שלומי

משתנים מתוקננים ומקדם המתאם

נדון בשני הנושאים האלה ובהמשך נראה שיש קשר ביניהם

הקדמה

$$V(X+Y)=V(X)+V(Y)+2\,Cov(X,Y)$$

ה- Cov מודד איזה קשר יש בין X ו Y. ראינו שכאשר X ו Y הם ב"ת אז $Cov(X,Y)=0$. כאשר $Cov(X,Y)>0$ אז הצירוף של שני המשתנים מגדיל את הפיזור מעבר לסכום השונויות. כאשר $Cov(X,Y)<0$ אז אחד מקזז את האחר או משלים אותו מקטינה מסוימת.

$Cov(X,Y)$ מודד קשר בין X ל Y, אך יש בעיה איתו בכך שהוא תלוי ביחידות, נניח שמכפילים את X בקבוע c אז
$$Cov(cX,Y)=E(cXY)-E(cX)\cdot E(Y)=c[E(XY)-E(X)E(Y)]=c\cdot Cov(X,Y)$$
לכן גודל היחידות משפיע על Cov. היינו רוצים למנוע זאת.

הגדרה

יהי X משתנה אז המשתנה המתוקנן של X, שיסומן $\hat{X}$? שווה ל $\frac{X-\mu_X}{\sigma_X}$ כאשר $V(X)\neq 0$.
כאשר $\mu_X=E(X)$, $\sigma_X=\sqrt{V(X)}$.

טענה

$E(\hat{X})=0$ לכל משתנה X.

הוכחה

$$E(\hat{X})=E\left(\frac{X-\mu_X}{\sigma_X}\right)=\frac{1}{\sigma_X}\left(E(X-\mu_X)\right)=0$$

טענה

$$V(\hat{X})=1$$

הוכחה

$$V(\hat{X})=V\left(\frac{X-\mu_X}{\sigma_X}\right)=\frac{1}{\sigma_X^2}V(X-\mu_X)=\frac{V(X)}{\sigma_X^2}=1$$

טענה

אם $Y=aX+b$ (טרנספורמציה לינארית) אז אם $a>0$
אז ל-X ול-Y יש אותו משתנה מתוקנן ואם $a<0$
אז המשתנה המתוקנן של אחד הוא מינוס המשתנה
המתוקנן של השני.

הסבר

$$V(Y)=a^2\cdot V(X)=a^2\cdot\sigma_X^2\ ,\ \mu_Y=E(Y)=a\cdot E(X)+b=a\mu_X+b$$

$$\hat{Y}=\frac{Y-\mu_Y}{\sigma_Y}$$

$$\hat{Y}=\frac{Y-\mu_Y}{\sigma_Y}=\frac{(aX+b)-(a\mu_X+b)}{\sqrt{a^2\cdot\sigma_X^2}}=\frac{a(X-\mu_X)}{|a|\sigma_X}$$

כיוון הפוך

אם $\hat{Y}=\pm\hat{X}$ אז Y הוא טרנספורמציה לינארית
של X

הסבר

$$\frac{Y-\mu_Y}{\sigma_Y}=\pm\frac{X-\mu_X}{\sigma_X}$$

ולכן
$$Y=\mu_Y\pm\frac{\sigma_Y}{\sigma_X}(X-\mu_X)$$

כאשר $\mu_X,\mu_Y,\sigma_X,\sigma_Y$ הם קבועים.

דוגמא/

$X\sim Bin\left(2,\frac{1}{2}\right)$. מצאו את המשתנה המתוקנן של X.

$\mu_X=2\cdot\frac{1}{2}=1$, $\sigma_X=\sqrt{2\cdot\frac{1}{2}\cdot\frac{1}{2}}=\sqrt{\frac{1}{2}}$

אם X מקבל את הערך 0 אז $\hat{X}=\frac{0-1}{\sqrt{\frac{1}{2}}}=-\sqrt{2}$

אם X מקבל את הערך 1 אז $\hat{X}=\frac{1-1}{\sqrt{\frac{1}{2}}}=0$

אם X מקבל את הערך 2 אז $\hat{X}=\frac{2-1}{\sqrt{\frac{1}{2}}}=\sqrt{2}$

לכן $P(\hat{X}=-\sqrt{2})=P(\hat{X}=+\sqrt{2})=\left(\frac{1}{2}\right)^2$

$P(\hat{X}=0)=2\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2}$

מקדם המתאם

הגדרה

מקדם המתאם בין X ל Y מסומן ב $r(X,Y)$ או $\rho(X,Y)$. מתקיים:

$$r(X,Y)=\frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}}$$

או

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

טענה

$$Cov(\hat{X},\hat{Y})=r(X,Y)$$

הסבר

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

$$=E(\hat{X}\hat{Y})-0\cdot 0=E(\hat{X}\cdot\hat{Y})=E\left(\frac{X-\mu_X}{\sigma_X}\cdot\frac{Y-\mu_Y}{\sigma_Y}\right)=$$

$$=\frac{E\left((X-\mu_X)(Y-\mu_Y)\right)}{\sigma_X\cdot\sigma_Y}=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}=r(X,Y)$$

הראנו קודם ש $\hat{X}$ ו $\hat{Y}$ לא תלויים ביחידות (בטרנספורמציה לינארית), לכן $r(X,Y)$ לא תלוי ביחידות.

טענות

א. $-1\le r(X,Y)\le 1$

ב. $r(X,Y)=1 \iff$ יש קשר לינארי עולה בין X ל Y.

ג. $r(X,Y)=-1 \iff$ יש קשר לינארי יורד בין X ל Y.

ד. $r(X,Y)=0 \iff Cov(X,Y)=0$

כל אלה בהנחה ש $\sigma_X, \sigma_X>0$ (השונויות לא שוות לאפס).

3. הוכחות

ג. אם $Cov(X,Y)=0$ אז $\frac{Cov(X,Y)}{\sigma_X \sigma_Y}=0$

שונות תמיד אי-שלילית

א.

$$0 \le V(\hat{X}+\hat{Y}) = V(\hat{X}) + V(\hat{Y}) + 2\cdot Cov(\hat{X},\hat{Y}) =$$

$$= 1+1+2r(X,Y) = 2(1+r(X,Y))$$

לכן $r(X,Y) \ge -1$

$$0 \le V(\hat{X}-\hat{Y}) = V(\hat{X}) + V(\hat{Y}) - 2\cdot r(X,Y) = 2(1-r(X,Y))$$

לכן $r(X,Y) \le 1$

ב. אם יש קשר לינארי עולה אז $\hat{Y}=\hat{X}$ ואז

$$r(X,Y) = Cov(\hat{X},\hat{Y}) = Cov(\hat{X},\hat{X}) = V(\hat{X}) = 1$$

אם יש קשר לינארי יורד אז $\hat{Y}=-\hat{X}$ ואז

$$r(X,Y) = Cov(\hat{X},\hat{Y}) = Cov(\hat{X},-\hat{X}) = -V(\hat{X}) = -1$$

בכיוון ההפוך: אם $r(X,Y)=\pm 1$ אז יש קשר לינארי, נטפל כעת למקרה ש $r(X,Y)=1$.

לכן $r(X,Y)=1$: $Cov(\hat{X},\hat{Y})=1$

$$V(\hat{X}-\hat{Y}) = \underbrace{V(\hat{X})}_{=1} + \underbrace{V(\hat{Y})}_{=1} - 2\cdot \underbrace{Cov(\hat{X},\hat{Y})}_{=1}$$

כלומר $V(\hat{X}-\hat{Y})=0$! $\hat{X}-\hat{Y}$ הוא משתנה מנוון – שמקבל רק ערך אחד: 0 ולכן $\hat{X}=\hat{Y}$ ולכן Y הוא טרנספורמציה לינארית של X.

שאלה

מה אפשר להגיד על מקדם המתאם בין גובה למשקל?

תשובה

יש קשר חיובי בין גובה למשקל, לכן כפי הנראה המתאם הוא חיובי. אבל אין קשר מוחלט ולכן בודאי אין קשר ליניארי. לכן המתאם לא שווה ל 1.

שאלה

האם קיים משתנה X, כך שהמתאם בין X ל X^2 הוא 1+?

האם קיים משתנה X, כך שהמתאם בין X ל X^2 הוא 1- ?

תשובה

באופן כללי אין קשר ליניארי בין X ל X^2. אבל אם המשתנה X מקבל רק שני ערכים אפשריים, אז יש קשר ליניארי. אם X מקבל את הערכים 0 ו 1 כל אחד בסיכוי חיובי, אז יש קשר ליניארי עולה. אם X מקבל את הערכים 1- ו 0 אז יש קשר ליניארי יורד בין X ל X^2 והמתאם הוא 1-.

שאלה

מטילים 3 פעמים מטבע הוגן. ההטלות הן ב"ת.

יהי Y – מספר "העצים" בשתי ההטלות הראשונות.

יהי Z – מספר "העצים" בשתי ההטלות האחרונות.

מצאו את $r(Y,Z)$.

תשובה

הקשר בין המשתנים הנו קשר עולה, לכן נצפה למתאם חיובי. אך הקשר אינו מוחלט ולכן בודאי לא לינארי. לכן המתאם קטן מ 1.

$$r(Y,Z)=\frac{Cov(Y,Z)}{\sqrt{V(Y)V(Z)}}$$

כל אחד מבין Y ו Z מתפלג $Bin(2,\frac{1}{2})$. לכן

$$\sqrt{V(Y)V(Z)}=\sqrt{2\cdot\frac{1}{2}\cdot\frac{1}{2}\cdot 2\cdot\frac{1}{2}\cdot\frac{1}{2}}=0.5$$

$$Cov(Y,Z)=Cov(X_1+X_2,\ X_2+X_3)=$$

$$=\underset{\text{(ב"ת)}}{Cov(X_1,X_2)}+\underset{\text{ב"ת}}{Cov(X_1,X_3)}+Cov(X_2,X_2)+\underset{\text{(ב"ת)}}{Cov(X_2,X_3)}=$$

$$=0+0+V(X_2)+0=0.5\cdot 0.5=0.25$$

$$r(Y,Z)=\frac{0.25}{0.5}=0.5$$ לכן נקבל

שלומי