הסקה סטטיסטית שיעור 4

הקדמה חומר קריאה 4א, חומר קריאה 4ב, חומר קריאה 4ג, חומר קריאה 4ד, חומר קריאה 4ה

1 התפלגות משותפת

מ"מ בדיד

עבור שני מ"מ בדידים X, Y כאשר X מקבל ערכים $\{x_1, \dots, x_n\}$, Y מקבל ערכים $\{y_1, \dots, y_m\}$. הזוג הסדור (X, Y) מקבל ערכים $\{(x_1, y_1), (x_1, y_2), \dots, (x_i, y_j), \dots, (x_n, y_m)\}$.
הPMF המשותף של X ו-Y הוא הפונקציה $p(x_i, y_j)$ שנותנת את ההסתברות למאורע המשותף: $P(X = x_i, Y = y_j)$.

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	$\cdots$	$p(x_1, y_j)$	$\cdots$	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	$\cdots$	$p(x_2, y_j)$	$\cdots$	$p(x_2, y_m)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	$\cdots$	$p(x_i, y_j)$	$\cdots$	$p(x_i, y_m)$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	$\cdots$	$p(x_n, y_j)$	$\cdots$	$p(x_n, y_m)$

דוגמה 1, עבור שתי קוביות:

$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

דוגמה 2, עבור X=תוצאת קוביה א, T=סכום שתי הקוביות:

$X \backslash T$	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

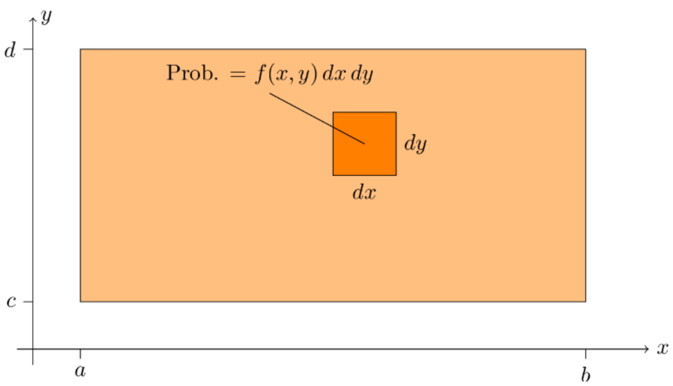
זאת בעצם פונקציית הסתברות, ולכן צריכה לקיים:

$$0 \le p(x_i, y_j) \le 1, \qquad \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) = 1$$

מ"מ רציף

המקרה הרציף שקול. משתמשים בPMF רציף ואינטגרלים:

אם X מקבל ערכים $[a, b]$, Y מקבל ערכים $[c, d]$. הזוג הסדור (X, Y) מקבל ערכים $[a, b] \times [c, d]$. הPDF המשותף של X ו-Y הוא הפונקציה $f(x, y)$ שנותנת את ההסתברות למאורע המשותף – כלומר, ההסתברות ש (X, Y) נמצא במלבן קטן ברוחב dx וגובה dy מסביב ל-(x, y) היא $f(x, y)\, dx\, dy$.



והיא צריכה לקיים:

$$0 \le f(x, y), \qquad \int_a^b \int_c^d f(x, y)\ dx\, dy = 1$$

נשים לב: בגלל שהיא PDF, היא יכולה לקבל ערכים גדולים מ-1. זאת פונקציית **צפיפות**, לא הסתברות.

מה צריך לדעת לגבי אינטגרלים כפולים בקורס:

- להבין את המשמעות של אינטגרל כפול בתור סכום כפול.
- לחשב אינטגרל כפול על מלבן.
- עבור שטח לא מלבני (עם שטח A), כאשר $f(x, y) = c$ הוא קבוע, לדעת שהאינטגרל הכפול זהה ל $c \times A$.

מאורעות

אפשר להשתמש במ"מ כדי לתאר זוגות של מ"מ או מאורעות משותפים.

דוגמה 3

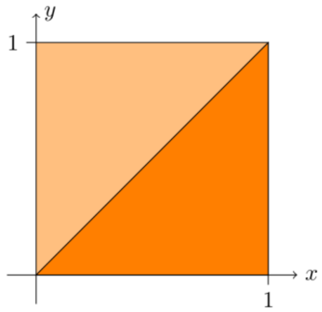
$X\backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

מתוך דוגמה 1, נתאר את המאורע $B = `Y - X \geq 2`$ ונמצא את ההסתברות שלו:

$$B = \{(1,3),(1,4),(1,5),(1,6),(2,4),(2,5),(2,6),(3,5),(3,6),(4,6)\}$$

כלומר, $P(B) = \frac{10}{36}$.

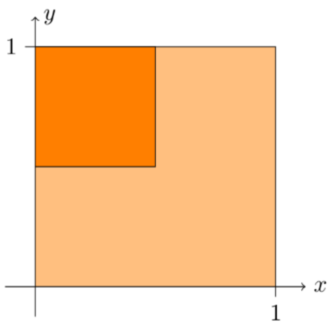
דוגמה 4



נניח ש X,Y מקבלים ערכים ב $[0,1]$ עם צפיפות אחידה $f(x,y) = 1$. נתאר את המאורע $`X > Y`$ ונמצא את ההסתברות:

ביחד, X ו-Y מקבלים ערכים בריבוע היחידה. המאורע $`X > Y`$ מתאים לכל מה שמתחת לישר $y = x$. מכיוון שפונקציית הצפיפות קבועה, ההסתברות היא פשוט החלק שהמאורע תופס מהשטח הכולל. במקרה הזה, חצי.

דוגמה 5



נניח ש X,Y מקבלים ערכים ב $[0,1]$ עם צפיפות אחידה $f(x,y) = 4xy$. צ"ל ש $f(x,y)$ היא PDF משותפת תקינה, נתאר את המאורע $A = `X < 0.5 \ and \ Y > 0.5`$, ונמצא את $P(A)$.

ביחד, X ו-Y מקבלים ערכים בריבוע היחידה, וניתן לראות את המאורע A. כדי להוכיח שהפונקציה היא PDF, נראה שהיא תמיד חיובית (טריוויאלי) ושההסתברות הכוללת היא 1:

$$\int_0^1 \int_0^1 4xy \, dx \, dy = \int_0^1 [2x^2y]_0^1 \, dy = \int_0^1 2y \, dy = 1$$

נמצא את ההסתברות:

$$P(A) = \int_0^{0.5} \int_{0.5}^1 4xy \, dy \, dx = \int_0^{0.5} [2xy^2]_{0.5}^1 \, dx = \int_0^{0.5} \frac{3x}{2} \, dx = \frac{3}{16}$$

CDF משותפת

נניח ש X, Y הם מ"מ עם התפלגות משותפת. הـCDF המשותפת מוגדרת: $F(x,y) = P(X \leq x, Y \leq y)$.

במקרה הרציף, אם X ו-Y בעלי צפיפות משותפת על $[a,b] \times [c,d]$, אז הـCDF המשותפת היא:

$$F(x,y) = \int_c^y \int_a^x f(u,v) \, du \, dv$$

כדי למצוא את הـPDF המשותפת, נגזור את הـCDF. בגלל שיש שני משתנים צריך לעשות נגזרת חלקית:

$$f(x,y) = \frac{\partial^2 F}{\partial x \partial y}(x,y)$$

במקרה הבדיד, אם ל-X, Y יש PMF משותפת $p(x_i, y_j)$ אז הـCDF המשותפת:

$$F(x,y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j)$$

תכונות של CDF משותפת

הCDF המשותפת של Y,X: $F(x,y)$, צריכה לקיים:

א) מונוטונית עולה.

ב) $F(x,y)=0$ נמצא בקצה השמאלי למטה. אם הקצה הוא $(-\infty,-\infty)$, זה אומר ש $\lim_{(x,y)\to(-\infty,-\infty)} F(x,y)=0$.

ג) $F(x,y)=1$ נמצא בקצה הימני למעלה. אם הקצה הוא (∞,∞), זה אומר ש $\lim_{(x,y)\to(\infty,\infty)} F(x,y)=1$.

דוגמה 6

נמצא את הCDF המשותפת של המ"מ מדוגמה 5. המאורע $X\le x, Y\le y$ הוא מלבן מוכל בריבוע היחידה. נחשב:

$$F(x,y)=\int_0^y 4uv\ \ du\,dv = x^2y^2$$

דוגמה 7

בדוגמה 1, נחשב את $F(3.5,4)$. זה המאורע $X\le 3.5, Y\le 4$. הסכום הוא $\frac{1}{3}$.

$X\backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

התפלגות שולית

חישוב אחד מתוך זוג של התפלגות משותפת.

PMF שולי – דוגמה 8

נחשב את הPMF השולי של T,X מדוגמה 2:
בטבלה, כל שורה מייצגת ערך X וכל עמודה ערך T. נחשב את הסכומים של כל שורה ועמודה.
הסכומים האלה הם ההתפלגות השולית.

$X\backslash T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(t_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

"התפלגות **שולית**" נובע מכך שההתפלגויות רשומות בשוללי הטבלה.

באופן כללי:

$$p_X(x_i)=\sum_j p(x_i,y_j)\,,\qquad p_Y(y_j)=\sum_i p(x_i,y_j)$$

PDF שולית

עבור התפלגות משותפת רציפה, בתחום $[a,b]\times[c,d]$:

$$f_X(x)=\int_c^d f(x,y)dy\,,\qquad f_Y(y)=\int_a^b f(x,y)dx$$

דוגמה 9

נניח ש (X,Y) מקבל ערכים בתוך $[0,1]\times[1,2]$, עם PDF משותפת $f(x,y)=\frac{8}{3}x^3y$. נמצא את הPDF השוליים:

כדי למצוא את $f_X(x)$ נעשה אינטגרל לפי y:

$$f_X(x)=\int_1^2 \frac{8}{3}x^3y\ dy=\left[\frac{4}{3}x^3y^2\right]_1^2=4x^3$$

והפוך בשביל y:

$$f_Y(y)=\int_0^1 \frac{8}{3}x^3y\ dx=\left[\frac{2}{3}x^4y^1\right]_0^1=\frac{2}{3}y$$

דוגמה 10

נניח ש (X,Y) מקבל ערכים בתוך $[0,1]\times[0,1]$, עם PDF משותפת $f(x,y)=\frac{3}{2}(x^2+y^2)$. נמצא את ה$PDF$ השולית $f_X(x)$, ונחשב את $P(X<0.5)$:

$$f_X(x)=\int_0^1 \frac{3}{2}(x^2+y^2)\,dy=\left[\frac{3}{2}x^2y+\frac{y^3}{2}\right]_0^1=\frac{3}{2}x^2+\frac{1}{2}$$

$$P(X<0.5)=\int_0^{0.5} f_X(x)dx=\int_0^{0.5}\frac{3}{2}x^2+\frac{1}{2}dx=\left[\frac{1}{2}x^3+\frac{1}{2}x\right]_0^{0.5}=\frac{5}{16}$$

***CDF* שולית**

אם Y,X מקבלים ערכים מ $[a,b]\times[c,d]$, אז:

$$F_X(x)=F(x,d),\qquad F_Y(y)=F(b,y)$$

אם $d=\infty$, זה הגבול: (והפוך עבור $b=\infty$)

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

דוגמה 11

הCDF המשותפת בדוגמה 10 הייתה $F(x,y)=\frac{1}{2}(x^3y+xy^3)$ על $[0,1]\times[0,1]$. נמצא את הCDF השוליים כדי לחשב את $P(X<0.5)$:

$$F_X(x)=F_X(x,1)=\frac{1}{2}(x^3+x),\qquad F_Y(y)=F_Y(1,y)=\frac{1}{2}(y+y^3)$$

אז, $P(X<0.5)=F_X(0.5)=\frac{1}{2}(0.5^2+0.5)=\frac{5}{16}$. כמו שראינו.

תלת מימד

תיארנו את $P(a<X<b)$ בתור שטח מתחת לגרף דו מימדי של הPDF $f(x)$ בתחום $[a,b]$. מכיוון שהתחום של (X,Y) הוא כבר שטח דו מימדי, הגרף של $f(x,y)$ הוא שטח מעל האיזור הזה (דמיינו מפה טופוגרפית עם בליטות). ואז ההסתברות היא ה**נפח**.

2 אי תלות

מ"מ בעלי התפלגות משותפת X,Y ייקראו בת"ל אם הCDF המשותפת שלהם היא מכפלת הCDF השוליים:

$$F(X,Y)=F_X(x)F_Y(y)$$

עבור מ"מ בדידים, זה שקול לכך שהPMF המשותפת היא מכפלת הPMF השוליים:

$$p(x_i,y_j)=p_X(x_i)p_Y(y_j)$$

עבור מ"מ רציף, זה שקול לכך שהPDF המשותפת היא מכפלת הPDF השוליים:

$$f(x,y)=f_X(x)f_Y(y)$$

דוגמה 12

עבור מ"מ בדידים, המשמעות של אי תלות היא שההסתברות בתא היא המכפלה של ההתפלגויות השוליות של העמודה והשורה. ניתן לראות שבטבלה הראשונה זה מתקיים ובשנייה לא.

$X\backslash Y$	1	2	3	4	5	6	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/6	1/6	1/6	1/6	1/6	1/6	1

$X\backslash T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

דוגמה 13

עבור מ"מ רציפים, המשמעות של אי תלות היא שאפשר לבטא את ה-PDF המשותפת כמכפלה של פונקציה של x ופונקציה של y.

א) נניח ש-X בתחום $[0,0.5]$, Y בתחום $[0,1]$, $f(x,y)=96x^2y^3$. אז X, Y בת"ל. הצפיפויות השוליות הן:

$$f_X(x)=24x^2,\qquad f_Y(y)=4y^3$$

ב) אם $f(x,y)=1.5(x^2+y^2)$ בריבוע היחידה, אז X,Y לא בת"ל כי אין איך לבטא את $f(x,y)$ כמכפלה $f_X(x)f_Y(y)$.

ג) אם $F(x,y)=\frac{1}{2}(x^3y+y^3x)$ בריבוע היחידה, אז X,Y לא בת"ל כי ה-CDF לא מכפלה של $F_X(x)F_Y(y)$.

3 שונות משותפת – *covariance* (חומר קריאה ב4)

שונות משותפת היא מדד לקורלציה בין מ"מ. הגדרה: $Cov(X,Y)=E\big((X-\mu_X)(Y-\mu_Y)\big)$. תכונות:

1) עבור קבועים a,b,c,d: $Cov(aX+b,cY+d)=ac\cdot Cov(X,Y)$
2) $Cov(W+X,Y)=Cov(W,Y)+Cov(X,Y)$
3) $Cov(X,X)=Var(X)$
4) $Cov(X,Y)=E(XY)-\mu_X\mu_Y$
5) $Var(X+Y)=Var(X)+Var(Y)+2Cov(X,Y)$
6) אם X,Y בת"ל אז $Cov(X,Y)=0$. **זה לא בהכרח נכון בכיוון ההפוך.**

נשים לב שהגדרה 4 דומה להגדרת שונות. ואכן, אם $X=Y$ אז:

$$Cov(X,X)=E(XX)-E(X)E(X)=E(X^2)-E(X)^2=Var(X)$$

נגזר מתכונה 6: אם X,Y בת"ל אז $Var(X+Y)=Var(X)+Var(Y)$.

סכומים ואינטגרלים לחישוב שונות משותפת

מכיוון ששונות משותפת מוגדרת לפי כתוחלת, אפשר לחשב אותה בתור סכום או אינטגרל (כמו כל תוחלת):

המקרה הבדיד: אם ל-X,Y יש PMF משותפת $p(x_i,y_j)$ אז:

$$Cov(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)(x_i-\mu_X)(y_j-\mu_Y)=\left(\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)x_iy_j\right)-\mu_X\mu_Y$$

המקרה הרציף: אם ל-X,Y יש PDF משותפת $f(x,y)$ על $[a,b]\times[c,d]$, אז:

$$Cov(X,Y)=\int_c^d\int_a^b (x-\mu_X)(y-\mu_Y)f(x,y)\,dx\,dy=\left(\int_c^d\int_a^b xyf(x,y)\,dx\ \,dy\right)-\mu_X\mu_Y$$

דוגמה 1

נטיל מטבע הוגן 3 פעמים. יהי X מספר העץ בפעמיים הראשונות ויהי Y מספר העץ בפעמיים האחרונות. נחשב את $Cov(X,Y)$.

דרך ראשונה – עם טבלת ההסתברויות.

$X\backslash Y$	0	1	2	$p(x_i)$
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
$p(y_j)$	1/4	1/2	1/4	1

$$\Omega=\{TTT,TTH,THT,THH,HTT,HTH,HHT,HHH\}$$

מהשוליים, נחשב את התוחלת: $E(X)=1=E(Y)$.

מהגדרת שונות משותפת:

$$Cov(X,Y)=E\big((X-\mu_X)(Y-\mu_Y)\big)=\sum_{i,j}p(x_i,y_j)(x_i-1)(y_j-1)$$

נחשב ונקבל:

$$Cov(X,Y)=\frac{1}{8}(0-1)(0-1)+\frac{1}{8}(2-1)(2-1)=\frac{1}{4}$$

או, לפי תכונה 4:

$$E(XY)=1\cdot\frac{2}{8}+2\cdot\frac{1}{8}+4\cdot\frac{1}{8}=\frac{5}{4}$$

$$Cov(X,Y)=E(XY)-\mu_X\mu_Y=\frac{5}{4}-1=\frac{1}{4}$$

דרך שנייה – לפי תכונות שונות משותפת.

אם X_i הוא תוצאת ההטלה ה-i, אז $X=X_1+X_2,\ \ Y=X_2+X_3$. אנחנו יודעים ש $E(X_i)=0.5, Var(X_i)=0.25$. אז מתכונה 2 נקבל:

$$Cov(X,Y)=Cov(X_1+X_2,X_2+X_3)=Cov(X_1,X_2)+Cov(X_1,X_3)+Cov(X_2,X_2)+Cov(X_2,X_3)$$

הטלות שונות הן בת"ל, אז:

$$Cov(X_1,X_2)=Cov(X_1,X_3)=Cov(X_2,X_3)=0$$

כלומר:

$$Cov(X,Y)=Cov(X_2,X_2)=Var(X_2)=1/4$$

דוגמה 2 – שונות משותפת 0 לא בהכרח גוררת אי תלות.

$Y\backslash X$	-2	-1	0	1	2	$p(y_j)$
0	0	0	1/5	0	0	1/5
1	0	1/5	0	1/5	0	2/5
4	1/5	0	0	0	1/5	2/5
$p(x_i)$	1/5	1/5	1/5	1/5	1/5	1

יהי X מ"מ שמקבל ערכים $-2,-1,\ 0,\ 1,\ 2$ בהתפלגות אחידה. יהי $Y=X^2$:

נחשב את התוחלת: $E(X)=0,\ \ E(Y)=2$. השונות המשותפת:

$Cov(X,Y)=\frac{1}{5}(-8-1+1+8)-0=0$, אבל יש תלות:

$$P(X=-2,Y=0)=0,\qquad P(X=-2)P(Y=0)=1/25$$

גם אינטואיטיבית ברור ש X,X^2 הם תלויים. הנקודה היא ששונות משותפת מודדת את היחס הלינארי בין מ"מ. היחס בין X,X^2 הוא ריבועי, והשונות המשותפת לא עולה עליו.

4 קורלציה

שונות משותפת מוגדרת לפי "יחידות של X כפול יחידות של Y". לכן קשה להשוות בין שונויות משותפות – השונות משתנה לפי הסקאלה. **קורלציה** היא דרך להוריד את הסקאלה מהשונות המשותפת. נגדיר את **מקדם המתאם** בין X ל-Y:

$$Cor(X,Y)=\rho=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

תכונות של קורלציה

1) מקדם המתאם ρ הוא השונות המשותפת של X,Y אחרי סטנדרטיזציה.
2) הוא חסר מימד.
3) $-1\le\rho\le 1$
   a. $\rho=+1$ אמ"מ $Y=aX+b$, עבור $a>0$
   b. $\rho=-1$ אמ"מ $Y=aX+b$, עבור $a<0$

בהמשך לדוגמה 2, נחשב את הקורלציה:

$$Cor(X,Y)=\rho=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=\frac{1/4}{1/2}=\frac{1}{2}$$

5 התפלגות רב-נורמלית

להתפלגות רב-נורמלית יש את פונקציית הצפיפות (הנוראית, יש לציין):

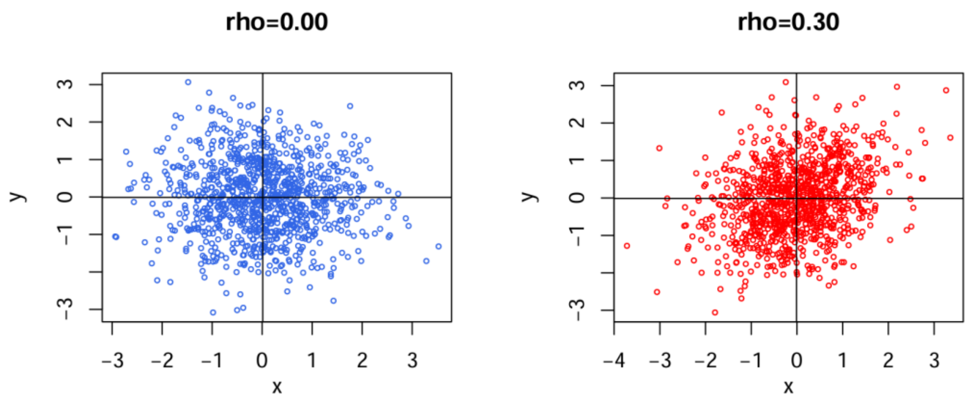
$$f(x,y)=\frac{e^{\left(\frac{-1}{2(1-\rho^2)}\right)\left[\frac{(x-\mu_X)^2}{\sigma_X^2}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}-\frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

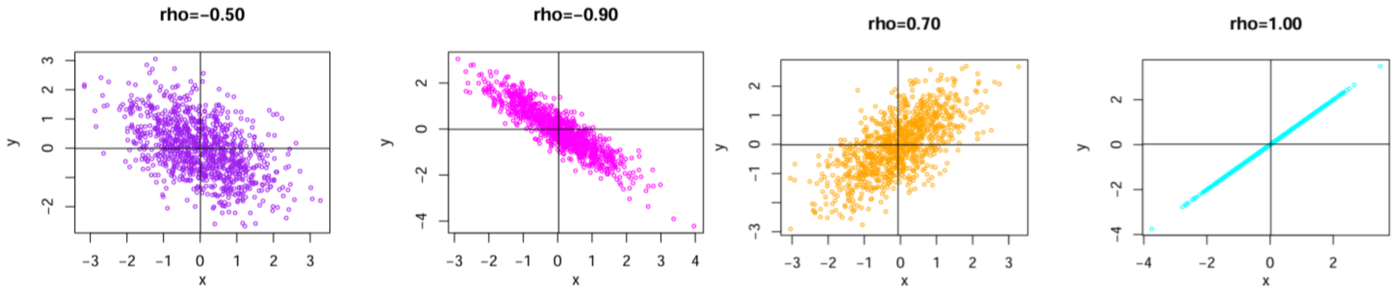
איכס.

בהתפלגות הזאת, ההתפלגויות השוליות של X, Y הן נורמליות, והקורלציה ביניהם היא ρ.

ניתן כמה דוגמאות עבור ערכין שונים של ρ. בנפרד, X ו-Y הם נורמלים סטנדרטים (כלומר $\sigma_X=\sigma_Y=1,\mu_X=\mu_Y=0$).

הגרפים ממחישים את החשיבות של הקורלציה. נשרטט קווים ישרים מקבילים לצירים דרך התוחלות. אם המתאם חיובי, רוב הנקודות יהיו ברבעים 1, 3. אם המתאם שלילי, הרוב יהיו ברובעים 2, 4. וככל ש ρ מתקרב ל ±1, הנקודות מסתדרות בקו ישר:





6 מבוא לסטטיסטיקה (חומר קריאה 4ג)

המטרה של סטטיסטיקה – להסיק דברים בהתבסס על דאטא. שלושה שלבים: איסוף דאטא, תיאור דאטא, ניתוח דאטא. זה מתאים לפרדיגמת השיטה המדעית – ננסח השערה לגבי האמת, נאסוף דאטא בניסויים, נתאר את הממצאים, ונסיק מהממצאים את חוזק הראיות לגבי ההשערה שלנו. חשוב לתכנן את הניסוי בצורה נכונה. "הכנסת זבל – קיבלת זבל". ניסוי לא טוב ייתן דאטא גרוע, שממנה אי אפשר להסיק. הדאטא עצמה בדרך כלל תגיע במנה רשימה, מערך או טבלה ענקיים. כדי להבין אותה, אפשר לחשב סטטיסטיקות סיכום כמו הממוצע, חציון, וטווחים וטווח בין-רבעוני. אפשר גם להציג את הדאטא בגרפים.

בפועל אנחנו רוצים להסיק מתוך הסטטיסטיקה משהו לגבי העולם. לעיים קרובות זה על ידי הגדרת מודל סטטיסטי עבור התהליך שממנו הדאטא הגיעה. לדוגמה, נניח שהדאטא הגיעה בצורת סדרת מדידות שאנחנו חושבים שהטעות שלה מתפלגת נורמלית (זה חייב להיות בערך כי הטעות כן סופית, והתחום של התפלגות נורמלית הוא אינסופי). נוכל להשתמש בדאטא כדי להוכיח או להפריך את ההשערה. בקורס הזה אנחנו נתמקד בשימוש בדאטא כדי להסיק דברים על הפרמטרים של המודל. ההשערה תמיד תהיה הסתברותית.

7 נראות מקסימלית (חומר קריאה 4ד)

חישוב ההסתברות של פרמטרים בהינתן דאטא. הנראות המקסימלית – *Maximum Likelihood Estimate*. איזה ערך של הפרמטר נותן את ההסתברות הכי גדולה לדאטא?

8 הסקה בייסיאנית (חומר קריאה 4ה)

נציג גרסה טבלאית לנוסחת בייס, בעזרת דוגמה: נניח שיש 3 סוגי מטבעות, כל אחד עם הסתברות שונה לעץ. סוג A הוגן, סוג B עם הסתברות 0.6, סוג C עם הסתברות 0.9. נניח שיש קופסה עם 5 מטבעות: 2 מסוג A, 2 מסוג B, ואחת מסוג C. נניח שבחרנו מטבע אחד באופן מקרי ואחיד, הטלנו וקיבלנו עץ. מה ההסתברות לסוג?

קצת טרמינולוגיה:

ניסוי: נבחר מטבע באופן מקרי ואחיד, נטיל אותו ונראה מה יצא.

דאטא: תוצאת הניסוי, במקרה שלנו ההטלה.

השערה: יש לנו 3 אפשרויות: A,B,C.

הסתברות פריורית (קודמת): ההסתברות לכל השערה, לפני הניסוי. במקרה שלנו:

$$P(A) = 0.4, \qquad P(B) = 0.4, \qquad P(C) = 0.2$$

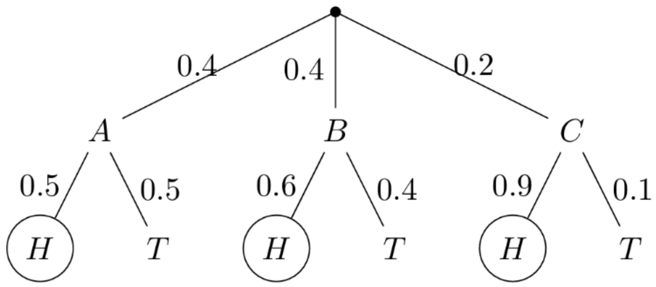
נראות (סיכוי): פונקציית הנראות (*likelihood*) היא $\mathbb{P}(\mathcal{D}|\mathcal{H})$ כלומר, הסבירות של הדאטא בהינתן השערה. במקרה שלנו:

$$P(D|A) = 0.5, \qquad P(D|B) = 0.6, \qquad P(D|C) = 0.9$$

הסתברות פוסטריורית (אוחרת): ההסתברות של כל השערה בהינתן הדאטא:

$$P(A|D), \qquad P(B|D), \qquad P(C|D)$$

נחשב את ההסתברות האוחרת. נשרטט את עץ ההסתברות:



אנחנו יודעים את $P(H)$, והעץ נותן לנו את $P(D|H)$. נחשב את $P(D)$ באמצעות נוסחת ההסתברות השלמה:

$$\begin{aligned} P(D) &= P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) \\ &= 0.5 \cdot 0.4 + 0.6 \cdot 0.4 + 0.9 \cdot 0.2 \\ &= 0.62 \end{aligned}$$

נחשב את ההסתברות האוחרת בעזרת נוסחת בייס:

$$P(A|D) = \frac{P(D|A)P(A)}{P(D)} = \frac{0.5 \cdot 0.4}{0.62} = \frac{0.20}{0.62}$$

$$P(B|D) = \frac{P(D|B)P(B)}{P(D)} = \frac{0.6 \cdot 0.4}{0.62} = \frac{0.24}{0.62}$$

$$P(C|D) = \frac{P(D|C)P(C)}{P(D)} = \frac{0.9 \cdot 0.2}{0.62} = \frac{0.18}{0.62}$$

נשים לב שההסתברות $P(D)$ זהה בכל אחד מהמכנים, והוא סכום המונים. נסדר את זה **בטבלת הסקה בייסיאנית**:

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
A	0.4	0.5	0.2	0.3226
B	0.4	0.6	0.24	0.3871
C	0.2	0.9	0.18	0.2903
total	1		0.62	1

כדי לחשב את העמודה האחרונה, בעצם נירמלנו את העמודה של המונה כדי שהסכום יהיה 1. כדי למצוא את ההשערה הכי סבירה, מספיק לחשב את העמודה של המונה.

סכום עמודת ה *likelihood* לא שווה 1 – זו לא פונקציית הסתברות. וזה לא מספיק כדי לדעת באמת – במקרה שלנו, למטבע C יש את ה *likelihood* הכי גדול אבל זו לא ההשערה הכי סבירה בפועל.

פונקציות צפיפות פריוריות ופוסטריוריות

- θ יהיה ערך ההשערה.
- $p(\theta)$ היא ה*PMF* הקודם של ההשערה.
- $p(\theta|D)$ היא ה*PMF* האוחר של ההשערה, בהינתן הדאטא.
- $p(D|\theta)$ היא פונקציית הנראות (לא פונקציית הסתברות!).

עדכונים מרובים

נניח שבחרנו מטבע, הטלנו וקיבלנו עץ. הטלנו שוב וקיבלו שוב עץ. מה ההסתברות עכשיו לכל סוג?

hypothesis	prior	likelihood 1	Bayes numerator 1	likelihood 2	Bayes numerator 2	posterior 2
θ	$p(\theta)$	$p(x_1=1\|\theta)$	$p(x_1=1\|\theta)p(\theta)$	$p(x_2=1\|\theta)$	$p(x_2=1\|\theta)p(x_1=1\|\theta)p(\theta)$	$p(\theta\|x_1=1,x_2=1)$
0.5	0.4	0.5	0.2	0.5	0.1	0.2463
0.6	0.4	0.6	0.24	0.6	0.144	0.3547
0.9	0.2	0.9	0.18	0.9	0.162	0.3990
total	1				0.406	1

הרצאה [מצגת שיעור 4א](), [מצגת שיעור 4ב]()

תזכורות:

חוק המספרים הגדולים – אם יש n דגימות (מ"מ) בת"ל בעלי אותה תוחלת ושונות סופית, אז הממוצע האמפירי שלהם בהסתברות גבוהה (עבור n גדול) יהיה מרוכז סביב התוחלת שלהם.

משפט הגבול המרכזי – אם יש n דגימות (מ"מ) בת"ל בעלי אותה תוחלת ושונות סופית, אז הממוצע האמפירי שלהם מתפלג קרוב לנורמלי ככל ש-n גדל.

זה שימושי להסקה סטטיסטית. כי בניסוי אמיתי, גם אם כל המ"מ באותה התפלגות, אנחנו לא יודעים מה ההתפלגות של כל משתנה אבל אנחנו כן יודעים מה הממוצע. ואם יש מספיק דאטא, אז הממוצע מתפלג קרוב לנורמלי. עדיין צריך לדעת את התוחלת והשונות, אבל זה יותר קל לשערך.

חוק בייס

$$\mathbb{P}(\mathcal{H}|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|\mathcal{H})\mathbb{P}(\mathcal{H})}{\mathbb{P}(\mathcal{D})}, \qquad \mathbb{P}(\text{hypothesis}|\text{data}) = \frac{\mathbb{P}(\text{data}|\text{hypothesis})\mathbb{P}(\text{hypothesis})}{\mathbb{P}(\text{data})}$$

1 הערכת פרמטר

דוגמה 1

נרצה לגלות מה אחוז האנשים p שעבורם יש לכוסברה טעם של סבון[1].

ניסוי: נבקש מ-n אנשים לטעום כוסברה.

מודל: המשתנה המקרי $X_i \sim Ber(p)$ הוא האינדיקטור למאורע "האדם ה-i אומר שיש לכוסברה טעם של סבון".

דאטא: $x_1, \dots, x_n$ הן התוצאות.

הסקה: נסיק מהו p מתוך הדאטא.

לדוגמה, שאלנו 100 אנשים ו-55 מהם אמרו סבון. נגדיר $\bar{X} = \sum_{i=1}^{n} X_i$ סכום האינדיקטורים. עבור p נתון, ההסתברות לקבל 55 היא ההתפלגות הבינומית:

$$P(\bar{X} = 55 \mid p) = \binom{100}{55} p^{55}(1-p)^{45}$$

נגדיר: ה"סיכוי" – *likelihood*:

$$\text{the likelihood: } P(\text{data} \mid p) = \binom{100}{55} p^{55}(1-p)^{45}$$

הסיכוי המקסימלי – *Maximum Likelihood Estimate (MLE)*

דרך לשערך את הפרמטר שמעניין אותנו. הMLE הוא הערך של p שממקסם את ה-*likelihood*. יש כמה שיטות:

1) חדו"א: כדי למצוא את הMLE, נגזור את נוסחת הסיכוי ונשווה לאפס:

$$\frac{d}{dp} P(\text{data} \mid p) = 0$$

וצריך לוודא שנקודת הקיצון הזו היא אכן נקודת מקסימום.

2) לפעמים הנגזרת אף פעם לא מתאפסת, וה-MLE הוא בקצה של תחום.

---

[1] מכון דוידסון - למה יש אנשים ששונאים כוסברה?

3) אם הפרמטר מקבל מספר סופי של ערכים, אפשר לחשב את הסיכוי לכל אחד ולקחת את הגדול.

בדוגמה שלנו, נגזור ונשווה לאפס:

$$\frac{d}{dp}P(\text{data} \mid p) = \binom{100}{55}\left(55p^{54}(1-p)^{45} - 45p^{55}(1-p)^{44}\right) = 0$$

$$55p^{54}(1-p)^{45} = 45p^{55}(1-p)^{44}$$

$$55(1-p) = 45p$$

$$55 = 100p$$

כלומר, ה-MLE הוא $\hat{p} = 55/100$.

באופן כללי, לפעמים יותר נוח לעבוד עם לוג. בגלל שלוג היא פונקציה מונוטונית עולה, נוכל לקחת לוג של שני הצדדים:

$$\text{log likelihood} = \ln(\text{likelihood}) = \ln\big(P(\text{data} \mid p)\big)$$

אז בדוגמה שלנו:

$$\ln\big(P(\text{data} \mid p)\big) = \ln\left(\binom{100}{55}p^{55}(1-p)^{45}\right) = \ln\binom{100}{55} + 55\ln(p) + 45\ln(1-p)$$

נגזור ונשווה לאפס:

$$0 = \frac{d}{dp}\ln\binom{100}{55} + 55\ln(p) + 45\ln(1-p) = \frac{55}{p} - \frac{45}{1-p}$$

$$\frac{45}{1-p} = \frac{55}{p} \implies 45p = 55 - 55p \implies 100p = 55 \implies p = \frac{55}{100}$$

תרגיל כיתה: מטבעות

יש קופסה עם 3 מטבעות, שלכל אחד יש הטיה לעץ: $2/3$, $1/2$, $1/3$. מוציאים מטבע אחד באופן מקרי ואחיד ומטילים 80 פעמים. נניח שקיבלנו 49 עץ.

א) מה הסיכוי של התוצאה הזאת עבור כל אחד מהמטבעות? איזה מטבע נותן את הסיכוי המקסימלי?

הדאטא D היא 49 עץ מתוך 80 הטלות. פונקציית הסיכוי $P(D|p)$ מקבלת 3 ערכים:

$$P\left(D|\frac{1}{3}\right) = \binom{80}{49}\left(\frac{1}{3}\right)^{49}\left(\frac{2}{3}\right)^{31} = 6.24 \cdot 10^{7}$$

$$P\left(D|\frac{1}{2}\right) = \binom{80}{49}\left(\frac{1}{2}\right)^{49}\left(\frac{2}{2}\right)^{31} = 0.024$$

$$P\left(D|\frac{2}{3}\right) = \binom{80}{49}\left(\frac{2}{3}\right)^{49}\left(\frac{1}{3}\right)^{31} = 0.082$$

המטבע שנותן את הסיכוי המקסימלי הוא השלישי, אז ננחש שזה המטבע שנבחר.

ב) נניח שיש מטבע יחיד עם p לא ידוע. נמצא את פונקציות הסיכוי עבור אותה הדאטא. מה הMLE עבור p?

p יכול להיות כל ערך בין 0 ל-1. כלומר הפונקציה:

$$P(D|p) = \binom{80}{49}p^{49}(1-p)^{31}$$

נוציא לוג, נגזור ונשווה לאפס:

$$0 = \frac{d}{dp}\ln\big(P(D|p)\big) = \frac{d}{dp}\left(\ln\binom{80}{49} + 49\ln p + 31\ln(1-p)\right)$$

$$\frac{49}{p} = \frac{31}{1-p} \implies 49 - 49p = 31p \implies 49 = 80p \implies p = \frac{49}{80}$$

תרגיל כיתה: נורות

עבור מ"מ רציף, נשתמש בPDF במקום הPMF:

נניח שזמן החיים של כל נורה מתפלג $Exp(\lambda)$. נבדוק 5 נורות: $x_1, \dots, x_5$.

נמצא את פונקציית הסיכוי: בהנחה שכל המשתנים המקריים בת"ל, הPDF המשותפת היא:

$$f(x_1, x_2, x_3, x_4, x_5 \mid \lambda) = \lambda^5 e^{-\lambda(x_1+x_2+x_3+x_4+x_5)}$$

נוציא לוג:

$$\ln\big(f(x_1, x_2, x_3, x_4, x_5 \mid \lambda)\big) = 5\ln\lambda - \lambda(x_1 + x_2 + x_3 + x_4 + x_5)\ln e$$

נגזור ונשווה לאפס:

$$0 = \frac{d}{d\lambda}\big(5\ln\lambda - \lambda(x_1 + x_2 + x_3 + x_4 + x_5)\big) = \frac{5}{\lambda} - (x_1 + x_2 + x_3 + x_4 + x_5)$$

$$\lambda = \frac{5}{x_1 + x_2 + x_3 + x_4 + x_5}$$

נניח שהזמנים שנמדדו הם: $x_1 = 2,\ x_2 = 3,\ x_3 = 1,\ x_4 = 3,\ x_5 = 4$. מה ה-$MLE$ של λ?

נציב את הערכים בפונקציה: $\lambda = 5/13$.

2 עדכון בייסיאני – פריורים בדידים (מצגת 4ב)

דוגמה: תרופות

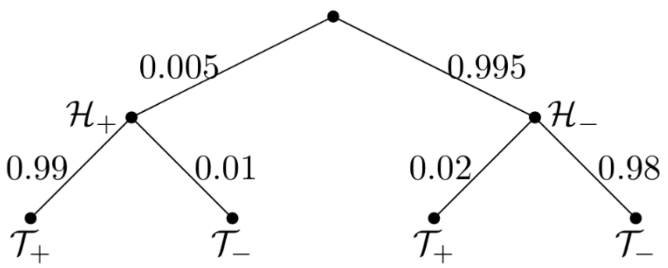
נתון: בניסוי, תרופה א' ריפאה 100% מהחולים, תרופה ב' 95%, תרופה ג' 90%. איזו נבחר?

נתון 2: באותו ניסוי, תרופה א' ריפאה 3 מתוך 3 מהחולים, תרופה ב' 19 מתוך 20, תרופה ג' 90,000 מתוך 100,000.

תרגיל כיתה: למידה מתוך הדאטא

למחלה מסויימת יש הימצאות (*prevalence*) של 0.005. לבדיקה יש 2% *false positive* ו-1% *false negative*.
נניח שאדם נבדק ויצא חיובי.

נייצג את הנתונים ע"י עץ: נגדיר $H_\pm$ את המאורע שהנבדק חולה או לא, $T_\pm$ את המאורע שהבדיקה חיובית או שלילית.



נוסחת ההסתברות השלמה:

$$\begin{aligned} P(T_+) &= P(T_+|H_+)P(H_+) + P(T_+|H_-)P(H_-) \\ &= 0.99 \cdot 0.005 + 0.02 \cdot 0.995 \\ &= 0.02485 \end{aligned}$$

לפי נוסחת בייס:

$$P(H_+|T_+) = \frac{P(T_+|H_+)P(H_+)}{P(T_+)} = \frac{0.99 \cdot 0.005}{0.02485} = 0.199$$

$$P(H_-|T_+) = \frac{P(T_+|H_-)P(H_-)}{P(T_+)} = \frac{0.02 \cdot 0.995}{0.02485} = 0.801$$

הבדיקה הגדילה בהרבה את ההסתברות שהנבדק חולה, אבל עדיין יותר סביר שהוא בריא.

הדאטא היא תוצאות הבדיקה, **ההשערות** הן האפשרויות H_+, H_-. ה ***likelihood*** הן ההסתברויות שהנבדק חולה בהינתן בדיקה:

$$P(T_+|H_+) = 0.99, \qquad P(H_+|T_+) = 0.02$$

נדגיש: *likelihood* היא ההסתברות **בהינתן** השערה, לא ההסתברות של ההשערה.

ההסתברויות הפריוריות של ההשערות הן ההסתברות ה"כללית" לכך שאדם חולה (בלי בדיקה).

$$P(H_+) = 0.005, \qquad P(H_-) = 0.995$$

ההסתברויות הפוסטריוריות של ההשערות הן ההסתברויות המעודכנות אחרי הבדיקה.

$$P(H_+|T_+) = 0.199, \qquad P(H_-|T_+) = 0.801$$

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{T}_+\|\mathcal{H})$	$P(\mathcal{T}_+\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{T}_+)$
$\mathcal{H}_+$	0.005	0.99	0.00495	0.199
$\mathcal{H}_-$	0.995	0.02	0.0199	0.801
total	1	NO SUM	$P(\mathcal{T}_+) = 0.02485$	1

$$\mathbb{P}(\mathcal{H}|\mathcal{D}) = \frac{\mathbb{P}(\mathcal{D}|\mathcal{H})\mathbb{P}(\mathcal{H})}{\mathbb{P}(\mathcal{D})}$$

$$\mathbb{P}(\text{hypothesis}|\text{data}) = \frac{\mathbb{P}(\text{data}|\text{hypothesis})\mathbb{P}(\text{hypothesis})}{\mathbb{P}(\text{data})}$$

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{total probability of data}}$$

תרגול

תרגיל 1

נתונים מ"מ $X_1, \dots, X_n \sim U(0,\theta)$ בת"ל. נרצה להעריך את θ.

$$\mathcal{L}(x_1, \dots, x_n|\theta) = P(X_1 = x_1, \dots, X_n = x_n|\theta) = \prod_{i=1}^{n} P(X_i = x_i|\theta) = \frac{1}{\theta^n}$$

θ חייב להיות גדול מ- $\max\{x_i\}$, כי אם ראינו מספר אז זה אומר שהוא היה בטווח.

תרגיל 2

נתון n_i – כמות האנשים בעלי i דירות. יש סה"כ m אנשים. נתון: $n = (n_0, n_1, n_2, n_3, 0 \dots)$ כלומר אין אנשים עם יותר מ-3 דירות. נניח שעבור אדם j, מתקיים:

$$P(X_j = i) \sim Poi(\lambda) = \frac{\exp(-\lambda)\,\lambda^i}{i!}$$

נתון מ"מ $N_i = \sum_{j=1}^{m} P(X_j = i)$ שסופר את מספר האנשים בעלי i דירות. כלומר $N = (N_0, N_1, N_2, N_3, 0, \dots)$.

נרצה להעריך את λ:

$$\begin{aligned}
\mathcal{L}(n_0, n_1, n_2, n_3|\lambda) &= P(N_0 = n_0, N_1 = n_1, N_2 = n_2, N_3 = n_3|\lambda) \\
&= \binom{m}{n_0}\left(\frac{e^{-\lambda}\lambda^0}{0!}\right)^{n_0}\binom{m-n_0}{n_1}\left(\frac{e^{-\lambda}\lambda^1}{1!}\right)^{n_1} \\
&\cdot \binom{m-n_1-n_0}{n_2}\left(\frac{e^{-\lambda}\lambda^2}{2!}\right)^{n_2}\binom{m-n_2-n_1-n_0}{n_3}\left(\frac{e^{-\lambda}\lambda^3}{3!}\right)^{n_3} \\
&= \binom{m}{n_0}\binom{m-n_0}{n_1}\binom{m-n_1-n_0}{n_2}\binom{m-n_2-n_1-n_0}{n_3}\prod_{i=0}^{3}\left(\frac{e^{-\lambda}\lambda^i}{i!}\right)^{n_i} \\
&= \frac{m!}{(m-n_0)!\,n_0!}\cdot\frac{(m-n_0)!}{(m-n_0-n_1)!\,n_1!}\cdot\frac{(m-n_1-n_0)!}{(m-n_1-n_0-n_2)!\,n_2!} \\
&\cdot \frac{(m-n_1-n_0-n_2)!}{(m-n_1-n_0-n_2-n_3)!\,n_3!}\cdot\prod_{i=0}^{3}\left(\frac{e^{-\lambda}\lambda^i}{i!}\right)^{n_i}
\end{aligned}$$

נשים לב שהמונים והמכנים מצטמצמים:

$$= \frac{m!}{n_0!\,n_1!\,n_2!\,n_3!}\prod_{i=0}^{3}\left(\frac{e^{-\lambda}\lambda^i}{i!}\right)^{n_i}$$

נגזור ונשווה ל-0 את ה-ln של הביטויים:

$$\begin{aligned}
\frac{d}{d\lambda}\ln\left(\frac{m!}{n_0!\,n_1!\,n_2!\,n_3!}\prod_{i=0}^{3}\left(\frac{e^{-\lambda}\lambda^i}{i!}\right)^{n_i}\right) &= \frac{d}{d\lambda}\ln\left(\frac{m!}{n_0!\,n_1!\,n_2!\,n_3!}\right) + \ln\left(\prod_{i=0}^{3}\left(\frac{e^{-\lambda}\lambda^i}{i!}\right)^{n_i}\right) \\
&= \frac{d}{d\lambda}\sum_{i=0}^{3} n_i\cdot\ln\left(\frac{e^{-\lambda}\lambda^i}{i!}\right) = \frac{d}{d\lambda}\sum_{i=0}^{3} n_i\cdot\left(\ln(e^{-\lambda}) + \ln(\lambda^i) - \ln(i!)\right) \\
&= \frac{d}{d\lambda}\sum_{i=0}^{3} n_i(-\lambda) + n_i\cdot i\cdot\ln(\lambda) - n_i\ln(i!) = \sum_{i=0}^{3} -n_i + \frac{n_i\cdot i}{\lambda} =
\end{aligned}$$

$$= -n_0 + \left(-n_1 + \frac{n_1}{\lambda}\right) + \left(-n_2 + \frac{2n_2}{\lambda}\right) + \left(-n_3 + \frac{3n_3}{\lambda}\right) = \frac{n_1 + 2n_2 + 3n_3}{\lambda} - n_0 - n_1 - n_2 - n_3 = 0$$

$$n_1 + 2n_2 + 3n_3 = \lambda(n_0 + n_1 + n_2 + n_3)$$

$$\lambda = \frac{n_1 + 2n_2 + 3n_3}{n_0 + n_1 + n_2 + n_3}$$

זאת נקודת המקסימום. כלומר, $\mathcal{L}(n_0, n_1, n_2, n_3|\lambda)$ (הסיכוי לקבל n_0, n_1, n_2, n_3 בהינתן λ) הוא מקסימלי עבור ה-λ הזה.