In lecture 5, we talked about two fundamental quantities to characterize the behavior of a random variable: the expected value and the variance. We now generalize these concepts to the situations we introduced in the previous lecture, namely when we are dealing with two (or more) random variables at the same time.

1 Expected value

1.1 Discrete random variables

Remember that if a discrete random variable X takes the values $\{a_1, a_2, \ldots, a_n\}$, with probabilities $p_X(a_1), p_X(a_2), \ldots, p_X(a_n)$, then

$$E[X] = \sum_{i=1}^{n} a_i p_X(a_i)$$

In other words, $E[X]$ is a weighted sum of the values X can take, where the weights are the probabilities $p_X(a_i)$. This motivates the general definition below, for a discrete random variable Z defined in terms of discrete random variables X and Y by $Z = g(X, Y)$, where g is a real-valued function of 2 variables, $g : \mathbb{R}^2 \to \mathbb{R}$.

Definition: Let X and Y be two discrete random variables with values $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$ respectively. Let $g : \mathbb{R}^2 \to \mathbb{R}$ be a function. Then $Z = g(X, Y)$ is a discrete random variable with expectation

$$\boxed{E[Z] = E[g(X,Y)] = \sum_j \sum_i g(a_i, b_j) p_{X,Y}(a_i, b_j) = \sum_j \sum_i g(a_i, b_j) P(X = a_i, Y = b_j)} \tag{1}$$

where $p_{X,Y}$ is the joint probability mass function of X and Y.

Example: Let X be a discrete random variable with Bernoulli distribution with parameter p, and Y be a discrete random variable with Bernoulli distribution with parameter q. Let $g(x, y) = xy$, and we consider the discrete random variables $Z = g(X, Y) = XY$.

$$\begin{aligned} E[Z] &= g(0,0)(1-p)(1-q) + g(1,0)p(1-q) + g(0,1)(1-p)q + g(1,1)pq \\ &= 0 \cdot (1-p)(1-q) + 0 \cdot p(1-q) + 0 \cdot (1-p)q + 1 \cdot pq \\ &= pq \end{aligned}$$

We note that for this example we implicitly assumed that X and Y were independent in order to be able to easily construct, in passing, the joint probability mass function $p_{X,Y}$.

1.2 Continuous random variables

As always in this course, the definition for the case of continuous random variables can be deduced from the case of discrete random variables, via the correspondence

$$\sum \;\leftrightarrow\; \int \qquad \text{and} \qquad p_{X,Y} \;\leftrightarrow\; f_{X,Y}$$

Definition: Let X and Y be two continuous random variables, and $g : \mathbb{R}^2 \to \mathbb{R}$ a real-valued function of two variables. $Z = g(X, Y)$ is a continuous random variable, with expectation

$$\boxed{E[Z] = E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{X,Y}(x, y) dx dy} \tag{2}$$

where $f_{X,Y}$ is the joint probability density function of X and Y.

Example: Let $P = (X, Y)$ be a point chosen at random in the unit square $[0, 1] \times [0, 1]$. The continuous random variable X is the x-coordinate of P, the continuous random variable Y its y-coordinate.

Since the point is chosen at random, the joint probability density function of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } 0 \le x \le 1\ ,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $Z = X^2 + Y^2$ the continuous random variable corresponding to the squared distance of the point P to the origin $(0,0)$.

$$\begin{aligned} E[Z] = E[X^2+Y^2] &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x^2+y^2) f_{X,Y}(x,y)dxdy \\ &= \int_0^1\int_0^1 (x^2+y^2)dxdy = \int_0^1 \left[\frac{x^3}{3} + xy^2\right]_0^1 dy \\ &= \int_0^1 \left(\frac{1}{3} + y^2\right) dy = \left[\frac{y}{3} + \frac{y^3}{3}\right]_0^1 \\ &= \frac{2}{3} \end{aligned}$$

1.3 Linearity of expectation

Let us here focus on the particular situation for which $g(x,y) = x + y$. If X and Y are continuous random variables and $Z = X + Y$, then

$$\begin{aligned} E[Z] = E[X+Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[x f_{X,Y}(x,y) + y f_{X,Y}(x,y)\right] dxdy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dydx + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy \\ &= \int_{-\infty}^{+\infty} x \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy\right) dx + \int_{-\infty}^{+\infty} y \left(\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx\right) dy \\ &= \int_{-\infty}^{+\infty} x f_X(x)dx + \int_{-\infty}^{+\infty} y f_Y(y)dy = E[X] + E[Y] \end{aligned}$$

We just proved that for any two continuous random variables X and Y,

$$\underline{E[X+Y] = E[X] + E[Y]}$$

The result also holds for *discrete* random variables, and the proof of it is very similar. You can find it in our textbook.

Remember that in lecture 5, we proved the linearity of expectation for a single random variable X, and showed that if a and b are real numbers, and X is a random variable, then

$$E[aX + b] = aE[X] + b$$

Combining the two results into one, we obtain the general form for the linearity of expectation, which we state explicitly below.

Let X and Y be two random variables, and a, b, and c real numbers. Then

$$\boxed{E[aX + bY + c] = aE[X] + bE[Y] + c} \tag{3}$$

Examples:

- We roll two fair dice, and let X_1 be the random variable corresponding to the result of the first dice, and X_2 be the random variable corresponding to the result of the second dice.

As we have already seen in Lecture 5 (page 2),

$$E[X_1] = E[X_2] = 3.5$$

Let

$$Y = X_1 + X_2$$

be the random variable corresponding to the sum of the results of the two dice. By linearity of the expectation,

$$E[Y] = E[X_1 + X_2] = E[X_1] + E[X_2] = 7$$

Using page 1 of Lecture 7, we can verify this result by the following direct computation:

$$E[Y] = 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36} + 7\cdot\frac{6}{36} + 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} + 10\cdot\frac{3}{36} + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} = \frac{252}{36} = 7$$

- Consider the example in pages 1 and 2, with the point $P = (X, Y)$ chosen at random.

  We have

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)dx$$

  To compute this, we need f_X. We construct f_X following what we learned in Lecture 7:

$$\forall\, x \in [0,1]\ ,\ f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \int_0^1 1dy = 1$$

  and

$$\forall\, x \notin [0,1],\ f_X(x) = 0$$

  since $f_{X,Y}(x,y) = 0$.

  Thus,

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)dx = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3}$$

  Since the roles of X and Y are interchangeable in $f_{X,Y}$, a similar calculation would yield

$$E[Y^2] = \frac{1}{3}$$

  Hence

$$E[X^2] + E[Y^2] = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

  In page 2, we found $E[X^2 + Y^2] = \frac{2}{3}$, so we indeed have

$$E[X^2 + Y^2] = E[X^2] + E[Y^2]$$

2 Covariance

2.1 Definition

A natural follow-up question to the previous section is whether one also has linearity of variances:

$$\mathrm{Var}(X+Y) \overset{?}{=} \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Let us see if this result can hold. For any two random variables X and Y,

$$\begin{aligned}
\mathrm{Var}(X+Y) &= E[((X+Y) - E[X+Y])^2] = E[((X+Y) - E[X] - E[Y])^2] = E[((X - E[X]) + (Y - E[Y]))^2] \\
&= E[(X - E[X])^2 + (Y - E[Y])^2 + 2(X - E[X])(Y - E[Y])] \\
&= \underbrace{E[(X - E[X])^2]}_{\mathrm{Var}(X)} + \underbrace{E[(Y - E[Y])^2]}_{\mathrm{Var}(Y)} + 2E[(X - E[X])(Y - E[Y])]
\end{aligned}$$

where we used the linearity of expectation for the last equality.

We therefore just derived the following result:

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2E[(X-E[X])(Y-E[Y])]$$

Let us now see if the third term is zero in general. Let us consider two discrete random variables X and Y with the following joint probability mass function:

X \ Y	-1	2	p_X
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
p_Y	$\frac{1}{2}$	$\frac{1}{2}$	1

We can readily compute the following quantities

$$E[X] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1 \qquad , \qquad E[Y] = (-1) \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{1}{2}$$

Then

$$\begin{aligned} E[(X-E[X])(Y-E[Y])] &= (0-1)(-1-\frac{1}{2}) \cdot \frac{1}{4} + (0-1)(2-\frac{1}{2}) \cdot 0 + (1-1)(-1-\frac{1}{2}) \cdot \frac{1}{4} \\ &+ (1-1)(2-\frac{1}{2}) \cdot \frac{1}{4} + (2-1)(-1-\frac{1}{2}) \cdot 0 + (2-1)(2-\frac{1}{2}) \cdot \frac{1}{4} = \frac{3}{8} + \frac{3}{8} = \frac{3}{4} \end{aligned}$$

We thus see that in general, $E[(X-E[X])(Y-E[Y])] \neq 0$. This quantity, which measures the amount to which $\mathrm{Var}(X+Y)$ differs from $\mathrm{Var}(X)+\mathrm{Var}(Y)$ plays an important role in probability and statistics, and is called the *covariance*.

Definition: Let X and Y be two random variables. The covariance between X and Y is defined by

$$\boxed{\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])]} \tag{4}$$

We thus see that we have the relation:

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)}$$

from which we understand that the covariance is a *measure of the joint variability* of the two random variables X and Y.

$\mathrm{Cov}(X,Y)$ will be positive if, on average, when X has a realization larger than $E[X]$, Y also has a realization larger than $E[Y]$; or when X has a realization lower than $E[X]$, Y has a realization lower than $E[Y]$. X and Y are then said to be *positively correlated*.

In the opposite situation (X larger likely implies Y lower, or X lower likely implies Y larger), the two random variables are said to be *negatively correlated*.

2.2 Alternative expression for the covariance

Using the linearity of expectations, you will derive in Recitation the following useful alternative expression for $\mathrm{Cov}(X,Y)$:

$$\boxed{\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]} \tag{5}$$

Illustration: For the discrete random variables X and Y discussed just above, we had $E[X]=1$, $E[Y]=\frac{1}{2}$, and $\mathrm{Cov}(X,Y)=\frac{3}{4}$. Let us now verify this with the alternative formula:

$$E[XY] = (-1) \cdot 0 \cdot \frac{1}{4} + 2 \cdot 0 \cdot 0 - 1 \cdot 1 \cdot \frac{1}{4} + 2 \cdot 1 \cdot \frac{1}{4} - 1 \cdot 2 \cdot 0 + 2 \cdot 2 \cdot \frac{1}{4} = \frac{5}{4}$$

$$E[XY] - E[X]E[Y] = \frac{5}{4} - 1 \cdot \frac{1}{2} = \frac{3}{4}$$

as desired.

2.3 Independent vs. uncorrelated

We have seen that in general, $\text{Cov}(X, Y) \neq 0$. Hence, **in general** $E[XY] \neq E[X]E[Y]$.

When X and Y are independent, however, equality holds.

This is easy to see, in the discrete case for example:

$$\begin{aligned}
E[XY] &= \sum_j \sum_i a_i b_j P(X = a_i, Y = b_j) \\
&= \sum_j \sum_i a_i b_j P(X = a_i) P(Y = b_j) \qquad \text{(because X and Y are independent)} \\
&= \underbrace{\left(\sum_j b_j P(Y = b_j) \right)}_{E[Y]} \cdot \underbrace{\left(\sum_i a_i P(X = a_i) \right)}_{E[X]} = E[Y]E[X]
\end{aligned}$$

Hence, **if X and Y are independent**, $\text{Cov}(X, Y) = 0$. We say **that X and Y are uncorrelated**. Independent random variables are uncorrelated, which makes intuitive sense.

Be careful: the reverse is not necessarily true: uncorrelated random variables can be dependent.

As an example, consider two discrete random variables X and Y with the following joint probability mass function:

X \ Y	-1	2	p_X
0	$\frac{1}{6}$	0	$\frac{1}{6}$
1	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{2}{3}$
2	$\frac{1}{6}$	0	$\frac{1}{6}$
p_Y	$\frac{1}{2}$	$\frac{1}{2}$	1

$P(X = 0, Y = -1) = \frac{1}{6} \neq p_X(0) \cdot p_Y(-1) = \frac{1}{12}$ so X and Y are not independent.

$$\begin{aligned}
E[X] &= 0 \cdot \frac{1}{6} + 1 \cdot \frac{2}{3} + 2 \cdot \frac{1}{6} = 1 \\
E[Y] &= (-1) \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{1}{2} \\
E[XY] &= 0 \cdot (-1) \cdot \frac{1}{6} + 0 \cdot 2 \cdot 0 + 1 \cdot (-1) \cdot \frac{1}{6} + 1 \cdot 2 \cdot \frac{1}{2} + 2 \cdot (-1) \cdot \frac{1}{6} + 2 \cdot 2 \cdot 0 = -\frac{1}{2} + 1 = \frac{1}{2}
\end{aligned}$$

Therefore,

$$\text{Cov}(X, Y) = \frac{1}{2} - 1 \cdot \frac{1}{2} = 0$$

The random variables X and Y are not independent, but they are uncorrelated.

2.4 Covariance under change of units

Let X and Y be two random variables, and a, b, c, and d be real numbers.

$$\begin{aligned}
\text{Cov}(aX + b, cY + d) &= E[(aX + b)(cY + d)] - E[aX + b]E[cY + d] \\
&= E[acXY + adX + bcY + db] - (aE[X] + b)(cE[Y] + d) \\
&= acE[XY] + adE[X] + bcE[Y] + bd - acE[X]E[Y] - adE[X] - bcE[Y] - bd \\
&= ac(E[XY] - E[X]E[Y]) = ac\text{Cov}(X, Y)
\end{aligned}$$

We just proved that

$$\boxed{\text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y)}$$

2.5 The correlation coefficient

The covariance provides a measure of whether two random variables X and Y are correlated, but a weakness of this measure is that its magnitude depends on the units chosen for X and Y, as can be seen from the relation derived in the previous section. According to this relation, air temperature in Manhattan and ice thickness on Turtle Pond in Central Park will have a larger covariance (if it is not zero) if they are measured in Fahrenheit and inches instead of Celsius and centimeters.

To address this weakness, we introduce the *correlation coefficient* $\rho(X,Y)$, which measures correlation in a way which is independent of the system of units chosen.

Definition: Let X and Y be two random variables. The **correlation coefficient** $\rho(X,Y)$ is defined by

$$\boxed{\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \qquad \text{if } \mathrm{Var}(X) \neq 0\,, \mathrm{Var}(Y) \neq 0 \qquad , \quad \rho(X,Y) = 0 \ \text{ otherwise}} \tag{6}$$

Example: For the discrete random variables X and Y discussed at the top of page 4, we have

$$\mathrm{Var}(X) = (0-1)^2 \cdot \frac{1}{4} + (1-1)^2 \cdot \frac{1}{2} + (2-1)^2 \cdot \frac{1}{4} = \frac{1}{2}$$
$$\mathrm{Var}(Y) = (-1-\frac{1}{2})^2 \cdot \frac{1}{2} + (2-\frac{1}{2})^2 \cdot \frac{1}{2} = \frac{9}{4}$$

We had also found that $\mathrm{Cov}(X,Y) = \frac{3}{4}$. Hence,

$$\rho(X,Y) = \frac{\frac{3}{4}}{\sqrt{\frac{9}{8}}} = \frac{\sqrt{2}}{2}$$

It can be shown, as is for example done in the class textbook, that one always has

$$\underline{-1 \leq \rho(X,Y) \leq 1}$$

Two random variables are most correlated if $\rho(X,Y) = 1$ or $\rho(X,Y) = -1$. This is the case when $X = Y$ or $X = -Y$, which makes intuitive sense.