

4 — MEANS AND VARIANCES

The term expectation (or mean) has so far been confined to a single random variable. Related to expectation is the term *variance* and this will be discussed before two important new distributions are introduced. Expectation and variance will then be applied to cases where there are two or more random variables.

Derived Random Variables

This course follows the common practice of using X and r to refer to the name and value of a single random variable. The most-frequently used example of a random variable has been the outcome of throwing a die.

If two people are betting on the outcome of throwing a die and wish to make matters (ever so slightly) more interesting they might decide to bet on some *function* of the outcome instead of the outcome itself. They might perhaps bet on the square of the outcome, the sine of the outcome, seven more than the outcome, two to the power of the outcome and so on.

Such a decision does not affect the number of possible outcomes or the values of the outcomes themselves. What is affected is the values the people bet on and, of much greater interest, the long-term average of the values bet on.

Any function of some random variable X is itself a random variable; call it Y where:

$$Y = f(X)$$

Subject to certain common-sense constraints f is an arbitrary function and Y is called a *derived random variable*. If f is like the square function it will lead to integer values. If it is like the sine function it will not and various issues arise which will not be addressed here.

Generalised Expectation

Fortunately the expectation of some function $f(X)$ of a random variable X is a trivial generalisation of (3.1):

$$E(f(X)) = \sum_r f(r) \cdot P(X = r) \quad (4.1)$$

This can be derived in exactly the same way as the simple expectation by considering the position of the centre of gravity of a light beam supporting weights. Suppose the bets are on the square of the outcome of throwing a die. The weights are just the same but they are placed at positions 0, 1, 4, 9, 16, 25 and 36 along the beam.

The analysis gives rise to the expectation:

$$E(X^2) = 0 \cdot \frac{1}{6} + 1 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 9 \cdot \frac{1}{6} + 16 \cdot \frac{1}{6} + 25 \cdot \frac{1}{6} + 36 \cdot \frac{1}{6} = \frac{91}{6} = 15 \frac{1}{6}$$

Many functions of random variables will be used in this course but the square function and one closely related to the square are undoubtedly the most important.

Variance

Suppose that, instead of simply squaring the outcome of each throw, the players first subtract the mean ($3\frac{1}{2}$ of course) and *then* square. The expectation of the values bet on now is:

$$E((X - 3\frac{1}{2})^2) = \sum_r (r - 3\frac{1}{2})^2 \cdot P(X = r)$$

This value is known as the variance (usually denoted by the letter V) or *mean squared deviation from the mean* usually denoted by σ^2 . In general:

$$\text{Variance} = \sigma^2 = V(X) = E((X - \mu)^2) = \sum_r (r - \mu)^2 \cdot P(X = r) \quad (4.2)$$

The item $V(X)$ is pronounced ‘the variance of X ’. The sum over r again refers to the range which is appropriate.

The variance gives a measure of the spread of values from the mean. Here is a table showing how the variance of the outcomes of throwing a die may be calculated:

| r | $r - \mu$ | $(r - \mu)^2$ | $P(X = r)$ | $(r - \mu)^2 \cdot P(X = r)$ |
|-----|----------------|----------------|---------------|------------------------------|
| 0 | $-\frac{7}{2}$ | $\frac{49}{4}$ | $\frac{0}{6}$ | 0 |
| 1 | $-\frac{5}{2}$ | $\frac{25}{4}$ | $\frac{1}{6}$ | $\frac{25}{24}$ |
| 2 | $-\frac{3}{2}$ | $\frac{9}{4}$ | $\frac{1}{6}$ | $\frac{9}{24}$ |
| 3 | $-\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{6}$ | $\frac{1}{24}$ |
| 4 | $\frac{1}{2}$ | $\frac{1}{4}$ | $\frac{1}{6}$ | $\frac{1}{24}$ |
| 5 | $\frac{3}{2}$ | $\frac{9}{4}$ | $\frac{1}{6}$ | $\frac{9}{24}$ |
| 6 | $\frac{5}{2}$ | $\frac{25}{4}$ | $\frac{1}{6}$ | $\frac{25}{24}$ |

The sum of the entries in the rightmost column is $\frac{70}{24}$ or $\frac{35}{12}$.

Standard Deviation

If expectation is related to the idea of a centre of gravity, the variance is related to the idea of moment of inertia.

To produce a measure of the spread of values from the mean which has the same dimension as values themselves it is common to use the square root of the variance and this is known as the *standard deviation* (denoted by σ , which is $\sqrt{\sigma^2}$ of course):

$$\text{Standard Deviation} = \sigma = \sqrt{\text{Variance}}$$

In the case of the die:

$$\text{Standard Deviation} = \sigma = \sqrt{\frac{70}{24}} \approx 1.71$$

The Geometric Distribution

A very common two-state entity to which the Binomial distribution is applied is any experiment or trial whose outcome is regarded as a success or a failure with no other possibility.

A special case of this is waiting for a No. 9 Bus. Each bus that stops at your stop is regarded as a trial. If the bus is a No. 9 the trial counts as a success and any bus which *isn't* a No. 9 is regarded as a failure.

This is a perfectly ordinary example of the Binomial distribution if you simply note success or failure for each of n buses. If you are hoping to *catch* a No. 9 Bus the circumstances are different. As soon as a No. 9 Bus comes to your stop you get on board and the sequence of trials abruptly terminates!

Let p be the probability of a single trial being a success and X be the random variable which represents the number of failures before the first success (at which point matters conclude). The probability of having to wait r trials before the first success is:

$$P(X = r) = (1 - p)^r p$$

In this, $1 - p$ is the probability of a trial being a failure and the probability of r such trials is $(1 - p)^r$. There has then to be a success and this explains the final p .

Here is an indexed set of probabilities. The sum happens to be 1 but that has yet to be demonstrated. This is known as the *Geometric distribution*.

Note that the range of r runs indefinitely upwards from zero. When $r = 0$ there are no failures before the first success. As r increases you have to sustain more and more failures before the first success and, in principle, there is no limit to the number of failures.

The sum of the probabilities is therefore a summation to infinity. The summation is valid because $(1 - p) < 1$:

$$\begin{aligned} \sum_{r=0}^{\infty} P(X = r) &= (1 - p)^0 p + (1 - p)^1 p + (1 - p)^2 p + \dots \\ &= ((1 - p)^0 + (1 - p)^1 + (1 - p)^2 + \dots) p = \frac{1}{1 - (1 - p)} p = 1 \end{aligned}$$

The Geometric distribution satisfies the informal requirement of being an indexed set of probabilities whose sum is 1. As with other distributions, the Geometric distribution is a family of distributions but this one has only one parameter. The description:

$$\text{Geometric}(p)$$

is used to refer to the general case.

Note that the sum to some lesser limit, k say, represents the probability of having *no more than* k failures before the first success:

$$\begin{aligned} \sum_{r=0}^k P(X = r) &= (1-p)^0 p + (1-p)^1 p + \cdots + (1-p)^k p \\ &= ((1-p)^0 + (1-p)^1 + \cdots + (1-p)^k) p \\ &= \frac{1 - (1-p)^{k+1}}{1 - (1-p)} p = 1 - (1-p)^{k+1} \end{aligned}$$

The Poisson Distribution

Suppose that over a long period of time a town has recorded an average of λ murders a year. Sometimes there is a run of good years in which there are no murders at all and then there may be a couple of bad years with several. How does one estimate the probability of there being exactly two murders next year?

This problem introduces the *Poisson distribution*. The analysis begins by taking λ as the expectation, the expected (or average) number of murders in a year.

Divide the year into n equal intervals (365 might be a sensible value for n) and assume that in any given interval the number of murders is zero or one. Given this assumption, the probability of there being a murder in any given interval is $\frac{\lambda}{n}$.

Having chosen n intervals in a full year, let X be the random variable which represents the number of murders in these n intervals. Subject to the assumption, the Binomial distribution gives the probability of there being r murders in the n intervals.

The assumption is important. If there were sometimes two murders in an interval the Trinomial distribution would be required. To ensure that the assumption is valid, n must be made sufficiently large that the intervals are sufficiently small for the possibility of two or more murders to be ignored.

Noting the assumption, the probability of there being r murders in the n intervals is:

$$\begin{aligned} P(X = r) &= \binom{n}{r} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \\ &= \frac{n!}{r!(n-r)!} \left(\frac{\lambda}{n}\right)^r \left(1 - \frac{\lambda}{n}\right)^{n-r} \\ &= \frac{n(n-1)\cdots(n-r+1)(n-r)\cdots 1}{r!(n-r)!} \frac{\lambda^r}{n^r} \left(1 - \frac{\lambda}{n}\right)^{n-r} \end{aligned}$$

Rearrange the expression. First, eliminate $(n-r)!$ from the numerator and the denominator of the term on the left and note that this leaves exactly r terms in the numerator. Then exchange the $r!$ in the denominator with the n^r in the denominator of the term in the middle. Expand n^r into r separate n s and expand $(1 - \lambda/n)^r$ as well:

$$\left(\frac{n}{n} \frac{n-1}{n} \cdots \frac{n-r+1}{n}\right) \frac{\lambda^r}{r!} \left(1 - \frac{n-r}{1!} \left(\frac{\lambda}{n}\right) + \frac{(n-r)(n-r-1)}{2!} \left(\frac{\lambda}{n}\right)^2 - \cdots\right)$$

Next, let n tend to infinity. This will reinforce the validity of the assumption made earlier and lead to a simplification of the expression. Given finite r , all the terms in the left-hand pair of brackets tend to 1. The terms in the right-hand pair of brackets can be simplified too; thus $(n-r)/n$ and $(n-r)(n-r+1)/n^2$ and so on all tend to 1. The expression reduces to:

$$\frac{\lambda^r}{r!} \left(1 - \frac{\lambda}{1!} + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \dots \right) = \frac{\lambda^r}{r!} e^{-\lambda}$$

The conclusion of this analysis is that the probability of there being r murders is:

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}$$

This is an indexed set of probabilities and its sum is readily shown to be 1:

$$\sum_{r=0}^{\infty} \frac{\lambda^r}{r!} e^{-\lambda} = \left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) e^{-\lambda} = e^{\lambda} \cdot e^{-\lambda} = 1$$

As with other distributions, the Poisson distribution is a family of distributions but, like the Geometric distribution, it has only one parameter. The description:

$$\text{Poisson}(\lambda)$$

is used to refer to the general case.

The expectation, $E(X)$, is readily calculated in the case of the Poisson distribution:

$$E(X) = \sum_{r=0}^{\infty} r \cdot P(X = r) = \sum_{r=0}^{\infty} r \cdot \frac{\lambda^r}{r!} e^{-\lambda} = \sum_{r=1}^{\infty} r \cdot \frac{\lambda^r}{r!} e^{-\lambda}$$

When $r = 0$ the term is zero so it is in order to begin the sum from $r = 1$. This means $r/r!$ can be treated as $1/(r-1)!$ and hence:

$$E(X) = \sum_{r=1}^{\infty} \frac{\lambda \cdot \lambda^{r-1}}{(r-1)!} e^{-\lambda} = \lambda \left(\sum_{r=1}^{\infty} \frac{\lambda^{r-1}}{(r-1)!} \right) e^{-\lambda} = \lambda \left(\frac{1}{0!} + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \dots \right) e^{-\lambda} = \lambda \cdot e^{\lambda} \cdot e^{-\lambda}$$

The e^{λ} and $e^{-\lambda}$ cancel so:

$$E(X) = \lambda$$

This result should have been obvious all along. The analysis of the Poisson distribution *began* by taking λ as the expectation (in the illustration this was the expected number of murders in a year). It is comforting to see that the result derived now is not in conflict with that original assumption.

The Prussian Cavalry Corps

An especially well-known illustration of the Poisson distribution was given by M.G. Bulmer. This example relates to 14 corps of Prussian cavalry. Over a period of 20 years, that is 280 corps-years, the number of deaths of cavalry officers from horse kicks was monitored. The mean number of such deaths per corps-year was 0.7 and this gives a value for λ .

Given Poisson(0.7), $P(X = r) = \frac{0.7^r}{r!}e^{-0.7}$, and a table of the probabilities for values of r from 0 to 4 is:

| | $r \rightarrow$ | | | | |
|------------|-----------------|-------|-------|-------|-------|
| X | 0 | 1 | 2 | 3 | 4 |
| $P(X = r)$ | 0.496 | 0.348 | 0.122 | 0.028 | 0.005 |

These values turn out to be very close to those determined from the observed data.

The five values sum to 0.999, slightly less than 1. With the Poisson distribution, r runs to infinity but probabilities for other than very small values of r are normally insignificant. In none of the 280 corps-years were there more than 4 deaths from horse kicks.

Three Rules for Expectation

I If a is a constant

$$E(a) = a$$

II If a is a constant and X is a random variable

$$E(aX) = aE(X)$$

III If X and Y are random variables

$$E(X + Y) = E(X) + E(Y)$$

These rules require formal proofs. First recall the definition of the expectation of an arbitrary function of a random variable (4.1):

$$E(f(X)) = \sum_r f(r) \cdot P(X = r)$$

In the first rule, the function is simply $f(X) = a$ and does not involve the random variable at all. Accordingly:

$$E(a) = \sum_r a \cdot P(X = r) = a \sum_r P(X = r) = a \cdot 1 = a$$

This proves the first rule. Note that it is assumed that the sum of the probabilities is 1. Note also that it is in order to take a constant outside a Σ sign.

In the second rule, the function is $f(X) = aX$ and:

$$E(aX) = \sum_r ar \cdot P(X = r) = a \sum_r r \cdot P(X = r) = aE(X)$$

The proof of the third rule is as follows:

$$\begin{aligned}
 E(X + Y) &= \sum_r \sum_s (r + s) \cdot P(X = r, Y = s) = \sum_r \sum_s (r + s) \cdot p_{r,s} \\
 &= \sum_r r \left(\sum_s p_{r,s} \right) + \sum_s s \left(\sum_r p_{r,s} \right) \\
 &= \sum_r r \cdot P(X = r) + \sum_s s \cdot P(Y = s) = E(X) + E(Y)
 \end{aligned}$$

Note that $\sum_s p_{r,s}$ and $\sum_r p_{r,s}$ are marginal sums (see page 2.3) being row and column totals respectively in an array representation of the probabilities of the elementary events associated with two random variables. These totals are $P(X = r)$ and $P(Y = s)$.

Application to Variance

The variance of a random variable X was defined in (4.2) as:

$$\text{Variance} = V(X) = E((X - \mu)^2)$$

The three rules for expectation can be used to derive an alternative expression for variance:

$$\begin{aligned}
 V(X) &= E((X - \mu)^2) = E(X^2 - 2\mu X + \mu^2) \\
 &= E(X^2) - E(2\mu X) + E(\mu^2) \\
 &= E(X^2) - 2\mu E(X) + \mu^2 \\
 &= E(X^2) - 2\mu^2 + \mu^2 \\
 &= E(X^2) - \mu^2 \\
 &= E(X^2) - (E(X))^2
 \end{aligned} \tag{4.3}$$

This expression for variance will be much used. The result can be expressed in words as:

Variance is the Expectation of the Square minus the Square of the Expectation

Warning

This expression for the variance often substantially reduces the amount of algebra required when analysing problems but there is an element of bad news. Although this expression is mathematically identical to $E((X - \mu)^2)$ there is a practical difference which Computer Scientists should readily appreciate...

The expression $E(X^2) - (E(X))^2$ may well involve taking a small difference between two large numbers, something best avoided if you are mindful of rounding errors. If you use a computer to process real data, stick to the expression $E((X - \mu)^2)$.

Three Rules for Variance

- I If a is a constant $V(a) = 0$
- II If a is a constant and X is a random variable $V(aX) = a^2V(X)$
- III If X and Y are *independent* random variables $V(X + Y) = V(X) + V(Y)$

Using the form in (4.3), both rule I and rule II are proved trivially.

$$V(a) = E(a^2) - (E(a))^2 = a^2 - a^2 = 0$$

Likewise:

$$V(aX) = E(a^2X^2) - (E(aX))^2 = a^2E(X^2) - a^2(E(X))^2 = a^2V(X)$$

Rule III is rather more cumbersome to prove. Begin thus:

$$\begin{aligned} V(X + Y) &= E((X + Y)^2) - (E(X + Y))^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - (E(X))^2 - 2E(X) \cdot E(Y) - (E(Y))^2 \end{aligned}$$

Now $V(X) = E(X^2) - (E(X))^2$ and $V(Y) = E(Y^2) - (E(Y))^2$ so:

$$V(X + Y) = V(X) + V(Y) + 2[E(XY) - E(X) \cdot E(Y)]$$

This is the general expression for the variance of the sum of two random variables whether or not they are independent. If they *are* independent, the item in square brackets, known as the *covariance* of X and Y , turns out to be zero. This will be demonstrated shortly.

Covariance — I

Covariance is usually denoted by the letter W and there is an alternative expression for $W(X, Y)$ which is directly analogous to that for the variance $V(X)$ derived as (4.3):

$$\text{Covariance} = W(X, Y) = E(XY) - E(X) \cdot E(Y) = E((X - \mu_X) \cdot (Y - \mu_Y))$$

See exercise 11. Notice that covariance may be negative whereas variance can never be.

Lemma

The determination of covariance requires some further consideration of the double-sigma notation. Suppose you wish to sum some function of r and s over r and s . If the function can be reduced to the product of two functions $f(r)$ and $g(s)$ such that f does not depend on s and g does not depend on r then:

$$\sum_r \sum_s f(r) \cdot g(s) = \left(\sum_r f(r) \right) \left(\sum_s g(s) \right)$$

An informal proof can be derived by consideration of the left-hand side:

$$\begin{aligned}
 \sum_r \sum_s f(r).g(s) &= f(0).g(0) + f(0).g(1) + f(0).g(2) \cdots \\
 &\quad + f(1).g(0) + f(1).g(1) + f(1).g(2) \cdots \\
 &\quad + f(2).g(0) + f(2).g(1) + f(2).g(2) \cdots \\
 &\quad \vdots \\
 &= f(0).[g(0) + g(1) + g(2) + \cdots] \\
 &\quad + f(1).[g(0) + g(1) + g(2) + \cdots] \\
 &\quad + f(2).[g(0) + g(1) + g(2) + \cdots] \\
 &\quad \vdots \\
 &= [f(0) + f(1) + f(2) + \cdots] [g(0) + g(1) + g(2) + \cdots] \\
 &= \left(\sum_r f(r) \right) \left(\sum_s g(s) \right)
 \end{aligned}$$

This is the right-hand side and the informal proof is concluded.

Covariance — II

The determination of covariance requires the evaluation of $E(XY)$:

$$E(XY) = \sum_r \sum_s r.s P(X = r, Y = s)$$

Now *if X and Y are independent*, $P(X = r, Y = s) = P(X = r) . P(Y = s)$, when:

$$\sum_r \sum_s r.s P(X = r, Y = s) = \sum_r \sum_s r.s P(X = r) . P(Y = s)$$

The function after the double-sigma sign can be separated into the product of two terms, the first of which does not depend on s and the second of which does not depend on r . Hence, by the lemma:

$$E(XY) = \left(\sum_r r.P(X = r) \right) \left(\sum_s s.P(Y = s) \right) = E(X) . E(Y)$$

Accordingly, if X and Y are independent, the covariance of X and Y :

$$W(X, Y) = E(XY) - E(X) . E(Y) = 0$$

and the variance of the sum of X and Y :

$$V(X + Y) = V(X) + V(Y)$$

Illustration

Various formulae for the expectation, variance and covariance have been presented and these formulae can now be applied to an illustrative case.

Recall the class of children who were classified boy-girl and fair-dark. Suppose that the information supplied consists only of a 2×2 table of the four elementary events:

| | | |
|------|---------------|----------------|
| | fair | dark |
| boy | $\frac{1}{3}$ | $\frac{1}{3}$ |
| girl | $\frac{1}{4}$ | $\frac{1}{12}$ |

A full analysis is presented on the following page. This takes the form of a completed and annotated proforma which can be used for analysing the data in any two-dimensional table of elementary events.

There are five exercises later which should all be carried out using this proforma as a guide. Note the following steps:

- First choose X and Y as the names of the two random variables. Let these have values r and s which can each be $0, 1, \dots$. Appropriate, and obvious, mappings have to be made to convert non-numerical events such as fair and dark into integers.
- Copy the table of elementary events and ornament it with X , Y , r and s and the chosen mappings. Compute the marginal sums and check that the row sums and column sums each total 1.
- Tabulate the probabilities $P(X = r)$ and $P(Y = s)$ separately and, using the values in the tables, compute the expectation, the expectation of the square, and the variance of X and Y .
- Set up the big table in the middle of the page. There will be one line of entries for each elementary event. The third column is simply a transcription of the probabilities of the elementary events and their sum should be 1. The totals of the fourth, fifth and sixth columns should give $E(X + Y)$, $E((X + Y)^2)$ and $E(XY)$. Check that $E(X + Y) = E(X) + E(Y)$ (these latter values being computed in the previous step).
- Compute the variance of the sum $V(X + Y)$ and check whether or not the result is the same as $V(X) + V(Y)$.
- Compute the covariance and add twice this to $V(X) + V(Y)$ and check that the result is the same as the value of $V(X + Y)$ computed in the previous step.
- Finally check each probability in the original table against the product of the two relevant marginal sums. If any probability is different from the relevant product, the two random variables are not independent.

A Full Analysis of a Pair of Random Variables

First copy the table of elementary events and compute the marginal sums...

| | | | | | |
|---|---|-----|----------------|----------------|---------------|
| | | Y | | | |
| | | s → | | | |
| | | 0 | 1 | | |
| X | r | 0 | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{2}{3}$ |
| | ↓ | 1 | $\frac{1}{4}$ | $\frac{1}{12}$ | $\frac{1}{3}$ |
| | | | $\frac{7}{12}$ | $\frac{5}{12}$ | |

Tabulate the probabilities $P(X = r)$ and $P(Y = s)$ separately and for each variable compute the expectation, the expectation of the square, and the variance...

| | | | | | |
|---|---------------|------------------------|---|----------------|-------------------------|
| r | P(X = r) | $E(X) = \frac{1}{3}$ | s | P(Y = s) | $E(Y) = \frac{5}{12}$ |
| 0 | $\frac{2}{3}$ | $E(X^2) = \frac{1}{3}$ | 0 | $\frac{7}{12}$ | $E(Y^2) = \frac{5}{12}$ |
| 1 | $\frac{1}{3}$ | $V(X) = \frac{2}{9}$ | 1 | $\frac{5}{12}$ | $V(Y) = \frac{35}{144}$ |

For each elementary event tabulate r , s , $p_{r,s}$, $(r + s)p_{r,s}$, $(r + s)^2 p_{r,s}$ and $(r \cdot s)p_{r,s}$ and then determine the sums of the four rightmost columns...

| r | s | $p_{r,s}$ | $(r + s)p_{r,s}$ | $(r + s)^2 p_{r,s}$ | $(r \cdot s)p_{r,s}$ |
|---|---|-------------------------|------------------|---------------------|----------------------|
| 0 | 0 | $\frac{1}{3}$ | 0 | 0 | 0 |
| 0 | 1 | $\frac{1}{3}$ | $\frac{1}{3}$ | $\frac{1}{3}$ | 0 |
| 1 | 0 | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | 0 |
| 1 | 1 | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{12}$ |
| | | 1 | $\frac{3}{4}$ | $\frac{11}{12}$ | $\frac{1}{12}$ |
| | | $\sum_r \sum_s p_{r,s}$ | $E(X + Y)$ | $E((X + Y)^2)$ | $E(XY)$ |

Now compute various values related to the pair X and Y ...

Variance of the sum (i): $V(X + Y) = E((X + Y)^2) - (E(X + Y))^2 = \frac{11}{12} - \frac{9}{16} = \frac{17}{48}$

Covariance: $W(X, Y) = E(XY) - E(X) \cdot E(Y) = \frac{1}{12} - \frac{1}{3} \times \frac{5}{12} = -\frac{1}{18}$

Variance of the sum (ii): $V(X + Y) = V(X) + V(Y) + 2W(X, Y) = \frac{2}{9} + \frac{35}{144} - \frac{1}{9} = \frac{17}{48}$

Finally check for independence; $P(X = r, Y = s)$ versus $P(X = r) \cdot P(Y = s)$...

$$\frac{1}{3} \neq \frac{2}{3} \times \frac{7}{12} \quad \frac{1}{3} \neq \frac{2}{3} \times \frac{5}{12} \quad \frac{1}{4} \neq \frac{1}{3} \times \frac{7}{12} \quad \frac{1}{12} \neq \frac{1}{3} \times \frac{5}{12}$$

Glossary

The following technical terms have been introduced:

| | | |
|-------------------------|------------------------|----------------------|
| variance | standard deviation | Poisson distribution |
| derived random variable | Geometric distribution | covariance |

Exercises — IV

Work in fractions.

1. Determine the variance of the Triangular distribution:

$$P(X = r) = \begin{cases} \frac{r}{21}, & \text{if } r \in \mathbb{N} \wedge 1 \leq r \leq 6 \\ 0, & \text{otherwise} \end{cases}$$

2. A random variable X is distributed Geometric(p) so $P(X = r) = (1 - p)^r p$. Show that $P(X \geq m + n \mid X \geq m) = P(X \geq n)$ for $m, n = 0, 1, 2, \dots$ [Thus X has the ‘lack of memory property’ since, given that $X - m \geq 0$, the distribution of $X - m$ is the same as the original distribution of X .]
3. As r takes the values $0, 1, 2, \dots$ show that the probabilities of the Poisson distribution $P(X = r) = \lambda^r e^{-\lambda} / r!$ initially increase monotonically then decrease monotonically. Additionally show that they reach their greatest value when r is the largest integer not exceeding λ .
4. A computer printout of n pages contains on average λ misprints per page. Estimate the probability that at least one page will contain more than k misprints.
5. Two random variables X and Y are independently distributed with means μ_x and μ_y and variances σ_x^2 and σ_y^2 respectively. Derive expressions for the mean and variance of XY .
6. Suppose that the data given in the 2×2 table on page 4.10 were replaced by the following values for the probabilities of the elementary events:

| | | |
|------|----------------|----------------|
| | fair | dark |
| boy | $\frac{4}{15}$ | $\frac{2}{15}$ |
| girl | $\frac{2}{5}$ | $\frac{1}{5}$ |

Complete a proforma like that shown on page 4.11 but based on these new values.

7. Suppose the hair colour is now classified in three ways and the results for a rather extraordinary class are recorded in the following 2×3 table:

| | | | |
|------|---------------|---------------|---------------|
| | fair | red | dark |
| boy | $\frac{1}{3}$ | 0 | $\frac{1}{3}$ |
| girl | 0 | $\frac{1}{3}$ | 0 |

Again complete a proforma. Note that s will now range 0, 1 and 2 and there will be an extra line of entries in the table headed s and $P(Y = s)$. There will be two extra lines in the main table.

8. Suppose that the data are now as in the following 7×7 table which is appropriate for two fair dice. Both r and s will now range from 0 to 6 and there will be seven lines of entries in the two short tables. In principle there will be 49 lines of entries in the main table but it is not necessary to fill this in. Only the four totals beneath the table are required and it will not be hard to see how to derive these by some obvious short cuts:

| | | | | | | | |
|---|---|----------------|----------------|----------------|----------------|----------------|----------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 2 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 3 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 4 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 5 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |
| 6 | 0 | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ | $\frac{1}{36}$ |

9. Complete a similar analysis using the data in the following table where the two dice always show the same value:

| | | | | | | | |
|---|---|---------------|---------------|---------------|---------------|---------------|---------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ |

10. To conclude this series of dice questions, use the data in the following table where the sum of the values shown by the two dice is always seven:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---------------|---------------|---------------|---------------|---------------|---------------|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ |
| 2 | 0 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 |
| 3 | 0 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 |
| 4 | 0 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 |
| 5 | 0 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 | 0 |
| 6 | 0 | $\frac{1}{6}$ | 0 | 0 | 0 | 0 | 0 |

11. By analogy with the method used to demonstrate that $E((X-\mu)^2) = E(X^2) - (E(X))^2$ (see (4.3) above) show that:

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

where $\mu_X = E(X)$ and $\mu_Y = E(Y)$.