LECTURE 13
COVARIANCE, CORRELATION, AND REGRESSION

1 Expectations with bivariate distributions

In the previous lecture, we extended the concept of distributions from the one-variable scenario to multivariate (bivariate) distributions. One can similarly extend the concept of the expected value or expectation from univariate distributions to multivariate.

Let X be a random variable with a PMF p_X and support S_X, or a PDF f_X. Recall that the expected value of $u(X)$, where $u(x)$ is some function is defined as

$$\begin{aligned} EX &= \sum_{x \in S_X} u(x) p_X(x), \text{ or} \\ EX &= \int_{-\infty}^{\infty} u(x) f_X(x) dx \end{aligned}$$

in the discrete and continuous cases respectively. In the bivariate case, the expectation is defined similarly, except the function u is a function with two arguments, and one must average across two variables using the joint PMF or PDF.

Definition 1. (Expected value) Let X and Y be two random variables with a joint PMF $p_{X,Y}(x,y)$ and supports S_X and S_Y, or a joint PDF $f_{X,Y}(x,y)$. The expected value of $U(X,Y)$ is defined as

$$EU(X,Y) = \sum_{x \in S_X} \sum_{y \in S_Y} U(x,y) p_{X,Y}(x,y)$$

in the discrete case, or

$$EU(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u(x,y) f_{X,Y}(x,y) dy dx$$

in the continuous case.

2 Covariance

One of the important examples is the covariance, which is a measure of association between two random variables.

Definition 2. (Covariance) Let X and Y be two random variables. The covariance between X and Y is defined as

$$Cov(X,Y) = E\left((X - EX)(Y - EY)\right).$$

Remark. 1. If X and Y are discrete with a joint PMF $p_{X,Y}$ and supports S_X, S_Y, then:

$$Cov(X,Y) = \sum_{x \in S_X} \sum_{y \in S_Y} (x - EX)(y - EY) p_{X,Y}(x,y).$$

Recall that EX and EY are two numbers. They must be computed before computing the covariance. In the continuous case, let $f_{X,Y}$ denote the joint PDF of X and Y:

$$Cov(X,Y) = \int \int (x - EX)(y - EY) f_{X,Y}(x,y) dy dx.$$

2. Covariance is a number that measures the co-variation or co-movement of two random variables. It is positive if the deviations of X and Y from their respective means tend to have the same sign: positive with positive and negative with negative. In this case, the two variables are positively related. The covariance is negative if the deviations of X and Y from their respective means tend to have opposite signs. In that case, the variables are negatively related.

Example. Consider the following example from Lecture 12.

Table 1: Example of a joint PMF

		Price of a share (S)		
		\$100	\$250	\$400
Earnings per share (E)	\$10	$\frac{2}{6}$	$\frac{1}{6}$	0
	\$20	0	$\frac{2}{6}$	$\frac{1}{6}$

The marginal PMF of the earnings per share (E) is $p_E(10) = p_E(20) = 0.5$, and therefore the mean of the earnings per share is $10 \times 0.5 + 20 \times 0.5 = 15$. The marginal PMF of the price (S) is $p_S(100) = 1/3$, $p_S(250) = 1/2$, $p_S(400) = 1/6$. Hence, the expected value of the price is

$$100 \times \frac{1}{3} + 250 \times \frac{1}{2} + 400 \times \frac{1}{6} = 225.$$

Next, the covariance between the two variables is

$$\begin{aligned}(10 - 15) \times (100 - 225) \times \frac{2}{6} + (10 - 15) \times (250 - 225) \times \frac{1}{6} \\ + (20 - 15) \times (250 - 225) \times \frac{2}{6} + (20 - 15) \times (400 - 225) \times \frac{1}{6} = 375.\end{aligned}$$

The covariance is positive and, therefore, the earnings per share and the price per share are positively associated.

Some properties of the covariance are given in the following theorem. Note that the theorem applies to both discrete and continuous cases, and the proof relies only on the linearity of the expectation.

Theorem 3. *(a)* $Cov(X,Y) = E(XY) - (EX)(EY)$.

(b) $Cov(X,X) = Var(X)$.

(c) $Cov(X,Y) = Cov(Y,X)$.

(d) Let a be a constant. $Cov(X,a) = 0$.

(e) Let a and b be two constants. $Cov(aX,bY) = abCov(X,Y)$.

(f) $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$.

(g) $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.

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

Proof. (a) First, consider

$$(X-EX)(Y-EY) = XY - X(EY) - (EX)(Y-EY).$$

Taking the expectations on both sides, we obtain

$$\begin{aligned} Cov(X,Y) &= E(XY) - E\left(X(EY)\right) - E\left((EX)(Y-EY)\right) \\ &= E(XY) - (EY)(EX) - (EX)E(Y-EY) \\ &= E(XY) - (EY)(EX), \end{aligned}$$

where the equality in the last line holds since $E(Y-EY) = 0$.

(b) Omitted.

(c) Omitted.

(d) Omitted.

(e) From the definition of the covariance,

$$\begin{aligned} Cov(aX,bY) &= E\left(aX - E(aX)\right)\left(bY - E(bY)\right) \\ &= E(aX - aEX)(bY - bEY) \\ &= abE(X-EX)(Y-EY) \\ &= abCov(X,Y). \end{aligned}$$

(f) From the definition of the covariance,

$$
\begin{aligned}
Cov(X+Y,Z) &= E\left[(X+Y-E(X+Y))\,(Z-EZ)\right] \\
&= E\left[(X-EX+Y-EY))\,(Z-EZ)\right] \\
&= E\left[(X-EX)(Z-EZ)+(Y-EY)(Z-EZ)\right] \\
&= E(X-EX)(Z-EZ)+E(Y-EY)(Z-EZ) \\
&= Cov(X,Z)+Cov(Y,Z),
\end{aligned}
$$

where the equality in the last line holds by the definition of the covariance.

(g) From the definition of the variance,

$$
\begin{aligned}
Var(X+Y) &= E\left[X+Y-E(X+Y)\right]^2 \\
&= E\left[X-EX+Y-EY\right]^2 \\
&= E\left[(X-EX)^2+(Y-EY)^2+2(X-EX)(Y-EY)\right] \\
&= E(X-EX)^2+E(Y-EY)^2+2E(X-EX)(Y-EY) \\
&= Var(X)+Var(Y)+2Cov(X,Y)
\end{aligned}
$$

where the equality in the last line holds by the definition of the covariance. □

3 Correlation coefficient

Covariance can be difficult to interpret since it can take any value from $-\infty$ to $+\infty$. For that reason and to obtain a unit-free measure of association[1], it is common to standardize the covariance by the product of standard deviations of the two variables. The resulting measure is called the *coefficient of correlation.*

Definition 4. (Correlation coefficient) The correlation coefficient between two random variables X and Y (denoted as $\rho_{X,Y}$) is defined as

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}.$$

The correlation coefficient is a unit-free measure. Hence, it is unaffected by the change of

[1] $Cov(X,Y)$ is measured in units of X $\times$ units of Y.

units of measurement. Consider the correlation between X and cY, where $c > 0$.

$$\begin{aligned}\rho_{X,cY} &= \frac{Cov(X, cY)}{\sqrt{Var(X)Var(cY)}} \\ &= \frac{cCov(X, Y)}{\sqrt{Var(X)c^2Var(Y)}} \\ &= \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \\ &= \rho_{X,Y}.\end{aligned}$$

More importantly, the correlation coefficient is restricted to the interval $[-1, 1]$ thus providing a very intuitive measure of the strength of association. A larger positive value of the correlation coefficient indicates a stronger positive association between two variables. Similarly, a smaller negative (closer to -1) value of the correlation coefficient indicates a stronger negative association between two variables.

Theorem 5. *The correlation coefficient between X and Y satistfies:*

(a) $-1 \leq \rho_{X,Y} \leq 1$.

(b) $\rho_{X,Y} = 1$ *if and only if* $Y = a + bX$ *for some constants* $a \in \mathbb{R}$ *and* $b > 0$.

(c) $\rho_{X,Y} = -1$ *if and only if* $Y = a + bX$ *for some constants* $a \in \mathbb{R}$ *and* $b < 0$.

Remark. When $\rho_{X,Y} = 1$ or -1, there is a *perfect linear* relationship between X and Y. Note that if there is a perfect nonlinear relationship between X and Y (such as $Y = X^2$), the correlation coefficient will be strictly between -1 and 1, i.e $-1 < \rho_{X,Y} < 1$. Thus, the correlation coefficient measures the degree of *linearity* in the relationship between two variables.

The proof of Theorem 5 requires another result.

Lemma 6. ***(Cauchy-Schwartz inequality)*** *Let X and Y be two random variables. Then,*

(a) $|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)}$.

(b) $|Cov(X, Y)| = \sqrt{Var(X)Var(Y)}$ *if and only if* $Y = a + bX$ *for some constants* a *and* b.

Proof. Define a new random variable

$$\begin{aligned}U &= Y - \frac{Cov(X, Y)}{Var(X)}X \\ &= Y - \beta X, \qquad (1)\end{aligned}$$

where

$$\beta = \frac{Cov(X, Y)}{Var(X)}.$$

By the results of Theorem 3(e) and (h),

$$\begin{aligned} Var(U) &= Var(Y - \beta X) \\ &= Var(Y) + \beta^2 Var(X) - 2\beta Cov(X, Y). \end{aligned}$$

Using the definition of β, we have:

$$\begin{aligned} Var(U) &= Var(Y) + \frac{(Cov(X,Y))^2}{(Var(X))^2} Var(X) - 2\frac{Cov(X,Y)}{Var(X)} Cov(X,Y) \\ &= Var(Y) - \frac{(Cov(X,Y))^2}{Var(X)}. \qquad (2) \end{aligned}$$

Since $Var(U) \geq 0$, it follows that

$$Var(Y) - \frac{(Cov(X,Y))^2}{Var(X)} \geq 0,$$

or

$$(Cov(X,Y))^2 \leq Var(X)Var(Y). \qquad (3)$$

Taking the square root of the both sides of the inequality in (3), we obtain:

$$|Cov(X,Y)| \leq \sqrt{Var(X)Var(Y)},$$

which completes the proof of part (a).

To show part (b), note that it follows from (2) that

$$(Cov(X,Y))^2 = Var(X)Var(Y)$$

if and only if $Var(U) = 0$. However, that implies that U is not a random variable, i.e. $U = a$ for some constant a. From the definition of U we obtain

$$a = Y - \beta X.$$

Hence, the second part of the lemma holds with $b = \beta = Cov(X,Y)/Var(X)$. □

The result of Theorem 5 follows immediately from the Cauchy-Schwartz inequality. The latter implies that

$$-\sqrt{Var(X)Var(Y)} \leq Cov(X,Y) \leq \sqrt{Var(X)Var(Y)}.$$

Dividing all sides of the inequality by $\sqrt{Var(X)Var(Y)}$, we obtain

$$-1 \leq \rho_{X,Y} \leq 1,$$

where the inequalities can hold as equalities if and only if $Y = a + \beta X$.

Example. Consider again the example in Table 1. The variance of earnings is

$$(10 - 15)^2 \times \frac{1}{2} + (20 - 15)^2 \times \frac{1}{2} = 25.$$

The variance of the price is

$$(100 - 225)^2 \times \frac{1}{3} + (250 - 225)^2 \times \frac{1}{2} + (400 - 225)^2 \times \frac{1}{6} = 10625.$$

We previously have computed that the covariance between the earnings and the price is 375. Hence, the correlation is equal to

$$\frac{375}{\sqrt{25 \times 10625}} \approx 0.73.$$

Hence, the two variables are positively associated.

Definition 7. (Uncorrelatedness) We say that X and Y are *uncorrelated* when $\rho_{X,Y} = 0$ ($Cov(X, Y) = 0$).

The meaning and implications of uncorrelatedness are discussed in the following section.

4 Regression and least squares.

We saw in Theorem 5 that the correlation coefficient equals 1 or -1 if and only if all possible realizations of the pair of random variables X and Y form a straight line with the slope $\beta = Cov(X, Y)/Var(X)$, i.e. with probability one, $Y = a + \beta X$. In that case, while still random, X and Y have a *perfect linear relationship*.

When the correlation coefficient is strictly between -1 and 1, it is impossible to draw a straight line connecting all possible realization of X and Y. As a matter of fact, there might not be a deterministic relationship between X and Y: there is no function $h(\cdot)$ such that $Y = h(X)$ with probability 1. Nevertheless, let's try to find a linear function that approximates the relationship between X and Y as good as possible (in the sense that will become clear immediately).

Let $a + bX$ be the approximating function, and therefore the approximation error is

$$\varepsilon = Y - a - bX.$$

Unless the correlation coefficient is exactly 1 or -1, it is impossible to find constants a and b so that ε is zero with probability one. Nevertheless, let's try to find a and b so that the variance of the approximation error is as small as possible. Note that small variability of the approximation error corresponds to a more accurate approximation. Hence, we are interested in solving the following problem:

$$\min_{a,b} E(Y-a-bX)^2.$$

This optimization problem is known as a *least squares* problem.

Definition 8. (Least squares and regression) The values of a and b that solve $\min_{a,b} E(Y-a-bX)^2$ are called the least squares coefficients. Let α and β denote the least squares coefficients (the values of a and b respectively that solve the least squares problem). A function $y=\alpha+\beta x$ is called the regression (or the regression line).

Remark. The regression line is the *best* linear approximation to the stochastic relationship between X and Y. It is best in the sense that the approximation error has the smallest variance.

Theorem 9. *The least squares coefficients are given by*

$$\begin{aligned} \beta &= \frac{Cov(X,Y)}{Var(X)}, \\ \alpha &= (EY)-\beta(EX). \end{aligned} \tag{4}$$

Proof. We can find the least squares coefficients by solving the first-order conditions for the least squares problem. Suppose that we can change the order of differentiation and integration.

$$\begin{aligned} \frac{\partial}{\partial b}E(Y-a-bX)^2 &= E\left[\frac{\partial}{\partial b}(Y-a-bX)^2\right] \\ &= -2\left[E(Y-a-bX)X\right]. \\ \frac{\partial}{\partial a}E(Y-a-bX)^2 &= E\left[\frac{\partial}{\partial a}(Y-a-bX)^2\right] \\ &= -2E(Y-a-bX). \end{aligned}$$

Hence, α and β must satisfy:

$$E\left[(Y-\alpha-\beta X)X\right] = 0, \tag{5}$$
$$E(Y-\alpha-\beta X) = 0. \tag{6}$$

From the second equation, it follows that

$$EY-\alpha-\beta EX=0,$$

which proves (4).

Next, substitute the expression for α into 5. We have

$$\begin{aligned}
0 &= E\left[(Y-\alpha-\beta X)\,X\right] \\
&= E\left[(Y-(EY-\beta EX)-\beta X)\,X\right] \\
&= E\left[((Y-EY)-\beta(X-EX))\,X\right] \\
&= E\left[(Y-EY)X-\beta(X-EX)X\right] \\
&= E\left[(Y-EY)X\right]-\beta E\left[(X-EX)X\right].
\end{aligned}$$

Hence,

$$\beta=\frac{E\left[(Y-EY)X\right]}{E\left[(X-EX)X\right]}.$$

However,

$$\begin{aligned}
E\left[(X-EX)X\right] &= EX^2-(EX)EX \\
&= Var(X).
\end{aligned}$$

Also,

$$\begin{aligned}
E\left[(Y-EY)X\right] &= EYX-(EY)EX \\
&= Cov(X,Y).
\end{aligned}$$

□

Remark. 1. The regression line goes through the point given by the expected values of X and Y, since by (4),

$$EY=\alpha+\beta EX.$$

2. The least squares error is

$$\varepsilon=Y-\alpha-\beta X.$$

Hence, we can write

$$Y=\alpha+\beta X+\varepsilon.$$

The first-order condition (6) implies that

$$E\varepsilon=0.$$

Hence, the random approximation error has zero mean. The first-order condition (5) implies that the approximation error is *uncorrelated* with X:

$$\begin{aligned} 0 &= E\varepsilon X \\ &= Cov(\varepsilon, X), \end{aligned}$$

where the second line follows from $Cov(\varepsilon, X) = E\varepsilon X - (E\varepsilon)(EX) = E\varepsilon X$, since the mean of ε is zero.

3. When X and Y are uncorrelated, $\rho_{X,Y} = 0$ and $Cov(X, Y) = 0$, and therefore,

$$\beta = \frac{Cov(X,Y)}{Var(X)} = 0.$$

In that case, the best linear approximation to the relationship between X and Y is a flat line (with zero slope). Note that uncorrelatedness does not necessarily mean that X and Y are unrelated. It is possible that the relationship between X and Ycan be described by a non-trivial nonlinear function (and therefore the two random variables are related). However, the best linear approximation to this relationship is a flat line.

4. The relationship between positively correlated variables is best approximated by a regression line with a positive slope. The relationship between negatively correlated variables is best approximated by a regression line with a negative slope.

5 Expectations and independence

Suppose that X and Y are continuously distributed, and let $f_{X,Y}$ be the joint PDF of X and Y. Recall that

$$EXY = \int\int xyf_{X,Y}(x,y)dydx.$$

Suppose that X and Y are statistically independent: for any x and y,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y).$$

In that case,

$$\begin{aligned} EXY &= \int\int xyf_X(x)f_Y(y)dydx \\ &= \int xf_X(x)\left[\int yf_Y(y)dy\right]dx \\ &= \int xf_X(x)E(Y)dx \\ &= (EY)\int xf_X(x)dx \\ &= (EY)(EX). \end{aligned}$$

Thus, under independence the expected value of a product of random variables is equal to the product of expectations. In particular, this implies that under independence,

$$Cov(X,Y) = EXY - (EX)(EY) = 0.$$

Theorem 10. *If X and Y are independent, then X and Y are uncorrelated.*

Remark. In general, uncorrelatedness does *not* imply independence.

The result can be extended to functions of random variables. Recall that by Theorem 9 in Lecture 12, functions of random variables are independent if the random variables are independent. Therefore, when X and Y are independent,

$$E[U(X)V(Y)] = E[U(X)]E[V(Y)].$$

The same results also hold for discrete random variables. The proof in the discrete case is identical after replacing integrals with sums and PDFs with PMFs.