25/12/19

[redacted] הסתברות:

שוויון וולד:

הכרחנו. אם $X_1, X_2, \dots$ מ"מ עם אותה תוחלת, N מ"מ שמקבל את הערכים $0,1,2,\dots$ ($N \le$ (מתנה)) וכולם ב"ת.

אינדקס עליון מ"מ מקרי

$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]\cdot E(x)$$

דוגמא:

חנות נעליים של רונית. נכנסים ביום N אנשים כאשר $N \sim Poi(100)$. כל אחד מהם מוציא בחנות בממוצע 200 ש"ח.
מה הפדיון היומי של החנות? (מ"מ)
מהו הפדיון הממוצע? (מספר)

← נכנסים N אנשים לחנות. האיש ה-i מוציא X_i ש"ח בחנות.
הפדיון הוא סכום כל ה-X_i של כל האנשים שנכנסו. הפדיון: $\sum_{i=1}^{N} X_i$
הפדיון הממוצע הוא התוחלת של הפדיון.

שוויון וולד

$$\Rightarrow E\left[\sum X_i\right] = E[N]\cdot E[x] \Rightarrow 100\cdot 200 = 20{,}000$$

שונות מותנית:

השונות של X בהינתן $y=k$

$$V[x|y=k] = E[x^2|y=k] - (E[x|y=k])^2$$

כמו שונות רגילה אבל כל התוחלות בהינתן $y=k$.

זהו מס' קבוע בדומה ל-$E[x|y=k]$ שהם מס' קבוע.

בדומה ל-$E(x|y)$ שזהו מ"מ אשר מקבל את הערך $E[x|y=k]$, כאשר $y=k$ (פונקציה של y).

פונקציה של y מ"מ.

⇐ מגדירים: $E[x|y] = E[x^2|y] - (E[x|y])^2$.

מה הקשר בין $V[x|y]$ לבין $V(x)$?

⊛ תזכורת: $E[E(x|y)] = E[x]$ (ונחפש קשר דומה בין $V[x]$ לבין $V[x|y]$).

נחשב $\Rightarrow E[V(x|y)] = E\left[E[x^2|y] - (E[x|y])^2\right] = E[E[x^2|y]] - E\left[(E[x|y])^2\right] =$

$E[g(y)]$ $\quad E[h(y) - F(y)]$ $\quad E[h(y)] - E[F(y)]$

חוק התוחלת השלמה $= \underbrace{E[x^2] - E\left[(E[x|y])^2\right]}_{⊛}$

מצטמצמים בהרכבה.

חוק התוחלת השלמה.

השונות של המ"מ $Z = E[X|Y]$

נחשב $\Rightarrow V[E[x|y]] = E\left[(E[x|y])^2\right] - (E[E[x|y]])^2 = \underbrace{E\left[(E[x|y])^2\right] - (E[x])^2}_{⊛⊛}$

$V[Z] = E[Z^2] - (E[Z])^2$

נחבר ⊛+⊛⊛ : $E[V[x|y]] + V[E[x|y]] = E[x^2] - (E[x])^2 = V[x]$. <= נוסחת פירוק השונות.

$E[(E[X|Y])^2]$ הצטמצם

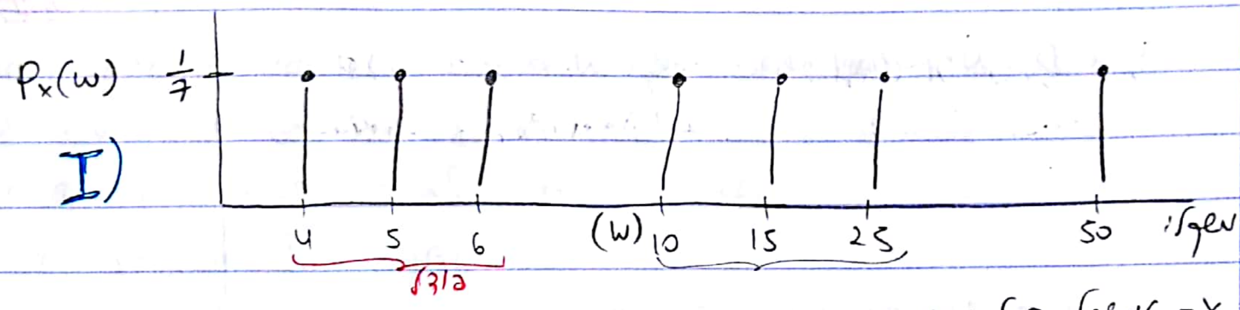
(זה הבסיס ל-ANOVA)

דוגמא!

עמית הולך לכלבייה בק"ש ומאמץ כלב באקראי. בכלבייה יש 7 כלבים.

שם	סוג	משקל
סימבה	פודל	4
טימון	פודל	5
פומבה	פודל	6
סקאר	פיטבול	10
שנזי	פיטבול	15
נלה	פיטבול	25
מופסה	דני ענק	50

I)



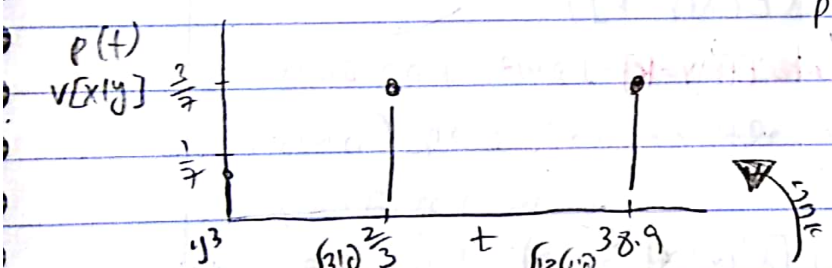
X - משקל הכלב שנבחר

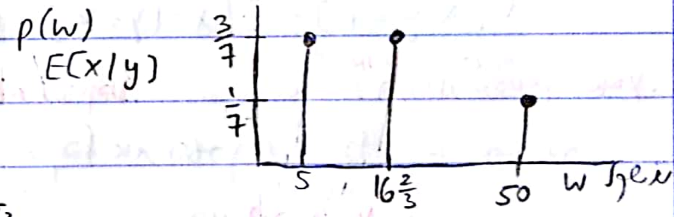
Y - סוג הכלב

המ"מ $E[X|Y]$ הוא מ"מ המקבל 3 ערכים:

- $E[X|y=\text{פודל}]=4\cdot\frac{1}{3}+5\cdot\frac{1}{3}+6\cdot\frac{1}{3}=5$
- $E[X|y=\text{פיטבול}]=10\cdot\frac{1}{3}+15\cdot\frac{1}{3}+25\cdot\frac{1}{3}=16\frac{2}{3}$
- $E[X|y=\text{דני}]=50\cdot 1=50$

$E[X|Y]$ הוא מ"מ המקבל את 3 הערכים הנ"ל וההסתברות שהוא מקבל $E(X|y=\text{פודל})$ היא $P(y=\text{פודל})$.





$V[X|Y]$ כנ"ל מ"מ המקבל 3 ערכים:

- $V[X|y=\text{פודל}]=E[X^2|y=\text{פודל}]-(E[X|y=\text{פודל}])^2=4^2\cdot\frac{1}{3}+5^2\cdot\frac{1}{3}+6^2\cdot\frac{1}{3}-5^2=\frac{2}{3}$
- $V[X|y=\text{פיטבול}]=10^2\cdot\frac{1}{3}+15^2\cdot\frac{1}{3}+25\cdot\frac{1}{3}-(16\frac{2}{3})^2=38.9$
- $V[X|y=\text{דני}]=50^2\cdot 1-(50)^2=0$

חשבו את $E[V(X|Y)]$ - התוחלת של המ"מ הנ"ל. זהו ממוצע השונות של הקבוצה {פיטבול לעומת פיטבול; פודל לעומת פודל}

להשוב זה, צריך להוסיף את $V[E[X|Y]]$ (השונות של המ"מ ב-I)

זו השונות בין סוגי כלבים: דני ממוצע מול פודל ממוצע מול פיטבול ממוצע. לפי נוסחאות פירוק השונות, סכום שני החלקים ייתן את השונות של משקל כלב, ללא קשר לסוג, $V[X]$.

הרצאה 13

25/12/17

המשך-

מ"מ רציפים:

מ"מ רציף הוא מ"מ שיכול לקבל "יותר מדי" ערכים - מה לא בן מניה ולכן לסכום $\sum_{k\in X(\Omega)}$ אין משמעות, כי יש יותר מדי איברים בסכום.

למשל: אורך חיי נורה, שעת הגעה של רכבת.

הגדרה:

פונקציית צפיפות של מ"מ X היא פונקציה $f_X(t):\mathbb{R}\to\mathbb{R}$, כך ש: $P(X\in A)=\int_A f_X(t)\,dt$ לכל קבוצה $A\subseteq\mathbb{R}$ מספרים ממשיים.

למשל:

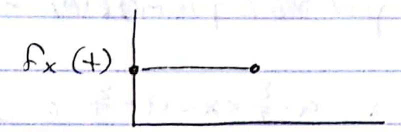
$$P(a\le X\le b)=\int_a^b f_X(t)\,dt$$
$$P(X\le b)=\int_{-\infty}^b f_X(t)\,dt$$
$$P(X\ge a)=\int_a^{\infty} f_X(t)\,dt$$

הצפיפות של מ"מ רציף מחליפה את ההתפלגות של בדיד.

דוגמאות:

X מ"מ אחיד בין 0 ל-1.

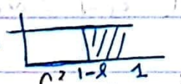
$$f_X(t)=\begin{cases}1 & ; 0\le t\le 1\\ 0 & ;\text{אחרת}\end{cases}$$



← X נקרא אחיד כי כל קטע באורך l בין 0 ל-1 - ההסתברות לקבל מה בתוכו היא l ללא קשר למיקום הקטע:

$$P(0<X<l)=\int_0^l f_X(t)\,dt\overset{l<1}{=}\int_0^l 1\,dt=l \Rightarrow$$

$$P(1-l<X<1)=\int_{1-l}^1 f_X(t)\,dt\overset{l<1}{=}\int_{1-l}^1 1\cdot dt=1-(1-l)=l \Rightarrow$$



$$P\left(\tfrac{3}{4}<X<2\right)=\int_{\frac34}^2 f_X(t)\,dt=\int_{\frac34}^1 f_X(t)\,dt+\int_1^2 f_X(t)\,dt=\int_{\frac34}^1 1\cdot dt+\int_1^2 0\cdot dt=\frac{1}{4}$$

מסקנות:

1. אם $f_X(t)=0$ לכל $t\in A$: $P(X\in A)=\int_A f_X(t)\,dt=\int_A 0\,dt=0$
2. $\overset{P(a\le X\le a)}{P(X=a)}=\int_a^a f_X(t)\,dt=0$. מ"מ רציף לא מקבל ערכים בודדים בהסתברות חיובית.
3. $\int_{-\infty}^{\infty} f_X(t)=P(-\infty<X<\infty)=P(\Omega)=1$
   X מ"מ לכן $X\in\mathbb{R}$

דוגמא:

נתון מ"מ רציף עם הצפיפות
$$f_X(t)=\begin{cases}c(4t-2t^2) & ; 0<t<2\\ 0 & ;\text{אחרת}\end{cases}$$

חשבו $P(X>1)=?$ (קודם כל נמצא את c).

נמצא את c לפי השוויון: $\int_{-\infty}^{\infty} f_X(t)\,dt=1$.

$$1=\int_{-\infty}^{\infty} f_X(t)\,dt=\int_0^2 f_X(t)\,dt+\int_{-\infty}^0 f_X(t)\,dt+\int_2^{\infty} f_X(t)\,dt=\int_0^2 c(4t-2t^2)\,dt+\cancel{\int_{-\infty}^0 0\,dt}+\cancel{\int_2^{\infty} 0\,dt}=$$
$$=c\left(\frac{4t^2}{2}-\frac{2t^3}{3}\right)\Big|_0^2=c\left(8-\frac{16}{3}\right)=c\left(\frac{8}{3}\right)\Rightarrow 1=c\frac{8}{3}\Rightarrow\boxed{c=\frac{3}{8}}$$

(חשב) $\Rightarrow P(X>1)=\int_1^{\infty} f_X(t)\,dt=\int_1^2 f_X(t)\,dt+\int_2^{\infty} f_X(t)\,dt=\int_1^2 \frac{3}{8}(4t-2t^2)\,dt+\int_2^{\infty} 0\,dt=\frac{1}{2}$ → בין 0 ל-1 הגיוני כי צורת הסתברות

פונקציית ההתפלגות המצטברת: (הצטברות)

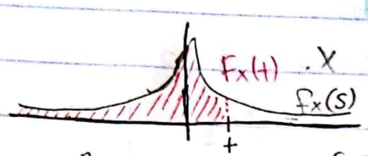
עבור מ"מ רציף X מגדירים פונקציה:

אותה נוסחא כמו הבדיד $F_X(t)=P(X\le t)=\int_{-\infty}^{t} f_X(s)\,ds$

<u>משפט</u>: הקשר בין $F_X(t)$ לבין $f_X(t)$? המצטברת לבין צפיפות.

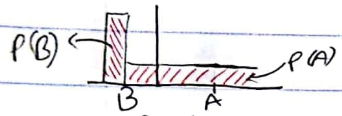
← המשפט היסודי של חדו"א:

$$F_X'(t)=\frac{d}{dt}\int_{-\infty}^{t} f_X(s)\,ds = f_X(t)$$



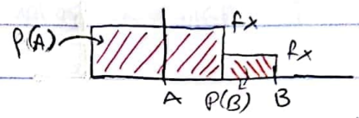
כלומר, הנגזרת של פונקציית המצטברת היא הצפיפות של X. המצטברת היא השטח מתחת ל- f_X עד לנקודה t.

← מה משמעות הצפיפות $f_X(t)$? אם הצפיפות בקבוצה A גדולה מהצפיפות ב- B.



בדוגמא זו $P(B)<P(A)$ למרות שהצפיפות ב-B גבוהה יותר בגלל שאורך הבסיס (אורך B) קטן יותר.

<u>מקרה אחר</u>:



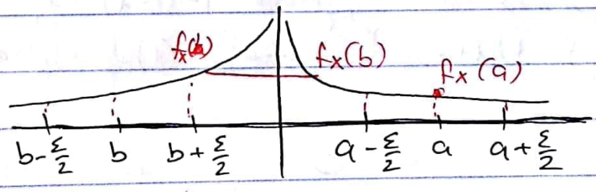
אז $P(A) > P(B)$ (אם אורך A שווה לאורך B).

ההסתברות ש- $a-\frac{\varepsilon}{2}<X<a+\frac{\varepsilon}{2}$, X בקטע באורך ε סביב a.

אם ε קטן אז $f_X(t)\approx f_X(a)$ כי f_X רציפה לרוב.

$$\Rightarrow P\left(a-\frac{\varepsilon}{2}<X<a+\frac{\varepsilon}{2}\right)=\int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_X(t)\,dt = \int_{a-\frac{\varepsilon}{2}}^{a+\frac{\varepsilon}{2}} f_X(a)\,dt = \varepsilon f_X(a)$$

$$\Rightarrow P(X \text{ קרוב ל-} a) \approx \varepsilon\cdot f_X(a)$$



בגלל ש- $f_X(b)>f_X(a)$ כי יותר סיכוי ש-X יקבל קרוב ל- b מאשר קרוב ל- a.