CHAPTER 10

Statistical Distributions

In this chapter, we shall present some probability distributions that play a central role in econometric theory. First, we shall present the distributions of some discrete random variables that have either a finite set of values or that take values that can be indexed by the entire set of positive integers. We shall also present the multivariate generalisations of one of these distributions.

Next, we shall present the distributions of some continuous random variables that take values in intervals of the real line or over the entirety of the real line. Amongst these is the normal distribution, which is of prime importance and for which we shall consider, in detail, the multivariate extensions.

Associated with the multivariate normal distribution are the so-called sampling distributions that are important in the theory of statistical inference. We shall consider these distributions in the final section of the chapter, where it will transpire that they are special cases of the univariate distributions described in the preceding section.

Discrete Distributions

Suppose that there is a population of N elements, Np of which belong to class $\mathcal{A}$ and $N(1-p)$ to class $\mathcal{A}^c$. When we select n elements at random from the population in n successive trials, we wish to know the probability of the event that x of them will be in $\mathcal{A}$ and that $n-x$ of them will be in $\mathcal{A}^c$.

The probability will be affected by the way in which the n elements are selected; and there are two ways of doing this. Either they can be put aside after they have been sampled, or else they can be restored to the population. Therefore we talk of sampling without replacement and of sampling with replacement.

If we sample with replacement, then the probabilities of selecting an element from either class will the same in every trial, and the size N of the population will have no relevance. In that case, the probabilities are governed by the binomial law. If we sample without replacement, then, in each trial, the probabilities of selecting elements from either class will depend on the outcomes of the previous trials and upon the size of the population; and the probabilities of the outcomes from n successive trials will be governed by the hypergeometric law.

Binomial Distribution

When there is sampling with replacement, the probability is p that an element selected at random will be in class $\mathcal{A}$, and the probability is $1-p$ that it will be in class $\mathcal{A}^c$. Moreover, the outcomes of successive trials will be statistically independent. Therefore, if a particular sequence has x elements in $\mathcal{A}$ in $n-x$ elements $\mathcal{A}^c$, then, as a statistical outcome, its probability will be $p^x(1-p)^{n-x}$.

There are altogether $nC_x = n!/\{(n-x)!x!\}$ such sequences, with x elements in $\mathcal{A}$ in $n-x$ elements in $\mathcal{A}^c$. These sequences represent a set of mutually exclusive ways in which the event in question can occur; and their probabilities can be added to give the probability of the event of which they are the particular instances. Therefore, the probability of the finding x elements in class $\mathcal{A}$ after n trials is given by the *binomial* probability function:

$$b(x; n, p) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}. \tag{1}$$

The number x of the elements in class $\mathcal{A}$ is commonly described as the number of successes, in which case $n-x$ is the number of failures. The archetypal example of a binomial random variable is the number of heads in n tosses of a coin.

The moment generating function of the binomial distribution is given by

$$\begin{aligned} M(x,t) = E(e^{xt}) &= \sum_{x=0}^{n} e^{xt}\frac{n!}{x!(n-x)!}p^x q^{n-x} \\ &= \sum_{x=0}^{n} \frac{n!}{x!(n-x)!}(pe^t)^x q^{n-x} \\ &= (q+pe^t)^n. \end{aligned} \tag{2}$$

By differentiating the function with respect t and then setting $t=0$, we can find the following moments:

$$\begin{aligned} E(x) &= np, \\ E(x^2) &= np(q+np), \\ V(y) &= E(x^2) - \{E(x)\}^2 = npq. \end{aligned} \tag{3}$$

Example. A standardised version of the binomial outcome is obtained by subtracting the mean np and by dividing by the standard deviation $\sqrt{npq}$. The result is

$$z = \frac{x-np}{\sqrt{npq}} = \frac{(x/n)-p}{\sqrt{pq/n}}. \tag{4}$$

It will be demonstrated later, in the context of our treatment of the normal distribution, that, as the number n of the trails increases, the distribution of the standardised binomial variable tends to the standard normal distribution $N(0,1)$, which has a mean of zero and a variance of unity. This theorem justifies the use of the normal distribution in approximating the binomial probabilities when the sample size is large.

One application of this result is in testing an hypothesis concerning the probability of the event that is designated a success. If n is sufficiently large, then the proportion of successes x/n will be approximately normally distributed with mean p and variance pq/n . The RHS of equation (4) shows how the standardised binomial can be expressed in term of proportions. If the value z, which is formed on the basis of an hypothesised value of p, is remote from zero, then the hypothesis will be called into question.

An hypothesis concerning a probability or a proportion within a population can also be tested in reference to the chi-square distribution of one degree of freedom, denoted $\chi^2(1)$, which is the distribution of the square of a standard normal variate. Let p_1 denote the probability of the event in question and let $p_2 = 1 - p_1$ denote the probability of its nonoccurrence. Then the following statistic has a limiting $\chi^2(1)$ distribution:

$$\begin{aligned} z^2 &= \frac{(x_1 - np_1)^2}{np_1} + \frac{(x_2 - np_2)^2}{np_2} \\ &= \frac{(x_1 - np_1)^2 p_2 + (\{x_2 - n\} + n\{1 - p_2\})^2 p_1}{np_1 p_2} \\ &= \frac{(x_1 - np_1)^2}{np_1(1 - p_1)}, \end{aligned} \tag{5}$$

where the final equality is in consequence of $n - x_2 = x_1$ and $1 - p_2 = p_1$. The first term on the RHS, which takes the form of the classical Pearson goodness-of-fit statistic, treats the categories of occurrence and non-occurrence (or of success and failure) in a symmetrical fashion. The final term of the RHS can be recognised as the square of the term on the RHS of (4). The latter uses an alternative notation in which $p_1 = p$ and $1 - p_1 = q$.

If the hypothesised values of p_1 and p_2 lead to a value of z that is improbably large, then it is likely that they differ significantly from the true parameter values of the population.

The Hypergeometric Distribution

There are $NC_n = N!/\{(N - n)!n!\}$ different ways of selecting n elements from N. Therefore, there are NC_n different samples of size n that may be selected without replacement from a population of size N.

We are supposing that there are Np element of the population in class $\mathcal{A}$ and $N(1 - p)$ in class $\mathcal{A}^c$. Therefore, there are

$$\{Np\}C_x \times \{N(1-p)\}C_{n-x} = \frac{Np!}{(Np - x)!x!} \times \frac{N(1-p)!}{\{N(1-p) - n - x\}!(n - x)!} \tag{6}$$

different sets of size n which can be selected from the population that contains x elements from $\mathcal{A}$ and $n - x$ elements from $\mathcal{A}^c$.

It follows that, in selecting n elements from the population without replacement, the probability of getting x in class $\mathcal{A}$ and $n - x$ in class $\mathcal{A}^c$ is

$$h(x; N, n, p) = \frac{NpC_x \times N(1-p)C_{n-x}}{NC_n}. \tag{7}$$

This is the *hypergeometric* probability function. In this case, the moment or cumulant generating functions are of little help in finding the moments, which becomes a somewhat tedious and intractable business.

If N is large and n and x are fixed, then selection without replacement is virtually the same as selection with replacement. Therefore, we should expect the hypergeometric distribution to converge to the binomial distribution as $N \to \infty$. From (7), we have

$$h(x) = \frac{(Np)!}{x!(Np-x)!} \times \frac{\{N(1-p)\}!}{(n-x)!\{N(1-p)-(n-x)\}!} \times \frac{(N-n)!n!}{N!}$$

(8)

$$= \frac{n!}{x!(n-x)!} \times \frac{\{Np(Np-1)\cdots(Np-x+1)\}\{Nq(Nq-1)\cdots(Nq-(n-x)+1)\}}{N(N-1)(N-2)\cdots(N-n+1)},$$

where $q = 1-p$. In the term on the RHS of the final expression, there are n factors in both the numerator and the denominator. Therefore, dividing numerator and denominator by N^n gives

$$(9)\qquad h(x) = nC_x\frac{\{p(p-\frac{1}{N})\cdots(p-\frac{x-1}{N})\}\{q(q-\frac{1}{N})\cdots(q-\frac{n+x-1}{N})\}}{\{(1-\frac{1}{N})(1-\frac{2}{N})\cdots(1-\frac{n-1}{N})\}}.$$

For any fixed x and n, there are

$$(10)\quad \lim(N\to\infty)\left\{p(p-\frac{1}{N})\cdots(p-\frac{x-1}{N})\right\} = p^x,$$

$$(11)\quad \lim(N\to\infty)\left\{q(q-\frac{1}{N})\cdots(q-\frac{n+x-1}{N})\right\} = q^{n-x} = (1-p)^{n-x},$$

$$(12)\quad \lim(N\to\infty)\left\{(1-\frac{1}{N})(1-\frac{2}{N})\cdots(1-\frac{x-1}{N})\right\} = 1.$$

Therefore,

$$(13)\qquad h(x;N,n,p) \to b(x;n,p) = nC_xp^x(1-p)^{n-x} \quad \text{as} \quad N\to\infty.$$

That is to say:

(14) If $x \sim h(p,N,n)$, then, when N is large, it is distributed approximately as $b(p,n)$.

The Poisson Distribution

The Poisson distribution may be derived directly as the probability of a rare event in a large number of trials, or else it may be derived as a limiting case of the binomial distribution. We shall begin by taking the latter approach.

Therefore, consider, $x \sim b(p;n)$, where $np = \mu$ is constant; and let $n \to \infty$, so that $p = \mu/n \to 0$. We can set $(1-p)^{n-x} = (1-p)^n(1-p)^{-x}$ and $p = \mu/n$ in equation (1) to give

$$(15)\qquad b(x) = \frac{n!}{(n-x)!x!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}.$$

The RHS of this equation may be re-arranged as follows:

$$(16)\qquad b(x) = \frac{\mu^x}{x!}\frac{n!}{(n-x)!n^x}\left(1-\frac{\mu}{n}\right)^n\left(1-\frac{\mu}{n}\right)^{-x}.$$

The expression may be disassembled for the purpose of taking limits in its component parts. The limits in question are

$$\lim(n \to \infty)\frac{n!}{(n-x)!n^x} = 1, \tag{17}$$

$$\lim(n \to \infty)\left(1 - \frac{\mu}{n}\right)^n = e^{-\mu}, \tag{18}$$

$$\lim(n \to \infty)\left(1 - \frac{\mu}{n}\right)^{-x} = 1. \tag{19}$$

The first of these becomes intelligible when the term in question is written as

$$\begin{aligned} \frac{n!}{(n-x)!n^x} &= \frac{n(n-1)\cdots(n-x+1)}{n^x} \\ &= 1\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{x+1}{n}\right); \end{aligned} \tag{20}$$

for each of the factors tends to unity as n increases. On reassembling the parts, it is found that the binomial function has a limiting form of

$$P(x;\mu) = \frac{\mu^x e^{-\mu}}{x!}. \tag{21}$$

This is the *Poisson* probability function.

The Poisson function can be derived by considering a specification for a so-called emission process or an arrival process. One can imagine a Geiger counter which registers the impacts of successive radioactive alpha particles upon a thin metallic film. Let $f(x,t)$ denote the probability of x impacts or arrivals in the time interval $(0,t]$. The following conditions are imposed:

(a) The probability of a single arrival in a very short time interval $(t, t + \Delta t]$ is $f(1, \Delta t) = a\Delta t$,

(b) The probability of more than one arrival during that time interval is negligible,

(c) The probability of an arrival during the time interval is independent of any occurrences in previous periods.

Certain consequences follow from these assumptions; and it can be shown that the Poisson distribution is the only distribution that fits the specification.

As a first consequence, it follows from the assumptions that the probability of there being x arrivals in the interval $(0, t + \Delta t]$ is

$$\begin{aligned} f(x, t + \Delta t) &= f(x,t)f(0,\Delta t) + f(x-1,t)f(1,\Delta t) \\ &= f(x,t)(1 - a\Delta t) + f(x-1,t)a\Delta t. \end{aligned} \tag{22}$$

This expression follows from the fact that there are two mutually exclusive ways in which the circumstance can arise. The first way is when all of the x arrivals occur in the interval $(0,t]$ and none occur in the interval $(t,t+\Delta t]$. The second way is when $x-1$ of the arrivals occur in the interval $(0,t]$ and one arrival occurs in the interval $(t,t+\Delta t]$. All other possibilities are ruled out by assumption (b). Assumption (c) implies that the probabilities of these two mutually exclusive or disjoint events are obtained by multiplying the probabilities of the events of the two sub-intervals.

The next step in the chain of deductions is to find the derivative of the function $f(x,t)$ with respect to t. From equation (22), it follows immediately that

$$\begin{aligned}\frac{df(x,t)}{dt} &= \lim(\Delta t\to 0)\frac{f(x,t+\Delta t)-f(x,t)}{\Delta t}\\ &= a\big\{f(x-1,t)-f(x,t)\big\}.\end{aligned} \tag{23}$$

The final step is to show that the function

$$f(x,t)=\frac{(at)^x e^{-at}}{x!} \tag{24}$$

satisfies the condition of equation (23). This is a simple matter of confirming that, according to the product rule of differentiation, we have

$$\begin{aligned}\frac{df(x,t)}{dt} &= \frac{ax(at)^{x-1}e^{-at}}{x!}-\frac{a(at)^x e^{-at}}{x!}\\ &= a\left\{\frac{(at)^{x-1}e^{-at}}{(x-1)!}-\frac{(at)^x e^{-at}}{x!}\right\}\\ &= a\left\{f(x-1,t)-f(x,t)\right\}.\end{aligned} \tag{25}$$

Another way of demonstrating the convergence of the binomial to the Poisson makes use of the cumulant generating functions which are just the logarithms of the moment generating functions. If $x\sim b(p;n)$, then its cumulant generating function can be written as

$$\begin{aligned}\kappa(x;t) &= n\log\big\{1-p(1-e^t)\big\}\\ &= n\left\{-p(1-e^t)-\frac{p^2(1-e^t)^2}{2}-\frac{p^3(1-e^t)^3}{3}\cdots\right\}.\end{aligned} \tag{26}$$

Here, the second equality follows from the Taylor series

$$\log(1+x)=\left\{x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots\right\}. \tag{27}$$

Setting $p=\mu/n$, where μ is fixed, gives

$$\begin{aligned}\kappa(x,t) &= n\left\{-\frac{\mu}{n}\left(1-e^t\right)-\frac{\mu^2}{2n^2}\left(1-e^t\right)^2-\cdots\right\}\\ &= -\mu\left(1-e^t\right)-\frac{\mu^2}{2n}\left(1-e^t\right)^2-\cdots.\end{aligned} \tag{28}$$

As $n \to \infty$, there is $p \to 0$, because μ is fixed, and the RHS tends to

$$\begin{aligned}-\mu\left(1-e^t\right) &= \mu\left(e^t-1\right)\\ &= \mu\left(t+\frac{t^2}{2!}+\frac{t^3}{3!}+\cdots\right),\end{aligned} \tag{29}$$

and each cumulant $\kappa_1, \kappa_2, \ldots = \mu$ has the value of the mean. Thus

$$\lim(n\to\infty)\kappa(x;t) = \mu\left(e^t-1\right). \tag{30}$$

Now consider $x \sim P(\mu)$. Then

$$\begin{aligned}\kappa(x,t) &= \log E\left(e^{tx}\right)\\ &= \log\left\{\sum_{x=0}^{\infty} e^{tx}\frac{e^{-\mu}\mu^x}{x!}\right\}\\ &= \log\left\{e^{-\mu}\sum\frac{e^{tx}\mu^x}{x!}\right\}.\end{aligned} \tag{31}$$

But, if $\lambda = e^t\mu$, then

$$\begin{aligned}\sum\frac{e^{tx}\mu^x}{x!} &= \sum\frac{\lambda^x}{x!} = 1+\lambda+\frac{\lambda^2}{2!}+\frac{\lambda^3}{3!}+\cdots\\ &= e^{\lambda},\end{aligned} \tag{32}$$

and therefore

$$\begin{aligned}\kappa(x;t) &= \log(e^{-\mu+\lambda})\\ &= \mu(e^t-1).\end{aligned} \tag{33}$$

Since (30) and (33) are identical, we can deduce the following:

(34) If $x \sim b(p,n)$, then $\mu = E(x) = np$ and, for fixed μ, there is $\lim(n\to\infty)\kappa(x;t) = \mu(e^t-1)$ which is the cumulant generating function of $x \sim P(\mu)$. Therefore, x is asymptotically distributed as the Poisson $P(\mu)$ for large n.

Finally, we have the following

(35) If $x \sim P(\mu_1)$ is independent of $x_2 \sim P(\mu_2)$, then $y = (x_1+x_2) \sim P(\mu_1+\mu_2)$.

Proof. For $x_1 \sim P(\mu_1)$, there is $\kappa(x_1;t) = \mu_1(e^t-1)$. For $x_2 \sim P(\mu_2)$, there is $\kappa(x_2;t) = \mu_2(e^t-1)$. Now, if x_1 and x_2 are independent, then

$$\begin{aligned}\kappa(x_1+x_2;t) &= \kappa(x_1;t)+\kappa(x_2;t)\\ &= (\mu_1+\mu_2)(e^t-1).\end{aligned} \tag{36}$$

This is the cumulant generating function of the $P(\mu_1+\mu_2)$ distribution.

Corollary. If $x_i \sim P(\mu_i); i = 1, \ldots, n$ is a sequence of n mutually independent Poisson variates, then $\sum x_i \sim P(\sum \mu_i)$. Moreover, and if $\mu_i = \mu$ for all i, then $n^{-1}\sum x_i = \bar{x} \sim P(\mu)$.

The Multinomial Distribution

In the binomial distribution, there are two mutually exclusive outcomes, $\mathcal{A}$ and $\mathcal{A}^c$. In the multinomial distribution, there are k mutually exclusive outcomes $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$, one of which must occur in any trial. The probabilities of these outcomes are denoted by $p_1, p_2, \ldots, p_k$.

In a sequence of n trials, there will be x_1 instances of $\mathcal{A}_1$, x_2 of $\mathcal{A}_2$ and so on, including x_k instances of $\mathcal{A}_k$; and the sum of these instances is $x_1+x_2+\cdots+x_k = n$. Since the outcomes of the trails are statistically independent, the probability that a particular sequence will arise which has these numbers of instances is given by the product $p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k}$. However, the number of such sequences is $n!/\{x_1!x_2!\cdots x_k!\}$; and, together, they represent the set of mutually exclusive ways in which the numbers can occur. It follows that the *multinomial* probability function is given by

$$M(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k} \tag{37}$$

The following results concerning the moments may be derived in a straightforward manner:

$$E(x_i) = \mu_i = np_i, \tag{38}$$

$$V(x_i) = \sigma_{ii} = np_i(1 - p_i), \tag{39}$$

$$C(x_i, x_j) = \sigma_{ij} = -np_ip_j, \quad i \neq j. \tag{40}$$

Example. The multinomial distribution provided the basis of one of the earliest statistical tests, which is Pearson's goodness-of-fit test. If the number n of the trails is large enough, then the distribution of the random variable

$$z^2 = \sum_{i=1}^{k} \frac{(x_i - np_i)^2}{np_i} \tag{41}$$

will be well approximated by a chi-square distribution of $k-1$ degrees of freedom. This statistic is an evident generalisation of the binomial statistic of (5).

If $p_1, \ldots, p_k$ are given a set of hypothesised probability values, then the validity of the hypothesis can be gauged via the resulting value of z^2. If z^2 exceeds a critical value, the hypothesis is liable to be rejected. By this mean, we can determine, for example, whether a sample comes from a multinomial distribution of known parameters.

We are also able to test whether two multinomial distributions are the same without any prior knowledge of their parameters

To understand how such comparisons may be conducted, let us begin by considering two independent multinomial distributions, the first of which has the parameters $n_1, p_{11}, p_{21}, \ldots, p_{k1}$, and the second of which has the parameters $n_2, p_{12}, p_{22}, \ldots, p_{k2}$. The observed frequencies may be denoted likewise by $x_{11}, x_{21}, \ldots, x_{k1}$ and $x_{12}, x_{22}, \ldots, x_{k2}$. For each distribution, we may consider forming a statistic in the manner of equation (41); and these statistics would be treated as if they were $\chi^2(k-1)$ variates. In view of their mutual independence, their sum, which is denoted by

$$\sum_{j=1}^{2}\sum_{i=1}^{k}\frac{(x_{ij}-n_jp_{ij})^2}{n_jp_{ij}}, \tag{42}$$

would be treated as a chi-square $\chi^2(2k-2)$ variate of $2k-2$ degrees of freedom—which is the sum of the degrees of freedom of its constituent parts. It should be recognised, however, that these variables embody the probability parameters p_{ij} about which we have no direct knowledge.

Now, consider the hypothesis that $p_{i1} = p_{i2} = p_i$ for all categories $A_i; i = 1, \ldots, k$, which is the hypothesis that the two multinomial distributions are the same. This hypothesis allows estimates of the probabilities to be obtained according to the formula

$$\hat{p}_i = \frac{x_{i1}+x_{i2}}{n_1+n_2}. \tag{43}$$

Given the constraint that $\hat{p}_1 + \cdots + \hat{p}_k = 1$, it follows that there are only $k-1$ independent parameters to be estimated, which consume altogether $k-1$ degrees of freedom. When the estimates are put in place of the unknown parameters, a statistic is derived in the form of

$$\sum_{j=1}^{2}\sum_{i=1}^{k}\frac{\Big[x_{ij}-n_j\big\{(x_{i1}+x_{i2})/(n_1+n_2)\big\}\Big]^2}{n_j\big\{(x_{i1}+x_{i2})/(n_1+n_2)\big\}} \tag{44}$$

which has an approximating $\chi^2(k-1)$ distribution. The statistic serves to test the hypothesis that the two multinomial distributions are the same, and the hypothesis will be rejected if the value of the statistic exceeds a critical number.

The Univariate Normal Distribution

The normal or Gaussian distribution is undoubtedly the most important of the continuous distributions. The *normal* probability density function is given by

$$N(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2}, \tag{45}$$

where μ and σ^2 are respectively the mean and the variance of the distribution, as will be demonstrated later. The standard normal distribution, which is denoted by $N(x;0,1)$, has $\mu = 0$ and $\sigma^2 = 1$.

We wish to show that the integral of the standard normal density function is unity over the real line. In fact, there is no closed-form analytic expression for this integral, and so we must work, instead, with the square of the integral. Consider, therefore, the product

$$(46)\qquad \begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}dx \times \int_{-\infty}^{\infty} e^{-y^2/2}dy \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy. \end{aligned}$$

The variables in this double integral can be changed to polar coordinates by the substitution of $x = \rho\sin\theta$ and $y = \rho\cos\theta$, which allows it to be written as

$$(47)\qquad \begin{aligned} I^2 &= \int_0^{\infty}\int_0^{2\pi} \rho e^{-\rho^2/2}d\theta d\rho = 2\pi\int_0^{\infty} \rho e^{-\rho^2/2}d\rho \\ &= 2\pi\int_0^{\infty} e^{-\omega}d\omega = 2\pi. \end{aligned}$$

The change-of-variable technique entails the Jacobian factor, which is the determinant of the matrix of the transformation from (θ, ρ) to (x, y). This takes the value of ρ. The final integral is obtained via a change of variables that sets $\rho^2/2 = \omega$; and it takes the value of unity—see equation (64). Since $I = (2\pi)^{1/2}$ is the integral of the function $\exp\{x^2/2\}$, the upshot is that the standrds normal density function $N(x; 0, 1) = (2\pi)^{-1/2}\exp\{x^2/2\}$ integrates to unity over the range of x.

The moment generating function of the standard normal $N(x; 0, 1)$ distribution is defined by

$$(48)\qquad \begin{aligned} M(x, t) = E(e^{xt}) &= \int_{-\infty}^{\infty} e^{xt}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-(x^2-2xt)/2}dx. \end{aligned}$$

By completing the square, it is found that

$$(49)\qquad \begin{aligned} x^2 - 2xt &= x^2 - 2xt + t^2 - t^2 \\ &= (x - t)^2 - t^2. \end{aligned}$$

Therefore, the moment generating function of the standard normal distribution is given by

$$(50)\qquad \begin{aligned} M(x, t) &= e^{t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-(x-t)^2/2}dx \\ &= e^{t^2/2}. \end{aligned}$$

If $y \sim N(\mu, \sigma^2)$ is a normal variate with parameters μ and σ^2, then it can be expressed as $y = \sigma x + \mu$, where $x \sim N(0, 1)$ is a standard normal variate. Then the moment generating function is

$$(51)\qquad M(y, t) = e^{\mu t}M(x, \sigma t) = e^{\mu t + (\sigma^2 t^2)/2}.$$

By differentiating the function with respect t and then setting $t = 0$, we can find the following moments:

$$\begin{aligned} E(y) &= \mu, \\ E(y^2) &= \sigma^2 + \mu^2, \\ V(y) &= E(x^2) - \{E(x)\}^2 = \sigma^2. \end{aligned} \tag{52}$$

Example. The normal distribution with $\mu = np$ and $\sigma^2 = npq$, where $q = 1 - p$, provides an approximation for the binomial $b(x; n, p)$ distribution when n is large. This result can be demonstrated using the moment generating functions. Let

$$z = \frac{x - \mu}{\sigma} = \frac{x - np}{\sqrt{npq}}. \tag{53}$$

The moment generating function for z is

$$\begin{aligned} M(z, t) &= e^{-\mu t/\sigma} M(x, t/\sigma) \\ &= e^{-\mu t/\sigma} (q + pe^{t/\sigma})^n. \end{aligned} \tag{54}$$

Taking logs and expanding the exponential term $pe^{t/\sigma}$ gives

$$\log M(z, t) = -\frac{\mu t}{\sigma} + n \log \left[1 + p \left\{ \left(\frac{t}{\sigma}\right) + \frac{1}{2!}\left(\frac{t}{\sigma}\right)^2 + \frac{1}{3!}\left(\frac{t}{\sigma}\right)^3 \cdots \right\}^2 \right], \tag{55}$$

where we have used $p + q = 1$. The logarithm on the RHS is in the form of $\log(1 + z)$, where z stands for the sum within the braces $\{,\}$ times p. This is amenable to the Maclaurin expansion of (27) on the condition that $|z| < 1$. Since $\sigma = \sqrt{npq}$ increases indefinitely with n, the condition is indeed fulfilled when n is sufficiently large. The Maclaurin expansion gives rise to the following expression:

$$\begin{aligned} \log M(z, t) = -\frac{\mu t}{\sigma} + n \Bigg[& p \left\{ \left(\frac{t}{\sigma}\right) + \frac{1}{2!}\left(\frac{t}{\sigma}\right)^2 + \cdots \right\} \\ & - \frac{p^2}{2} \left\{ \left(\frac{t}{\sigma}\right)^2 + \frac{1}{2!}\left(\frac{t}{\sigma}\right) + \cdots \right\}^2 + \cdots \Bigg]. \end{aligned} \tag{56}$$

Collecting terms in powers of t gives

$$\begin{aligned} \log M(z, t) &= \left(-\frac{\mu}{\sigma} + \frac{np}{\sigma} \right) t + n \left(-\frac{p}{\sigma} + \frac{p^2}{\sigma^2} \right) \frac{t^2}{2!} + \cdots \\ &= \frac{t^2}{2} + \cdots. \end{aligned} \tag{57}$$

Here, the second equality follows because the coefficient of t is zero and that of $t^2/2!$ is unity. Moreover, the coefficients associated with $\{t^3, t^4, \ldots\}$ all tend to zero as n increases. Thus $\lim(n \to \infty) \log M(z, t) = t^2/2$; from which it follows that

$$\lim(n \to \infty) M(z, t) = e^{t^2/2} \tag{58}$$

This is the moment generating function of the standard normal distribution, given already by (50); and thus the convergence of z in distribution to a standard normal is demonstrated.

The Gamma Distribution

Consider the function

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx. \tag{59}$$

This can be evaluated using integration by parts. The formula is

$$\int u \frac{dv}{dx} dx = uv - \int v \frac{du}{dx} dx, \tag{60}$$

which can be seen as a consequence of the product rule of differentiation. Let $u = x^{n-1}$ and $dv/dx = e^{-x}$. Then $v = -e^{-x}$ and

$$\begin{aligned} \int_0^\infty e^{-x} x^{n-1} dx &= \left[-x^{n-1} e^{-x} + \int e^{-x} (n-1) x^{n-2} dx \right]_0^\infty \\ &= (n-1) \int_0^\infty e^{-x} x^{n-2} dx. \end{aligned} \tag{61}$$

We may express the result above by writing $\Gamma(n) = (n-1)\Gamma(n-1)$, from which it follows, by iteration, that

$$\Gamma(n) = (n-1)(n-2)\cdots\Gamma(\delta), \tag{62}$$

where $0 < \delta < 1$. Examples of the gamma function are

$$\Gamma(1/2) = \sqrt{\pi}, \tag{63}$$

$$\Gamma(1) = \int_0^\infty e^{-x} dx = \Big[-e^{-x} \Big]_0^\infty = 1, \tag{64}$$

$$\Gamma(n) = (n-1)(n-2)\cdots\Gamma(1) = (n-1)!. \tag{65}$$

Here, in (65), it is assumed that n is an integer. The first of these results can be verified by confirming the following identities:

$$\begin{aligned} \sqrt{2\pi} &= \int_{-\infty}^\infty e^{-z^2/2} dz \\ &= \frac{\sqrt{2}}{2} \int_{-\infty}^\infty e^{-x} x^{-1/2} dx = \sqrt{2}\Gamma(1/2). \end{aligned} \tag{66}$$

The first equality is familiar from the integration of the standard normal density function. The second equality follows when the variable of integration is changed from z to $x = z^2/2$, and the final equality invokes the definition of the gamma function, which is provided by equation (59).

Using the gamma function, we may define a probability density function known as the *gamma type 1:*

$$\gamma_1(x; n) = \frac{e^{-x} x^{n-1}}{\Gamma(n)}; \qquad 0 < x < \infty. \tag{67}$$

For an integer value of n, the gamma type 1 gives the probability distribution of the waiting time to the nth event in a Poisson arrival process of unit mean. When $n = 1$, it becomes the exponential distribution, which relates to the waiting time for the first event.

To define the type 2 gamma function, we consider the transformation $z = \beta x$. Then, by the change-of-variable technique, we have

$$\begin{aligned} \gamma_2(z) &= \gamma_1\{x(z)\}\left|\frac{dx}{dz}\right| \\ &= \frac{e^{-z/\beta}(z/\beta)^{\alpha-1}}{\Gamma(\alpha)}\frac{1}{\beta}. \end{aligned} \tag{68}$$

Here we have changed the notation by setting $\alpha = n$. The probability function of the *type 2 gamma* distribution is written more conveniently as

$$\gamma_2(z;\alpha,\beta) = \frac{e^{-z/\beta}z^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}}. \tag{69}$$

An important special case of the γ_2 distribution is when $\alpha = r/2$ with $r \in \{0, 1, 2, \ldots\}$ and $\beta = 2$. This is the so-called *chi-square* distribution of r degrees of freedom:

$$\chi^2(x;r) = \frac{e^{-x/2}x^{(r/2)-1}}{\Gamma(r/2)2^{r/2}}. \tag{70}$$

Now let us endeavour to find the moment generating function of the γ_1 distribution. We have

$$\begin{aligned} M_x(t) &= \int e^{xt}\frac{e^{-x}x^{n-1}}{\Gamma(n)}dx \\ &= \int \frac{e^{-x(1-t)}x^{n-1}}{\Gamma(n)}dx. \end{aligned} \tag{71}$$

Now let $w = x(1-t)$. Then, by the change-of-variable technique,

$$\begin{aligned} M_x(t) &= \int \frac{e^{-w}w^{n-1}}{(1-t)^{n-1}\Gamma(n)}\frac{1}{(1-t)}dw \\ &= \frac{1}{(1-t)^n}\int \frac{e^{-w}w^{n-1}}{\Gamma(n)}dw \\ &= \frac{1}{(1-t)^n}. \end{aligned} \tag{72}$$

Also, the cumulant generating function is

$$\begin{aligned} \kappa(x;t) &= -n\log(1-t) \\ &= n\left(t + \frac{t^2}{2} + \frac{t^3}{3} + \cdots\right). \end{aligned} \tag{73}$$

We find, in particular, that

$$E(x) = V(x) = n. \tag{74}$$

We have defined the γ_2 distribution by

$$\gamma_2 = \frac{e^{-x/\beta}x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}; \qquad 0 \leq x < \infty. \tag{75}$$

Hence the moment generating function is defined by

$$\begin{aligned} M_x(t) &= \int_0^\infty \frac{e^{tx}e^{-x/\beta}x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}dx \\ &= \int_0^\infty \frac{e^{-x(1-\beta t)/\beta}x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}dx. \end{aligned} \tag{76}$$

Let $y = x(1-\beta t)/\beta$, which gives $dy/dx = (1-\beta t)/\beta$. Then, by the change-of-variable technique we get

$$\begin{aligned} M_x(t) &= \int_0^\infty \frac{e^{-y}}{\Gamma(\alpha)\beta^\alpha}\left(\frac{\beta y}{1-\beta t}\right)^{\alpha-1}\frac{\beta}{(1-\beta t)}dy \\ &= \frac{1}{(1-\beta t)^\alpha}\int \frac{y^{\alpha-1}e^{-y}}{\Gamma(\alpha)}dy \\ &= \frac{1}{(1-\beta t)^\alpha}. \end{aligned} \tag{77}$$

It follows that the cumulant generating function is

$$\begin{aligned} \kappa(x;t) = -\alpha\log(1-\beta t) &= -\alpha\left(-\beta t - \frac{\beta^2t^2}{2} - \frac{\beta^3t^3}{3} - \cdots\right) \\ &= \alpha\left(\beta t + \frac{\beta^2t^2}{2} + \frac{\beta^3t^3}{3} + \cdots\right). \end{aligned} \tag{78}$$

We find, in particular, that

$$\begin{aligned} E(x) &= \alpha\beta \\ V(x) &= \alpha\beta^2 \end{aligned} \tag{79}$$

Now consider two independent gamma variates of type 2: $x_1 \sim \gamma_2(\alpha_1, \beta_1)$ and $x_2 \sim \gamma_2(\alpha_2, \beta_2)$. Since x_1 and x_2 are independent, the cumulant generating function of their sum is the sum of their separate generating functions:

$$\begin{aligned} &\kappa(x_1 + x_2; t) = \kappa(x_1; t) + \kappa(x_2; t) \\ &\quad = \alpha_1\left(\beta_1 t + \frac{\beta_1^2t^2}{2} + \frac{\beta_1^3t^3}{3} + \cdots\right) + \alpha_2\left(\beta_2 t + \frac{\beta_2^2t^2}{2} + \frac{\beta_2^3t^3}{3} + \cdots\right). \end{aligned} \tag{80}$$

If $\beta_1 = \beta_2 = \beta$, then

$$\kappa(x_1 + x_2; t) = (\alpha_1 + \alpha_2)\left(\beta t + \frac{\beta^2 t^2}{2} + \frac{\beta^3 t^3}{3} + \cdots\right); \tag{81}$$

and so $y \sim \gamma_2(\alpha_1 + \alpha_2, \beta)$. In particular, if $\beta = 1$, then $y \sim \gamma_1(\alpha_1 + \alpha_2)$. This result may be generalised for $n > 2$ independent variables:

(82) If $x_1, x_2, \ldots, x_n$ is a set of independent $\gamma_2(\alpha_i, \beta)$ variates, then $\sum x_i \sim \gamma_2(\sum \alpha_i, \beta)$ and, in particular, if $\beta = 1$ and $\alpha_i = \alpha$ for all $i = 1, \ldots, n$, then $\sum x_i \sim \gamma_2(n\alpha, 1)$, or, equally, $\sum x_i \sim \gamma_1(n\alpha)$.

The Beta Distribution

Consider the integral

$$B(m, n) = \int_0^1 x^{m-1}(1 - x)^{n-1} dx. \tag{83}$$

This converges for $m, n > 0$. An alternative form of the $B(m, n)$ function is obtained from the transformation $x = 1/(1 + y)$, of which the derivative is $dx/dy = -1/(1 + y)^2$, and the inverse is $y = (1 - x)/x$. As $x \to 0$, it is found that $y \to \infty$ and, when $x = 1$, there is $y = 0$. Hence (83) becomes

$$\begin{aligned} B(m, n) &= \int_\infty^0 \frac{1}{(1 + y)^{m-1}} \frac{y^{n-1}}{(1 + y)^{n-1}} \frac{-1}{(1 + y)^2} dy \\ &= \int_\infty^0 \frac{y^{n-1}}{(1 + y)^{n+m}} dy. \end{aligned} \tag{84}$$

In (83), the argument obeys the inequality $0 < x < 1$, whereas, in (84), the argument obeys the inequality $0 < y < \infty$.

The *type 1 beta distribution,* which is taken from the integral of (83) is defined by the following density function:

$$\beta_1(x; m, n) = \frac{x^{m-1}(1 - x)^{n-1}}{B(m, n)}; \qquad 0 < x < 1, \quad m, n > 0. \tag{85}$$

The *type 2 beta distribution,* which is taken from the integral (84), is defined by the density function

$$\beta_2(y; m, n) = \frac{y^{n-1}}{(1 + y)^{m+n} B(m, n)}; \qquad y > 0, \quad m, n > 0. \tag{86}$$

In fact, the beta integral of (84) and (85) is related to the the gamma integral of (59):

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(p + q)}. \tag{87}$$

This formula will be used later.

Next, we shall prove an important theorem.

(88) If $x \sim \gamma_2(\alpha, \lambda)$ and $y \sim \gamma_2(\theta, \lambda)$ are independent random variables which have the gamma type 2 distribution, then $x/y = z \sim \beta_2(\alpha, \theta)$ is distributed as a beta type 2 variable.

Proof. When $x \sim \gamma_2(\alpha, \lambda)$ and $y \sim \gamma_2(\theta, \lambda)$ are independently distributed, there is

(89)
$$f(x, y) = e^{-(x+y)/\lambda} \frac{x^{\alpha-1} y^{\theta-1}}{\lambda^{\alpha+\theta}\Gamma(\alpha)\Gamma(\theta)}.$$

Let $v = x/y$ and $w = x + y$, whence

(90)
$$y = \frac{w}{v+1}, \quad x = \frac{vw}{v+1},$$

and $w, v > 0$. Also, let

(91)
$$J = \begin{bmatrix} \dfrac{\partial x}{\partial w} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial w} & \dfrac{\partial x}{\partial v} \end{bmatrix} = \begin{bmatrix} \dfrac{v}{v+1} & \dfrac{w}{(v+1)^2} \\ \dfrac{1}{v+1} & \dfrac{-w}{(v+1)^2} \end{bmatrix}$$

be the matrix of partial derivatives of the mapping from (v, w) to (x, y). The Jacobian of this transformation, which is absolute value of the determinant of the matrix, is

(92)
$$\|J\| = \frac{w}{(1+v)^2}.$$

It follows that the joint distribution of (v, w) is

(93)
$$\begin{aligned} g(w, v) &= \exp\left\{\frac{-w}{\lambda}\right\} \left\{\frac{vw}{v+1}\right\}^{\alpha-1} \left\{\frac{w}{v+1}\right\}^{\theta-1} \frac{w}{(v+1)^2} \frac{1}{\lambda^{\alpha+\theta}\Gamma(\alpha)\Gamma(\theta)} \\ &= \frac{\exp(-w/\lambda) w^{\alpha+\theta-1}}{\lambda^{\alpha+\theta}\Gamma(\alpha+\theta)} \times \frac{v^{\alpha-1}}{(1+v)^{\alpha+\theta} B(\alpha, \theta)} \\ &= \gamma_2(w) \times \beta_2(v). \end{aligned}$$

Here $w \sim \gamma_2(\alpha + \theta, \lambda)$ and $v \sim \beta_2(\beta, \theta)$ are independent random variables; and, moreover, v has the required beta type 2 distribution.

The Multivariate Normal Distribution

Let $\{x_1, x_2, \ldots, x_n\}$ be a sequence of independent random variables each distributed as $N(0, 1)$. Then the joint density function $f(x_1, x_2, \ldots, x_n)$ is given by the product of the individual density functions:

(94)
$$f(x_1, x_2, \ldots, x_n) = (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2}(x_1^2 + x_2^2 + \cdots + x_n^2)\right\}.$$

The sequence of n independent $N(0,1)$ variates constitutes a vector x of order n, and the sum of squares of the elements is the quadratic $x'x$. The zero valued expectations of the elements of the sequence can be gathered in a zero vector $E(x) = 0$ of order n, and their unit variances can be represented by the diagonal elements of an identity matrix of order n which constitutes the variance–covariance or dispersion matrix of x:

$$\begin{aligned} D(x) &= E[\{x - E(x)\}\{x - E(x)\}'] \\ &= E(xx') = I. \end{aligned} \tag{95}$$

In this notation, the probability density function of the vector x is the *n-variate standard normal* function

$$N(x; 0, I) = (2\pi)^{-n/2} \exp\left\{-\tfrac{1}{2}x'x\right\}. \tag{96}$$

Next, we consider a more general normal density function which arises from a linear transformation of x followed by a translation of the resulting vector. The combined transformations give rise to the vector $y = Ax + \mu$. It is reasonable to require that A is a matrix of full row rank, which implies that the dimension of y can be no greater than the dimension of x and that none of the variables within Ax is expressible as a linear combination of the others. In that case, there is

$$\begin{aligned} E(y) &= AE(x) + \mu = \mu \quad \text{and} \\ D(y) &= E[\{y - E(y)\}\{y - E(y)\}'] \\ &= E[\{Ax - E(Ax)\}\{Ax - E(Ax)\}'] \\ &= AD(x)A' = AA' = \Sigma. \end{aligned} \tag{97}$$

The density function of y is found via the change-of-variable technique. This involves expressing x in terms of the inverse function $x(y) = A^{-1}(y - \mu)$ of which the Jacobian is

$$\left\| \frac{\partial x}{\partial y} \right\| = \|A^{-1}\| = |\Sigma|^{-1/2}. \tag{98}$$

The resulting probability density function of the *multivariate normal* distribution is

$$N(y; \mu, \Sigma) = (2\pi)^{-n/2}|\Sigma|^{-1/2} \exp\{-\tfrac{1}{2}(y - \mu)'\Sigma^{-1}(y - \mu)\}. \tag{99}$$

We shall now consider the relationships which may subsist between groups of elements within the vector x. Let the vectors x and $E(x) = \mu$ and the dispersion matrix Σ be partitioned conformably to yield

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{100}$$

where x_1 is of order p, x_2 is of order q and $p + q = n$.

Consider, first, the case where $\Sigma_{ij} = 0$ if $i \neq j$. Then

$$\Sigma = \begin{bmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{bmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} & 0 \\ 0 & \Sigma_{22}^{-1} \end{bmatrix}. \tag{101}$$

It follows that, if $x \sim N(\mu, \Sigma)$, then

$$(x - \mu)'\Sigma^{-1}(x - \mu) = (x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1) + (x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2) \tag{102}$$

and

$$|\Sigma| = |\Sigma_{11}| \times |\Sigma_{22}|. \tag{103}$$

Therefore, the density function of x becomes

$$\begin{aligned} N(x; \mu, \Sigma) &= (2\pi)^{n/2}|\Sigma|^{-1/2} \exp\{-\frac{1}{2}(x - \mu)'\Sigma^{-1}(x - \mu)\}dx \\ &= (2\pi)^{p/2}|\Sigma_{11}|^{-1/2} \exp\{-\frac{1}{2}(x_1 - \mu_1)'\Sigma_{11}^{-1}(x_1 - \mu_1)\} \\ &\quad \times (2\pi)^{q/2}|\Sigma_{22}|^{-1/2} \exp\{-\frac{1}{2}(x_2 - \mu_2)'\Sigma_{22}^{-1}(x_2 - \mu_2)\}. \end{aligned} \tag{104}$$

which can be written in summary notation as

$$N(x; \mu, \Sigma) = N_1(x_1; \mu_1, \Sigma_{11})N_2(x_2; \mu_2, \Sigma_{22}). \tag{105}$$

Thus the marginal distribution of x_1 is independent of the marginal distribution of x_2.

A summary of these findings is as follows:

(106) If $x \sim N(\mu, \Sigma)$ can be partitioned as $x = [x_1', x_2']'$ and if $\Sigma_{12} = 0$, which is to say that x_1 and x_2 are uncorrelated, then the density function of x is the product of the marginal density functions $N(x_1, \mu_1, \Sigma_{11})$ and $N(x_2; \mu_2, \Sigma_{22})$; which is to say that x_1 and x_2 are statistically independent.

Next, we shall show how $x \sim N(\mu, \Sigma)$ can be expressed as the product of a marginal distribution and a conditional distribution. This is a matter of transforming x into two uncorrelated parts. The relevant transformation, which affects only the leading subvector of $x = [x_1', x_2']'$, takes the form of

$$\begin{aligned} y_1 &= x_1 - B'x_2, \\ y_2 &= x_2, \end{aligned} \tag{107}$$

where B' is a matrix of order $p \times q$.

The object is to discover a value of B' for which the following condition of non-correlation will be satisfied:

$$E[\{y_1 - E(y_1)\}\{y_2 - E(y_2)\}'] = 0. \tag{108}$$

When it is written in terms of x_1 and x_2, the condition becomes

$$\begin{aligned} 0 &= E[\{[x_1 - E(x_1)] - B'[x_2 - E(x_2)]\}\{x_2 - E(x_2)\}'] \\ &= \Sigma_{12} + B'\Sigma_{22}. \end{aligned} \tag{109}$$

The solution is $B' = \Sigma_{12}\Sigma_{22}^{-1}$; and thus the transformation is given by

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} I_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_q \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = Qx. \tag{110}$$

Now, if $x \sim N(\mu, \Sigma)$, then it follows that $y \sim N(Q\mu, Q\Sigma Q')$, where

$$\begin{aligned} Q\Sigma Q' &= \begin{bmatrix} I_p & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I_p \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} I_p & 0 \\ -\Sigma_{22}^{-1}\Sigma_{21} & I_p \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & 0 \\ 0 & \Sigma_{22} \end{bmatrix}. \end{aligned} \tag{111}$$

The condition of non-correlation implies that y_1 and y_2 are statistically independent. Therefore, the joint distribution of y_1 and y_2 is the product of their marginal distributions: $N(y) = N(y_1) \times N(y_2)$.

Next, we use the change-of-variable technique to recover the distribution of x from that of y. We note that the Jacobian of the transformation from x to y is unity (since its matrix is triangular with units on the principal diagonal). Thus, by using the inverse transformation $x = x(y)$, we can write the distribution of x as

(112)

$$\begin{aligned} N(x; \mu, \Sigma) &= N\{y(x); E(y), Q\Sigma Q'\} \\ &= N\{y_1(x); E(y_1), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\} \times N\{y_2(x); E(y_2), \Sigma_{22}\}, \end{aligned}$$

wherein, there are

$$\begin{aligned} y_1(x) &= x_1 - B'x_2, \\ E(y_1) &= \mu_1 - B'\mu_2, \\ y_2(x) &= x_2 \quad \text{and} \\ E(y_2) &= \mu_2. \end{aligned} \tag{113}$$

The second of the factors on the RHS of (112) is the marginal distribution of $x_2 = y_2(x)$. Since the product of the two factors is the joint distribution of x_1 and x_2, the first of the factors on the RHS must be the conditional distribution of x_1 given x_2.

A summary of these results is as follows:

(114) If $x \sim N(\mu, \Sigma)$, is partitioned as $x = [x_1', x_2']'$ with $\mu = [\mu_1', \mu_2']'$ partitioned conformably, then the marginal distribution of x_2 is $N(x_2; \mu_2, \Sigma_{22})$ and the conditional distribution of x_1 given x_2, is $N(y_1; E\{y_1\}, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$, where $y_1 = x_1 - B'x_2$ with $B' = \Sigma_{12}\Sigma_{22}^{-1}$, and where $E\{y_1\} = \mu_1 - B'\mu_2$.

Within the conditional distribution, there is the quadratic form

$$(115) \qquad [y_1 - E(y_1)]'[\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}]^{-1}[y_1 - E(y_1)].$$

This contains the term

$$(116) \qquad \begin{aligned} \varepsilon &= y_1 - E(y_1) \\ &= x_1 - \mu_1 - B'(x_2 - \mu_2) \\ &= x_1 - E(x_1|x_2). \end{aligned}$$

The conditional expectation, which is

$$(117) \qquad E(x_1|x_2) = E(x_1) + B'(x_2 - \mu_2),$$

is commonly described as the equation of the regression of x_1 on x_2, whilst B' is the matrix of the regression parameters. The quantity denoted by ε is described as the vector of prediction errors or regression disturbances.

The Chi-square Distribution

Chi-square distribution is a special case of a type 2 gamma distribution. The type 2 gamma has been denoted by $\gamma_2(\alpha, \beta)$, and its functional form is given by equation (69). When $\alpha = r/2$ and $\beta = 2$, the γ_2 density function becomes the probability density function of a *chi-square* distribution of r *degrees of freedom*:

$$(118) \qquad \chi^2(x; r) = \frac{e^{-x/2}x^{(r/2)-1}}{\Gamma(r/2)2^{r/2}}.$$

The importance of the chi-square is in its relationship with the normal distribution: the chi-square of one degree of freedom represents the distribution of the quadratic exponent $(x-\mu)^2/\sigma^2$ of the univariate normal $N(x; \mu, \sigma^2)$ function.

To demonstrate the relationship, let us consider the integral of a univariate standard normal $N(z; 0, 1)$ function, over the interval $[-\theta, \theta]$ together with the integral of the density function of $v = z^2$ over the interval $[0, \theta^2]$. We can use the change-of-variable technique to find the density function of v. The following relationship must hold:

$$(119) \qquad \int_{-\theta}^{\theta} N(z)dx = 2\int_0^{\theta^2} N\{z(v)\}\left|\frac{dz}{dv}\right| dv.$$

To be more explicit, we can use $z^2 = v$ and $dz/dv = v^{-1/2}/2$ in writing the following version of the equation:

$$(120) \qquad \begin{aligned} \int_{-\theta}^{\theta} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dy &= \int_0^{\theta^2} \frac{1}{\sqrt{2\pi}} e^{-v/2} v^{-1/2} dv \\ &= \int_0^{\theta^2} \frac{e^{-v/2}v^{-1/2}}{\Gamma(1/2)\sqrt{2}} dv. \end{aligned}$$

The derivative of the integral on the RHS is the density function of a $\gamma_2(v; 1/2, 2)$ variate evaluated at $v = \theta^2$. Thus, if $z \sim N(0,1)$, then $z^2 = v \sim \gamma_2(1/2, 2) = \chi^2(1)$.

The cumulant generating function of the $\chi^2(n)$ distribution is the cumulant generating function of the $\gamma_2(n/2, 2)$, which is obtained from equation (78). Thus,

$$\kappa(x;t) = \frac{n}{2}(2t + 4t^2/2 + 8t^3/3 + \cdots). \tag{121}$$

For this, it can be deduced that

$$\begin{aligned} E(x) &= n, \\ V(x) &= 2n. \end{aligned} \tag{122}$$

The outstanding feature of the chi-square distribution is its reproducing property:

(123) Let $x_1 \sim \chi^2(q)$ and $x_2 \sim \chi^2(r)$ be independent chi-square variates of q and r degrees of freedom respectively. Then $x_1 + x_2 = y \sim \chi^2(q + r)$ is a chi-square variate of $q + r$ degrees of freedom.

Proof. Given that x_1 and x_2 are statistically independent, the moment generating function of their sum is the product of their separate moment generating functions. Thus

$$\begin{aligned} M(y,t) &= E\{e^{(x_1+x_2)t}\} \\ &= E\{e^{x_1 t}\}E\{e^{x_2 t}\}. \end{aligned} \tag{124}$$

We know that a $\chi^2(r)$ variate has a $\gamma_2(r/2, 2)$ distribution, and we know that the moment generating function of the latter is $(1-2t)^{-(r/2)}$. Hence

$$M(y,t) = (1-2t)^{-(q/2)}(1-2t)^{-(r/2)} = (1-2t)^{-(q+r)}. \tag{125}$$

This is the moment generation function of the $\chi^2(q + r)$ distribution. Therefore, $y \sim \chi^2(q + r)$.

The chi-square is the essential part of the distribution of the maximum-likelihood estimator of the variance σ^2 of a normal distribution. We have already shown that, if $x \sim N(\mu, \sigma^2)$, then $\{(x - \mu)/\sigma\}^2 \sim \chi^2(1)$. From this it follows, in view of the preceding result, that

(126) If $\{x_1, x_2, \ldots, x_n\}$ is a random sample with $x_i \sim N(\mu, \sigma^2)$ for all i, then

$$y = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2$$

has a chi-square distribution with n degrees of freedom, which can be expressed by writing $y \sim \chi^2(n)$.

There is a straightforward extension of this result which makes use of matrix notation. Thus, if $y \sim N(\mu, \Sigma)$ has multivariate normal distribution, and if

$C'C = \Sigma^{-1}$, then $x = C(y-\mu) \sim N(0, I_n)$. In that case, $x'x \sim \chi^2(n)$. But $x'x = (y-\mu)'C'C(y-\mu) = (y-\mu)'\Sigma^{-1}(y-\mu)$, so we can make the following statement:

(127) If, $y \sim N(\mu, \Sigma)$ is an normal vector with n elements, then the quadratic product $(y-\mu)'\Sigma^{-1}(y-\mu) \sim \chi^2(n)$ is a chi-square variate on n degrees of freedom.

Example. Consider a multinomial distribution with probabilities $p_1, \ldots, p_k$. The numbers $x_1, \ldots, x_k$ of the sample points that fall into each of the k categories have binomial marginal distributions which tend to normal distributions as the numbers increase. The joint distribution of the numbers will tend to a multivariate normal distribution with a mean vector μ_k and a dispersion matrix Σ_k whose elements are specified under (38)–(40):

$$\mu_k = \begin{bmatrix} np_1 \\ np_2 \\ \vdots \\ np_k \end{bmatrix}, \quad \Sigma_k = n\begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \ldots & -p_1p_k \\ -p_2p_1 & p_2(1-p_2) & \ldots & -p_2p_k \\ \vdots & \vdots & & \vdots \\ -p_kp_1 & -p_kp_2 & \ldots & p_k(1-p_k) \end{bmatrix}. \tag{128}$$

Since $x_1 + x_2 + \cdots + x_k = n$, there are are only $k-1$ degrees of freedom amongst the k variables and, therefore, in describing the joint distribution, we may omit the last of these variables which is the kth. Accordingly, we may omit the last element of μ_k and the last row and column of Σ_k. The resulting matrix and vector may be denoted simply by $\mu = \mu_{k-1}$ and $\Sigma = \Sigma_{k-1}$.

The inverse of Σ is

$$\Sigma^{-1} = \frac{1}{n}\begin{bmatrix} \dfrac{1}{p_1} + \dfrac{1}{p_k} & \dfrac{1}{p_k} & \ldots & \dfrac{1}{p_k} \\ \dfrac{1}{p_k} & \dfrac{1}{p_2} + \dfrac{1}{p_k} & \ldots & \dfrac{1}{p_k} \\ \vdots & \vdots & & \vdots \\ \dfrac{1}{p_k} & \dfrac{1}{p_k} & \ldots & \dfrac{1}{p_{k-1}} + \dfrac{1}{p_k} \end{bmatrix}. \tag{129}$$

This can be verified directly by forming the product of Σ and Σ^{-1} and by noting that the result is an identity matrix. For this, it is necessary to observe that $p_k = 1 - \sum_{i=1}^{k-1} p_i$.

The quadratic function, which is the exponent of the limiting multivariate normal distribution, is as follows:

(130)

$$\begin{aligned}(x-\mu)'\Sigma^{-1}(x-\mu) &= \sum_{i=1}^{k-1}\frac{(x_i - np_i)^2}{np_i} + \frac{1}{np_k}\sum_{i=1}^{k-1}\sum_{j=1}^{k-1}(x_i - np_i)(x_j - np_j) \\ &= \sum_{i=1}^{k-1}\frac{(x_i - np_i)^2}{np_i} + \frac{1}{np_k}\left\{\sum_{i=1}^{k-1}(x_i - np_i)\right\}^2 \\ &= \sum_{i=1}^{k}\frac{(x_i - np_i)^2}{np_i}.\end{aligned}$$

Here, the final equality arises when, in the second term of the penultimate expression, $\sum_{i=1}^{k-1} np_i = n - np_k$ is subtracted from $\sum_{i=1}^{k-1} x_i = n - x_k$. The final expression on the RHS is the conventional form of Pearson's goodness-of-fit statistic which has been presented already under (41). The quadratic is distributed as a $\chi^2(k-1)$ in the limit as $n \to \infty$.

The F Distribution

If $x \sim \chi^2(n)$ and $y \sim \chi^2(n)$ are two independent chi-square variates, then the F statistic, which is also called the variance ratio in the appropriate contexts, is defined as

$$F = \frac{x}{n} \Big/ \frac{y}{m}. \tag{131}$$

We proceed to derive the density function of F.

First, if $x \sim \gamma_2(n/2, 2)$ and $y \sim \gamma_2(m/2, 2)$ are independent, then according to (75), the ratio $z = x/y$ has the following type 2 beta distribution:

$$\beta_2(z; n/1, m/2) = \frac{z^{(n/2)-1}}{(1+z)^{(m+n)/2} B(n/2, m/2)}. \tag{132}$$

Now consider the transformation $v = nF/m$, for which $dv/dF = n/m$.

$$\begin{aligned} h(F) &= \frac{n}{m} \frac{nF/m^{(n/2)-1}}{\{(m+nF)/m\}^{(m+n)/2} B(n/2, m/2)} \\ &= \frac{n^{n/2} m^{m/2} F^{(n/2)-1}}{(m+nF)^{(m+n)/2} B(n/2, m/2)}, \end{aligned} \tag{133}$$

which is the probability density function of the F distribution with n and m degrees of freedom. If R is distributed as F with n and m degrees of freedom, we may write $R \sim F(n, m)$.

"Student's" t Distribution

Let $z \sim N(0, 1)$ be a standard normal variate and let $u \sim \chi^2(n)$ be a chi-square variate of n degrees of freedom, and let the two variables be statistically independent. Then Student's t ratio may be defined as the quotient

$$z = \frac{x}{\sqrt{u/n}}. \tag{134}$$

Since $x^2 \sim \chi^2(1)$ has a chi-square distribution of one degree of freedom, it is clear that $z^2 = v \sim F(1, n)$ has an F distribution of 1 and n degrees of freedom. This relationship allows us to derive the probability density function of Student's t in straightforward manner.

To begin, let us observe that the intervals $[-\tau, 0]$ and $[0, \tau]$, which are on the two branches of the t distribution, both map into the interval $[0, \tau^2]$ in the domain of the F distribution. Thus, integrating the t along the positive branch gives

$$\begin{aligned}\int_0^\tau t(z)dz &= \frac{1}{2}\int_0^{\tau^2} F(v)dv \\ &= \int_0^\tau \frac{1}{2}F\{v(z)\}\left|\frac{dv}{dz}\right| dz.\end{aligned} \tag{135}$$

An expression for the t density function can be obtained by evaluating the expression under integral on the RHS. First there is the F density function:

$$F(v;1,n) = \frac{n^{n/2}v^{-1/2}}{(n+v)^{(n+1)/2}B(1/2,n/2)}. \tag{136}$$

Then there is the function $v(z) = z^2$, and, finally, there is the derivative $dv/dz = 2z$. Putting these together gives the density function of *Student's* t distribution with n degrees of freedom:

$$t(z;n) = \frac{1}{n^{1/2}(1+z^2/n)^{(n+1)/2}B(1/2,n/2)}. \tag{137}$$

An alternative expression is available for the density function which relies upon the fact that

$$B(1/2,m/2) = \frac{\Gamma(1/2)\Gamma(m/2)}{\Gamma\{(m+1)/2\}} = \frac{\sqrt{\pi}\Gamma(m/2)}{\Gamma\{(m+1)/2\}}. \tag{138}$$

The alternative form is

$$t(z;n) = \frac{\Gamma\{(n+1)/2\}}{\Gamma(n/2)\sqrt{n\pi}\{1+z^2/n\}^{(n+1)/2}}. \tag{139}$$

and since n is an integer, the gamma functions can be written, alternatively, as factorials.