5.7 The central limit theorem

We already discussed approximating the binomial and Poisson distributions by the Gaussian distribution when the number of trials is large. We now discuss why the *Gaussian distribution* is so common and therefore so important.

Central limit theorem

Suppose that X_i, $i = 1, 2, \ldots, n$, are *independent* random variables, each of which is described by a probability density function $f_i(x)$ (these may all be different) with a mean μ_i and a variance σ_i^2. The random variable

$$Z = \frac{1}{n}\sum_i X_i \tag{5.39}$$

i.e. the 'mean' of the X_i, has the following properties:

(i) its *expectation* value is given by $E[Z] = \frac{1}{n}\sum_i \mu_i$ (5.40)

(ii) its *variance* is given by $V[Z] = \frac{1}{n^2}\sum_i \sigma_i^2$ (5.41)

(iii) as $n \to \infty$ the *probability function* of *Z* tends to a *Gaussian* with corresponding mean and variance.

The theorem holds if the probability density functions $f_i(x)$ possess formal means and variances.

Properties (i) and (ii) of the theorem are easily proved, as follows.

$$E[Z] = \frac{1}{n}\left(E[X_1] + E[X_2] + \ldots + E[X_n]\right) = \frac{1}{n}\left(\mu_1 + \mu_2 + \ldots + \mu_n\right) = \frac{1}{n}\sum_i \mu_i$$

a result which does *not* require that the X_i are *independent* random variables. If $\mu_i = \mu$ for all *i* then this becomes

$$E[Z] = \frac{n\mu}{n} = \mu$$

If the X_i *are* independent, we know that $V[aX + bY + c] = a^2V[X] + b^2V[Y]$, so

$$V[Z] = V\left[\frac{1}{n}\left(X_1 + X_2 + \ldots + X_n\right)\right]$$

$$= \frac{1}{n^2}\left(V[X_1]+V[X_2]+\ldots+V[X_n]\right) = \frac{1}{n^2}\sum_i \sigma_i^2$$

The property (iii) is the *reason for the ubiquity of the Gaussian distribution* and is most easily proved by considering the moment generating function $M_Z(t)$ of *Z*.

The MGF of the sum of *N* independent random variables is the product of their individual MGFs. If we remember the general result (chapter 3) that the MGF of $S_N = c_1X_1 + c_2X_2 + \ldots + c_NX_N$ (where the c_i are constants) is given by

$$M_{S_N}(t) = \prod_{i=1}^{N} M_{X_i}(c_i t)$$

Then, the MGF for *Z* is given by

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \qquad (5.42)$$

where $M_{X_i}(t)$ is the MGF of $f_i(x)$.

$$M_{X_i}\left(\frac{t}{n}\right) = E\left[e^{\frac{t}{n}X_i}\right] = E\left[1 + \frac{t}{n}X_i + \frac{1}{2}\frac{t^2}{n^2}X_i^2 + \ldots\right] = 1 + \frac{t}{n}E[X_i] + \frac{1}{2}\frac{t^2}{n^2}E\left[X_i^2\right] + \ldots$$

$$= 1 + \mu_i\frac{t}{n} + \frac{1}{2}\left(\sigma_i^2 + \mu_i^2\right)\frac{t^2}{n^2} + \ldots$$

and as *n* becomes large

$$M_{X_i}\left(\frac{t}{n}\right) \approx \exp\left(\frac{\mu_i t}{n} + \frac{1}{2}\sigma_i^2\frac{t^2}{n^2}\right)$$

as may be verified by expanding the exponential up to terms including $(t/n)^2$. Therefore

$$M_Z(t) = \prod_{i=1}^{n}\exp\left(\frac{\mu_i t}{n} + \frac{1}{2}\sigma_i^2\frac{t^2}{n^2}\right) = \exp\left(\frac{\sum_i \mu_i}{n}t + \frac{1}{2}\frac{\sum_i \sigma_i^2}{n^2}t^2\right)$$

Comparing this with the form of the MGF for a Gaussian distribution,

$$M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

we can see that the probability density function $g(z)$ of *Z* tends to a *Gaussian distribution* with mean $\frac{1}{n}\sum_i \mu_i$ and variance $\frac{1}{n^2}\sum_i \sigma_i^2$.

In particular, *if we consider Z to be the mean of n independent measurements of the same random variable X (so that* $X_i = X$ *for* $i = 1, 2, \ldots, n$ *) then, as* $n \to \infty$*, Z has a Gaussian distribution with mean μ and variance* σ^2 / n.

If $X_1, X_2, \ldots, X_n$, is a random sample of size *n* taken from a population with mean μ and finite variance σ^2 and if $\bar{X}$ is the sample mean, the limiting form of the distribution of $\bar{X}$ is *Gaussian distribution with mean μ and variance* σ^2 / n. Moreover, $Z = \dfrac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ as $n \to \infty$, has the standard normal distribution.

The normal approximation for $\bar{X}$ depends on the sample size *n*. Figure 7-3(a) is the distribution obtained for throws of a single, six-sided true die. The probabilities are equal (1/6) for all the values obtained: 1, 2, 3, 4, 5, or 6. Figure 7-3(b) is the distribution of the average score obtained when tossing two dice, and Fig. 7-3(c), 7-3(d), and 7-3(e) show the distributions of average scores obtained when tossing 3, 5, and 10 dice, respectively. Notice that, although the population (one die) is relatively far from normal, the distribution of averages is approximated reasonably well by the normal distribution for sample sizes as small as five.

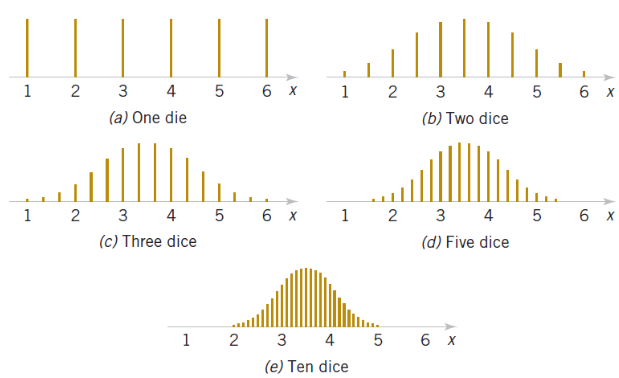

FIGURE 7-3 Distributions of average scores from throwing dice. *Source*: [Adapted with permission from Box, Hunter, and Hunter (1978).]

The central limit theorem is the underlying reason why many of the random variables encountered in engineering and science are normally distributed.

Chapter 6 Joint distributions

Bibliografie: Riley et al. (2006), Montgomery (2011)

It is common in the physical sciences to consider simultaneously two or more random variables that are not independent, in general, and are thus described by *joint probability density functions*. We will concentrate mainly on *bivariate* distributions, i.e. *distributions of only two random variables*, though the results may be extended to *multivariate* distributions.

When dealing with *bivariate* distributions, the random variables can both be discrete, or both continuous, or one discrete and the other continuous. In general, for the random variables *X* and *Y*, the joint distribution will take an infinite number of values unless both *X* and *Y* have only a finite number of values. We will consider only the cases where *X* and *Y* are either both discrete or both continuous random variables.

6.1 Discrete bivariate distributions

If *X* is a discrete random variable that takes the values $\{x_i\}$ and *Y* one that takes the values $\{y_j\}$ then the probability function of *the joint distribution* is defined as

$$f(x,y)=\begin{cases}\Pr\left(X=x_i,\ Y=y_j\right) & \text{for } x=x_i, y=y_j\\ 0 & \text{otherwise}\end{cases} \tag{6.1}$$

We may therefore think of $f(x,y)$ as a set of *spikes* at valid points in the *xy*-plane, whose height at (x_i,y_j) represents the probability of obtaining $X=x_i$ and $Y=y_j$. The normalisation of $f(x,y)$ implies

$$\sum_i\sum_j f\left(x_i,y_j\right)=1 \tag{6.2}$$

where the sums over *i* and *j* take all valid pairs of values. We can also define the cumulative probability function

$$F(x,y)=\sum_{x_i\le x}\sum_{y_j\le y} f\left(x_i,y_j\right) \tag{6.3}$$

from which it follows that the probability that *X* lies in the range $[a_1,a_2]$ and *Y* lies in the range $[b_1,b_2]$ is given by

$$\Pr\left(a_1 < X \le a_2,\ b_1 < Y \le b_2\right) = F\left(a_2, b_2\right) - F\left(a_1, b_2\right) - F\left(a_2, b_1\right) + F\left(a_1, b_1\right) \quad (6.4)$$

Finally, we define *X* and *Y* to be *independent* if we can write their joint distribution in the form

$$f\left(x, y\right) = f_X\left(x\right) f_Y\left(y\right) \quad (6.5)$$

i.e. as the product of two univariate distributions.

6.2 Continuous bivariate distributions

In the case where both *X* and *Y* are continuous random variables, the PDF of the joint distribution is defined by

$$f\left(x, y\right) dx\ dy = \Pr\left(x < X \le x + dx,\ \ y < Y \le y + dy\right) \quad (6.6)$$

so $f\left(x, y\right) dx\ dy$ is the probability that *X* lies in the range $\left[x, x + dx\right]$ and *Y* lies in the range $\left[y, y + dy\right]$. The two-dimensional function $f\left(x, y\right)$ must be everywhere non-negative and that normalization requires

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f\left(x, y\right) dx\ dy = 1$$

It follows further that

$$\Pr\left(a_1 < X \le a_2,\ b_1 < Y \le b_2\right) = \int_{b_1}^{b_2} \int_{a_1}^{a_2} f\left(x, y\right) dx\ dy \quad (6.7)$$

We can also define the *cumulative* probability function by

$$F\left(x, y\right) = \Pr\left(X \le x,\ Y \le y\right) = \int_{-\infty}^{x} \int_{-\infty}^{y} f\left(u, v\right) du\ dv \quad (6.8)$$

from which we see that (as for the discrete case),

$$\Pr\left(a_1 < X \le a_2,\ b_1 < Y \le b_2\right) = F\left(a_2, b_2\right) - F\left(a_1, b_2\right) - F\left(a_2, b_1\right) + F\left(a_1, b_1\right)$$

Finally, we note that the definition of *independence* (6.5) for discrete bivariate distributions also applies to continuous bivariate distributions.

Exercise 1. A flat table is ruled with parallel straight lines a distance *D* apart, and a thin needle of length $l < D$ is tossed onto the table at random. What is the probability that the needle will cross a line?

Let θ be the angle that the needle makes with the lines, and let *x* be the distance from the center of the needle to the nearest line. Since the needle is tossed 'at random' onto the table, the angle θ is uniformly distributed in the interval $[0, \pi]$, and the distance *x* is uniformly distributed in the interval $[0, D/2]$. Assuming that θ and *x* are *independent*, their *joint distribution* is just the product of their individual distributions, and is given by

$$f(\theta, x) = \frac{1}{\pi} \frac{1}{D/2} = \frac{2}{\pi D}$$

The needle will cross a line if the distance *x* of its center from that line is less than $\frac{1}{2} l \sin\theta$. Thus the required probability is

$$\frac{2}{\pi D} \int_0^{\pi} \int_0^{\frac{1}{2} l \sin\theta} dx\, d\theta = \frac{2}{\pi D} \frac{l}{2} \int_0^{\pi} \sin\theta\, d\theta = \frac{2l}{\pi D}$$

6.3 Marginal and conditional distributions

Given a bivariate distribution $f(x, y)$, we may be interested only in the probability function for *X* irrespective of the value of *Y*. This *marginal* distribution of *X* is obtained by summing or integrating the joint probability distribution over all allowed values of *Y*. Thus, the *marginal distribution of X* is given by

$$f_X(x) = \begin{cases} \sum_j f(x, y_j) & \text{for a discrete distribution} \\ \int f(x, y)\, dy & \text{for a continuous distribution} \end{cases} \tag{6.9}$$

It is clear that an analogous definition exists for the marginal distribution of *Y*.

Alternatively, one might be interested in the probability function of *X given that Y takes some specific value of* $Y = y_0$, i.e. $\Pr(X = x | Y = y_0)$. This *conditional distribution* of *X* is given by

$$g(x) = \frac{f(x, y_0)}{f_Y(y_0)} \tag{6.10}$$

where $f_Y(y)$ is the marginal distribution of *Y*.

Exercise 2. Suppose that 3 batteries are randomly chosen from a group of 3 new, 4 used but still working, and 5 defective batteries. If we let X and Y denote, respectively, the number of new and used but still working batteries that are chosen, then the joint probability function of X and Y, $p(i,j)=P(X=i,Y=j)$, is given by

$$p(i,j)=\frac{C_3^iC_4^jC_5^{3-i-j}}{C_{12}^3}$$

Because of the C_{12}^3 equally likely outcomes, there are, by the basic principle of counting, $C_3^iC_4^jC_5^{3-i-j}$ possible choices that contain exactly i new, j used, and $3-i-j$ defective batteries.

$$p(0,0)=\frac{C_5^3}{C_{12}^3}=\frac{10}{220}\quad p(0,1)=\frac{C_4^1C_5^2}{C_{12}^3}=\frac{40}{220}\quad p(0,2)=\frac{C_4^2C_5^1}{C_{12}^3}=\frac{30}{220}\quad p(0,3)=\frac{C_4^3}{C_{12}^3}=\frac{4}{220}$$

$$p(1,0)=\frac{C_3^1C_5^2}{C_{12}^3}=\frac{30}{220}\quad p(1,1)=\frac{C_3^1C_4^1C_5^1}{C_{12}^3}=\frac{60}{220}\quad p(1,2)=\frac{C_3^1C_4^2}{C_{12}^3}=\frac{18}{220}$$

$$p(2,0)=\frac{C_3^2C_5^1}{C_{12}^3}=\frac{15}{220}\quad p(2,1)=\frac{C_3^2C_4^1}{C_{12}^3}=\frac{12}{220}$$

$$p(3,0)=\frac{C_3^3}{C_{12}^3}=\frac{1}{220}$$

i\j	0	1	2	3	Row Sum = $P(X=i)$
0	$\frac{10}{220}$	$\frac{40}{220}$	$\frac{30}{220}$	$\frac{4}{220}$	$\frac{84}{220}$
1	$\frac{30}{220}$	$\frac{60}{220}$	$\frac{18}{220}$	0	$\frac{108}{220}$
2	$\frac{15}{220}$	$\frac{12}{220}$	0	0	$\frac{27}{220}$
3	$\frac{1}{220}$	0	0	0	$\frac{1}{220}$
Column Sums = $P(Y=j)$	$\frac{56}{220}$	$\frac{112}{220}$	$\frac{48}{220}$	$\frac{4}{220}$	

The marginal probability function of X is obtained by computing the row sums, in accordance with the eq. (6.9), whereas the marginal probability function of Y is obtained by computing the column sums. Because the individual probability functions of X and Y thus appear in the margin of such a table, they are often referred to as being the *marginal probability functions* of X and Y, respectively. It should be noted that to check the correctness of such a table we could sum the marginal row (or the marginal column) and verify that its sum is 1.

6.4 Properties of joint distributions

The probability density function $f(x,y)$ contains all the information on the joint probability distribution of two random variables *X* and *Y*. It is conventional to characterize $f(x,y)$ by certain of its *properties*. These properties are based on the concept of *expectation* values. The expectation value of any function $g(X,Y)$ of the random variables *X* and *Y* is given by

$$E\left[g(X,Y)\right]=\begin{cases}\sum_i\sum_j g(x_i,y_j)\,f(x_i,y_j) & \text{for the discrete case}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,f(x,y)\,dxdy & \text{for the continuous case}\end{cases} \qquad (6.11)$$

1. *Means*

The means of *X* and *Y* are defined respectively as the expectation values of the variables *X* and *Y*. Thus, the mean of *X* is given by

$$E[X]=\mu_X=\begin{cases}\sum_i\sum_j x_i f(x_i,y_j) & \text{for the discrete case}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dxdy & \text{for the continuous case}\end{cases} \qquad (6.12)$$

$E[Y]$ is obtained in a similar manner.

Exercise 3. Show that if *X* and *Y* are *independent* random variables then $E[XY]=E[X]E[Y]$.

Let us consider the case where *X* and *Y* are continuous random variables. Since *X* and *Y* are independent $f(x,y)=f_X(x)\,f_Y(y)$, so that

$$\begin{aligned}E[XY]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_X(x)\,f_Y(y)\,dxdy\\ &=\int_{-\infty}^{\infty} xf_X(x)\,dx\int_{-\infty}^{\infty} yf_Y(y)\,dy=E[X]E[Y]\end{aligned} \qquad (6.13)$$

An analogous proof exists for the discrete case.

2. *Variances*

The definitions of the variances of *X* and *Y* are analogous to those for the single-variable case, i.e. the variance of *X* is given by

$$V\left[X\right]=E\left[\left(X-\mu_X\right)^2\right]=\sigma_X^2=\begin{cases}\sum_i\sum_j\left(x_i-\mu_X\right)^2 f\left(x_i,y_j\right) & \text{for the discrete case}\\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-\mu_X\right)^2 f\left(x,y\right)dxdy & \text{for the continuous case}\end{cases} \tag{6.14}$$

Equivalent definitions exist for the variance of *Y*.

3. *Covariance and correlation*

Means and variances of joint distributions provide useful information about their marginal distributions, but we have not yet given any indication of how to measure the relationship between the two random variables. Of course, it may be that the two random variables are independent, but often this is not so. For example, if we measure the heights and weights of a sample of people we would not be surprised to find a tendency for tall people to be heavier than short people and vice versa. The *covariance* and the *correlation*, can be defined for a bivariate distribution and are useful in characterizing the relationship between the two random variables.

The *covariance* of two random variables *X* and *Y* is defined by

$$Cov\left[X,Y\right]=\sigma_{XY}=E\left[\left(X-\mu_X\right)\left(Y-\mu_Y\right)\right]\ , \tag{6.15}$$

where μ_X and μ_Y are the expectation values of *X* and *Y* respectively. Related to the covariance is the *correlation* of the two random variables, defined by

$$Corr\left[X,Y\right]=\rho_{XY}=\frac{Cov\left[X,Y\right]}{\sigma_X\sigma_Y}\ , \tag{6.16}$$

where σ_X and σ_Y are the standard deviations of *X* and *Y* respectively. The correlation function lies between −1 and +1. If the value assumed is negative, *X* and *Y* are said to be *negatively correlated*, if it is positive they are said to be *positively correlated* and if it is zero they are said to be *uncorrelated*.

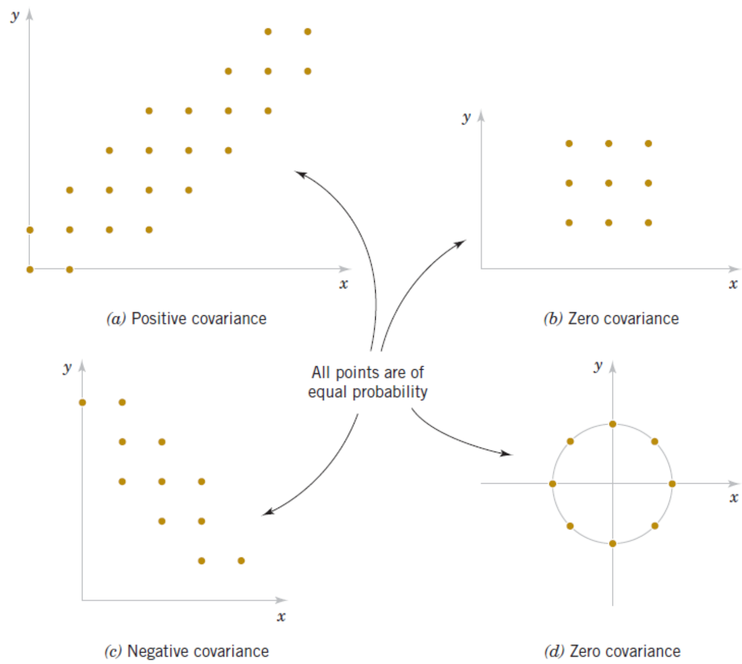

FIGURE 5-12 Joint probability distributions and the sign of covariance between *X* and *Y*.

If the points in the joint probability distribution of *X* and *Y* that receive positive probability tend to fall along a line of positive (or negative) slope, σ_{XY}, is positive (or negative). If the points tend to fall along a line of positive slope, *X* tends to be greater than μ_X when *Y* is greater than μ_Y. Therefore, the product of the two terms $x-\mu_X$ and $y-\mu_Y$ tends to be positive. However, if the points tend to fall along a line of negative slope, $x-\mu_X$ tends to be positive when $y-\mu_Y$ is negative, and vice versa. Therefore, the product of $x-\mu_X$ and $y-\mu_Y$ tends to be negative. In this sense, the covariance between *X* and *Y* describes the variation between the two random variables. Figure 5-12 assumes all points are equally likely and shows examples of pairs of random variables with positive, negative, and zero covariance.

The covariance of two *independent* variables, *X* and *Y*, is zero. It immediately follows from (6.16) that their correlation is also zero, and this justifies the use of the term '*uncorrelated*' for two such variables.

$$\begin{aligned} Cov[X,Y] &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E[XY-\mu_X Y-\mu_Y X+\mu_X\mu_Y] \\ &= E[XY]-\mu_X E[Y]-\mu_Y E[X]+\mu_X\mu_Y \\ &= E[XY]-\mu_X\mu_Y \qquad (6.17) \end{aligned}$$

Now, if X and Y are *independent* then $E[XY]=E[X]E[Y]=\mu_X\mu_Y$ and so $Cov[X,Y]=0$. It is important to note that the converse of this result is not necessarily true; two variables dependent on each other can still be uncorrelated. In other words, it is possible (and not uncommon) for two variables X and Y to be described by a joint distribution $f(x,y)$ that *cannot* be factorized into a product of the form $g(x)h(y)$, but for which $Corr[X,Y]=0$.

We have already asserted that if the correlation of two random variables is positive (negative) they are said to be positively (negatively) correlated. We have also stated that the correlation lies between −1 and +1. The terminology suggests that if the two RVs are identical (i.e. $X=Y$) then they are completely correlated and that their correlation should be +1. Likewise, if $X=-Y$ then the functions are completely negative correlated and their correlation should be −1. Values of the correlation function between these extremes show the existence of some degree of correlation.

In fact it is not necessary that $X=Y$ for $Corr[X,Y]=1$; it is sufficient that Y is a linear function of X, i.e. $Y=aX+b$ (with a positive). If a is negative then $Corr[X,Y]=-1$. To show this we first note that $\mu_Y=a\mu_X+b$.

$$Y=aX+b=aX+\mu_Y-a\mu_X \quad \Rightarrow \quad Y-\mu_Y=a(X-\mu_X)$$

and using the definition of the covariance $Cov[X,Y]=E[(X-\mu_X)(Y-\mu_Y)]$:

$$Cov[X,Y]=aE\left[(X-\mu_X)^2\right]=a\sigma_X^2$$

It follows from the properties of the variance that $V[Y]=a^2V[X]$, so $\sigma_Y=|a|\sigma_X$ and so, using the definition of the correlation,

$$Corr[X,Y]=\frac{Cov[X,Y]}{\sigma_X\sigma_Y}=\frac{a\sigma_X^2}{|a|\sigma_X^2}=\frac{a}{|a|}=\pm 1 \ ,$$

which is the stated result.

Exercise 4. A biased die gives probabilities $\frac{1}{2}p, p, p, p, p, 2p$ of throwing 1, 2, 3, 4, 5, 6 respectively. If the random variable X is the number shown on the die and the random variable Y is defined as X^2, calculate the covariance and correlation of X and Y.

We have already calculated that

$$p=\frac{2}{13}\ ,\ E[X]=\frac{53}{13}\ ,\ E\left[X^2\right]=\frac{253}{13},\ V[X]=\frac{480}{169}$$

Using $Cov[X,Y]=E[XY]-E[X]E[Y]$ we obtain

$$Cov[X,Y]=Cov\left[X,X^2\right]=E\left[X^3\right]-E[X]E\left[X^2\right]$$

$$E\left[X^3\right]=1^3\times\frac{1}{2}p+\left(2^3+3^3+4^3+5^3\right)p+6^3\times 2p=\frac{1313}{2}p=101$$

and the covariance of X and Y is given by

$$Cov[X,Y]=101-\frac{53}{13}\times\frac{253}{13}=\frac{3660}{169}$$

The correlation is defined by $Corr[X,Y]=\dfrac{Cov[X,Y]}{\sigma_X\sigma_Y}$. The standard deviation of Y may be calculated from the definition of the variance. Letting $\mu_Y=E\left[X^2\right]=\dfrac{253}{13}$ gives

$$\sigma_Y^2=E\left[\left(Y-\mu_Y\right)^2\right]=\frac{p}{2}\left(1^2-\mu_Y\right)^2+p\left(2^2-\mu_Y\right)^2+p\left(3^2-\mu_Y\right)^2+$$
$$+p\left(4^2-\mu_Y\right)^2+p\left(5^2-\mu_Y\right)^2+2p\left(6^2-\mu_Y\right)^2=\frac{28824}{169}$$

We deduce that

$$Corr[X,Y]=\frac{3660}{169}\sqrt{\frac{169}{28824}}\sqrt{\frac{169}{480}}\approx 0.984$$

Thus the random variables X and Y display a strong degree of positive correlation, as we would expect.

We note that the covariance of X and Y occurs in various expressions. For example, if X and Y are *not* independent then

$$V[X+Y]=E\left[(X+Y)^2\right]-\left(E[X+Y]\right)^2$$
$$=E\left[X^2\right]+2E[XY]+E\left[Y^2\right]-\left(E[X]+E[Y]\right)^2$$
$$=E\left[X^2\right]+2E[XY]+E\left[Y^2\right]-\left(E[X]\right)^2-2E[X]E[Y]-\left(E[Y]\right)^2$$

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

$$V[X+Y] = V[X] + V[Y] + 2Cov[X,Y] \qquad (6.18)$$

More generally, for *a*, *b* and *c* constant

$$V[aX + bY + c] = a^2V[X] + b^2V[Y] + 2abCov[X,Y] \qquad (6.19)$$

Note that if *X* and *Y* are in fact *independent* then $Cov[X,Y]=0$ and we recover the expression: $V[aX + bY + c] = a^2V[X] + b^2V[Y]$.

We may use (6.19) to obtain an approximate expression for $V\left[f(X,Y)\right]$ for any arbitrary function *f*, even when the random variables *X* and *Y* are correlated. Approximating $f(X,Y)$ by the linear terms of its Taylor expansion about the point (μ_X, μ_Y), we have

$$f(X,Y) \approx f(\mu_X, \mu_Y) + \left(\frac{\partial f}{\partial X}\right)(X - \mu_X) + \left(\frac{\partial f}{\partial Y}\right)(Y - \mu_Y) \qquad (6.20)$$

where the partial derivatives are evaluated at $X = \mu_X$ and $Y = \mu_Y$. Taking the variance of both sides, and using (6.19), we find

$$V\left[f(X,Y)\right] \approx \left(\frac{\partial f}{\partial X}\right)^2 V[X] + \left(\frac{\partial f}{\partial Y}\right)^2 V[Y] + 2\left(\frac{\partial f}{\partial X}\right)\left(\frac{\partial f}{\partial Y}\right)Cov[X,Y] \qquad (6.21)$$

Clearly, if $Cov[X,Y]=0$, we recover the result (3.50). We note that (6.21) is exact if $f(X,Y)$ is linear in *X* and *Y*.

For several variables X_i, $i = 1, 2, \ldots, n$, we can define the symmetric *covariance matrix* whose elements are

$$V_{ij} = Cov\left[X_i, X_j\right] \qquad (6.22)$$

and the symmetric *correlation matrix*

$$\rho_{ij} = Corr\left[X_i, X_j\right] \qquad (6.23)$$

The diagonal elements of the covariance matrix are the *variances of the variables*, whilst those of the correlation matrix are unity.

Exercise 5. A card is drawn at random from a normal 52-card pack and its identity noted. The card is replaced, the pack shuffled and the process repeated. Random variables *W*, *X*, *Y*, *Z* are defined as follows:

$W = 2$ if the drawn card is a heart; $W = 0$ otherwise.
$X = 4$ if the drawn card is an ace, king, or queen; $X = 2$ if the card is a jack or ten; $X = 0$ otherwise.
$Y = 1$ if the drawn card is red; $Y = 0$ otherwise.
$Z = 2$ if the drawn card is black and an ace, king or queen; $Z = 0$ otherwise.

Establish the correlation matrix for *W*, *X*, *Y*, *Z*.

The means of the variables are given by

$$\mu_W = 2\times\frac{1}{4} = \frac{1}{2} \qquad \mu_X = 4\times\frac{3}{13} + 2\times\frac{2}{13} = \frac{16}{13}$$
$$\mu_Y = 1\times\frac{1}{2} = \frac{1}{2} \qquad \mu_Z = 2\times\frac{6}{52} = \frac{3}{13}$$

The variances, calculated from $\sigma_U^2 = V[U] = E\left[U^2\right] - \left(E[U]\right)^2$, where $U = W, X, Y$ or Z, are:

$$\sigma_W^2 = \left(4\times\frac{1}{4}\right) - \left(\frac{1}{2}\right)^2 = \frac{3}{4}$$
$$\sigma_X^2 = \left(16\times\frac{3}{13}\right) + \left(4\times\frac{2}{13}\right) - \left(\frac{16}{13}\right)^2 = \frac{472}{169}$$
$$\sigma_Y^2 = \left(1\times\frac{1}{2}\right) - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$$
$$\sigma_Z^2 = \left(4\times\frac{6}{52}\right) - \left(\frac{3}{13}\right)^2 = \frac{69}{169}$$

The covariance are found by first calculating $E[WX]$ etc. and then forming $E[WX] - \mu_W\mu_X$ etc.

$$E[WX] = 2\times4\times\frac{3}{52} + 2\times2\times\frac{2}{52} = \frac{8}{13}\ , \qquad Cov[W,X] = \frac{8}{13} - \frac{1}{2}\frac{16}{13} = 0$$

$$E[WY] = 2\times1\times\frac{1}{4} = \frac{1}{2}\ , \qquad Cov[W,Y] = \frac{1}{2} - \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

$$E[WZ] = 0\ , \qquad Cov[W,Z] = 0 - \frac{1}{2}\frac{3}{13} = -\frac{3}{26}$$

$$E[XY] = 4\times1\times\frac{6}{52} + 2\times1\times\frac{4}{52} = \frac{8}{13}\ , \qquad Cov[X,Y] = \frac{8}{13} - \frac{16}{13}\frac{1}{2} = 0$$

$$E[XZ] = 4\times2\times\frac{6}{52} = \frac{12}{13}\ , \qquad Cov[X,Z] = \frac{12}{13} - \frac{16}{13}\frac{3}{13} = \frac{108}{169}$$

$$E[YZ] = 0\ , \qquad Cov[Y,Z] = 0 - \frac{1}{2}\frac{3}{13} = -\frac{3}{26}$$

The correlations $Corr[W,X]$ and $Corr[X,Y]$ are clearly zero; the remainder are:

$$Corr[W,Y] = \frac{1}{4}\left(\frac{3}{4}\times\frac{1}{4}\right)^{-1/2} = 0.577$$

$$Corr[W,Z] = -\frac{3}{26}\left(\frac{3}{4}\times\frac{69}{169}\right)^{-1/2} = -0.209$$

$$Corr[X,Z] = \frac{108}{169}\left(\frac{472}{169}\times\frac{69}{169}\right)^{-1/2} = 0.598$$

$$Corr[Y,Z] = -\frac{3}{26}\left(\frac{1}{4}\times\frac{69}{169}\right)^{-1/2} = -0.361$$

Finally, then, we can write down the correlation matrix:

$$\rho = \begin{pmatrix} 1 & 0 & 0.58 & -0.21 \\ 0 & 1 & 0 & 0.60 \\ 0.58 & 0 & 1 & -0.36 \\ -0.21 & 0.60 & -0.36 & 1 \end{pmatrix}$$

As would be expected, *X* is uncorrelated with either *W* or *Y*, colour and face-value being two independent characteristics. Positive correlations are to be expected between *W* and *Y* and between *X* and *Z*; both correlations are fairly strong. Moderate anticorrelations exist between *Z* and both *W* and *Y*.

6.5 Important joint distributions

In this section we will examine an important multivariate distribution, the *multinomial distribution*, which is an extension of the binomial distribution.
The binomial distribution describes the probability of obtaining *x* 'successes' from *n* independent trials, where each trial has only two possible outcomes. This may be generalized to the case where each trial has *k* possible outcomes with respective probabilities $p_1, p_2, \ldots, p_k$. If we consider the random variables X_i, $i = 1, 2, \ldots, k$ to be the number of outcomes of type *i* in *n* trials then we may calculate their joint probability function

$$f\left(x_1, x_2, \ldots, x_k\right) = \Pr\left(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k\right) \qquad (6.24)$$

where we must have $\sum_{i=1}^{k} x_i = n$. In *n* trials the probability of obtaining x_1 outcomes of type 1, followed by x_2 outcomes of type 2 etc. is given by

$$p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \qquad (6.25)$$

However, the number of distinguishable permutations of this result is

$$\frac{n!}{x_1! x_2! \cdots x_k!}$$

and thus

$$f\left(x_1, x_2, \ldots, x_k\right) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} \qquad (6.26)$$

This is the *multinomial probability distribution*.
If $k = 2$ then the multinomial distribution reduces to the familiar binomial distribution. Although in this form the binomial distribution appears to be a function of two random variables, it must be remembered that, in fact, since $p_2 = 1 - p_1$ and $x_2 = n - x_1$, the distribution of X_1 is entirely determined by the parameters *p* and *n*. X_1 has a binomial distribution. In fact, any of the random variables X_i has a binomial distribution, i.e. the *marginal distribution* of each X_i is binomial with parameters *n* and p_i. It immediately follows that

$$E\left[X_i\right] = np_i \quad \text{and} \quad V\left[X_i\right] = np_i\left(1 - p_i\right) \qquad (6.27)$$

Exercise 7. At a village fête patrons were invited, for a 10 p entry fee, to pick without looking six tickets from a drum containing equal large numbers of red, blue and green tickets. If five or more of the tickets were of the same color a prize of 100 p was awarded. A consolation award of 40 p was made if two tickets of each color were picked. Was a good time had by all?

In this case, all types of outcome (red, blue and green) have the same probabilities. The probability of obtaining any given combination of tickets is given by the multinomial distribution $f\left(x_1, x_2, \ldots, x_k\right) = \dfrac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$ with $n = 6$, $k = 3$ and $p_i = \dfrac{1}{3}$, $i = 1, 2, 3$.

(i) The probability of picking six tickets of the same color is given by

$$\Pr\left(\text{six of the same colour}\right) = 3 \times \frac{6!}{6!0!0!}\left(\frac{1}{3}\right)^6\left(\frac{1}{3}\right)^0\left(\frac{1}{3}\right)^0 = \frac{1}{243}$$

The factor of 3 is present because there are three different colors.

(ii) The probability of picking five tickets of one color and one ticket of another colour is

$$\Pr\left(\text{five of one colour; one of another}\right) = 3 \times 2 \times \frac{6!}{5!1!0!}\left(\frac{1}{3}\right)^5\left(\frac{1}{3}\right)^1\left(\frac{1}{3}\right)^0 = \frac{4}{81}$$

The factors of 3 and 2 are included because there are three ways to choose the color of the five matching tickets, and then two ways to choose the color of the remaining ticket.

(iii) Finally, the probability of picking two tickets of each color is

$$\Pr\left(\text{two of each colour}\right) = \frac{6!}{2!2!2!}\left(\frac{1}{3}\right)^2\left(\frac{1}{3}\right)^2\left(\frac{1}{3}\right)^2 = \frac{10}{81}$$

Thus the expected return to any patron was, in pence,

$$100\left(\frac{1}{243} + \frac{4}{81}\right) + \left(40 \times \frac{10}{81}\right) = 10.29$$

A good time was had by all but the stallholder!