דגימה עם החזרה</u>.

VI <u>משתנה מקרי רציף</u>

כאשר המ"מ מתאר פונקציה רציפה כלשהי (כמו מס' הנקודות בהיקף של מעגל) אין משמעות לנקודות בודדות, ואפשר לדבר רק על "שטחים". במעגל למשל, לכל נקודה יש הסתברות אפס (כי יש אינסוף נקודות), אבל לאפשרות ש-x הוא בין שתי נקודות במעגל יש הסתברות כלשהי.

ההסתברות הזו נקבעת לפי "צפיפות" ההסתברות של הנקודות באותו קטע -

<u>פונקציית ההסתברות $f(x)$ ($f(x)\geq 0$)</u>:

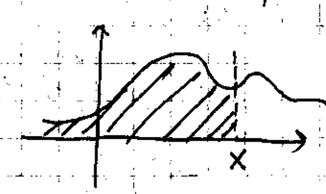
$$\boxed{P(a\le X\le b)=\int_a^b f(x)\,dx}$$

וכיוון שאלו מתארים הסתברות, סה"כ השטח=1:

$$\boxed{\int_{-\infty}^{\infty} f(x)\,dx = 1}$$

<u>פונקציית התפלגות מצטברת</u>: נותנת את ההסתברות המצטברת עד נקודה x:

$$\boxed{F_X(x)=P(X\le x)=\int_{-\infty}^{x} f(x)\,dx}$$



VII <u>משתנים מקריים דו-מימדיים</u>

כאשר מס"ן מתאר מאורע שמורכב משני משתנים מקריים, קוראים למאורע המשולב:

משתנה מקרי דו-מימדי:

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

I) אם X,Y ב"ת לכל זוג ערכים x_n, y_m אזי:

$$\boxed{P\{X=x_n, Y=y_m\}=P\{X=x_n\}P\{Y=y_m\}}$$

II) אם X,Y משתנים מקריים רציפים ב"ת לכל $-\infty<x,y<\infty$ אזי:

$$\boxed{P\{X\le x,\ Y\le y\}=P(X\le x)P(Y\le y)=F_X(x)F_Y(y)}$$

ראו בעמוד 8

III) אם קיימת תלות בין X,Y אזי:

$$\boxed{P\{X=x_n, Y=y_m\}=P(X=x_n)P(Y=y_m\,|\,X=x_n)}$$

התפלגות שולית

- משתנים מקריים דו-מימדיים נהוגים לסדר בטבלאות:
בסוף כל עמודה/שורה רושמים את סכום הערכים שלה.
כך, בשורת הסה"כ מקבלים את התפלגות X
ובעמודת הסה"כ - התפלגות Y.

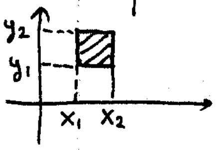
$Y \backslash X$	x_1	x_2	$x_3 \cdots$	x_n	סה"כ
y_1					התפלגות של Y
y_2					
y_3					
$\vdots$					
y_n					
סה"כ	התפלגות של X				1

$$\boxed{\begin{aligned} P_X(x) &= P(X=x) = \sum_y P(x,y) \\ P_Y(y) &= P(Y=y) = \sum_x P(x,y) \end{aligned}}$$

להתפלגויות האלה קוראים <u>התפלגויות שוליות</u> וסכום החלקים שלהן שווה 1.

הסתברות במלבן -

אפשר לומר שמ"מ דו-מימדי מתאר מלבן בעל צלעות x ו-y. שטח המלבן, שהוא אינטגרל על פונקציית הצפיפות שלו, יתן את ההסתברות שלו.
אם רוצים לבדוק את ההסתברות של "תת-מלבן" עושים:



$$\Rightarrow \boxed{P(x_1 \le X \le x_2,\ y_1 \le Y \le y_2) = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1)}$$

ופונקציית הצפיפות של המלבן תהיה הנגזרת השניה המשותפת של פונקציית ההתפלגות המשותפת:

$$\Rightarrow \boxed{f(x,y) = \frac{\partial^2}{\partial x \partial y} P(x,y)}$$

ואם X,Y ב"ת:

$$\Rightarrow \boxed{f(x,y) = f_X(x) f_Y(y)}$$

משתנה פואסון דו-מימדי

X,Y משתנה פואסון דו-מימדי עם פרמטרים λ ו-μ: $\langle X,Y \rangle \sim P(\lambda,\mu)$ אם:
I) $X \sim P(\lambda)$.
II) $Y \sim P(\mu)$.
III) X,Y ב"ת.

משתנה מקרי דו-מימדי מחובר

$$\Rightarrow \boxed{P(X+Y=n) = \sum_k P(X=k,\ Y=n-k) = \sum_k P(X=k)P(Y=n-k)}$$

תוחלת - (Expectation) שונות - (Variance)

VIII

התוחלת של מ"מ X מתארת את ערך ההסתברות הממוצעת בהתפלגות, ומסומנת ע"י $E(X)$, או ע"י μ. את חישוב התוחלת עושים בצורת ממוצע משוקלל:

עבור מ"מ בדיד: $\boxed{E(X)=\sum_k x_k P(X=x_k)}$

עבור מ"מ רציף: $\boxed{E(X)=\int_{-\infty}^{\infty} x f(x)dx}$

תכונות של תוחלת

1) $E(cX)=cE(X)$

2) $E(X+c)=E(X)+c$

3) $E(aX+b)=aE(X)+b$

4) $E(X+Y \mid x,y\in \mathbb{Z})=E(X)+E(Y)$; אם X,Y ב"ת: $E(XY)=E(X)E(Y)$

5) תוחלת הסטיית מהתוחלת: $E(X-EX)=E(X)-E(X)=0$

6) אם התפלגות X סימטרית סביב ערך a, אז: $E(X)=a$

7) $E(g(x))=\sum_k g(k)P(X=k)$

חציון - Median

מסדרים את ערכי המדגם בסדר עולה ובוחרים את הפרט האמצעי. אם מס' הפרטים הוא זוגי, עושים ממוצע בין שני פרטי האמצע.

החציון מסומן M ומקיים: לפחות עבור חצי מהפרטים באוכלוסיה ערך המשתנה אינו עולה על M וכן גם לפחות עבור חצי מהפרטים ערך המשתנה אינו עולה על M:

$$\boxed{P(X\le M)\ge \tfrac{1}{2} \text{ וגם } P(X\le M)\ge \tfrac{1}{2}}$$

שונות - Variance

השונות מבדיקה את פיזור הערכים סביב התוחלת:

עובדים עם ריבוע תוחלת $\boxed{Var(X)=\sigma^2=E(X-EX)^2=E(X^2)-(E(X))^2}$

הסטייה בגלל שלפי תכונה 5 של התוחלת שצריך משהו "אחר".

צריך לשים לב שהשונות היא ריבוע התוחלת, ולכן גם המימדים שלה מרובעים. מכאן:

סטיית תקן - Standard Deviation

מוגדרת ע"י שורש השונות, כדי לקבל את מימדי התוחלת: $\boxed{\sigma=\sqrt{Var(X)}}$

אור שגיא ©

תכונות של שונות

המשך תכונות בעמוד הבא

1) $Var(X) \geq 0$ ($Var\,X=0$ אם ורק כאשר X הוא קבוע)

2) $Var(X+a) = Var(X)$

3) $Var(bX) = b^2 Var(X)$

4) אם $X_1,\dots,X_n$ ב"ת: $Var(X_1+\dots+X_n) = Var(X_1)+\dots+Var(X_n)$

5) $Var(X+Y) = Var(X)+Var(Y)+2Cov(X,Y)$

התפלגויות מיוחדות - שונות ותוחלת

1) התפלגות אחידה בין a,b: $X \sim U(a,b)$: $E(X) = \frac{a+b}{2}$

$X \sim U[1,\dots,N]$: $EX = \frac{N+1}{2}$, $Var\,X = \frac{N^2-1}{12}$ $\{P(X=k) = \frac{1}{N}\}$

2) התפלגות בינומית: $X \sim B(n,p)$: $E(X) = np$, $Var(X) = np(1-p)$

3) התפלגות גיאומטרית: $X \sim G(p)$: $E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$

4) התפלגות פואסון: $X \sim P(\lambda)$: $E(X) = \lambda$, $Var(X) = \lambda$

5) התפלגות מעריכית: $X \sim e(\lambda)$: $E(X) = \frac{1}{\lambda}$, $Var(x) = \frac{1}{\lambda^2}$

IX התפלגות נורמלית

אם לוקחים את ההתפלגות הבינומית $X \sim B(n,p)$ כאשר $p=\frac{1}{2}$ ו-n שואף לגדול

מקבלים עקומת התפלגות שלא שטוחה. DeMoivre מצא שאם מזיזים את הגרף

כך ש: $E(X)=0$, $\sigma=1$ מקבלים עקומת פעמון בעלת הפונקציה: $\boxed{f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}}$

הפונקציה הזו נקראת התפלגות נורמלית סטנדרטית: $\boxed{X \sim N(0,1)}$

הפונקציה הזו, כמו במקרים אחרים, היא פונקציית צפיפות של ההתפלגות.

כל משתנה מקרי בינומי, עם הסתברות p להצלחה ותוחלת וסטיית תקן μ ו-σ בהתאם

אפשר להמיר למשתנה מקרי נורמלי סטנדרטי:

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

הטרנספורמציה $\boxed{Z = \frac{X-\mu}{\sigma}}$ משנה מ"מ נורמלי עם μ ו-σ בהתאם

למ"מ נורמלי סטנדרטי עם $\mu=0$ ו-$\sigma=1$.

ואז אפשר לחשב הסתברויות רציפות: $P(a \leq X \leq b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$

תיקון רציפות © ובכדי להיות יותר מדויקים מסתכלים על קצוות המרווחים בגרף: לא על a, אלא על $a-\frac{1}{2}$:

$$\boxed{P(a \leq X \leq b) \simeq P\left(\frac{a-\frac{1}{2}-\mu}{\sigma} \leq Z \leq \frac{b+\frac{1}{2}-\mu}{\sigma}\right)}$$

5) $Var(X+Y) = Var(X)+Var(Y)\pm 2Cov(X,Y)$

שונות משותפת - Covariance

$$\boxed{Cov(X,Y) = E(XY)-E(X)E(Y)}$$

$$\boxed{Cov(X,Y) = E(X-E(X))(Y-E(Y))}$$

X ניבוי לינארי

נניח לנו מ"מ דו-מימדי $\langle X,Y\rangle$ ואנחנו רוצים להעריך את X בהינתן Y (או הפוך), ואנחנו משתמשים בקירוב לינארי (כי הקשר ביניהם הוא דווקא לינארי) מצפים:
$\hat{Y} = aX+b$ (ניבוי Y). בכדי שנתמודד לעשות קירובים צריך להתחשב בטעות שאנחנו עושים. טעות זו (Standard Error) מוגדרת ע"י ריבוע הסטיה בין הערך האמיתי לקירוב שאנחנו נותנים:

$$\boxed{S^2 = E(\hat{Y}-Y)^2}$$

כיון שאנחנו רוצים למזער את הטעות, אנחנו רוצים למצוא a, b שיתנו לנו מינימום. נמצא שהמשוואה הבאה מספקת את הקירוב הטוב ביותר:

$$\hat{Y} = \hat{a}X+\hat{b} = \frac{Cov(X,Y)}{Var(X)}X + [E(Y)-\hat{a}E(X)]$$

או:

$$\hat{Y}-E(Y) = \hat{a}[X-E(X)]$$

$\boxed{\hat{a} = \frac{Cov(X,Y)}{Var(X)}}$ נותן לנו את שיפוע קו הניבוי - אם $\hat{a}=0$, כלומר $Cov(X,Y)=0$, אז המשתנים ב"ת. לקו הניבוי קוראים גם <u>קו הרגרסיה הלינארית</u>.

כיון שהגענו את הטעות למינימום, נרצה לראות במה טעות חסכנו:

ליחס זה קוראים <u>מקדם המתאם</u>:

$$\boxed{r = \frac{Var(Y)-S^2}{Var(Y)} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}}$$

<u>תכונות מקדם המתאם (קורלציה)</u>

1) $0\le r^2\le 1 \Leftarrow -1\le r\le 1$

② I) ידיעת ערך אחד משני המשתנים וניבוי לינארי חוסכים r^2 משונות הניבוי של השני.

II) r חיובי מבטיח קשר חיובי, r שלילי - קשר שלילי.

③ $r=\pm1$: <u>מתאם מלא</u> - חסכנו 100% מהטעות.

$r=0$: <u>חוסר מתאם</u> - המ"מ ב"ת, ולכן אין בעצם קורלציה.

אחרת: <u>מתאם חלקי</u>.

$\Leftarrow$ מתאם חיובי או שלילי - האם הקשר "מצבר" או "מציק".

אור בגלור ©

- אי-שוויון צ'בישב

במקרה שפונקצית ההסתברות של מ"מ אינה ידועה, ניתן למצוא הערכה להסתברות,
בעזרת אי-שוויון צ'בישב שנותן **חסם עליון** להסתברות (וזו הערכה מדויקת מאוד):
יהי X מ"מ עם תוחלת μ ושונות σ^2. אזי לכל $\varepsilon>0$ מתקיים:

$$\boxed{P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}}$$

- חוק המספרים הגדולים

אם ניקח את אי-שוויון צ'בישב ונשתמש בו לבדיקת הסטייה בין **ממוצע** תצפיות לבין
התוחלת שלהן (כשהן ב"ת כמובן), נקבל את המשפט הבא:

$\forall(x_1,\dots,x_n)$ תצפיות ב"ת בעלות תוחלת μ וסטיית-תקן σ

$$\bar{X}_n=\frac{1}{n}\sum_{k=1}^{n}X_k \quad\Rightarrow\quad E(\bar{X}_n)=\frac{1}{n}\sum_{k=1}^{n}E(X_k)=\mu$$

$$Var(\bar{X}_n)=\frac{1}{n^2}\sum_{k=1}^{n}Var(X_k)=\frac{\sigma^2}{n}$$

$$P\{|\bar{X}_n-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{n\varepsilon^2}$$

ואם נעשה הרבה מאוד תצפיות, נקבל ש:

$$\boxed{\lim_{n\to\infty}P\{|\bar{X}_n-\mu|\geq\varepsilon\}=0}$$

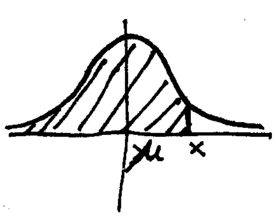
או במלים אחרות, אם עושים הרבה מאוד תצפיות, הסטייה בין ממוצע ערכי התצפיות
לבין התוחלת, שואף לאפס - כלומר $\bar{X}_n=\mu$, עפ"י ההסתברות. בקיצור רושמים:

$$\boxed{\bar{X}_n\xrightarrow[n\to\infty]{P}\mu}$$

- משפט הגבול המרכזי

כזכור, כשרצינו לתקנן משתנה כלשהו משתמשים בנוסחה: $\frac{X-\mu}{\sigma}$. ולכן
כדי לתקנן ממוצע תצפיות נשתמש ב: $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$. משפט הגבול המרכזי אומר:
יהיו $x_1,x_2,\dots$ מ"מ ב"ת ושווי התפלגות עם תוחלת $\mu<\infty$ ושונות $\sigma^2<\infty$.
אזי המ"מ $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$ שואף להתפלגות נורמלית סטנדרטית כאשר $n\to\infty$:

$$\forall(-\infty<x<\infty):\ P\left\{\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\leq x\right\}\xrightarrow[n\to\infty]{}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du=\Phi(x)$$



$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

$\Phi(x)$

ומסמנים בקיצור:

$$\boxed{\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\sim N(0,1)}$$

ארנבאום ©

הערות על משפט הגבול המרכזי

1) ישנו משפט חזק יותר, למקרה שהמ"מ אינם שווי-התפלגות:

יהיו $X_1, X_2, \dots$ מ"מ ב"ת (לאו דווקא שווי התפלגות), עם תוחלת סופית $E(X_k)=\mu_k$

ושונויות סופיות $Var(X_k)=\sigma_k^2$, כך שלכל k: $\frac{\sigma_k^2}{\sum_{m=1}^{n}\sigma_m} \xrightarrow[n\to\infty]{} 0$.

אזי:
$$\boxed{Y = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n}\mu_k}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}} \rightsquigarrow N(0,1)}$$

כלומר: גם אם יש הרבה גורמים קטנים שמשפיעים על המ"מ, אם מצליחים לבודד אותם ההשפעה שלהם מתבטלת ומקבלים התפלגות נורמלית סטנדרטית.

2)
$$\boxed{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}}}$$

3) $\forall (t \geq 3): \Phi(t) \simeq 1$

4) $\Phi(-t) = 1 - \Phi(t)$

5) $P(X \leq c) \rightsquigarrow \Phi(c)$

$P(X \geq c) \rightsquigarrow 1 - \Phi(c)$

סטטיסטיקה - סיכום

I אומדים

כיוון שבעולם האמיתי עובדים לפי ניסויים, אין סיבה שנדע את הערכים האמיתיים של משתנים כגון p, μ, σ, r ואחרים. לכן עובדים עם **אומדים** - פונקציות שאמורות להעריך עבורנו את גודל המשתנה. יש שני סוגים של אומדים:

1) אומד חסר הטייה: אומד שמקיים את התנאי: $\boxed{E(\hat{\theta}) = \theta}$

2) אומד עקבי: סדרת אומדים שמקיימים את התנאי: $\boxed{Y_i \xrightarrow[n\to\infty]{P} \theta}$

בד"כ נעדיף לעבוד עם אומדים חסרי הטייה (אח"ה).

1) אומד אח"ה עבור μ: $\boxed{\hat{\mu} = \bar{X}_n}$ $\{ E(\bar{X}_n) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \mu \}$

2) אח"ה עבור σ^2: במקרה זה צריך להתייחס לשני מקרים שונים:

I) μ ידוע: $\boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} (X_k - \mu)^2}$

II) μ לא ידוע: $\boxed{\hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n} (X_k - \bar{X}_n)^2}$

3) אח"ה עבור המתאם r: $\boxed{\hat{r} = \frac{1}{n}\sum_{k=1}^{n} (X_k - \bar{X}_n)(Y_k - \bar{Y}_n) = \frac{1}{n} \cdot \hat{\sigma}_x \cdot \hat{\sigma}_y}$

- האומד הוא בעצם מ"מ שמתאר ממוצע של אומדנים שונים שקיבלנו בתצפיות השונות.

II בדיקת השערות

כשאנחנו רוצים לבדוק השערה מסוימת (H_1) אנחנו צריכים לבדוק אותה כנגד מודל קיים - השערת אפס (H_0). אנחנו לא "מקבלים" השערות, אלא "דוחים" את השערת האפס.

כדי לדחות את השערת האפס אנחנו צריכים לקבוע אזור דחיה (R): אוסף כל התוצאות האפשריות לניסוי שאנחנו מחליטים מראש שישכנעו אותנו ש-H_0 מוטעה.

- רמת מובהקות: הסיכוי לדחות את השערת האפס אם היא נכונה: $\boxed{\alpha = P\{R | H_0\}}$
בפרקטיקה משתמשים ב- $\boxed{\alpha = 0.05 \text{ או } \alpha = 0.01}$.

- טעות סטטיסטית מסוג ראשון: דחייה של השערת האפס אם היא נכונה.

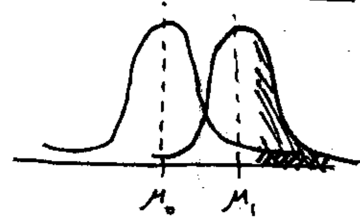
מכאן נובע שרמת המובהקות היא בעצם הסיכוי לעשות טעות סטטיסטית מסוג ראשון, ולכן חייבים לקחת α קטנה מאוד.

- <u>טעות סטטיסטית מסוג שני</u>: אי-דחייה של השערת האפס כאשר ההשערה הנגדית נכונה.
- <u>עוצמת מבחן</u>: הסיכוי <u>לא</u> לעשות טעות סטטיסטית מסוג שני (אם H_1 נכונה) - כלומר, הסיכוי לדחות את השערת האפס כשהיא לא נכונה. $\boxed{(1-\beta)=P\{R|H_1\}}$
- <u>הערות</u>

1) אין סימטריה בין שני סוגי הטעויות: טעות מסוג ראשון "מסוכנת" יותר - היא גורמת לתשובה של תיאוריה מוטעית. טעות מסוג שני רק דוחה השערה חדשה, היא לא מטעה את הציבור.
2) באם רמת המובהקות הינה סטנדרט קבוע מראש, עוצמת המבחן תלויה ב-n בלבד, כאשר כמובן נרצה שהיא תהיה כמה שיותר גבוהה.
3) ההבדל בין תצפית אחת להרבה תצפיות: כפי שאפשר לראות מהשרטוטים, ברגע שמגדילים את מס' התצפיות מגדילים את עוצמת המבחן.

<u>n תצפיות</u>

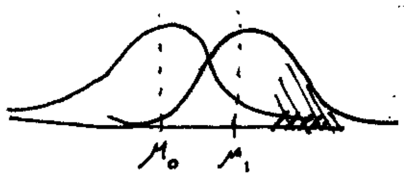
$\sigma^{*2}=\frac{\sigma^2}{n}$



μ_0 μ_1

- אזור הדח"ה
- עצמת המבחן

<u>תצפית אחת</u>



μ_0 μ_1

- אזור הדח"ה
- עוצמת המבחן

- <u>חישוב אזור הדח"ה</u>

1) $$P(\bar{X}\geq c|H_0)=P\left(\frac{\bar{X}-\mu_0}{\sqrt{\sigma^2/n}}\geq\frac{c-\mu_0}{\sqrt{\sigma^2/n}}\right)=1-\Phi\left(\frac{c-\mu_0}{\sqrt{\sigma^2/n}}\right)=\alpha$$

$$\Rightarrow \boxed{c=\mu_0+\sqrt{\sigma^2/n}\,\Phi^{-1}(1-\alpha)}$$

עבור השערות מסוג $H_1:\mu>\mu_0$

2) $$P(\bar{X}\leq c|H_0)=\Phi\left(\frac{c-\mu_0}{\sqrt{\sigma^2/n}}\right)\Rightarrow\boxed{c=\mu_0-\sqrt{\sigma^2/n}\,\Phi^{-1}(1-\alpha)}$$

עבור השערות מסוג $H_1:\mu<\mu_0$

- <u>חישוב עוצמת המבחן</u>

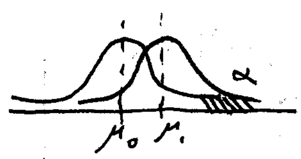
$$P\{R|H_1\}=P\{\bar{X}\geq c|\mu=\mu_1\}=P\{\bar{X}\geq c|\bar{X}\sim N(\mu_1,\tfrac{\sigma^2}{n})\}=$$
$$=\boxed{1-\Phi\left(\frac{c-\mu_1}{\sqrt{\sigma^2/n}}\right)}$$

p-value: ההסתברות, בהינתן H_0, לתוצאה קיצונית לפחות כמו זו שהתקבלה.

$$\boxed{p\text{-value} = P(\bar{X} \geq b \mid H_0) = 1 - \Phi\left(\frac{b-\mu_0}{\sqrt{\sigma^2/n}}\right)}$$

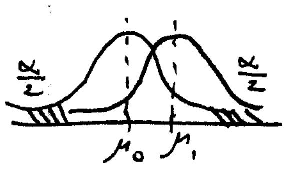
III סוגי מבחנים

1) השערות פשוטות ⇐ מבחנים חד-צדדיים: במקרה ובודקים השערות מסוג $\mu > \mu_0$ או $\mu_0 > \mu$ מסתפקים במבחן חד-צדדי - אזור דחייה אחד:
עובדים לפי רמת מובהקות α.

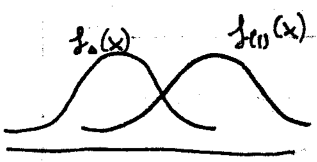


- אזור דחייה

2) השערות מורכבות ⇐ מבחן דו-צדדי: במקרים ובודקים השערה מסוג $\mu \neq \mu_0$
צריכים להתחשב בשני צדי ההתפלגות. ולכן חייבים לדאוג להשתמש ברמת מובהקות $\frac{\alpha}{2}$.
אחרת מקבלים רמת מובהקות גבוהה משרצינו והסיכוי לטעות
טעות מסוג ראשון עולה.



3) מבחן יחס נראות מקסימלי



יש לנו שתי התפלגויות שמתוארות ע"י פונקציות הצפיפות
שלהן: $f_0(x)$ ו- $f_1(x)$.
הגודל $\frac{f_1(x)}{f_0(x)}$ נקרא יחס הנראות של H_1 ל- H_0.
למת Neiman-Pearson קובעת שמבחן יחס הנראות (עם רמת מובהקות נתונה) הוא
מבחן העוצמה המקסימלי. אנחנו נדחה את H_0 כשהיחס $\frac{f_1(x)}{f_0(x)}$ גדול מאוד מאחד -
כלומר כשהתוצאות מוטות מאוד ל- H_1.
כאשר $\alpha = P\{R \mid H_0\}$ וכאשר C_α נקבע על-ידינו, קובעת הלמה:

$$\boxed{R_\alpha = P\left\{X = x \;\middle|\; \frac{f_1(x)}{f_0(x)} > C_\alpha\right\}}$$

- חישוב אזור דחייה במבחן דו-צדדי

$$\boxed{P(|\bar{X}_n - \mu_0| \geq c \mid H_0) = \alpha}$$

$$\Rightarrow \Phi\left(\frac{c}{\sqrt{\sigma^2/n}}\right) = 1 - \frac{\alpha}{2} \quad \Rightarrow \quad \boxed{c = \sqrt{\sigma^2/n}\;\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)}$$

$$P(|\bar{X}_n - \mu_0| \leq c \mid H_0) = \alpha$$

$$2\Phi\left(\frac{c}{\sqrt{\sigma^2/n}}\right) - 1 = \alpha \quad \Rightarrow \quad \boxed{c = \sqrt{\sigma^2/n}\;\Phi^{-1}\left(\frac{1+\alpha}{2}\right)}$$

אור בכור ©

IV התפלגות חי-בריבוע χ^2

המבחנים הקודמים שהוצגו רלוונטיים למקרים בהם μ ו-σ קבועים. אך לפעמים נרצה לבדוק השערות מסוג שונה. למשל, ש-σ משתנה. נסתכל על מקרה שבו σ גדלה ונבדיל לשני מקרים שונים:

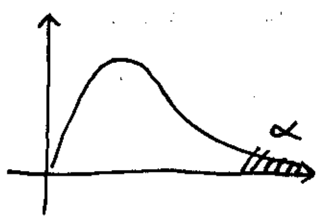
1) μ ידוע: נשתמש בהערכה $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2$

קל לראות שהאומד הזה לא יכול להתפלג נורמלית, היות ו- $\hat{\sigma}_n^2 \geq 0$ תמיד.

כדי לבחון את השערת האפס נעבוד עם הגודל הבא:

$$\boxed{H_0:\ \frac{n\hat{\sigma}_n^2}{\sigma^2} = \sum_{k=1}^{n}\left(\frac{X_k-\mu}{\sigma}\right)^2 = \sum_{k=1}^{n} Z_k^2 \ , \quad Z_k \sim N(0,1)}$$

נזכיר: התפלגות סכום ריבועי מ"מ נורמליים סטנדרטיים נקראת התפלגות חי-בריבוע עם n דרגות חופש: χ_n^2.



2) μ לא ידוע: נשתמש בהערכה $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X}_n)^2$

במקרה זה נשתמש בקירוב להתפלגות חי-בריבוע עם $n-1$ דרגות חופש, בהתאם ל-$(n-1)$ דרגות החופש של $\hat{\sigma}_n^2$.

$$H_0:\ \boxed{\frac{(n-1)\hat{\sigma}_n^2}{\sigma^2} \sim \chi_{n-1}^2}$$

- חישוב אזור דחייה

בודקים בטבלת התפלגות χ^2 ומוצאים את הערך של c עבור α עם n דרגות חופש. מציבים את c לתוך המשוואה ומקבלים:

$$\boxed{P\left(\frac{n\hat{\sigma}_n^2}{\sigma^2} \geq c\right) = \alpha}$$

$$\boxed{\hat{\sigma}_n \geq \sigma \cdot \sqrt{\frac{c}{n}}}$$

V - מבחן χ^2 לטיב-ההתאמה

המבחן הזה נועד כדי לבדוק האם סט מסו"ים של נתונים מתפלג בצורה מסו"ימת.

השערת האפס מניחה התפלגות מסו"ימת וה-p_i הנלווים לה.

<u>הגדרה</u>: לניסוי (מולטינומי) יש m תוצאות אפשריות.

השערת האפס: הסיכוי לתוצאה ה-k הוא $p_k \geq 0$ לכל $k=1,\dots,m$, כאשר $\sum_{k=1}^{m} p_k = 1$.

עורכים n תצפיות ב"ת. תחת השערת האפס, בקירוב "טוב", הגודל:

$$\boxed{X^2=\sum_{k=1}^{n} \frac{(n_k - np_k)^2}{np_k} \sim \chi^2_{n-1}}$$

כש-n_k הוא מס' הפעמים שהתקבל k.

נדחה את השערת האפס עבור ערכי χ^2

גדולים באופן מובהק: $\boxed{X^2 > \chi^2_{n-1}(\alpha)}$

- מבחן χ^2 לאי-תלות

נסתכל על מ"מ דו-מימדי $\langle X,Y \rangle$, כאשר: $j=1,\dots,l$; $i=1,\dots,k$; $p_{ij}=P(X=x_i, Y=y_j)$.

ההתפלגויות השוליות, כזכור, מוגדרות: $p_{i\cdot} = \sum_{j=1}^{l} p_{ij} = \sum_{j=1}^{l} P(X=x_i, Y=y_j) = P(X=x_i)$

$$p_{\cdot j} = \sum_{i=1}^{k} p_{ij} = P(Y=y_j)$$

הפעם אנחנו רוצים לבדוק האם קיימת תלות בין שני המ"מ.

כזכור, אם המ"מ ב"ת, אז: $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$

נסתכל על שני מקרים:

1) <u>התפלגויות שוליות ידועות</u>: זהו מבחן אי-התלות הוא מקרה פרטי של מבחן טיב-ההתאמה:

$$\boxed{X^2 = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(n_{ij} - np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}} \sim \chi^2_{(kl-1)}}$$

2) <u>התפלגויות שוליות לא ידועות</u>: אז משתמשים באומדנים להתפלגויות השוליות

(כאשר האומדן עבור הסתברות הוא השכיחות היחסית: $\hat{p}_{ij} = \frac{n_{ij}}{n}$)

ודרגות החופש הן $(k-1)(l-1)$:

$$\boxed{X^2 = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{\left(n_{ij} - \frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{\frac{n_{i\cdot}n_{\cdot j}}{n}} = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(n_{ij} - n\hat{p}_{i\cdot}\hat{p}_{\cdot j})^2}{n\hat{p}_{i\cdot}\hat{p}_{\cdot j}} \sim \chi^2_{(k-1)(l-1)}}$$

דף נוסחאות: הסתברות וסטטיסטיקה לפיזיקאים

הסתברות מותנית: $\Rightarrow p(B|A) = \frac{p(A\cap B)}{p(A)}$

הסתברות א-פריורית: $\Rightarrow p(A) = p(B)p(A|B) + p(\bar{B})p(A|\bar{B})$

נוסחת ההסתברות השלמה: $p(A) = \sum_{k=1}^{n} p(B_k)\,p(A|B_k)$

הסתברות אפוסטריורית: $\Rightarrow p(B|A) = \frac{p(A\cap B)}{p(A)} = \frac{p(B)p(A|B)}{p(A)}$

נוסחת בייס: $p(B_k|A) = \frac{p(B_k\cap A)}{p(A)} = \frac{p(B_k)p(A|B_k)}{\sum_{m=1}^{n} p(B_m)p(A|B_m)}$

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

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

- מ"מ רציף

פונקציית צפיפות: $\Rightarrow P(a \le X \le b) = \int_a^b f(x)\,dx \quad , \int_{-\infty}^{\infty} f(x)\,dx = 1 =$ שטח

פונקציית הסתברות מצטברת: $\Rightarrow F_X(x) = P(X \le x) = \int_{-\infty}^{x} f(x)\,dx$

- מ"מ דו-מימדי

תלות בין X,Y: $\Rightarrow P\{X = x_n, Y = y_m\} = P(X=x_n)P(Y=y_m \mid X=x_n)$

$\Rightarrow P\{X+Y = n\} = \sum_k P(X=k, Y=n-k) = \sum_k P(X=k)P(Y=n-k)$

התפלגויות שוליות: $\Rightarrow P_X(x) = \sum_y P(x,y) = P(X=x) \quad , P_Y(y) = \sum_x P(x,y) = P(Y=y)$

- תוחלת, שונות, cov

$\Rightarrow E(x) = \sum_k k\,P(X=k) \quad , E(x^2) = \sum_k k^2 P(X=k) \quad , E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$

$\Rightarrow Var(x) = E(x^2) - [E(x)]^2 \quad , \sigma = \sqrt{Var(x)}$

$\Rightarrow Var(X,Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$

$\Rightarrow Cov(X,Y) = E(X\cdot Y) - E(X)E(Y)$

$E(X\cdot Y) = \sum_x \sum_y x\cdot y\cdot P(x,y)$

- רגרסיה לינארית

$\Rightarrow r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \quad , \hat{a} = \frac{Cov(X,Y)}{Var(x)}$

$\hat{Y} = \hat{a}[X - E(x)] + E(Y) \quad , S^2 = Var(Y)(1-r)$

$\Rightarrow E(\tau) = E(T) + \frac{Var(T)}{E(T)}$

התפלגויות מיוחדות

1) $X\sim U[1,\dots,n]$; $p(X=k)=\frac{1}{n}$, $E(X)=\frac{n+1}{2}$, $Var(X)=\frac{n^2-1}{12}$

$X\sim U(a,b)$; $p(X=k)=\frac{1}{b-a}$, $E(X)=\frac{b-a}{2}$, $Var(X)=\frac{(b-a)^2}{12}$

$X\sim B(n,p)$; $p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$, $E(X)=np$, $Var(X)=np(1-p)$

$X\sim G(p)$; $p(X=k)=p(1-p)^{k-1}$, $E(X)=\frac{1}{p}$, $Var(X)=\frac{1-p}{p^2}$

$X\sim P(\lambda)$: $p(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$, $E(X)=\lambda$, $Var(X)=\lambda$

$\sim P(\lambda t)$: $p(X=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t}$, $E(X)=\lambda$, $Var(X)=\lambda$

$X\sim e(\lambda)$: $p(X\le t)=1-e^{-\lambda t}$, $E(X)=\frac{1}{\lambda}$, $Var(X)=\frac{1}{\lambda^2}$

$p(X>t)=e^{-\lambda t}$

$X\sim I$ $E(X)=p$, $Var(X)=p(1-p)$

סטטיסטיקה

מס' גדולים: $\lim_{n\to\infty}P\{|\bar{X}_n-\mu|\ge\varepsilon\}=0$ צ'בישב: $P\{|X-\mu|\ge\varepsilon\}\le\frac{\sigma^2}{\varepsilon^2}$

גבול מרכזי: $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sigma\sqrt{n}}\sim N(0,1)$, $Z=\frac{X-\mu}{\sigma}$

מובהקות: $\alpha=P\{R|H_0\}$: $P(\bar{X}_n\ge c)=1-\Phi(\alpha)$, $P(\bar{X}_n\le c)=\Phi(\alpha)$

$c=\mu_0+\sqrt{\sigma^2/n}\,\Phi^{-1}(1-\alpha)$ $c=\mu_0-\sqrt{\sigma^2/n}\,\Phi^{-1}(1-\alpha)$

: $P(|\bar{X}_n-\mu_0|\ge c)=2\left(1-\Phi\left(\frac{c}{\sqrt{\sigma^2/n}}\right)\right)=\alpha$

$P(|\bar{X}_n-\mu_0|\le c)=2\Phi\left(\frac{c}{\sqrt{\sigma^2/n}}\right)-1=\alpha$

עוצמת מבחן: $(1-\beta)=P\{R|H_1\}$: $P(\bar{X}_n\ge c|H_1)=1-\Phi\left(\frac{c-\mu_1}{\sqrt{\sigma^2/n}}\right)$

p-value: $p\text{-value}=P(\bar{X}_n\ge b|H_0)=1-\Phi\left(\frac{b-\mu_0}{\sqrt{\sigma^2/n}}\right)$

אמד לשונות: $\hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\mu)^2$, $\hat{\sigma}^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X}_n)^2$

שונות משתנה: $\frac{n\hat{\sigma}_n^2}{\sigma^2}\ge\chi^2_{n}(\alpha)$ $\frac{(n-1)\hat{\sigma}_n^2}{\sigma^2}\ge\chi^2_{n-1}(\alpha)$

טיב התאמה: $X^2=\sum_{k=1}^{n}\frac{(n_k-np_k)^2}{np_k}\ge\chi^2_{n-1}(\alpha)$

$X^2=\sum_{i=1}^{k}\sum_{j=1}^{\ell}\frac{(n_{ij}-np_{i\cdot}p_{\cdot j})^2}{np_{i\cdot}p_{\cdot j}}\ge\chi^2_{k\ell-1}(\alpha)$, $X^2=\sum_{i=1}^{k}\sum_{j=1}^{\ell}\frac{\left(n_{ij}-\frac{n_{i\cdot}n_{\cdot j}}{n}\right)^2}{n_{i\cdot}n_{\cdot j}/n}\ge\chi^2_{(k-1)(\ell-1)}(\alpha)$

$\hat{\mu}=\bar{X}_n$, $\hat{r}=\frac{1}{n}\cdot\hat{\sigma}_X\cdot\hat{\sigma}_Y$

1) <u>מציאת Cov(X,Y) ועוד דברים נחמדים</u>

$$\boxed{Cov(X,Y) = E(X\cdot Y) - E(X)E(Y)}$$

הכי פשוט לעבוד עם טבלת התפלגויות משותפות של שני המשתנים. את $E(X)$ ו-$E(Y)$ מוצאים מההתפלגויות השוליות, לפי הנוסחה:

$$\boxed{\begin{aligned} E(x) &= \sum_k k\,P(X=k) \\ E(X\cdot Y) &= \sum_x \sum_y x\cdot y\cdot P(X,Y) \end{aligned}}$$

את $E(X\cdot Y)$ מוצאים לפי הערכים בתוך הטבלה:

נניח שזו הטבלה:

	1	2	3	4	P(Y=y)
1	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{7}{16}$
2	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{5}{16}$
3	0	0	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$
4	0	0	0	$\frac{1}{16}$	$\frac{1}{16}$
P(X=x)	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{5}{16}$	$\frac{7}{16}$	

$$E(X) = 1\cdot\frac{1}{16} + 2\cdot\frac{3}{16} + 3\cdot\frac{5}{16} + 4\cdot\frac{7}{16} = \frac{25}{8}$$

$$E(Y) = 1\cdot\frac{7}{16} + 2\cdot\frac{5}{16} + 3\cdot\frac{3}{16} + 4\cdot\frac{1}{16} = \frac{15}{8}$$

$$E(X\cdot Y) = \frac{1}{16}(1\cdot1 + 2\cdot2 + 3\cdot3 + 4\cdot4) + \frac{2}{16}(2\cdot1 + 3\cdot1 + 4\cdot1 + 3\cdot2 + 4\cdot2 + 4\cdot3) = \frac{30}{16} + \frac{70}{16} = \frac{100}{16}$$

$$Cov(X,Y) = E(X\cdot Y) - E(X)E(Y) = \frac{100}{16} - \frac{25}{8}\cdot\frac{15}{8} = \frac{400-375}{64} = \frac{25}{64}$$

ואז קל למצוא:

$$\boxed{Var(X,Y) = Var(X) + Var(Y) + 2Cov(X,Y)}$$

$$\boxed{Var(X) = E(X^2) - (E(X))^2} = \left[1\cdot\frac{1}{16} + 4\cdot\frac{3}{16} + 9\cdot\frac{5}{16} + 16\cdot\frac{7}{16}\right] - \frac{625}{64} = \frac{680-625}{64} = \frac{55}{64}$$

$$\boxed{Var(Y) = E(Y^2) - (E(Y))^2} = \left[1\cdot\frac{7}{16} + 4\cdot\frac{5}{16} + 9\cdot\frac{3}{16} + 16\cdot\frac{1}{16}\right] - \frac{225}{64} = \frac{280-225}{64} = \frac{55}{64}$$

$$Var(X,Y) = \frac{110}{64} + \frac{200}{16} = \frac{910}{64}$$

ואז גם קל למצוא מקדם מתאם:

$$\boxed{r = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}}$$

$$r = \frac{\frac{25}{64}}{\frac{55}{64}} = \frac{25}{55} = \frac{5}{11}.$$

ואפילו נוכל למצוא קו רגרסיה ל-Y:

$$\boxed{\hat{Y} = \hat{a}\,[X - E(X)] + E(Y) = \frac{Cov(X,Y)}{Var(X)}[X - E(X)] + E(Y)}$$

$$\hat{Y} = \frac{5}{11}X + \frac{5}{11}$$