הרצאה 13-ב

18/12/17

המשך משבוע שעבר:

$$(X_1,\dots,X_r)\sim multi(n,(p_1,\dots,p_r))$$

מבצעים n ניסויים בהם יש r תוצאות אפשריות.

p_1 - ההסתברות לתוצאה 1

⋮

p_r - ההסתברות לתוצאה r.

X_i - מס' הניסויים שבהם התקבלה התוצאה i.

רצינו לחשב $cov(X_i,X_j)$. האינטואיציה - כאשר X_i גדול - הרבה ניסויים יצאו לי i ⇐ פחות ניסויים יכולים לצאת על j ⇐ X_j קטן.

השונות המשותפת צריכה להיות שלילית.

הגדרנו מ"מ:

$$I_{lm}=\begin{cases}1 & \text{בניסוי } l \text{ יצאה תוצאה } m\\ 0 & \text{אחרת}\end{cases}$$

לפי הסימונים הנ"ל:

$$X_i=\sum_{l=1}^{n} I_{li}$$ ← מס' הניסויים שבהם התקבלה התוצאה i

$$X_j=\sum_{l=1}^{n} I_{lj}$$

⇐ אם $i=j$ ⇐ $cov(X_i,X_j)=V[X_i]$ (קל לחשוב, שונות מ"מ בינומי).

⇐ אם $i\neq j$ ⇐ $cov(X_i,X_j)=cov\left(\sum_{l=1}^{n} I_{li},\sum_{k=1}^{n} I_{kj}\right)=\sum_{l=1}^{n}\sum_{k=1}^{n} cov(I_{li},I_{kj})$ ⊛

⇐ אם $k\neq l$, I_{li} תלוי בניסוי שונה מ- I_{kj}

⇐ I_{li}, I_{kj} ב"ת

⇐ $cov(I_{li},I_{kj})=0$

לכן: $⊛=\sum_{l=1}^{n} cov(I_{li},I_{lj})=\sum_{l=1}^{n}\left(E[I_{li}I_{lj}]-E[I_{li}]\cdot E[I_{lj}]\right)$

$$I_{li}I_{lj}=\begin{cases}1 & \text{שניהם } 1\\ 0 & \text{אחרת}\end{cases}=\begin{cases}1 & \text{תוצאה } i \text{ וגם תוצאה } j \text{ בניסוי } l\\ 0 & \text{אחרת}\end{cases}=0$$ ← לא קורה אף-פעם

$$=\sum_{l=1}^{n}\left(0-p(\text{תוצאה } i)\,p(\text{תוצאה } j)\right)=\sum_{l=1}^{n}-p_ip_j=-np_ip_j.$$

הגדרה: מקדם המתאם של שני מ"מ X,Y הוא:

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{V[X]V[Y]}}$$

טענה 1: לכל X,Y ; $-1\le \rho(X,Y)\le 1$.

טענה 2: אם $\rho(X,Y)=1$ אז $Y=aX+b$ עם $a>0$

אם $\rho(X,Y)=-1$ אז $Y=aX+b$ עם $a<0$

דוגמא:

חשבו את $\rho(X_i,X_j)$ של X_i,X_j מהדוגמא הקודמת:

תזכורת: $X_j \sim Bin(n,P_j)$; $X_i \sim Bin(n,P_i)$.

עבור X_i - הצלחה = תוצאה i, כשלון = כל השאר.

$q_i=1-P_i$

← הצבה בנוסחאות: $V[X_i]=n\cdot P_iq_i$

$V[X_j]=nP_jq_j$

$$\rho(X,Y)=\frac{-nP_iP_j}{\sqrt{nP_iq_i\cdot nP_jq_j}}=-\frac{P_iP_j}{\sqrt{P_iq_iP_jq_j}}$$

← למשל אם $P_i=P_j=\frac{1}{2}$ (התוצאות סבירות באותו סכוי וכן התוצאות היחידות האפשריות) כמו עץ או פלי במטבע הוגן.

$\rho(X_i,X_j)=\frac{\frac{1}{2}\cdot\frac{1}{2}}{\sqrt{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}}}=-1$ וזה אכן תואם את האינטואיציה: $X_i=n-X_j$

מס׳ עצים, מס׳ ניסויים, מס׳ פלי

הוכחת טענה 1:

נסמן: $Z=\frac{X}{\sqrt{V[X]}}+\frac{Y}{\sqrt{V[Y]}}$ ונחשב את $V[Z]$:

שונות תמיד חיובית

נוסחא משבוע שעבר

$$0\le V[Z]=V\left[\frac{X}{\sqrt{V[X]}}+\frac{Y}{\sqrt{V[Y]}}\right]=V\left[\frac{X}{\sqrt{V[X]}}\right]+V\left[\frac{Y}{\sqrt{V[Y]}}\right]+2Cov\left(\frac{X}{\sqrt{V[X]}},\frac{Y}{\sqrt{V[Y]}}\right)$$

$V[aX]=a^2V[X]$

$$=\frac{1}{\sqrt{V[X]}}\cdot V[X]+\frac{1}{V[Y]}V[Y]+2\frac{Cov(X,Y)}{\sqrt{V[X]}\sqrt{V[Y]}}=\underbrace{1+1}_{\text{עובר אגף}}+2\rho(X,Y)$$

$$-2\le 2\rho(X,Y) \Rightarrow -1\le \rho(X,Y)$$

הצד השני $\rho(X,Y)\le 1$ מתקבל באותה דרך עבור: $Z=\frac{X}{\sqrt{V[X]}}-\frac{Y}{\sqrt{V[Y]}}$

הוכחת טענה 2: שוויון קושי שוורץ. (יחד עם זה - מקבלים גם טענה 1
כי - שוויון קושי שוורץ).

הרצאה 6
18/12/17

-המשך-

תוחלת מותנית:

הגדרנו התפלגות מותנית

$$P_{X|Y}(k|l)=P(X=k|Y=l)=\frac{P_{X,Y}(k,l)}{P_Y(l)}$$

זו פונקציית התפלגות של המ"מ: $Z=X|Y=l$.

ניתן לחשב את התוחלת של Z: $E[Z]=E[X|Y=l]=\sum_k kP_{X|Y}(k|l)$

זו בהגדרה התוחלת של X בהינתן $Y=l$

דוגמא:

$X,Y\sim Bin(n,p)$ מ"מ ב"ת. נחשב את $E[X|X+Y=m]$.

מבצעים 2n ניסויי ברנולי.

X - מס' ההצלחות מתוך n הראשונים.

Y - מס' ההצלחות מתוך n האחרונים.

X+Y - מס' ההצלחות סה"כ. הטענה ש- $X+Y\sim Bin(2n,p)$.

ראשית נחשב את ההתפלגות המותנית:

{המאורע: k הצלחות מתוך n ראשונים ו-m סה"כ, הוא המאורע k הצלחות מתוך n ראשונים ו-m-k מתוך n אחרונים}

$$P_{X|X+Y}(k,m)=\frac{P(X=k,X+Y=m)}{P(X+Y=m)}=\frac{1}{\binom{2n}{m}p^m(1-p)^{2n-m}}\cdot P(X=k,Y=m-k)=\frac{1}{\binom{2n}{m}p^m(1-p)^{2n-m}}\cdot\binom{n}{k}p^k(1-p)^{n-k}\binom{n}{m-k}p^{m-k}(1-p)^{n-(m-k)}$$

$P(X=k)\cdot P(Y=m-k)$

מציבים $X=k$ במשוואה השנייה.

$$=\frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}\cdot\frac{\cancel{p^k\cdot p^{m-k}}\cdot\cancel{(1-p)^{n-k}\cdot(1-p)^{n-m+k}}}{\cancel{p^m}\cdot\cancel{(1-p)^{2n-m}}}$$

$$=\frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}\Rightarrow$$

קיבלנו $X|X+Y=m$ מתפלג $Hyp(2n,n,m)$.

תוחלת של $Hyp(2n,n,m)$

מסקנה: $E[X|X+Y=m]=\frac{n}{2n}\cdot m=\frac{m}{2}$.

⇐ כל המשפטים לגבי תוחלת של מ"מ נכונים גם לגבי תוחלת מותנית:

למשל: (1) $E[g(X)|Y=y]=\sum g(k)P_{X|Y}(k|y)$

(2) ליניאריות: $E[\sum_{i=1}^n X_i|Y=y]=\sum_{i=1}^n E[X_i|Y=y]$

אפשר לדבר על המ"מ $Z=E[X|Y]$. זו התוחלת המותנית של X בהינתן Y,

אם ידוע $Y=y$ אז $Z=E[X|Y=y]$.

לדוגמא: צמח הולך לכלבייה ובוחר כלב באקראי. בכלבייה כלבים מ-3 סוגים.

פודל-קטן, פיטבול-בינוני, דני ענק-ענק.

נסמן ב-X-משקל הכלב, Y-סוג הכלב.

בסימונים הנ"ל: $E[X]$ - מס' קבוע שערכו ממוצע המשקל של הכלבים בכלביה.

$E[X|Y]$ - מ"מ שיכול לקבל רק 3 ערכים וערכו נקבע לחלוטין ע"י Y.

← אם עמית בחר פודל אז: $E[X|Y]=E[X|Y=\text{פודל}]$ = ממוצע המשקל של הפודלים בכלביה. אם עמית בחר פיטבול אז $E[X|Y]=E(X|Y=\text{פיטבול}]$ = ממוצע המשקל של הפיטבולים בכלביה. אותו דבר גם לבחירת דני ענק.

(!) **חשוב**: כל עוד לא ידוע Y (הסוג הנבחר) אי-אפשר לקבוע איזה משלושת הערכים $E(X|Y=\text{פודל}]$, $E[X|Y=\text{פיטבול}]$, $E[X|Y=\text{דני}]$ יקבל המ"מ $Z=E[X|Y]$.

(*) **הערה**: יש לחשוב על $Z=E(X|Y]$ כפונקציה של המ"מ Y כמו $g(y)=y^2$, $g(y)=\sin y$.

משפט: חוק התוחלת השלמה: $E[X]=E[E[X|Y]]$.

⇐ בחזרה לדוגמא: $E[X]=E[\boxed{E[X|Y]}]$.

ממוצע משוקלל של 3 ערכים: ממוצע משקל הפודלים, משקל הפיטבולים ומשקל הדנים

ממוצע המשקל של כל הכלבים

(*) **הערה**: בביטוי $E(E(X|Y)]$ התוחלת הפנימית היא ממוצע של X, התוחלת החיצונית היא ממוצע לפי Y, משמעות הנוסחה היא חלוקת חישוב התוחלת לשלבים: קודם X בהינתן Y ורק אח"כ לפי Y.

⇐ **הוכחה**:

$$E[\underbrace{E[X|Y]}_{g(Y)}]\underset{=\sum g(k)P_Y(k)}{=}\sum_k E[X|Y=k]\cdot P_Y(k)\overset{\text{הגדרת } E[X|Y=k]}{=}\sum_k\left(\sum_m m\cdot\boxed{P_{X|Y}(m|k)}\right)\cdot P_Y(k)=$$

$E_Y(\cdot)$ $E_X(\cdot)$

$$\sum_k\sum_m m\boxed{\frac{P_{X,Y}(m,k)}{\cancel{P_Y(k)}}}\cdot\cancel{P_Y(k)}=\sum_m\sum_k m\cdot P_{X,Y}(m,k)=\sum_m m\Big(\underbrace{\sum_k P_{X,Y}(m,k)}_{\text{סכום שורה ש' בטבלה}}\Big)=$$

$$\sum m\cdot P_X(m)\overset{\text{הגדרה}}{=}E[X]$$

דוגמא: יהיו ... X_1,X_2,X_3 מ"מ ב"ת אשר לא ידוע עליהם דבר מלבד $E[X_i]=t$ לכל i, (כלומר יש להם אותה תוחלת). יהי $N\sim Poi(\lambda)$ ב"ת בקודמים.

נסמן $Y=\sum_{i=1}^{N}X_i=X_1+X_2+\dots+X_N$ {מס' המחוברים בסכום תלוי ב-N "סכום מקרי של מ"מ"}.

למשל: N - מס' האנשים שנכנסים לחנות ביום. X_i הסכום שהאיש ה-i מוציא בחנות. ואז Y - הפדיון (סכום שהרוויחו) בחנות באותו יום. חשבו $E[Y]$.

← אינטואיציה: אם נכנסים בממוצע $\underbrace{\lambda}_{E[N]}$ אנשים וכל אחד מוציא בממוצע t, אז הפדיון הממוצע הוא $\lambda\cdot t$.

⇐ **הוכחה**: נתנה ב-N.

$$E[Y]\overset{\text{לפי } N}{=}E[E[Y|N]]=\sum_k E\Big[\sum_{i=1}^{N}X_i\Big|N=k\Big]\cdot P_N(k)\overset{\text{הצבה}}{=}\sum_k E\Big[\sum_{i=1}^{k}X_i\Big|N=k\Big]P_N(k)=\{N,X_i \text{ ב"ת ולכן אפשר להוריד את ההתניה}\}$$

$$\sum_k E\Big[\sum_{i=1}^{k}X_i\Big]P_N(k)\overset{\text{לינאריות}}{=}\sum_k\Big(\sum_{i=1}^{k}E(X_i)\Big)P_N(k)\underset{E(X_i)=t}{=}\sum_k\sum t\,P_N(k)=\sum_k kt\,P_N(k)=E[t\cdot N]=tE[N]$$

מס' קבוע תלוי ב-k

$\sum g(k)P_N(k)=E[g(N)]$