Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT19

Föreläsning 4

Nancy Abdallah

Chalmers - Göteborgs Universitet

April 4, 2019

Tvådimensionella stokastiska variabler

- En **tvådimensionell stokastisk variabel** är en funktion (X, Y) definierad på ett utfallsrum S.
- En två dimensionell s.v. kan vara **diskret** (om både X och Y är diskreta) eller **kontinuerlig** (om både X och Y är kontinuerliga).

Tvådimensionella diskreta s.v.

Låt X och Y vara två diskreta s.v.

- Sannolikhetsfunktionen f_{XY} av den tvådimensionell s.v (X, Y) är definierad som

$$f_{XY}(x, y) = P(X = x \text{ och } Y = y)$$

- En funktion $f_{XY}(x, y)$ är en sannolikhetsfunktion om och endast om för varje (x, y)
    - $f_{XY}(x, y) \geq 0$
    - $\sum_{\text{alla } x} \sum_{\text{alla } y} f_{XY}(x, y) = 1$
- Fördelningsfunktionen $F(x, y)$ är definierad som

$$F(x, y) = P(X \leq x \text{ och} Y \leq y) = \sum_{j \leq x} \sum_{k \leq y} f_{XY}(j, k)$$

Exempel

Låt X och Y vara antalet pojker resp flickor i en slumpmässigt vald svensk familk. Sannolikhetsfunktion $f_{XY}(x, y)$ ges i tabellen nedan

Y \ X	0	1	2	3	4
0	0.38	0.16	0.04	0.01	0.01
1	0.17	0.08	0.02		
2	0.05	0.02	0.01		
3	0.02	0.01			
4	0.02				

$F_{XY}(1, 2) = P(X \leq 1 \text{ och } Y \leq 2) =$
$f_{XY}(0, 0)+f_{XY}(0, 1)+f_{XY}(0, 2)+f_{XY}(1, 0)+f_{XY}(1, 1)+f_{XY}(1, 2) =$
$0.38 + 0.17 + 0.05 + 0.16 + 0.08 + 0.02 = 0.86$

Marginella sannolikhetsfunktionen

Låt (X, Y) vara en tvådimensionell diskret s.v. med täthetsfunktion f_{XY}. De **marginella** sannolikhetsfunktionerna f_X och f_Y för X och Y respektiv, ges av

$$f_X(x) = \sum_{alla\ y} f_{XY}(x, y) = P(X = x)$$

och

$$f_Y(y) = \sum_{alla\ x} f_{XY}(x, y) = P(X = y)$$

Exempel

(föregående exempel)

Y \ X	0	1	2	3	4	f_Y
0	0.38	0.16	0.04	0.01	0.01	0.60
1	0.17	0.08	0.02			0.27
2	0.05	0.02	0.01			0.08
3	0.02	0.01				0.03
4	0.02					0.02
f_X	0.64	0.27	0.07	0.01	0.01	1

Tvådimensionella kontinuerliga s.v

Låt X och Y vara två kontinuerliga s.v.

- En funktion f_{XY} som uppfyller
  1. $f_{XY}(x, y) \geq 0$ för alla reella värden x och y
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1$
  3. $P(a \leq X \leq b \text{ och } c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy\, dx$

  kallas täthetsfunktionen för (X, Y)
- De marginella täthetsfunktionerna f_X och f_Y för X och Y respektiv, ges av

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad \text{och} \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$$

Exempel (Ex.8 i boken)

Låt X representera temperaturen och Y tiden i minuter för hur lång tid en dieselmotor tar på sig för att värma upp. Täthetsunktionen för (X, Y) ges av

$$f_{XY}(x, y) = c(4x + 2y + 1) \qquad 0 \leq x \leq 40, \quad 0 \leq y \leq 2$$

1. *Beräkna c.*
2. *Beräkna sannolikheten att temperaturen blir större än 20 och det tar minst 1 minut för att motoren värmer upp.*
3. *Beräkna marginella täthetsfunktioner f_X och f_Y.*
4. *Beräkna sannolikheten att det tar minst en minut för att motoren värmer upp.*
5. *Beräkna sannolikheten att temparaturen är större än 20.*

Exempel (Lösning)

1.

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dx\,dy &= \int_0^{40}\int_0^2 c(4x+2y+1)dy\,dx \\
&= \int_0^{40} c[4xy+y^2+y]_0^2 dx \\
&= \int_0^{40} c(8x+6)dx \\
&= c[4x^2+6x]_0^{40} = 6640c = 1
\end{aligned}$$

Så är $c = \frac{1}{6640}$

Exempel (lösning)

2. $P(X > 20 \text{ och } Y \geq 1) = \int_{20}^{40} \int_{1}^{2} \frac{1}{6640}(4x + 2y + 1)dydx \approx 0.3735$
3. $f_X = \int_{-\infty}^{\infty} f_{XY}(x, y)dy = \int_{0}^{2} \frac{1}{6640}(4x + 2y + 1)dy = \frac{8x+6}{6640}$
   $f_Y = \int_{-\infty}^{\infty} f_{XY}(x, y)dx = \int_{0}^{40} \frac{1}{6640}(4x + 2y + 1)dx = \frac{3240+80y}{6640}$
4. $P(Y \geq 1) = \int_{1}^{2} f_Y(y)dt = \frac{1}{6640}[3240y + 40y^2]_1^2 \approx 0.506$
5. $P(X > 20) = \int_{20}^{40} f_X(x)dx = \frac{1}{6640}[4x^2 + 6x]_{20}^{40} \approx 0.741$

Oberoende stokastiska variabler

- *(kapitel 1)* Två händelser A och B är oberoende om $P(A \cap B) = P(A)P(B)$.
- Två stokastiska variabler X och Y är oberoende om

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

  för alla x och y.
- I det diskreta exemplet (s.3) är X och Y inte oberoende eftersom $0 = f_{XY}(3, 1) \neq f_X(3)f_Y(1) = 0.01 \cdot 0.27 = 0.0027$
- I det kontinuerliga exemplet (s.7) är X och Y inte oberoende eftersom $f_{XY}(x, y) \neq f_X(x)f_Y(y)$ $(P(X > 20 \text{ och } Y \geq 1) \neq P(X > 20)P(Y \geq 1))$

Väntevärde

Låt (X, Y) vara en tvådimensionell s.v. med täthetsfunktion $f_{XY}(x, y)$ och $H(X, Y)$ en funktion. Så definierar vi

- **Väntevärdet** $E[H(X, Y)]$ med

$$E[H(X, Y)] = \sum_{alla\ x} \sum_{alla\ y} H(x, y) f_{XY}(x, y)$$

  om (X, Y) är diskret och, om (X, Y) är kontinuerlig,

$$E[H(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) f_{XY}(x, y) dxdy$$

- Vi får räkneregler som $E[cX] = cE[X]$ och $E[X + Y] = E[X] + E[Y]$.
- Om X och Y är oberoende, så får vi

$$E[XY] = E[X]E[Y]$$

Kovarians

Låt X och Y vara två s.v. med $E[X] = \mu_X$ och $E[Y] = \mu_Y$.

- **Kovariansen** $Cov[X, Y]$ eller σ_{XY} är ett mått på samvariation mellan X och Y.
- $Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$.
- Med räkneregler $Cov[X, Y] = E[XY] - E[X]E[Y]$.
- Konsekvens: Om X och Y är oberoende så är $Cov[X, Y] = 0$.
- Räknergler:
  1. $Cov[cX, Y] = Cov[X, cY] = cCov[X, Y]$.
  2. $Cov[X, Y] = Cov[Y, X]$.

Korrelation

- Pearson korrelationskoefficient ρ_{XY} är ett mått som beskriver om det finns en linjär relation mellan två stokastiska variabler X och Y.
- ρ_{XY} definieras som

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov[X, y]}{\sqrt{Var[X] Var[Y]}}$$

- $-1 \leq \rho_{XY} \leq 1$.
- $|\rho_{XY}| = 1$ om och endast om $Y = a + bX$ där a och b är reella tal, $b \neq 0$.
- $\rho_{XY} = 0$ betyder att det finns ingen linjär relation mellan X och Y.

Betingade fördelningar

- *(Kapitel 2)* Om A och B är händelser och $P(B) \neq 0$, har vi

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- Den betingade täthetsfunktionen (sannolikhetsfunktionen) för den stokastiska variabeln X givet $Y = y$ definieras av:

$$f_{X|Y} = \frac{f_{XY}(x, y)}{f_Y(y)}$$

 och betingade täthetsfunktionen (sannolikhetsfunktionen) för Y givet $X = x$ definieras av:

$$f_{Y|X} = \frac{f_{XY}(x, y)}{f_X(x)}$$