3.4.4 Moments

The mean (or expectation) of X is sometimes called the *first moment* of X, since it is defined as the sum or integral of the probability density function multiplied by the first power of x. By a simple extension the k-th moment of a distribution is defined by

$$\mu_k = E\left[X^k\right] = \begin{cases} \sum_j x_j^k f\left(x_j\right) \text{ for a discrete distribution} \\ \int x^k f\left(x\right) dx \text{ for a continuous distriution} \end{cases} \tag{3.22}$$

The symbol μ_k denote $E\left[X^k\right]$, the k-th moment of the distribution. The mean of the distribution is then denoted by μ_1, often abbreviated simply to μ.

A useful result that relates the *second moment*, the *mean* and the *variance* of a distribution is proved using the properties of the expectation operator:

$$\begin{aligned} V\left[X\right] &= E\left[\left(X-\mu\right)^2\right] = E\left[X^2 - 2\mu X + \mu^2\right] = \\ &= E\left[X^2\right] - 2\mu E\left[X\right] + \mu^2 = \\ &= E\left[X^2\right] - 2\mu^2 + \mu^2 = E\left[X^2\right] - \mu^2 \end{aligned}$$

$$V\left[X\right] = E\left[X^2\right] - \mu^2 \tag{3.23}$$

In alternative notations, this result can be written

$$\left\langle \left(x-\mu\right)^2 \right\rangle = \left\langle x^2 \right\rangle - \left\langle x \right\rangle^2 \quad \text{or} \quad \sigma^2 = \mu_2 - \mu_1^2 \tag{3.24}$$

Exercise 1. A biased die has probabilities $p/2, p, p, p, p, 2p$ of showing 1, 2, 3, 4, 5, 6 respectively. Find (i) the mean, (ii) the second moment and (iii) the variance of this probability distribution.

By demanding that the sum of the probabilities equals unity we require $p = 2/13$. Using the definition of the mean (3.16) for a discrete distribution,

$$\begin{aligned} E\left[X\right] &= \sum_j x_j f\left(x_j\right) = 1\times\frac{1}{2}p + 2\times p + 3\times p + 4\times p + 5\times p + 6\times 2p \\ &= \frac{53}{2}p = \frac{53}{2}\times\frac{2}{13} = \frac{53}{13} \approx 4.07 \end{aligned}$$

Using the definition of the second moment (3.22),

$$E\left[X^2\right]=\sum_j x_j^2 f\left(x_j\right)=1^2\times\frac{1}{2}p+2^2\times p+3^2\times p+4^2\times p+5^2\times p+6^2\times 2p$$
$$=\frac{253}{2}p=\frac{253}{13}$$

Using the definition of the variance (3.19), with $\mu=53/13$, we obtain

$$V[X]=\sum_j\left(x_j-\mu\right)^2 f\left(x_j\right)$$
$$=(1-\mu)^2\frac{1}{2}p+(2-\mu)^2 p+(3-\mu)^2 p+(4-\mu)^2 p+(5-\mu)^2 p+(6-\mu)^2 2p$$
$$=\frac{1320}{169}p=\frac{480}{169}$$

It is easy to verify that $V[X]=E\left[X^2\right]-\left(E[X]\right)^2$.

In practice, to calculate the moments of a distribution it is often simpler to use the moment generating function discussed later.

3.4.5 Central moments

The variance $V[X]$ is sometimes called the *second central moment* of the distribution, since it is defined as the sum or integral of the probability density function multiplied by the second power of $x-\mu$. The origin of the term 'central' is that by subtracting μ from *x* before squaring we are considering the moment about the mean of the distribution, rather than about $x=0$. Thus the *k*-th central moment of a distribution is defined as

$$\nu_k=E\left[(X-\mu)^k\right]=\begin{cases}\sum_j\left(x_j-\mu\right)^k f\left(x_j\right) & \text{for a discrete distribution}\\ \int(x-\mu)^k f(x)\,dx & \text{for a continuous distriution}\end{cases}\qquad(3.25)$$

ν_k is the notation for the *k*-th central moment. Thus $V[X]=\nu_2$ and we may write (3.23) as $\nu_2=\mu_2-\mu_1^2$. The first central moment of a distribution is always zero since, for example in the continuous case,

$$\nu_1=\int(x-\mu)f(x)\,dx=\int xf(x)\,dx-\mu\int f(x)\,dx=\mu-\mu=0$$

We can write the *k*-th central moment of a distribution in terms of its *k*-th and lower-order moments by expanding $(X-\mu)^k$ in powers of *X*. We have already noted that $\nu_2=\mu_2-\mu_1^2$, and similar expressions may be obtained for higher-order central moments. For example,

$$\begin{aligned} \nu_3 = E\left[(X-\mu_1)^3\right] &= E\left[X^3 - 3\mu_1 X^2 + 3\mu_1^2 X - \mu_1^3\right] \\ &= E\left[X^3\right] - 3\mu_1 E\left[X^2\right] + 3\mu_1^2 E\left[X\right] - \mu_1^3 \\ &= \mu_3 - 3\mu_1\mu_2 + 3\mu_1^2\mu_1 - \mu_1^3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \end{aligned} \tag{3.26}$$

$$\begin{aligned} \nu_4 = E\left[(X-\mu_1)^4\right] &= E\left[X^4 - 4\mu_1 X^3 + 6\mu_1^2 X^2 - 4\mu_1^3 X + \mu_1^4\right] \\ &= E\left[X^4\right] - 4\mu_1 E\left[X^3\right] + 6\mu_1^2 E\left[X^2\right] - 4\mu_1^3 E[X] + \mu_1^4 \\ \nu_4 &= \mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4 \end{aligned} \tag{3.27}$$

In general, it is straightforward to show that

$$\nu_k = \mu_k - C_k^1\mu_{k-1}\mu_1 + \ldots + (-1)^r C_k^r \mu_{k-r}\mu_1^r + \ldots + (-1)^{k-1}\left(C_k^{k-1} - 1\right)\mu_1^k \tag{3.28}$$

Direct evaluation of the sum or integral in (3.25) can be rather tedious for higher moments, and it is usually quicker to use the moment generating function, from which the central moments can be easily evaluated as well.

Exercise 2. The PDF for a Gaussian distribution with mean μ and variance σ^2 is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Obtain an expression for the *k*-th central moment of this distribution.

The *k*-th central moment of $f(x)$ is given by

$$\begin{aligned} \nu_k &= \int_{-\infty}^{\infty} (x-\mu)^k f(x)\,dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^k e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \end{aligned}$$

We made the substitution,

$$y = x - \mu \qquad dy = dx$$

$$\nu_k = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} y^k e^{-\frac{y^2}{2\sigma^2}} dy \tag{3.29}$$

It is clear that if *k* is *odd* then the integrand is an odd function of *y* and hence the integral equals zero. Thus, $\nu_k = 0$ if *k* is odd $k = 2n+1$. When *k* is even, we could calculate ν_k by integrating by parts to obtain a reduction formula, but it is more elegant to consider instead the standard integral

$$I = \int_{-\infty}^{\infty} e^{-\alpha y^2} dy = \sqrt{\frac{\pi}{\alpha}}$$

$$I = \int_{-\infty}^{\infty} e^{-\alpha y^2} dy = \pi^{\frac{1}{2}} \alpha^{-\frac{1}{2}} \tag{3.30}$$

and differentiate it repeatedly with respect to α . Thus, we obtain

$$\frac{dI}{d\alpha} = -\int_{-\infty}^{\infty} y^2 e^{-\alpha y^2} dy = -\frac{1}{2}\pi^{\frac{1}{2}}\alpha^{-\frac{3}{2}}$$

$$\frac{d^2 I}{d\alpha^2} = \int_{-\infty}^{\infty} y^4 e^{-\alpha y^2} dy = \frac{1}{2}\frac{3}{2}\pi^{\frac{1}{2}}\alpha^{-\frac{5}{2}}$$

$$\frac{d^3 I}{d\alpha^3} = -\int_{-\infty}^{\infty} y^6 e^{-\alpha y^2} dy = -\frac{1}{2}\frac{3}{2}\frac{5}{2}\pi^{\frac{1}{2}}\alpha^{-\frac{7}{2}}$$

$$\vdots$$

$$\frac{d^n I}{d\alpha^n} = (-1)^n \int_{-\infty}^{\infty} y^{2n} e^{-\alpha y^2} dy = (-1)^n \frac{1}{2}\frac{3}{2}\frac{5}{2}\cdots\frac{2n-1}{2}\pi^{\frac{1}{2}}\alpha^{-\frac{2n+1}{2}}$$

Setting $\alpha = \frac{1}{2\sigma^2}$ and substituting the above result into (3.29), we find (for *k* even)

$$\nu_k = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{2}\frac{3}{2}\frac{5}{2}\cdots\frac{k-1}{2}\pi^{\frac{1}{2}}\left(\frac{1}{2\sigma^2}\right)^{-\frac{k+1}{2}}$$

$$\nu_k = 1\cdot 3\cdot 5\cdots(k-1)\sigma^k \qquad \text{k par} \tag{3.31}$$

Generally, one may also characterise a probability distribution $f(x)$ using the closely related *normalised* and *dimensionless central moments*

$$\gamma_k = \frac{\nu_k}{\nu_2^{k/2}} = \frac{\nu_k}{\sigma^k} \tag{3.32}$$

γ_3 and γ_4 are more commonly called, respectively, the *skewness* and *kurtosis* of the distribution. The skewness γ_3 of a distribution is zero if it is symmetrical about its mean. If the distribution is skewed to values of *x* smaller than the mean then $\gamma_3 < 0$. Similarly $\gamma_3 > 0$ if the distribution is skewed to higher values of *x*.

From the above example, we see that the *kurtosis* of the Gaussian distribution is given by

$$\gamma_4 = \frac{\nu_4}{\nu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3 \tag{3.33}$$

It is therefore common practice to define the *excess kurtosis* of a distribution as $\gamma_4 - 3$. A positive value of the excess kurtosis implies a relatively narrower peak and wider wings than the Gaussian distribution with the same mean and variance. A negative excess kurtosis implies a wider peak and shorter wings.

3.4.5 Functions of random variables

Suppose *X* is some random variable for which the probability density function $f(x)$ is known. In many cases, we are more interested in a related random variable $Y = Y(X)$, where $Y(X)$ is some function of *X*. What is the probability density function $g(y)$ for the new random variable *Y* ? We now discuss how to obtain this function.

1. Discrete random variables

If *X* is a discrete RV that takes only the values x_i, $i = 1, 2, \dots, n$, then *Y* must also be discrete and takes the values $y_i = Y(x_i)$, although some of these values may be identical. The probability function for *Y* is given by

$$g(y) = \begin{cases} \sum_j f(x_j) & \text{if } y = y_i \\ 0 & \text{otherwise} \end{cases} \tag{3.34}$$

where the sum extends over those values of *j* for which $y_i = Y(x_j)$. The simplest case arises when the function $Y(X)$ possesses a single-valued inverse $X(Y)$. In this case, only one *x*-value corresponds to each *y*-value, and we obtain a closed-form expression for $g(y)$ given by

$$g(y) = \begin{cases} f(x(y_i)) & \text{if } y = y_i \\ 0 & \text{otherwise} \end{cases} \tag{3.35}$$

If $Y(X)$ does not possess a single-valued inverse then the situation is more complicated and it may not be possible to obtain a simple expression for $g(y)$. Nevertheless, whatever the form of $Y(X)$, one can always use (3.34) to obtain the numerical values of the probability function $g(y)$ at $y = y_i$.

2. Continuous random variables

If *X* is a continuous RV, then so too is the new random variable $Y = Y(X)$. The probability that *Y* lies in the range *y* to $y + dy$ is given by

$$g(y)\,dy = \int_{dS} f(x)\,dx \tag{3.36}$$

where dS corresponds to all values of x for which Y lies in the range y to $y+dy$. Once again the simplest case occurs when $Y(X)$ possesses a single-valued inverse $X(Y)$. In this case, we may write

$$g(y)dy = \left| \int_{x(y)}^{x(y+dy)} f(x')dx' \right| = \int_{x(y)}^{x(y)+\left|\frac{dx}{dy}\right|dy} f(x')dx' = \left|\frac{dx}{dy}\right| dy \cdot f(x(y))$$

From which we obtain

$$g(y) = f(x(y))\left|\frac{dx}{dy}\right| \qquad (3.37)$$

Exercise 3. A lighthouse is situated at a distance L from a straight coastline, opposite a point O, and sends out a narrow continuous beam of light simultaneously in opposite directions. The beam rotates with constant angular velocity. If the random variable *Y* is the distance along the coastline, measured from O, of the spot that the light beam illuminates, find its probability density function.

The situation is illustrated in figure 3.3. Since the light beam rotates at a constant angular velocity, θ is distributed uniformly between −π/2 and π/2, and so $f(\theta)=1/\pi$. Now $y = Ltg\theta$, which possesses the single-valued inverse $\theta = arctg\frac{y}{L}$, provided that θ lies between −π/2 and π/2. Since

$$\frac{dy}{d\theta} = L\frac{1}{\cos^2\theta} = L(1+tg^2\theta) = L\left(1+\left(\frac{y}{L}\right)^2\right),$$

From (3.37) we find

$$g(y) = f(x(y))\left|\frac{dx}{dy}\right| = \frac{1}{\pi}\left|\frac{d\theta}{dy}\right| = \frac{1}{\pi L\left(1+(y/L)^2\right)} \quad \text{for } -\infty < y < \infty$$

A distribution of this form is called a *Cauchy distribution*.

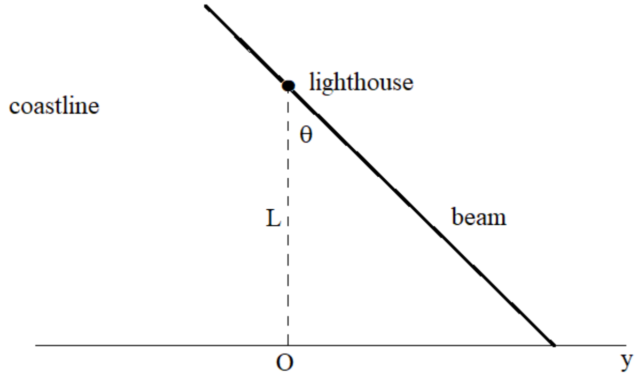

Figure 3.3 The illumination of a coastline by the beam from a lighthouse.

If $Y(X)$ does not possess a single-valued inverse then we encounter complications, since there exist several intervals in the *X*-domain for which *Y* lies between y and $y+dy$. This is illustrated in figure 3.4, which shows a function $Y(X)$ such that $X(Y)$ is a double-valued function of *Y*. The range *y* to $y+dy$ corresponds to *X*'s being either in the range x_1 to x_1+dx_1 or in the range x_2 to x_2+dx_2. In general, it may not be possible to obtain an expression for $g(y)$ in closed form, although the distribution may always be obtained numerically using (3.36).

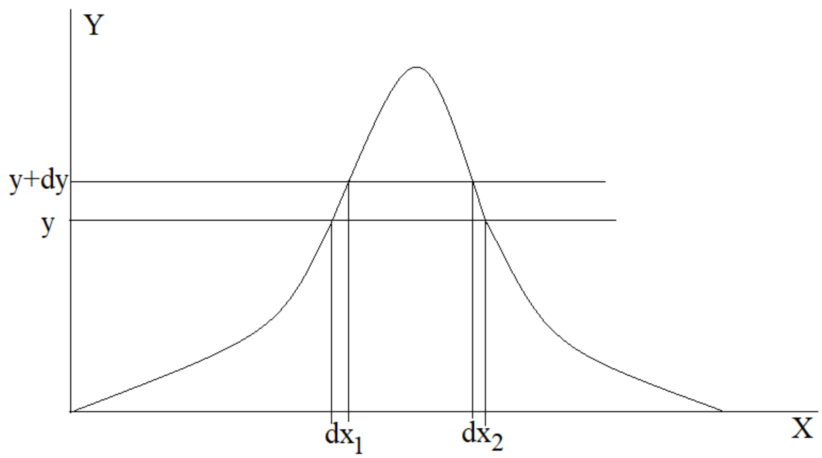

Figure 3.4

However, a closed-form expression may be obtained in the case where there exist single-valued functions $x_1(y)$ and $x_2(y)$ giving the two values of *x* that correspond to any given value of *y*. In this case,

$$g(y)dy = \left|\int_{x_1(y)}^{x_1(y+dy)} f(x)dx\right| + \left|\int_{x_2(y)}^{x_2(y+dy)} f(x)dx\right|,$$

from which we obtain

$$g(y) = f(x_1(y))\left|\frac{dx_1}{dy}\right| + f(x_2(y))\left|\frac{dx_2}{dy}\right| \qquad (3.38)$$

This result may be generalised straightforwardly to the case where the range *y* to $y+dy$ corresponds to more than two *x*-intervals.

Exercise 4. The random variable *X* is Gaussian distributed with mean μ and variance σ^2. Find the PDF of the new variable $Y=(X-\mu)^2/\sigma^2$.

It is clear that $X(Y)$ is a double-valued function of *Y*. However, in this case, it is straightforward to obtain single-valued functions giving the two values of *x* that correspond to a given value of *y*; these are

$$y = \frac{(x-\mu)^2}{\sigma^2} \quad \Rightarrow \quad y\sigma^2 = (x-\mu)^2 \quad \Rightarrow \quad \pm\sigma\sqrt{y} = x - \mu$$
$$x = \mu \pm \sigma\sqrt{y} \quad \Rightarrow \quad x_1 = \mu - \sigma\sqrt{y}\,, \quad x_2 = \mu + \sigma\sqrt{y}$$

where $\sqrt{y}$ is taken to mean the positive square root.

The PDF of *X* is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (3.39)$$

Since $\frac{dx_1}{dy} = -\frac{\sigma}{2\sqrt{y}}$ and $\frac{dx_2}{dy} = \frac{\sigma}{2\sqrt{y}}$, from (3.38) we obtain

$$g(y) = f(x_1(y))\left|\frac{dx_1}{dy}\right| + f(x_2(y))\left|\frac{dx_2}{dy}\right|$$

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}y} \left|\frac{-\sigma}{2\sqrt{y}}\right| + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}y} \left|\frac{\sigma}{2\sqrt{y}}\right|$$

$$g(y) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}y} \frac{\sigma}{2\sqrt{y}}$$

$$g(y) = \frac{1}{2\sqrt{\pi}} \left(\frac{1}{2}y\right)^{-1/2} e^{-\frac{1}{2}y} \qquad (3.40)$$

This is the gamma distribution $\gamma\left(\frac{1}{2},\frac{1}{2}\right)$.

3.4.6. Functions of several random variables

We may extend our discussion further, to the case in which the new random variable is a function of *several* other random variables. Let us consider the random variable $Z = Z(X,Y)$, which is a function of two other RVs *X* and *Y*. Given that these variables are described by the joint probability density function $f(x,y)$, we wish to find the probability density function $p(z)$ of the variable Z.

If *X* and *Y* are both discrete RVs then

$$p(z) = \sum_{i,j} f(x_i, y_j) \quad (3.41)$$

where the sum extends over all values of *i* and *j* for which $Z(x_i, y_j) = z$. Similarly, if *X* and *Y* are both continuous RVs then $p(z)$ is found by requiring that

$$p(z)dz = \iint_{dS} f(x, y)\,dx\,dy \quad (3.42)$$

where *dS* is the infinitesimal area in the *xy*-plane lying between the curves $Z(x, y) = z$ and $Z(x, y) = z + dz$.

3.4.7 Expectation values and variances

In some cases, one is interested only in the expectation value or the variance of the new variable *Z* rather than in its full probability density function. Let us consider the random variable $Z = Z(X, Y)$, which is a function of two RVs *X* and *Y* with a known joint distribution $f(x, y)$.

It is clear that $E[Z]$ and $V[Z]$ can be obtained, in principle, by first using the methods discussed above to obtain $p(z)$ and then evaluating the appropriate sums or integrals. The intermediate step of calculating $p(z)$ is not necessary, however, since it is straightforward to obtain expressions for $E[Z]$ and $V[Z]$ in terms of the variables *X* and *Y*. For example, if *X* and *Y* are continuous RVs then the expectation value of *Z* is given by

$$E[Z] = \int zp(z)dz = \iint Z(x, y) f(x, y)\,dx\,dy \quad (3.43)$$

An analogous result exists for discrete random variables.

Integrals of the form (3.43) are often difficult to evaluate. Nevertheless, we may use (3.43) to derive an important general result concerning expectation values. If *X* and *Y* are any two random variables and *a* and *b* are arbitrary constants then by letting $Z = aX + bY$ we find

$$E[aX + bY] = aE[X] + bE[Y] \quad (3.44)$$

Furthermore, we may use this result to obtain an approximate expression for the expectation value $E[Z(X, Y)]$ of any arbitrary function of *X* and *Y*. Letting $\mu_X = E[X]$ and $\mu_Y = E[Y]$, and provided $Z(X, Y)$ can be reasonably approximated by the linear terms of its Taylor expansion about the point (μ_X, μ_Y), we have

$$Z(X, Y) \approx Z(\mu_X, \mu_Y) + \left(\frac{\partial Z}{\partial X}\right)(X - \mu_X) + \left(\frac{\partial Z}{\partial Y}\right)(Y - \mu_Y) \quad (3.45)$$

where the partial derivatives are evaluated at $X = \mu_X$ and $Y = \mu_Y$. Taking the expectation values of both sides, we find

$$E\left[Z(X,Y)\right] \approx Z(\mu_X,\mu_Y) + \left(\frac{\partial Z}{\partial X}\right)\left(E[X]-\mu_X\right) + \left(\frac{\partial Z}{\partial Y}\right)\left(E[Y]-\mu_Y\right) = Z(\mu_X,\mu_Y)$$

which gives the approximate result:

$$E\left[Z(X,Y)\right] \approx Z(\mu_X,\mu_Y) \tag{3.46}$$

By analogy with (3.43), the variance of $Z = Z(X,Y)$ is given by

$$V[Z] = \int (z-\mu_Z)^2\, p(z)\,dz = \iint \left[Z(x,y)-\mu_Z\right]^2 f(x,y)\,dx\,dy \tag{3.47}$$

where $\mu_Z = E[Z]$. We may use this expression to derive a second useful result. If X and Y are two *independent* random variables, so that $f(x,y) = g(x)h(y)$, and a, b and c are constants then by setting $Z = aX + bY + c$ in (3.47) we obtain

$$V[aX+bY+c] = a^2V[X] + b^2V[Y] \tag{3.48}$$

From (3.48) we also obtain the important special case

$$V[X+Y] = V[X-Y] = V[X] + V[Y] \tag{3.49}$$

Provided X and Y are indeed *independent* random variables, we may obtain an approximate expression for $V\left[Z(X,Y)\right]$, for any arbitrary function $Z(X,Y)$. Taking the variance of both sides of (3.45), and using (3.48), we find

$$V\left[Z(X,Y)\right] \approx \left(\frac{\partial Z}{\partial X}\right)^2 V[X] + \left(\frac{\partial Z}{\partial Y}\right)^2 V[Y] \tag{3.50}$$

the partial derivatives are evaluated at $X = \mu_X$ and $Y = \mu_Y$.

3.5 Generating functions

When dealing with particular sets of functions f_n, each member of the set being characterised by a different non-negative integer n, it is sometimes possible to summarise the whole set by a single function of a dummy variable (say t), called a *generating function*. The relationship between the generating function and the n-th member f_n of the set is that if the generating function is expanded as a power series in t then f_n is the coefficient of t^n. For example, in the expansion of the generating function $G(z,t) = \left(1-2zt+t^2\right)^{-1/2}$, the coefficient of t^n is the n-th Legendre polynomial $P_n(z)$, i.e.

$$G(z,t) = \left(1 - 2zt + t^2\right)^{-1/2} = \sum_{n=0}^{\infty} P_n(z)t^n \qquad (3.51)$$

Many useful properties of, and relationships between, the members of a set of functions could be established using the generating function and other functions obtained from it, e.g. its derivatives.

Similar ideas can be used in the area of probability theory, and two types of generating function can be usefully defined: *probability generating function* and *moment generating function*.

3.5.1 Probability generating functions

Probability generating functions are restricted in applicability to integer distributions, of which the most common (the binomial, the Poisson and the geometric) are considered in this and later subsections. In such distributions a random variable may take only non-negative integer values. The actual possible values may be finite or infinite in number, but, for formal purposes, all integers, $0,1,2,\ldots$ are considered possible. If only a finite number of integer values can occur in any particular case then those that cannot occur are included but are assigned zero probability.

If the probability that the random variable *X* takes the value x_n is $f(x_n)$, then

$$\sum_n f(x_n) = 1$$

However, only non-negative integer values of x_n are possible, and we can, without ambiguity, write the probability that *X* takes the value *n* as f_n, with

$$\sum_{n=0}^{\infty} f_n = 1 \qquad (3.52)$$

We define the *probability generating function* $\Phi_X(t)$ by

$$\Phi_X(t) = \sum_{n=0}^{\infty} f_n\, t^n \qquad (3.53)$$

It is apparent that $\Phi_X(t) = E\left[t^X\right]$ and that, by (3.52), $\Phi_X(1) = 1$.

Probably the simplest example of a probability generating function (PGF) is provided by the random variable *X* defined by (Bernoulli)

$$X = \begin{cases} 1 & \text{if the outcome of a single trial is a 'success'} \\ 0 & \text{if the trial ends in 'failure'} \end{cases}$$

If the probability of success is *p* and that of failure $q = 1 - p$ then

$$\Phi_X(t) = qt^0 + pt^1 + 0 + 0 + \ldots = q + pt \tag{3.54}$$

A Poisson-distributed integer variable with mean λ has a PGF

$$\Phi_X(t) = \sum_{n=0}^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} t^n = e^{-\lambda} e^{\lambda t} \tag{3.55}$$

We note that, as required, $\Phi_X(1) = 1$ in both cases.

Useful results will be obtained by differentiating (3.53) with respect to *t*. Taking the first derivative we find

$$\frac{d\Phi_X(t)}{dt} = \sum_{n=0}^{\infty} n\, f_n\, t^{n-1} \quad \Rightarrow \quad \Phi_X'(1) = \sum_{n=0}^{\infty} n f_n = E[X] \tag{3.56}$$

and differentiating once more we obtain

$$\frac{d^2\Phi_X(t)}{dt^2} = \sum_{n=0}^{\infty} n(n-1)\, f_n\, t^{n-2} \quad \Rightarrow \quad \Phi_X''(1) = \sum_{n=0}^{\infty} n(n-1)\, f_n = E\left[X(X-1)\right] \tag{3.57}$$

Equation (3.56) shows that $\Phi_X'(1)$ gives the mean of *X*. Using both (3.57) and (3.23) $V[X] = E\left[X^2\right] - \mu^2$, allows us to write

$$\begin{aligned} \Phi_X''(1) + \Phi_X'(1) - \left[\Phi_X'(1)\right]^2 &= E\left[X(X-1)\right] + E[X] - \left(E[X]\right)^2 \\ &= E\left[X^2\right] - E[X] + E[X] - \left(E[X]\right)^2 \\ &= E\left[X^2\right] - \left(E[X]\right)^2 = V[X] \end{aligned} \tag{3.58}$$

and so express the variance of *X* in terms of the derivatives of its probability generating function.

Exercise 5. A random variable *X* is given by the number of trials needed to obtain a first success when the chance of success at each trial is constant and equal to *p*. Find the probability generating function for *X* and use it to determine the mean and variance of *X*.

Clearly, at least one trial is needed, and so $f_0 = 0$. If *n* (≥ 1) trials are needed for the first success, the first $n-1$ trials must have resulted in failure. Thus

$$\Pr(X = n) = q^{n-1} p \;\; , \quad n \geq 1 \tag{3.59}$$

where $q = 1 - p$ is the probability of failure in each individual trial. The corresponding probability generating function is thus

$$\Phi_X(t) = \sum_{n=0}^{\infty} f_n \, t^n = \sum_{n=1}^{\infty} \left(q^{n-1} p\right) t^n$$
$$= \frac{p}{q} \sum_{n=1}^{n} (qt)^n = \frac{p}{q} \frac{qt}{1-qt} = \frac{pt}{1-qt} \qquad (3.60)$$

where we have used the result for the sum of a geometric series, to obtain a closed-form expression for $\Phi_X(t)$. Again, as must be the case, $\Phi_X(1) = 1$.

To find the mean and variance of *X* we need to evaluate $\Phi'_X(1)$ and $\Phi''_X(1)$. Differentiating (3.60) gives

$$\Phi'_X(t) = \frac{p}{(1-qt)^2} \qquad \Rightarrow \qquad \Phi'_X(1) = \frac{1}{p}$$
$$\Phi''_X(t) = \frac{2pq}{(1-qt)^3} \qquad \Rightarrow \qquad \Phi''_X(1) = \frac{2q}{p^2}$$

Thus

$$E[X] = \Phi'_X(1) = \frac{1}{p}$$
$$V[X] = \Phi''_X(1) + \Phi'_X(1) - \left[\Phi'_X(1)\right]^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}$$

A distribution with probabilities of the general form (3.59) is known as a *geometric distribution* and is discussed later. This form of distribution is common in 'waiting time' problems.

Sums of random variables

We now turn to considering the sum of two or more *independent* random variables, say *X* and *Y*, and denote by S_2 the random variable

$$S_2 = X + Y$$

If $\Phi_{S_2}(t)$ is the PGF for S_2, the coefficient of t^n in its expansion is given by the probability that $X + Y = n$ and is thus equal to the sum of the probabilities that $X = r$ and $Y = n - r$ for all values of *r* in $0 \le r \le n$. Since such outcomes for different values of *r* are mutually exclusive, we have

$$\Pr(X + Y = n) = \sum_{r=0}^{n} \Pr(X = r) \Pr(Y = n - r) \qquad (3.61)$$

Multiplying both sides by t^n and summing over all values of *n* enables us to express this relationship in terms of probability generating functions as follows:

$$\Phi_{X+Y}(t)=\sum_{n=0}^{\infty}\Pr(X+Y=n)t^{n}=\sum_{n=0}^{\infty}\sum_{r=0}^{n}\Pr(X=r)t^{r}\Pr(Y=n-r)t^{n-r}$$

$$=\sum_{r=0}^{\infty}\sum_{n=r}^{\infty}\Pr(X=r)t^{r}\Pr(Y=n-r)t^{n-r}$$

The change in summation order is justified by reference to figure 3.5, which illustrates that the summations are over exactly the same pairs of values of *n* and *r*, but with the first (inner) summation over the points in a column rather than over the points in a row.

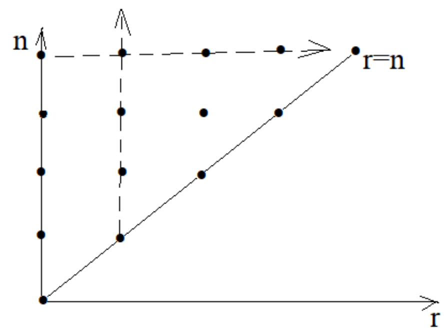

Figure 3.5

Now, setting $n=r+s$ gives the final result,

$$\Phi_{X+Y}(t)=\sum_{r=0}^{\infty}\Pr(X=r)t^{r}\sum_{s=0}^{\infty}\Pr(Y=s)t^{s}$$

$$=\Phi_X(t)\Phi_Y(t) \qquad (3.62)$$

i.e. the PGF of the sum of two *independent* random variables is equal to the product of their individual PGFs.

$$E\left[t^{X+Y}\right]=E\left[t^{X}\right]E\left[t^{Y}\right] \qquad (3.63)$$

Clearly result (3.62) can be extended to more than two random variables:

$$\Phi_{\sum_{i=1}^{n}X_i}(t)=\prod_{i=1}^{n}\Phi_{X_i}(t) \qquad (3.64)$$

and, further, if all the X_i have the same probability distribution,

$$\Phi_{\sum_{i=1}^{n}X_i}(t)=\left[\Phi_X(t)\right]^{n} \qquad (3.65)$$

This latter result has immediate application in the deduction of the PGF for the binomial distribution from that for a single trial, equation (3.54).

3.5.2 Moment generating functions

For a random variable *X*, and a real number *t*, the *moment generating function* (MGF) is defined by

$$M_X(t) = E\left[e^{tX}\right] = \begin{cases} \sum_i e^{tx_i} f(x_i) & \text{for a discrete distribution} \\ \int e^{tx} f(x)\,dx & \text{for a continuous distribution} \end{cases} \tag{3.66}$$

$M(0) = E[1] = 1$

The PGF and the MGF for a random variable *X* are closely related:

$$\Phi_X(t) = E\left[t^X\right], \qquad M_X(t) = E\left[e^{tX}\right] \tag{3.67}$$

The MGF can thus be obtained from the PGF by replacing *t* by e^t, and vice versa. The MGF has more general applicability, however, since it can be used with both continuous and discrete distributions whilst the PGF is restricted to non-negative integer distributions.

As its name suggests, the MGF is particularly useful for obtaining the moments of a distribution, as is easily seen by noting that

$$\begin{aligned} M_X(t) = E\left[e^{tX}\right] &= E\left[1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \dots\right] \\ &= 1 + E[X]t + E\left[X^2\right]\frac{t^2}{2!} + E\left[X^3\right]\frac{t^3}{3!} + \dots \end{aligned} \tag{3.68}$$

Assuming that the MGF exists for all *t* around the point $t = 0$, we can deduce that the moments of a distribution are given in terms of its MGF by

$$E\left[X^n\right] = \left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0} \tag{3.69}$$

Similarly, by substitution in $V[X] = E\left[X^2\right] - \left(E[X]\right)^2$, the variance of the distribution is given by

$$V[X] = M''_X(0) - \left[M'_X(0)\right]^2 \tag{3.70}$$

where the prime denotes differentiation with respect to *t*.

Exercise 6. The MGF for the Gaussian distribution is given by

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \tag{3.71}$$

Find the expectation and variance of this distribution.

Using (3.69),

$$M'_X(t) = \left(\mu + \sigma^2 t\right) e^{\mu t + \frac{1}{2}\sigma^2 t^2} \qquad E[X] = M'_X(0) = \mu$$

$$M''_X(t) = \left[\sigma^2 + \left(\mu + \sigma^2 t\right)^2\right] e^{\mu t + \frac{1}{2}\sigma^2 t^2} \qquad E\left[X^2\right] = M''_X(0) = \sigma^2 + \mu^2$$

Thus, using (3.70),

$$V[X] = M''_X(0) - \left[M'_X(0)\right]^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

The mean is found to be μ and the variance σ^2 and justifies the use of these symbols in the Gaussian distribution.

The moment generating function has several useful properties that follow from its definition and can be employed in simplifying calculations.

Scaling and shifting

If $Y = aX + b$, where *a* and *b* are arbitrary constants, then

$$M_Y(t) = E\left[e^{tY}\right] = E\left[e^{t(aX+b)}\right] = e^{tb} E\left[e^{taX}\right] = e^{bt} M_X(at)$$

$$M_Y(t) = e^{bt} M_X(at) \tag{3.72}$$

This result is often useful for obtaining the central moments of a distribution. If the MFG of *X* is $M_X(t)$ then the variable $Y = X - \mu$ has the MGF $M_Y(t) = e^{-\mu t} M_X(t)$, which clearly generates the *central moments* of *X*, i.e.

$$E\left[(X-\mu)^n\right] = E\left[Y^n\right] = M_Y^{(n)}(0) = \left(\frac{d^n}{dt^n}\left[e^{-\mu t} M_X(t)\right]\right)_{t=0} \tag{3.73}$$

Sums of random variables

If $X_1, X_2, \ldots, X_N$ are *independent* random variables and $S_N = X_1 + X_2 + \ldots + X_N$ then

$$M_{S_N}(t) = E\left[e^{tS_N}\right] = E\left[e^{t(X_1 + X_2 + \ldots + X_N)}\right] = E\left[\prod_{i=0}^{N} e^{tX_i}\right]$$

Since the X_i are *independent*,

$$M_{S_N}(t) = \prod_{i=1}^{N} E\left[e^{tX_i}\right] = \prod_{i=1}^{N} M_{X_i}(t) \tag{3.74}$$

In words, the MGF of the sum of *N* independent random variables is the product of their individual MGFs. By combining (3.74) with (3.72), we obtain the more general result 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

$$M_{S_N}(t) = \prod_{i=1}^{N} M_{X_i}(c_i t) \tag{3.75}$$

Uniqueness

If the MGF of the random variable X_1 is identical to that for X_2 then the probability distributions of X_1 and X_2 are identical. This is intuitively reasonable although a rigorous proof is complicated, and beyond the scope of this course.