F6, Tvådimensionell slumpvariabel

Christian Tallberg

Statistiska institutionen

Stockholms universitet

Simultan sannolikhetsfördelning

Två stokastiska variabler X och Y är *oberoende* om

$$\begin{aligned} p(x,y) &= p(x)\,p(y) \text{ (diskreta s.v.)} \\ f(x,y) &= f(x)\,f(y) \text{ (kontinuerliga s.v.)} \end{aligned}$$

för *alla* x och y, där $p(x,y) = \Pr(X = x \text{ och } Y = y)$ är den simultana sannolikhetsfördelningen för x och y.

Omvänt gäller:

Om två stokastiska variabler X och Y är *oberoende* är

$$\begin{aligned} p(x,y) &= p(x)\,p(y) \\ f(x,y) &= f(x)\,f(y)\,. \end{aligned}$$

Två stokastiska variabler X och Y är *okorrelerade* om kovariansen mellan X och Y är lika med noll.

Kovariansen beräknas som

$$\begin{aligned} Cov(X,Y) &= E(X-\mu_x)\left(Y-\mu_y\right) \\ &= E(XY) - E(X)\,E(Y) \text{ där} \\ E(XY) &= \sum xyp(x,y)\,. \end{aligned}$$

Omvänt gäller:

Om två stokastiska variabler X och Y är *okorrelerade* är $Cov(X,Y) = 0$.

Oberoende stokastiska variabler är okorrelerade men okorrelerade stokastiska variabler behöver inte vara oberoende.

I en population med familjer har $X =$ Antal pojkar och $Y =$ Antal flickor följande simultana fördelning:

	$p(x,y)$		
$x \backslash y$	0	1	$p(x)$
0	0.25	0	0.25
1	0	0.50	0.50
2	0.25	0	0.25
$p(y)$	0.50	0.50	1

Produkten xy ges av

xy		
$x \backslash y$	0	1
0	0	0
1	0	1
2	0	2

och sannolikhetsfördelningen för xy ges av

xy	$p(xy)$
0	0.50
1	0.50
2	0

OBS! $p(xy) = \Pr(XY = xy)$

men $p(x, y) = \Pr(X = x \text{ och } Y = y)$.

Är de s.v. X och Y

1. okorrelerade?

2. oberoende?

Lösning 1)

$$\begin{aligned} E(XY) &= \sum xyp(x,y) \\ &= 0 \cdot 0.50 + 1 \cdot 0.50 + 2 \cdot 0 = 0.5 \\ E(X) &= 0 \cdot 0.25 + 1 \cdot 0.50 + 2 \cdot 0.25 = 1 \\ E(Y) &= 0 \cdot 0.50 + 1 \cdot 0.50 = 0.5 \\ Cov(X,Y) &= E(XY) - E(X)E(Y) \\ &= 0.5 - 1 \cdot 0.5 = 0 \end{aligned}$$

dvs de stokastiska variablerna X och Y är okorrelerade.

Lösning 2).

$$\begin{aligned} p(0,0) &= \Pr(X = 0 \text{ och } Y = 0) \\ &= 0.25 \text{ men} \\ \Pr(X=0)\Pr(Y=0) &= 0.25 \cdot 0.50 \\ &= 0.125 \neq 0.25 \end{aligned}$$

dvs de stokastiska variablerna X och Y är inte oberoende.

Linjära kombinationer av stokastiska variabler

För alla stokastiska variabler X och Y gäller att

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) = \mu_x + \mu_y \text{ och} \\ E(X-Y) &= E(X) - E(Y) = \mu_x - \mu_y \end{aligned}$$

samt

$$\begin{aligned} V(X+Y) &= E(X+Y)^2 - [E(X+Y)]^2 \\ &= E(X+Y)^2 - [E(X) + E(Y)]^2 \\ &= E\left(X^2\right) + 2E(XY) + E\left(Y^2\right) \\ &\quad - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2 \\ &= E\left(X^2\right) - [E(X)]^2 + E\left(Y^2\right) - [E(Y)]^2 \\ &\quad + 2[E(XY) - E(X)E(Y)] \\ &= V(X) + V(Y) + 2Cov(X,Y). \end{aligned}$$

På liknande sätt kan man visa att

$$V(X-Y) = V(X) + V(Y) - 2Cov(X,Y).$$

Om X och Y är oberoende gäller att

$$\begin{aligned} V(X+Y) &= V(X) + V(Y) \text{ och} \\ V(X-Y) &= V(X) + V(Y). \end{aligned}$$

Exempel 4 och 5. Körner (sid 103)

Den simultana sannolikhetsfördelningen $p(x,y)$ för de s.v. X och Y.

$y \backslash x$	0	1	2	$p(y)$
1	0.08	0.12	0.30	0.50
2	0.12	0.18	0.20	0.50
$p(x)$	0.20	0.30	0.50	1.00

Bilda den nya variabeln $S = X+Y$. Vad är väntevärde och varians för S?

s	1	2	3	4
$p(s)$	0.08	0.24	0.48	0.20

Väntevärdet och variansen blir

$$\begin{aligned} E(S) &= \sum_{alla\ s} sp(s) = 2.8 \\ E\left(S^2\right) &= \sum_{alla\ s} s^2 p(s) = 8.56 \\ V(S) &= 8.56 - 2.8^2 = 0.72 \end{aligned}$$

Alternativ lösning: Beräkna väntevärde och varians på vanligt sätt. De blir

$$E(X) = 1.3 \quad V(X) = 0.61 \quad E(Y) = 1.5 \quad V(Y) = 0.2$$

$$\begin{aligned} E(XY) &= \sum xyp(x,y) = 0 \cdot 1 \cdot 0.08 + 1 \cdot 1 \cdot 0.12 \\ &\quad +2 \cdot 1 \cdot 0.30 + 0 \cdot 2 \cdot 0.12 + 1 \cdot 2 \cdot 0.18 \\ &\quad +2 \cdot 2 \cdot 0.20 \\ &= 1.88 \end{aligned}$$

Kovariansen blir

$$Cov(X,Y) = 1.88 - 1.3 \cdot 1.5 = -0.77$$

$$\begin{aligned} E(S) &= E(X+Y) = E(X) + E(Y) = 1.3 + 1.5 \\ V(S) &= V(X+Y) = V(X) + V(Y) + 2Cov(X,Y) \\ &= 0.61 + 0.25 + 2 \cdot (-0.07) = 0.72. \end{aligned}$$

För summan av n stokastiska variabler $X_1 + X_2 + ... + X_n$ gäller

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

Om variablerna är parvis oberoende gäller dessutom

$$V(X_1 + X_2 + ... + X_n) = V(X_1) + V(X_2) + ... + V(X_n)$$

Om X_1 och X_2 är oberoende och normalfördelade variabler, $X_1 \sim N\left(\mu_1; \sigma_1^2\right)$ och $X_2 \sim N\left(\mu_2; \sigma_2^2\right)$, gäller att

$$\begin{aligned} E(X_1 + X_2) &= E(X_1) + E(X_2) = \mu_1 + \mu_2 \\ V(X_1 + X_2) &= V(X_1) + V(X_2) = \sigma_1^2 + \sigma_2^2. \end{aligned}$$

Det vill säga linjärkombinationerna

$$\begin{aligned} (X_1 + X_2) &\sim N\left(\mu_1 + \mu_2; \sigma_1^2 + \sigma_2^2\right) \text{ och} \\ (X_1 - X_2) &\sim N\left(\mu_1 - \mu_2; \sigma_1^2 + \sigma_2^2\right). \end{aligned}$$

Exempel:

X_1 och X_2 är oberoende stokastiska variabler som har följande fördelningar: $X_1 \sim N\left(10; 3^2\right)$ och $X_2 \sim N\left(9; 4^2\right)$. Vad är sannolikheten att $X_1 > X_2$?

$$\Pr\left(X_1 > X_2\right) = \Pr\left(\left(X_1 - X_2\right) > 0\right)$$

Låt $Y = X_1 - X_2$, $Y \sim N\left(10 - 9; 3^2 + 4^2\right) = N\left(1; 5^2\right)$

$$\begin{aligned} \Pr\left(X_1 > X_2\right) &= \Pr\left(Y > 0\right) = \Pr\left(\frac{Y-1}{5} > \frac{0-1}{5}\right) \\ &= \Pr\left(Z > -\frac{1}{5}\right) = \Phi\left(0.2\right) = 0.5793 \end{aligned}$$