Questions 7: Expectation and Correlation

Roman Belavkin

Middlesex University

Question 1

Let variable x can have values 1, 2 and 3 with probabilities $P(1) = 1/5$, $P(2) = 3/5$ and $P(3) = 1/5$. What is the expected value of x? Compare it with mean value of $(1, 2, 2, 2, 3)$?

Answer: *The expected value is*

$$E\{x\} = 1\frac{1}{5} + 2\frac{3}{5} + 3\frac{1}{5} = 2$$

The mean value of $(1, 2, 2, 2, 3)$ *is*

$$\frac{1+2+2+2+3}{5} = 2$$

The mean value here is the same as the expected value. In fact, we could interpret the mean value of $(1, 2, 2, 2, 3)$ *as the expected value of* x *with values 1, 2, 2, 2 and 3 with all probabilities equal 1/5.*

Question 2

Consider the following sets of values for variables x and y:

x	y
-1	-2
0	0
1	2

Compute the expected values and variances of x and y. You can compute them as the means and the mean square deviations. Compare the results. Which variable is more uncertain (risky)?

Answer: *First, we compute the expected values as the means:*

$$E\{x\} = Mean(x) = \frac{-1+0+1}{3} = 0$$
$$E\{y\} = Mean(y) = \frac{-2+0+2}{3} = 0$$

Thus, x and y have equal expected values. Next, we compute the deviations from the mean

$$\text{for } x: \quad -1-0=-1\,, \quad 0-0=0\,, \quad 1-0=1$$
$$\text{for } y: \quad -2-0=-2\,, \quad 0-0=0\,, \quad 2-0=2$$

The variances we compute as the means of squared deviations

$$Var\{x\} = \frac{(-1)^2+0^2+1^2}{3} = \frac{2}{3}$$
$$Var\{y\} = \frac{(-2)^2+0^2+2^2}{3} = \frac{8}{3}$$

Thus, y has greater variance than x.

Although both variables have the same expected values (the means), their variances are different. Variable y is more uncertain (i.e. more risky) because it has a greater variance (higher dispersion).

Question 3

Compute the covariance and correlation between x and y from Question 9:

x	y
-1	-2
0	0
1	2

Are these variables correlated, uncorrelated or anticorrelated?

Answer: *From Question 9, the expected values are $E\{x\} = 0$ and $E\{y\} = 0$. The deviations are*

$$\text{for } x: \quad -1-0=-1\,, \quad 0-0=0\,, \quad 1-0=1$$
$$\text{for } y: \quad -2-0=-2\,, \quad 0-0=0\,, \quad 2-0=2$$

We compute the covariance as the mean of deviations multiplied together:

$$Cov(x,y) = \frac{(-1)(-2)+0\times 0+1\times 2}{3} = \frac{4}{3}$$

To find the correlation, we need simply to divide the covariance by the square root of the product of the variances (using results from Question 3.9, $Var(x) = \frac{2}{3}$ and $Var(y) = \frac{8}{3}$):

$$Corr(x, y) = \frac{4/3}{\sqrt{\frac{2}{3} \times \frac{8}{3}}} = \frac{4/3}{4/3} = 1$$

Thus, x and y are correlated.

Question 4

Suppose the database contains data for m independent variables. What should the covariance and the correlation matrices look like?

Answer: *Because all m variables are independent, their covariances as well as correlations should be zero. Thus, the covariance matrix will be*

$$\begin{pmatrix} Var(x_1) & 0 & \dots & 0 \\ 0 & Var(x_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & Var(x_m) \end{pmatrix}$$

and the correlation matrix will be

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Thus, all the elements apart from the diagonal will be zero.