Many Random Variables

So far we have mostly dealt with a single real-valued random variable. We now look at the study of a collection of random variables. We can think of this as a single vector valued random variable $\mathbf{X} = (X_1, \ldots, X_n)$. However, what is the distribution of this vector valued random variable and how is this related to the distribution of the individual real valued random variables?

In what follows, we will give proofs using discrete random variables but the proofs can be taken over to arbitrary random variables using limiting arguments.

Expectation

We calculate the expecation of a sum $X + Y$ of two random variables X and Y.

$$\begin{aligned}
E(X+Y) &= \sum_{a\in D;b\in D'} (a+b)P(X=a;Y=b) \\
&= \sum_{a\in D;b\in D'} aP(X=a;Y=b) + \sum_{a\in D;b\in D'} bP(X=a;Y=b)
\end{aligned}$$

The first sum on the right hand side can be written as a pair of summations

$$\sum_{a\in D;b\in D'} aP(X=a;Y=b) = \sum_{a\in D} a \left(\sum_{b\in D'} P(X=a;Y=b) \right)$$

Note that $Y=b$ are mutually exclusive events so that *exactly* one of them must occur. Hence

$$(X=a) = \vee_{b\in D} \left((X=a) \cap (Y=b) \right)$$

is a disjoint union. It follows that

$$P(X=a) = \sum_{b\in D'} P(X=a;Y=b)$$

A different way to see this is as follows. Now, $Y = b$ are mutually exclusive events for different $b \in D'$. Moreover, $1 = \sum_{b\in D'} P(Y=b)$. The decomposition law states that for mutually exclusive events B_b so that $\sum_b P(B_b) = 1$, we have $P(A) = \sum_n P(A \cap B_n)$. We then obtain

$$\sum_{a\in D} a \left(\sum_{b\in D'} P(X=a;Y=b) \right) = \sum_{a\in D} aP(X=a) = E(X)$$

Similarly, the second sum gives us $E(Y)$.

We deduce that for any (finite) collection of random variables $E(X_1+\cdots+X_n) = E(X_1) + \cdots + E(X_n)$; the expectation of the sum is the sum of the expecations.

Covariance

We can try to apply a similar idea to the problem of determining $E(XY)$.

$$E(X \cdot Y) = \sum_{a\in D; b\in D'} (ab)P(X=a; Y=b)$$

$$= \sum_{a\in D} a \left(\sum_{b\in D'} bP(X=a; Y=b) \right)$$

The problem is that we do not appear to have any control over the latter sum since the "b is inside"! So we seem to need the identity

$$P(X=a; Y=b) = P(X=a)P(Y=b)$$

Recall that this will follow if $X=a$ and $Y=b$ are *independent* events. *If* this is the case, then the above sum simplifies

$$\sum_{a\in D} a \left(\sum_{b\in D'} bP(X=a; Y=b) \right) =$$

$$\sum_{a\in D} a \left(\sum_{b\in D'} bP(X=a)P(Y=b) \right) =$$

$$\sum_{a\in D} aP(X=a) \left(\sum_{b\in D'} bP(Y=b) \right) =$$

$$E(X)E(Y)$$

However, *without* that crucial ingredient, we do not have the the identity $E(XY) = E(X)E(Y)$. More generally, the difference $\mathrm{C}ov(X,Y) = E(XY) - E(X)E(Y)$ is called the *Covariance* of the random variables X and Y.

For any constants a and b we can apply the additivity of expectations to obtain

$$E((aX+bY)^2)) = a^2E(X^2) + 2abE(XY) + b^2E(Y^2)$$

On the other hand

$$E((aX+bY))^2 = (aE(X)+bE(Y))^2 = a^2E(X)^2 + 2abE(X)E(Y) + b^2E(Y)^2$$

Hence,

$$\sigma^2(aX+bY) = a^2\sigma^2(X) + 2ab\mathrm{C}ov(X,Y) + b^2\sigma^2(Y)$$

Since $\sigma^2(Z) \geq 0$ for any real valued random variable, we see that $\sigma^2(aX+bY) \geq 0$ for all a and b. It follows (by completing the square) that

$$\mathrm{C}ov(X,Y)^2 \leq \sigma^2(X)\sigma^2(Y)$$

It is thus useful to think of the *correlation* which is defined as

$$\rho(X,Y) = \frac{\mathrm{C}ov(X,Y)}{\sigma(X)\sigma(Y)}$$

as the *cosine* of an "angle" between X and Y (this makes sense only if $\sigma(X)$ and $\sigma(Y)$ are non-zero!). If the angle is *acute* then X "pulls Y towards it" and otherwise, it "pushes it away". In both cases, X and Y are "correlated". On the other hand, if X and if the covariance is 0, then the "angle" is a right angle and we can say that X and Y are "un-correlated".

Independence

Two random variables X and Y are said to be independent of each other if the events $X \leq x$ and $Y \leq y$ are independent:

$$P(X \leq x; Y \leq y) = P(X \leq x)P(Y \leq y)$$

We note that the usual decomposition of probabilities gives us

$$\begin{aligned} &P(a < X \leq b; c < Y \leq d) = \\ &\qquad P(X \leq b; Y \leq d) - P(X \leq a; Y \leq d) \\ &\qquad\qquad - P(X \leq b; Y \leq c) + P(X \leq a; Y \leq c) \end{aligned}$$

If the random variables are independent, we see that this gives

$$
\begin{aligned}
P(a < X \leq b; c < Y \leq d) = \\
&P(X \leq b)P(Y \leq d) - P(X \leq a)P(Y \leq d) \\
&\qquad - P(X \leq b)P(Y \leq c) + P(X \leq a)P(Y \leq c)
\end{aligned}
$$

Now the right-hand side is the same as

$$
\begin{aligned}
(P(X \leq b) - P(X \leq a)) \cdot (P(Y \leq d) - P(Y \leq c)) \\
= P(a < X \leq b) \cdot P(c < Y \leq d)
\end{aligned}
$$

So we can re-state the independence of X and Y as

$$
P(a < X \leq b; c < Y \leq d) = P(a < X \leq b) \cdot P(c < Y \leq d)
$$

By the result on Covariance above, we see that if X and Y are independent, then C$ov(X,Y)=0$. *Warning*: The converse is not necessarily true!

More generally, we can define a finite collection of random variables X_i for $i=1,\ldots,n$ to be independent of

$$
P(X_1 \leq a_1; \ldots; X_n \leq a_n) = P(X_1 \leq a_1) \cdots P(X_n \leq a_n)
$$

Warning: Note that if X is idependent of Y and Y is independent of Z, then it does not follow that X is independent of Z; for example, X and Z could be the same variable in which case they are correlated!

Warning: Just because the *distribution* of two variables is different, it does *not* mean that they are independent. In many cases, the distributions of X and X^2 are quite different, however, they are not independent.

Warning: Just because the *distribution* of two variables is the same it does not mean that they are *not* independent. In fact, when we try to carry out the same experiment a number of times, we(often) *want* the result of each experiment to be independent *and* identically distributed. A number of questions in probability deal with Independent, Identically Distributed (or i.i.d.) random variables.

An important consequence of independence of the random variables X_i is that in this case

$$
\sigma^2(X_1 + \cdots + X_n) = \sigma^2(X_1) + \cdots + \sigma^2(X_n)
$$

We will use this identity in what follows.