Chapter 6.4 Independence and Measuring Association

Jim Albert and Monika Hu

Chapter 6 Joint Probability Distributions

Example

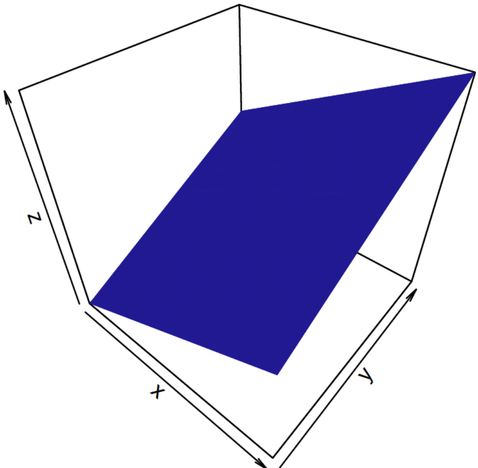
- Suppose one has two random variables (X, Y) that have the joint density

$$f(x,y) = \begin{cases} x + y, & 0 < x < 1, 0 < y < 1; \\ 0, & \text{elsewhere.} \end{cases}$$

- This density is positive over the unit square, but the value of the density increases in X (for fixed y) and also in Y (for fixed x).

Graph of joint density function

- Below is displayed a graph of this joint pdf – the density is a section of a plane that reaches its maximum value at the point (1, 1).



Marginal density

- From this density, one computes the marginal pdfs of X and Y.
- For example, the marginal density of X is given by

$$f_X(x) = \int_0^1 x + y dy$$

$$= x + \frac{1}{2}, \ 0 < x < 1.$$

- Similarly, one can show that the marginal density of Y is given by $f_Y(y) = y + \frac{1}{2}$ for $0 < y < 1$.

Independence

- Two random variables X and Y are said to be *independent* if the joint pdf factors into a product of their marginal densities, that is

$$f(x,y) = f_X(x)f_Y(y).$$

for all values of X and Y.

Example

- Are X and Y independent in our example? We look at the product
$$f_X(x)f_Y(y) = (x + \frac{1}{2})(y + \frac{1}{2})$$
which is clearly not equal to the joint pdf $f(x, y) = x + y$ for values of x and y in the unit square. So X and Y are not independent in this example.

Measuring association by covariance

- In the situation like this one where two random variables are not independent, it is desirable to measure the association pattern.
- A standard measure of association is the covariance defined as the expectation

$$Cov(X, Y) = E\left((X - \mu_X)(Y - \mu_Y)\right)$$

$$= \int\int (x - \mu_X)(y - \mu_Y) f(x, y) dxdy.$$

- For computational purposes, one writes the covariance as

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

$$= \int\int (xy) f(x, y) dxdy - \mu_X \mu_Y.$$

Example

- For our example, one computes the expectation $E(XY)$ from the joint density:

$$E(XY) = \int_0^1 \int_0^1 (xy)(x+y)dxdy$$
$$= \int \frac{y}{3} + \frac{y^2}{2} dy = \frac{1}{3}.$$

- One can compute that the means of X and Y are given by $\mu_X = 7/12$ and $\mu_Y = 7/12$, respectively.

Example

- So then the covariance of X and Y is given by

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

$$= \frac{1}{3} - \left(\frac{7}{12}\right)\left(\frac{7}{12}\right)$$

$$= -\frac{1}{144}.$$

Interpreting a covariance

- It can be difficult to interpret a covariance value since it depends on the scale of the support of the X and Y variables.
- One standardizes this measure of association by dividing by the standard deviations of X and Y resulting in the correlation measure ρ:

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

Interpreting a covariance

- One can find the variances of X and Y to be $\sigma_X^2 = 11/144$ and $\sigma_Y^2 = 11/144$.
- Then the correlation is given by

$$\rho = \frac{-1/144}{\sqrt{11/144}\sqrt{11/144}}$$

$$= -\frac{1}{11}.$$

Interpreting a correlation

- It can be shown that the value of the correlation ρ falls in the interval $(-1, 1)$
- A value of $\rho = -1$ or $\rho = 1$ indicates that Y is a linear function of X with probability 1.
- Here the correlation value is a small negative value indicates weak negative association between X and Y.