Matematisk Statistik och Disktret Matematik, MVE051/MSG810, VT20

Föreläsning 5

Nancy Abdallah

Chalmers - Göteborgs Universitet

Two-dimensional discrete random variables

Let X and Y be two discrete random variables.

- The ordered pair (X, Y) is called a two-dimensional discrete random variable.
- A function f_{XY} such that

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

  is called the joint density for (X, Y).
- A function $f_{XY}(x, y)$ is a joint density function for (X, Y) if and only if
  - $f_{XY}(x, y) \geq 0$
  - $\sum_{\text{all } x} \sum_{\text{all } y} f_{XY}(x, y) = 1$

Example

Let X and Y be the number of girls, respectively boys in a randomly chosen Swedish family. The joint density function $f_{XY}(x, y)$ is given in the table below.

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				

$$\sum_{x=0}^{4} \sum_{y=0}^{4} f_{XY}(x, y) = 1.$$

$$P(X = 0 \text{ and } Y = 1) = f_{XY}(0, 1) = 0.16.$$

$$P(X = 2) = f_{XY}(2, 0) + f_{XY}(2, 1) + f_{XY}(2, 2) = 0.08$$

Marginal density functions

Let (X, Y) be a two-dimensional discrete random variable with joint density function f_{XY}. The marginal density for X is given by

$$f_X(x) = \sum_{all\ y} f_{XY}(x, y)$$

and the marginal density for Y is given by

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

Continuation of previous example

Y / X	0	1	2	3	4	f_X
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_Y	0.64	0.27	0.07	0.01	0.01	1

Two-dimensional continuous random variables

Let X and Y be two continuous random variables.

- The ordered pair (X, Y) is called a two-dimensional continuous random variable.
- A function f_{XY} such that
  1. $f_{XY}(x, y) \geq 0$ for all $x \in \mathbb{R}$ and $y \in \mathbb{R}$
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dy dx = 1$
  3. $P(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy\, dx$

  is called the joint density function for (X, Y).
- The marginal density f_X and f_Y for X, respectively $Y$are given by

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

Ex.8 in the book

Let X denote the temperature and Y denote the time in minutes that it takes for the diesel engine on an automobile to get ready to start. Assume that the joint density for (X, Y) is given by $f_{XY}(x, y) = c(4x + 2y + 1)$ for $0 \leq x \leq 40$ and $0 \leq y \leq 2$.

1. Find c.
2. Find the probability that the temperature will exceed 20° and it will take at least 1 minute for the car to be ready to start.
3. Find the marginal density functions for X and Y.
4. Find the probability that it will take at least one minute for the car to be ready to start.
5. Find the probability that the temperature will exceed 20°.

Solution

1.

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

Hence, $c = \frac{1}{6640}$

Solution

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. $P(X > 20) = \int_{20}^{40} f_X(x)dx = \frac{1}{6640}[4x^2 + 6x]_{20}^{40} \approx 0.741$

Independent Random Variables

- *(From chapter 2)* Two events A and B are independent if and only if $P(A \cap B) = P(A)P(B)$.
- Two random variables X and Y are independent if and only if

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

  for all x and y
- In the example on p.2 (discrete case) the variables X are Y are not independent since $0 = f_{XY}(3, 1) \neq f_X(3)f_Y(1) = 0.01 \cdot 0.27 = 0.0027$
- In the example on p.6 (continuous case), X and Y are not independent since $f_{XY}(x, y) \neq f_X(x)f_Y(y)$ $(P(X > 20 \text{ and } Y \geq 1) \neq P(X > 20)P(Y \geq 1))$

Expected Value

Let (X, Y) be a two-dimensional random variable with density function $f_{XY}(x, y)$ and $H(X, Y)$ a random variable.The expected value $E[H(X, Y)]$ is given by

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

if (X, Y) is discrete and

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

if (X, Y) is continuous.

In particular, if X and Y are discrete, then

- $E[X] = \sum_{\text{all } x} \sum_{\text{all } y} x f_{XY}(x, y)$.
- $E[Y] = \sum_{\text{all } x} \sum_{\text{all } y} y f_{XY}(x, y)$.

If X and Y are continuous, then,

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

Theorem

If X and Y are independent then $E[XY] = E[X]E[Y]$.

Example p.2

- $E[X] = 0f_{XY}(0,0) + 0f_{XY}(0,1) + 0f_{XY}(0,2) + 0f_{XY}(0,3) + 0f_{XY}(0,4) + 1f_{XY}(1,0) + \cdots + 4f_{XY}(4,0) = 0.6.$
- $E[X]$ can also be computed as a one-dimensional random variable using the marginal density function: $E[X] = 0(0.6) + 1(0.27) = 2(0.16) + 3(0.03) + 4(0.02) = 0.6.$
- $E[Y] = 0(0.64) + 0.27 + 2(0.07) + 3(0.01) + 4(0.01) = 0.48.$
- $E[XY] = 0.08 + 2(0.02) + 2(0.02) + 2(2)(0.01) + 3(0.01) = 0.23 \neq 0.6(0.48) = 0.288$, so X and Y are dependent.

Kovarians

Let X and Y be two random variable with $E[X] = \mu_X$ and $E[Y] = \mu_Y$.

- The **Covariance** $Cov[X, Y]$ or σ_{XY} is given by

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

- The convariance measures how X and Y vary relative to one another.
- $Cov[X, Y] = E[XY] - E[X]E[Y]$.
- If X and Y are independent then $Cov[X, Y] = 0$.
- Räknergler:
  1. $Cov[cX, Y] = Cov[X, cY] = cCov[X, Y]$, where c is a constant.
  2. $Cov[X, Y] = Cov[Y, X]$.

Correlation

- Pearson correlation coefficient ρ_{XY} is a measure that indicates whether X and Y are *linearly* related.
- ρ_{XY} is defined by

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

- $-1 \leq \rho_{XY} \leq 1$.
- $|\rho_{XY}| = 1$ if and only if $Y = a + bX$ where a and b are real numbers $b \neq 0$.
- $\rho_{XY} = 0$ means that there is no linear relation between X and Y *(there might be another type of relation)*.

Conditional densities

- *(Chapter 2)* If A and B are two events and $P(B) \neq 0$, then
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
- The conditional density function for the random variable X given $Y = y$ is given by
$$f_{X|Y} = \frac{f_{XY}(x, y)}{f_Y(y)}$$
and the conditional density function for the random variable Y given $X = x$ is given by
$$f_{Y|X} = \frac{f_{XY}(x, y)}{f_X(x)}$$