Chapter 10

Generating Functions

10.1 Generating Functions for Discrete Distributions

So far we have considered in detail only the two most important attributes of a random variable, namely, the mean and the variance. We have seen how these attributes enter into the fundamental limit theorems of probability, as well as into all sorts of practical calculations. We have seen that the mean and variance of a random variable contain important information about the random variable, or, more precisely, about the distribution function of that variable. Now we shall see that the mean and variance do *not* contain *all* the available information about the density function of a random variable. To begin with, it is easy to give examples of different distribution functions which have the same mean and the same variance. For instance, suppose X and Y are random variables, with distributions

$$p_X = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 1/4 & 1/2 & 0 & 0 & 1/4 \end{pmatrix},$$

$$p_Y = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/4 & 0 & 0 & 1/2 & 1/4 & 0 \end{pmatrix}.$$

Then with these choices, we have $E(X) = E(Y) = 7/2$ and $V(X) = V(Y) = 9/4$, and yet certainly p_X and p_Y are quite different density functions.

This raises a question: If X is a random variable with range $\{x_1, x_2, \ldots\}$ of at most countable size, and distribution function $p = p_X$, and if we know its mean $\mu = E(X)$ and its variance $\sigma^2 = V(X)$, then what else do we need to know to determine p completely?

Moments

A nice answer to this question, at least in the case that X has finite range, can be given in terms of the *moments* of X, which are numbers defined as follows:

$$\begin{aligned}
\mu_k &= k\text{th moment of } X \\
&= E(X^k) \\
&= \sum_{j=1}^{\infty} (x_j)^k p(x_j)\ ,
\end{aligned}$$

provided the sum converges. Here $p(x_j) = P(X = x_j)$.

In terms of these moments, the mean μ and variance σ^2 of X are given simply by

$$\begin{aligned}
\mu &= \mu_1, \\
\sigma^2 &= \mu_2 - \mu_1^2\ ,
\end{aligned}$$

so that a knowledge of the first two moments of X gives us its mean and variance. But a knowledge of *all* the moments of X determines its distribution function p completely.

Moment Generating Functions

To see how this comes about, we introduce a new variable t, and define a function $g(t)$ as follows:

$$\begin{aligned}
g(t) &= E(e^{tX}) \\
&= \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} \\
&= E\left(\sum_{k=0}^{\infty} \frac{X^k t^k}{k!}\right) \\
&= \sum_{j=1}^{\infty} e^{tx_j} p(x_j)\ .
\end{aligned}$$

We call $g(t)$ the *moment generating function* for X, and think of it as a convenient bookkeeping device for describing the moments of X. Indeed, if we differentiate $g(t)$ n times and then set $t = 0$, we get μ_n:

$$\begin{aligned}
\left.\frac{d^n}{dt^n} g(t)\right|_{t=0} &= g^{(n)}(0) \\
&= \left.\sum_{k=n}^{\infty} \frac{k!\,\mu_k t^{k-n}}{(k-n)!\,k!}\right|_{t=0} \\
&= \mu_n\ .
\end{aligned}$$

It is easy to calculate the moment generating function for simple examples.

Examples

Example 10.1 Suppose X has range $\{1, 2, 3, \ldots, n\}$ and $p_X(j) = 1/n$ for $1 \le j \le n$ (uniform distribution). Then

$$\begin{aligned} g(t) &= \sum_{j=1}^{n} \frac{1}{n} e^{tj} \\ &= \frac{1}{n}(e^t + e^{2t} + \cdots + e^{nt}) \\ &= \frac{e^t(e^{nt}-1)}{n(e^t-1)} \ . \end{aligned}$$

If we use the expression on the right-hand side of the second line above, then it is easy to see that

$$\begin{aligned} \mu_1 &= g'(0) = \frac{1}{n}(1 + 2 + 3 + \cdots + n) = \frac{n+1}{2}, \\ \mu_2 &= g''(0) = \frac{1}{n}(1 + 4 + 9 + \cdots + n^2) = \frac{(n+1)(2n+1)}{6} \ , \end{aligned}$$

and that $\mu = \mu_1 = (n+1)/2$ and $\sigma^2 = \mu_2 - \mu_1^2 = (n^2 - 1)/12$. □

Example 10.2 Suppose now that X has range $\{0, 1, 2, 3, \ldots, n\}$ and $p_X(j) = \binom{n}{j} p^j q^{n-j}$ for $0 \le j \le n$ (binomial distribution). Then

$$\begin{aligned} g(t) &= \sum_{j=0}^{n} e^{tj} \binom{n}{j} p^j q^{n-j} \\ &= \sum_{j=0}^{n} \binom{n}{j} (pe^t)^j q^{n-j} \\ &= (pe^t + q)^n \ . \end{aligned}$$

Note that

$$\begin{aligned} \mu_1 = g'(0) &= n(pe^t + q)^{n-1} pe^t \big|_{t=0} = np \ , \\ \mu_2 = g''(0) &= n(n-1)p^2 + np \ , \end{aligned}$$

so that $\mu = \mu_1 = np$, and $\sigma^2 = \mu_2 - \mu_1^2 = np(1-p)$, as expected. □

Example 10.3 Suppose X has range $\{1, 2, 3, \ldots\}$ and $p_X(j) = q^{j-1}p$ for all j (geometric distribution). Then

$$\begin{aligned} g(t) &= \sum_{j=1}^{\infty} e^{tj} q^{j-1} p \\ &= \frac{pe^t}{1 - qe^t} \ . \end{aligned}$$

Here

$$\begin{aligned}\mu_1 &= g'(0) = \left.\frac{pe^t}{(1-qe^t)^2}\right|_{t=0} = \frac{1}{p}\ ,\\ \mu_2 &= g''(0) = \left.\frac{pe^t + pqe^{2t}}{(1-qe^t)^3}\right|_{t=0} = \frac{1+q}{p^2}\ ,\end{aligned}$$

$\mu = \mu_1 = 1/p$, and $\sigma^2 = \mu_2 - \mu_1^2 = q/p^2$, as computed in Example 6.26. □

Example 10.4 Let X have range $\{0, 1, 2, 3, \ldots\}$ and let $p_X(j) = e^{-\lambda}\lambda^j/j!$ for all j (Poisson distribution with mean λ). Then

$$\begin{aligned}g(t) &= \sum_{j=0}^{\infty} e^{tj}\frac{e^{-\lambda}\lambda^j}{j!}\\ &= e^{-\lambda}\sum_{j=0}^{\infty}\frac{(\lambda e^t)^j}{j!}\\ &= e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}\ .\end{aligned}$$

Then

$$\begin{aligned}\mu_1 &= g'(0) = \left.e^{\lambda(e^t-1)}\lambda e^t\right|_{t=0} = \lambda\ ,\\ \mu_2 &= g''(0) = \left.e^{\lambda(e^t-1)}(\lambda^2e^{2t} + \lambda e^t)\right|_{t=0} = \lambda^2 + \lambda\ ,\end{aligned}$$

$\mu = \mu_1 = \lambda$, and $\sigma^2 = \mu_2 - \mu_1^2 = \lambda$.

The variance of the Poisson distribution is easier to obtain in this way than directly from the definition (as was done in Exercise 6.2.29). □

Moment Problem

Using the moment generating function, we can now show, at least in the case of a discrete random variable with finite range, that its distribution function is completely determined by its moments.

Theorem 10.1 Let X be a discrete random variable with finite range $\{x_1, x_2, \ldots, x_n\}$, distribution function p, and moment generating function g. Then g is uniquely determined by p, and conversely.

Proof. We know that p determines g, since

$$g(t) = \sum_{j=1}^{n} e^{tx_j}p(x_j)\ .$$

Conversely, assume that $g(t)$ is known. We wish to determine the values of x_j and $p(x_j)$, for $1 \le j \le n$. We assume, without loss of generality, that $p(x_j) > 0$ for $1 \le j \le n$, and that

$$x_1 < x_2 < \ldots < x_n\ .$$

We note that $g(t)$ is differentiable for all t, since it is a finite linear combination of exponential functions. If we compute $g'(t)/g(t)$, we obtain

$$\frac{x_1p(x_1)e^{tx_1} + \ldots + x_np(x_n)e^{tx_n}}{p(x_1)e^{tx_1} + \ldots + p(x_n)e^{tx_n}} \ .$$

Dividing both top and bottom by e^{tx_n}, we obtain the expression

$$\frac{x_1p(x_1)e^{t(x_1-x_n)} + \ldots + x_np(x_n)}{p(x_1)e^{t(x_1-x_n)} + \ldots + p(x_n)} \ .$$

Since x_n is the largest of the x_j's, this expression approaches x_n as t goes to ∞. So we have shown that

$$x_n = \lim_{t\to\infty} \frac{g'(t)}{g(t)} \ .$$

To find $p(x_n)$, we simply divide $g(t)$ by e^{tx_n} and let t go to ∞. Once x_n and $p(x_n)$ have been determined, we can subtract $p(x_n)e^{tx_n}$ from $g(t)$, and repeat the above procedure with the resulting function, obtaining, in turn, $x_{n-1}, \ldots, x_1$ and $p(x_{n-1}), \ldots, p(x_1)$. $\square$

If we delete the hypothesis that X have finite range in the above theorem, then the conclusion is no longer necessarily true.

Ordinary Generating Functions

In the special but important case where the x_j are all nonnegative integers, $x_j = j$, we can prove this theorem in a simpler way.

In this case, we have

$$g(t) = \sum_{j=0}^{n} e^{tj}p(j) \ ,$$

and we see that $g(t)$ is a *polynomial* in e^t. If we write $z = e^t$, and define the function h by

$$h(z) = \sum_{j=0}^{n} z^j p(j) \ ,$$

then $h(z)$ is a polynomial in z containing the same information as $g(t)$, and in fact

$$\begin{aligned} h(z) &= g(\log z) \ , \\ g(t) &= h(e^t) \ . \end{aligned}$$

The function $h(z)$ is often called the *ordinary generating function* for X. Note that $h(1) = g(0) = 1$, $h'(1) = g'(0) = \mu_1$, and $h''(1) = g''(0) - g'(0) = \mu_2 - \mu_1$. It follows from all this that if we know $g(t)$, then we know $h(z)$, and if we know $h(z)$, then we can find the $p(j)$ by Taylor's formula:

$$\begin{aligned} p(j) &= \text{coefficient of } z^j \text{ in } h(z) \\ &= \frac{h^{(j)}(0)}{j!} \ . \end{aligned}$$

For example, suppose we know that the moments of a certain discrete random variable X are given by

$$\begin{aligned}\mu_0 &= 1\ ,\\ \mu_k &= \frac{1}{2} + \frac{2^k}{4}\ , \qquad \text{for } k \ge 1\ .\end{aligned}$$

Then the moment generating function g of X is

$$\begin{aligned}g(t) &= \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!}\\ &= 1 + \frac{1}{2}\sum_{k=1}^{\infty} \frac{t^k}{k!} + \frac{1}{4}\sum_{k=1}^{\infty} \frac{(2t)^k}{k!}\\ &= \frac{1}{4} + \frac{1}{2}e^t + \frac{1}{4}e^{2t}\ .\end{aligned}$$

This is a polynomial in $z = e^t$, and

$$h(z) = \frac{1}{4} + \frac{1}{2}z + \frac{1}{4}z^2\ .$$

Hence, X must have range $\{0, 1, 2\}$, and p must have values $\{1/4, 1/2, 1/4\}$.

Properties

Both the moment generating function g and the ordinary generating function h have many properties useful in the study of random variables, of which we can consider only a few here. In particular, if X is any discrete random variable and $Y = X + a$, then

$$\begin{aligned}g_Y(t) &= E(e^{tY})\\ &= E(e^{t(X+a)})\\ &= e^{ta}E(e^{tX})\\ &= e^{ta}g_X(t)\ ,\end{aligned}$$

while if $Y = bX$, then

$$\begin{aligned}g_Y(t) &= E(e^{tY})\\ &= E(e^{tbX})\\ &= g_X(bt)\ .\end{aligned}$$

In particular, if

$$X^* = \frac{X - \mu}{\sigma}\ ,$$

then (see Exercise 11)

$$g_{x^*}(t) = e^{-\mu t/\sigma} g_X\left(\frac{t}{\sigma}\right)\ .$$

If X and Y are *independent* random variables and $Z = X + Y$ is their sum, with p_X, p_Y, and p_Z the associated distribution functions, then we have seen in Chapter 7 that p_Z is the *convolution* of p_X and p_Y, and we know that convolution involves a rather complicated calculation. But for the generating functions we have instead the simple relations

$$\begin{aligned} g_Z(t) &= g_X(t)g_Y(t) , \\ h_Z(z) &= h_X(z)h_Y(z) , \end{aligned}$$

that is, g_Z is simply the *product* of g_X and g_Y, and similarly for h_Z.

To see this, first note that if X and Y are independent, then e^{tX} and e^{tY} are independent (see Exercise 5.2.38), and hence

$$E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) .$$

It follows that

$$\begin{aligned} g_Z(t) &= E(e^{tZ}) = E(e^{t(X+Y)}) \\ &= E(e^{tX})E(e^{tY}) \\ &= g_X(t)g_Y(t) , \end{aligned}$$

and, replacing t by $\log z$, we also get

$$h_Z(z) = h_X(z)h_Y(z) .$$

Example 10.5 If X and Y are independent discrete random variables with range $\{0, 1, 2, \ldots, n\}$ and binomial distribution

$$p_X(j) = p_Y(j) = \binom{n}{j} p^j q^{n-j} ,$$

and if $Z = X + Y$, then we know (cf. Section 7.1) that the range of X is

$$\{0, 1, 2, \ldots, 2n\}$$

and X has binomial distribution

$$p_Z(j) = (p_X * p_Y)(j) = \binom{2n}{j} p^j q^{2n-j} .$$

Here we can easily verify this result by using generating functions. We know that

$$\begin{aligned} g_X(t) = g_Y(t) &= \sum_{j=0}^{n} e^{tj} \binom{n}{j} p^j q^{n-j} \\ &= (pe^t + q)^n , \end{aligned}$$

and

$$h_X(z) = h_Y(z) = (pz + q)^n .$$

Hence, we have

$$g_Z(t) = g_X(t)g_Y(t) = (pe^t + q)^{2n} ,$$

or, what is the same,

$$\begin{aligned} h_Z(z) &= h_X(z)h_Y(z) = (pz+q)^{2n} \\ &= \sum_{j=0}^{2n} \binom{2n}{j} (pz)^j q^{2n-j} , \end{aligned}$$

from which we can see that the coefficient of z^j is just $p_Z(j) = \binom{2n}{j} p^j q^{2n-j}$. □

Example 10.6 If X and Y are independent discrete random variables with the non-negative integers $\{0, 1, 2, 3, \ldots\}$ as range, and with geometric distribution function

$$p_X(j) = p_Y(j) = q^j p ,$$

then

$$g_X(t) = g_Y(t) = \frac{p}{1 - qe^t} ,$$

and if $Z = X + Y$, then

$$\begin{aligned} g_Z(t) &= g_X(t)g_Y(t) \\ &= \frac{p^2}{1 - 2qe^t + q^2e^{2t}} . \end{aligned}$$

If we replace e^t by z, we get

$$\begin{aligned} h_Z(z) &= \frac{p^2}{(1-qz)^2} \\ &= p^2 \sum_{k=0}^{\infty} (k+1) q^k z^k , \end{aligned}$$

and we can read off the values of $p_Z(j)$ as the coefficient of z^j in this expansion for $h(z)$, even though $h(z)$ is not a polynomial in this case. The distribution p_Z is a negative binomial distribution (see Section 5.1). □

Here is a more interesting example of the power and scope of the method of generating functions.

Heads or Tails

Example 10.7 In the coin-tossing game discussed in Example 1.4, we now consider the question "When is Peter first in the lead?"

Let X_k describe the outcome of the kth trial in the game

$$X_k = \begin{cases} +1, & \text{if } k\text{th toss is heads,} \\ -1, & \text{if } k\text{th toss is tails.} \end{cases}$$

Then the X_k are independent random variables describing a Bernoulli process. Let $S_0 = 0$, and, for $n \geq 1$, let

$$S_n = X_1 + X_2 + \cdots + X_n \ .$$

Then S_n describes Peter's fortune after n trials, and Peter is first in the lead after n trials if $S_k \leq 0$ for $1 \leq k < n$ and $S_n = 1$.

Now this can happen when $n = 1$, in which case $S_1 = X_1 = 1$, or when $n > 1$, in which case $S_1 = X_1 = -1$. In the latter case, $S_k = 0$ for $k = n - 1$, and perhaps for other k between 1 and n. Let m be the *least* such value of k; then $S_m = 0$ and $S_k < 0$ for $1 \leq k < m$. In this case Peter loses on the first trial, regains his initial position in the next $m - 1$ trials, and gains the lead in the next $n - m$ trials.

Let p be the probability that the coin comes up heads, and let $q = 1 - p$. Let r_n be the probability that Peter is first in the lead after n trials. Then from the discussion above, we see that

$$\begin{aligned}
r_n &= 0 \ , \qquad \text{if } n \text{ even}, \\
r_1 &= p \qquad (= \text{probability of heads in a single toss}), \\
r_n &= q(r_1 r_{n-2} + r_3 r_{n-4} + \cdots + r_{n-2} r_1) \ , \qquad \text{if } n > 1,\ n \text{ odd}.
\end{aligned}$$

Now let T describe the time (that is, the number of trials) required for Peter to take the lead. Then T is a random variable, and since $P(T = n) = r_n$, r is the distribution function for T.

We introduce the generating function $h_T(z)$ for T:

$$h_T(z) = \sum_{n=0}^{\infty} r_n z^n \ .$$

Then, by using the relations above, we can verify the relation

$$h_T(z) = pz + qz(h_T(z))^2 \ .$$

If we solve this quadratic equation for $h_T(z)$, we get

$$h_T(z) = \frac{1 \pm \sqrt{1 - 4pqz^2}}{2qz} = \frac{2pz}{1 \mp \sqrt{1 - 4pqz^2}} \ .$$

Of these two solutions, we want the one that has a convergent power series in z (i.e., that is finite for $z = 0$). Hence we choose

$$h_T(z) = \frac{1 - \sqrt{1 - 4pqz^2}}{2qz} = \frac{2pz}{1 + \sqrt{1 - 4pqz^2}} \ .$$

Now we can ask: What is the probability that Peter is *ever* in the lead? This probability is given by (see Exercise 10)

$$\begin{aligned}
\sum_{n=0}^{\infty} r_n &= h_T(1) = \frac{1 - \sqrt{1 - 4pq}}{2q} \\
&= \frac{1 - |p - q|}{2q} \\
&= \begin{cases} p/q, & \text{if } p < q, \\ 1, & \text{if } p \geq q, \end{cases}
\end{aligned}$$

so that Peter is sure to be in the lead eventually if $p \geq q$.

How long will it take? That is, what is the expected value of T? This value is given by

$$E(T) = h_T'(1) = \begin{cases} 1/(p-q), & \text{if } p > q, \\ \infty, & \text{if } p = q. \end{cases}$$

This says that if $p > q$, then Peter can expect to be in the lead by about $1/(p-q)$ trials, but if $p = q$, he can expect to wait a long time.

A related problem, known as the Gambler's Ruin problem, is studied in Exercise 23 and in Section 12.2. □

Exercises

1 Find the generating functions, both ordinary $h(z)$ and moment $g(t)$, for the following discrete probability distributions.

(a) The distribution describing a fair coin.

(b) The distribution describing a fair die.

(c) The distribution describing a die that always comes up 3.

(d) The uniform distribution on the set $\{n, n+1, n+2, \ldots, n+k\}$.

(e) The binomial distribution on $\{n, n+1, n+2, \ldots, n+k\}$.

(f) The geometric distribution on $\{0, 1, 2, \ldots,\}$ with $p(j) = 2/3^{j+1}$.

2 For each of the distributions (a) through (d) of Exercise 1 calculate the first and second moments, μ_1 and μ_2, directly from their definition, and verify that $h(1) = 1$, $h'(1) = \mu_1$, and $h''(1) = \mu_2 - \mu_1$.

3 Let p be a probability distribution on $\{0, 1, 2\}$ with moments $\mu_1 = 1$, $\mu_2 = 3/2$.

(a) Find its ordinary generating function $h(z)$.

(b) Using (a), find its moment generating function.

(c) Using (b), find its first six moments.

(d) Using (a), find p_0, p_1, and p_2.

4 In Exercise 3, the probability distribution is completely determined by its first two moments. Show that this is always true for any probability distribution on $\{0, 1, 2\}$. *Hint*: Given μ_1 and μ_2, find $h(z)$ as in Exercise 3 and use $h(z)$ to determine p.

5 Let p and p' be the two distributions

$$p = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1/3 & 0 & 0 & 2/3 & 0 \end{pmatrix},$$

$$p' = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 2/3 & 0 & 0 & 1/3 \end{pmatrix}.$$

(a) Show that p and p' have the same first and second moments, but not the same third and fourth moments.

(b) Find the ordinary and moment generating functions for p and p'.

6 Let p be the probability distribution

$$p = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 1/3 & 2/3 \end{pmatrix} ,$$

and let $p_n = p * p * \cdots * p$ be the n-fold convolution of p with itself.

(a) Find p_2 by direct calculation (see Definition 7.1).

(b) Find the ordinary generating functions $h(z)$ and $h_2(z)$ for p and p_2, and verify that $h_2(z) = (h(z))^2$.

(c) Find $h_n(z)$ from $h(z)$.

(d) Find the first two moments, and hence the mean and variance, of p_n from $h_n(z)$. Verify that the mean of p_n is n times the mean of p.

(e) Find those integers j for which $p_n(j) > 0$ from $h_n(z)$.

7 Let X be a discrete random variable with values in $\{0, 1, 2, \ldots, n\}$ and moment generating function $g(t)$. Find, in terms of $g(t)$, the generating functions for

(a) $-X$.

(b) $X + 1$.

(c) $3X$.

(d) $aX + b$.

8 Let $X_1, X_2, \ldots, X_n$ be an independent trials process, with values in $\{0, 1\}$ and mean $\mu = 1/3$. Find the ordinary and moment generating functions for the distribution of

(a) $S_1 = X_1$. *Hint*: First find X_1 explicitly.

(b) $S_2 = X_1 + X_2$.

(c) $S_n = X_1 + X_2 + \cdots + X_n$.

9 Let X and Y be random variables with values in $\{1, 2, 3, 4, 5, 6\}$ with distribution functions p_X and p_Y given by

$$\begin{aligned} p_X(j) &= a_j , \\ p_Y(j) &= b_j . \end{aligned}$$

(a) Find the ordinary generating functions $h_X(z)$ and $h_Y(z)$ for these distributions.

(b) Find the ordinary generating function $h_Z(z)$ for the distribution $Z = X + Y$.

(c) Show that $h_Z(z)$ cannot ever have the form

$$h_Z(z) = \frac{z^2 + z^3 + \cdots + z^{12}}{11} .$$

Hint: h_X and h_Y must have at least one nonzero root, but $h_Z(z)$ in the form given has no nonzero real roots.

It follows from this observation that there is no way to load two dice so that the probability that a given sum will turn up when they are tossed is the same for all sums (i.e., that all outcomes are equally likely).

10 Show that if

$$h(z) = \frac{1 - \sqrt{1 - 4pqz^2}}{2qz} ,$$

then

$$h(1) = \begin{cases} p/q, & \text{if } p \le q, \\ 1, & \text{if } p \ge q, \end{cases}$$

and

$$h'(1) = \begin{cases} 1/(p-q), & \text{if } p > q, \\ \infty, & \text{if } p = q. \end{cases}$$

11 Show that if X is a random variable with mean μ and variance σ^2, and if $X^* = (X - \mu)/\sigma$ is the standardized version of X, then

$$g_{X^*}(t) = e^{-\mu t/\sigma} g_X\left(\frac{t}{\sigma}\right) .$$

10.2 Branching Processes

Historical Background

In this section we apply the theory of generating functions to the study of an important chance process called a *branching process.*

Until recently it was thought that the theory of branching processes originated with the following problem posed by Francis Galton in the *Educational Times* in 1873.[1]

> Problem 4001: A large nation, of whom we will only concern ourselves with the adult males, N in number, and who each bear separate surnames, colonise a district. Their law of population is such that, in each generation, a_0 per cent of the adult males have no male children who reach adult life; a_1 have one such male child; a_2 have two; and so on up to a_5 who have five.
>
> Find (1) what proportion of the surnames will have become extinct after r generations; and (2) how many instances there will be of the same surname being held by m persons.

[1] D. G. Kendall, "Branching Processes Since 1873," *Journal of London Mathematics Society,* vol. 41 (1966), p. 386.

The first attempt at a solution was given by Reverend H. W. Watson. Because of a mistake in algebra, he incorrectly concluded that a family name would always die out with probability 1. However, the methods that he employed to solve the problems were, and still are, the basis for obtaining the correct solution.

Heyde and Seneta discovered an earlier communication by Bienaymé (1845) that anticipated Galton and Watson by 28 years. Bienaymé showed, in fact, that he was aware of the correct solution to Galton's problem. Heyde and Seneta in their book *I. J. Bienaymé: Statistical Theory Anticipated,*[2] give the following translation from Bienaymé's paper:

> If . . . the mean of the number of male children who replace the number of males of the preceding generation were less than unity, it would be easily realized that families are dying out due to the disappearance of the members of which they are composed. However, the analysis shows further that when this mean is equal to unity families tend to disappear, although less rapidly . . . .
>
> The analysis also shows clearly that if the mean ratio is greater than unity, the probability of the extinction of families with the passing of time no longer reduces to certainty. It only approaches a finite limit, which is fairly simple to calculate and which has the singular characteristic of being given by one of the roots of the equation (in which the number of generations is made infinite) which is not relevant to the question when the mean ratio is less than unity.[3]

Although Bienaymé does not give his reasoning for these results, he did indicate that he intended to publish a special paper on the problem. The paper was never written, or at least has never been found. In his communication Bienaymé indicated that he was motivated by the same problem that occurred to Galton. The opening paragraph of his paper as translated by Heyde and Seneta says,

> A great deal of consideration has been given to the possible multiplication of the numbers of mankind; and recently various very curious observations have been published on the fate which allegedly hangs over the aristocrary and middle classes; the families of famous men, etc. This fate, it is alleged, will inevitably bring about the disappearance of the so-called *families fermées.*[4]

A much more extensive discussion of the history of branching processes may be found in two papers by David G. Kendall.[5]

---

[2] C. C. Heyde and E. Seneta, *I. J. Bienaymé: Statistical Theory Anticipated* (New York: Springer Verlag, 1977).

[3] ibid., pp. 117–118.

[4] ibid., p. 118.

[5] D. G. Kendall, "Branching Processes Since 1873," pp. 385–406; and "The Genealogy of Genealogy: Branching Processes Before (and After) 1873," *Bulletin London Mathematics Society,* vol. 7 (1975), pp. 225–253.

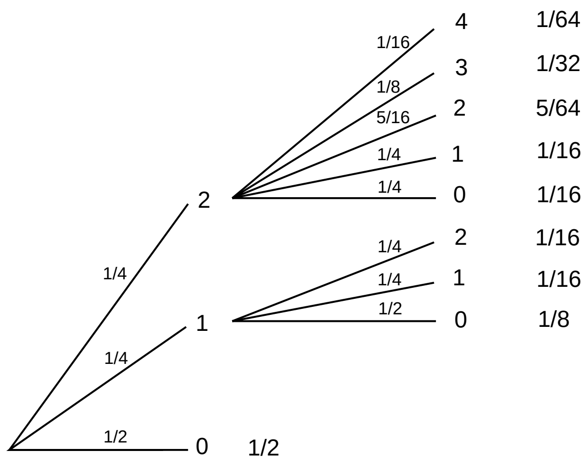

Figure 10.1: Tree diagram for Example 10.8.

Branching processes have served not only as crude models for population growth but also as models for certain physical processes such as chemical and nuclear chain reactions.

Problem of Extinction

We turn now to the first problem posed by Galton (i.e., the problem of finding the probability of extinction for a branching process). We start in the 0th generation with 1 male parent. In the first generation we shall have 0, 1, 2, 3, ... male offspring with probabilities p_0, p_1, p_2, p_3, ... If in the first generation there are k offspring, then in the second generation there will be $X_1 + X_2 + \cdots + X_k$ offspring, where X_1, X_2, ..., X_k are independent random variables, each with the common distribution p_0, p_1, p_2, ... This description enables us to construct a tree, and a tree measure, for any number of generations.

Examples

Example 10.8 Assume that $p_0 = 1/2$, $p_1 = 1/4$, and $p_2 = 1/4$. Then the tree measure for the first two generations is shown in Figure 10.1.

Note that we use the theory of sums of independent random variables to assign branch probabilities. For example, if there are two offspring in the first generation, the probability that there will be two in the second generation is

$$\begin{aligned} P(X_1 + X_2 = 2) &= p_0p_2 + p_1p_1 + p_2p_0 \\ &= \frac{1}{2} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{2} = \frac{5}{16} \ . \end{aligned}$$

We now study the probability that our process dies out (i.e., that at some generation there are no offspring).

Let d_m be the probability that the process dies out by the mth generation. Of course, $d_0 = 0$. In our example, $d_1 = 1/2$ and $d_2 = 1/2 + 1/8 + 1/16 = 11/16$ (see Figure 10.1). Note that we must add the probabilities for all paths that lead to 0 by the mth generation. It is clear from the definition that

$$0 = d_0 \leq d_1 \leq d_2 \leq \cdots \leq 1 \ .$$

Hence, d_m converges to a limit d, $0 \leq d \leq 1$, and d is the probability that the process will ultimately die out. It is this value that we wish to determine. We begin by expressing the value d_m in terms of all possible outcomes on the first generation. If there are j offspring in the first generation, then to die out by the mth generation, each of these lines must die out in $m-1$ generations. Since they proceed independently, this probability is $(d_{m-1})^j$. Therefore

$$d_m = p_0 + p_1 d_{m-1} + p_2 (d_{m-1})^2 + p_3 (d_{m-1})^3 + \cdots \ . \tag{10.1}$$

Let $h(z)$ be the ordinary generating function for the p_i:

$$h(z) = p_0 + p_1 z + p_2 z^2 + \cdots \ .$$

Using this generating function, we can rewrite Equation 10.1 in the form

$$d_m = h(d_{m-1}) \ . \tag{10.2}$$

Since $d_m \to d$, by Equation 10.2 we see that the value d that we are looking for satisfies the equation

$$d = h(d) \ . \tag{10.3}$$

One solution of this equation is always $d = 1$, since

$$1 = p_0 + p_1 + p_2 + \cdots \ .$$

This is where Watson made his mistake. He assumed that 1 was the only solution to Equation 10.3. To examine this question more carefully, we first note that solutions to Equation 10.3 represent intersections of the graphs of

$$y = z$$

and

$$y = h(z) = p_0 + p_1 z + p_2 z^2 + \cdots \ .$$

Thus we need to study the graph of $y = h(z)$. We note that $h(0) = p_0$. Also,

$$h'(z) = p_1 + 2p_2 z + 3p_3 z^2 + \cdots \ , \tag{10.4}$$

and

$$h''(z) = 2p_2 + 3 \cdot 2p_3 z + 4 \cdot 3p_4 z^2 + \cdots \ .$$

From this we see that for $z \geq 0$, $h'(z) \geq 0$ and $h''(z) \geq 0$. Thus for nonnegative z, $h(z)$ is an increasing function and is concave upward. Therefore the graph of

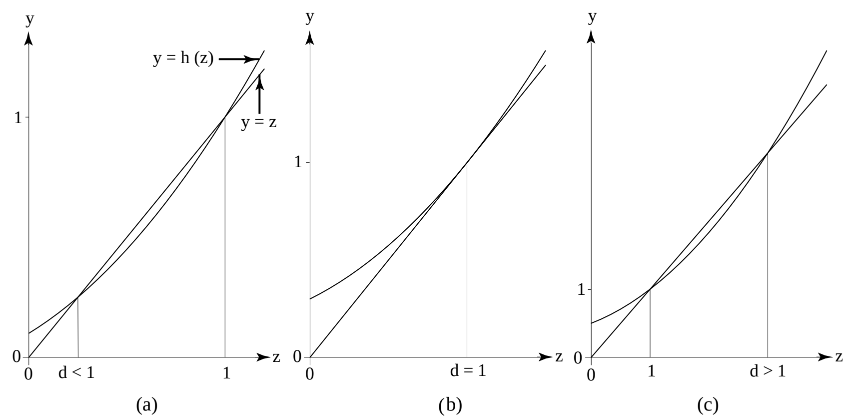

Figure 10.2: Graphs of $y = z$ and $y = h(z)$.

$y = h(z)$ can intersect the line $y = z$ in at most two points. Since we know it must intersect the line $y = z$ at $(1, 1)$, we know that there are just three possibilities, as shown in Figure 10.2.

In case (a) the equation $d = h(d)$ has roots $\{d, 1\}$ with $0 \leq d < 1$. In the second case (b) it has only the one root $d = 1$. In case (c) it has two roots $\{1, d\}$ where $1 < d$. Since we are looking for a solution $0 \leq d \leq 1$, we see in cases (b) and (c) that our only solution is 1. In these cases we can conclude that the process will die out with probability 1. However in case (a) we are in doubt. We must study this case more carefully.

From Equation 10.4 we see that

$$h'(1) = p_1 + 2p_2 + 3p_3 + \cdots = m \ ,$$

where m is the expected number of offspring produced by a single parent. In case (a) we have $h'(1) > 1$, in (b) $h'(1) = 1$, and in (c) $h'(1) < 1$. Thus our three cases correspond to $m > 1$, $m = 1$, and $m < 1$. We assume now that $m > 1$. Recall that $d_0 = 0$, $d_1 = h(d_0) = p_0$, $d_2 = h(d_1)$, …, and $d_n = h(d_{n-1})$. We can construct these values geometrically, as shown in Figure 10.3.

We can see geometrically, as indicated for d_0, d_1, d_2, and d_3 in Figure 10.3, that the points $(d_i, h(d_i))$ will always lie above the line $y = z$. Hence, they must converge to the first intersection of the curves $y = z$ and $y = h(z)$ (i.e., to the root $d < 1$). This leads us to the following theorem. □

Theorem 10.2 Consider a branching process with generating function $h(z)$ for the number of offspring of a given parent. Let d be the smallest root of the equation $z = h(z)$. If the mean number m of offspring produced by a single parent is ≤ 1, then $d = 1$ and the process dies out with probability 1. If $m > 1$ then $d < 1$ and the process dies out with probability d. □

We shall often want to know the probability that a branching process dies out by a particular generation, as well as the limit of these probabilities. Let d_n be

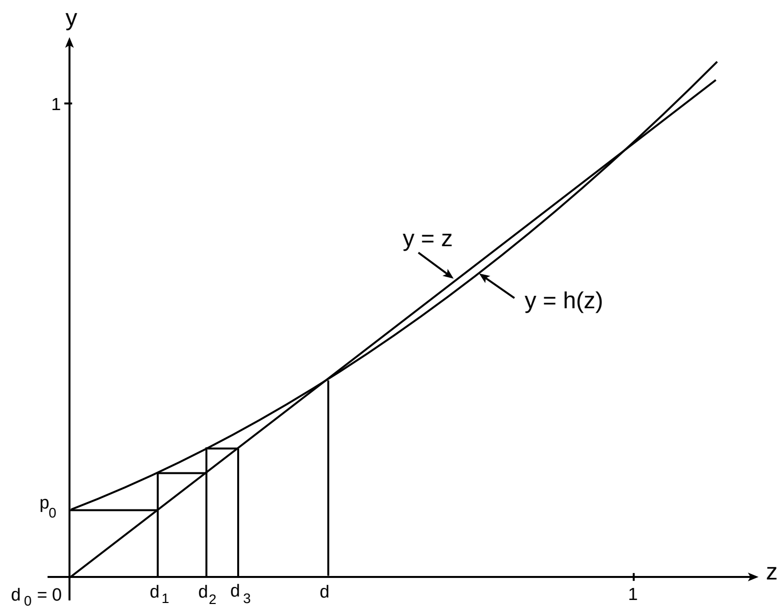

Figure 10.3: Geometric determination of d.

the probability of dying out by the nth generation. Then we know that $d_1 = p_0$. We know further that $d_n = h(d_{n-1})$ where $h(z)$ is the generating function for the number of offspring produced by a single parent. This makes it easy to compute these probabilities.

The program **Branch** calculates the values of d_n. We have run this program for 12 generations for the case that a parent can produce at most two offspring and the probabilities for the number produced are $p_0 = .2$, $p_1 = .5$, and $p_2 = .3$. The results are given in Table 10.1.

We see that the probability of dying out by 12 generations is about .6. We shall see in the next example that the probability of eventually dying out is 2/3, so that even 12 generations is not enough to give an accurate estimate for this probability.

We now assume that at most two offspring can be produced. Then

$$h(z) = p_0 + p_1 z + p_2 z^2 \ .$$

In this simple case the condition $z = h(z)$ yields the equation

$$d = p_0 + p_1 d + p_2 d^2 \ ,$$

which is satisfied by $d = 1$ and $d = p_0/p_2$. Thus, in addition to the root $d = 1$ we have the second root $d = p_0/p_2$. The mean number m of offspring produced by a single parent is

$$m = p_1 + 2p_2 = 1 - p_0 - p_2 + 2p_2 = 1 - p_0 + p_2 \ .$$

Thus, if $p_0 > p_2$, $m < 1$ and the second root is > 1. If $p_0 = p_2$, we have a double root $d = 1$. If $p_0 < p_2$, $m > 1$ and the second root d is less than 1 and represents the probability that the process will die out.

Generation	Probability of dying out
1	.2
2	.312
3	.385203
4	.437116
5	.475879
6	.505878
7	.529713
8	.549035
9	.564949
10	.578225
11	.589416
12	.598931

Table 10.1: Probability of dying out.

$$
\begin{aligned}
p_0 &= .2092 \\
p_1 &= .2584 \\
p_2 &= .2360 \\
p_3 &= .1593 \\
p_4 &= .0828 \\
p_5 &= .0357 \\
p_6 &= .0133 \\
p_7 &= .0042 \\
p_8 &= .0011 \\
p_9 &= .0002 \\
p_{10} &= .0000
\end{aligned}
$$

Table 10.2: Distribution of number of female children.

Example 10.9 Keyfitz[6] compiled and analyzed data on the continuation of the female family line among Japanese women. His estimates at the basic probability distribution for the number of female children born to Japanese women of ages 45–49 in 1960 are given in Table 10.2.

The expected number of girls in a family is then 1.837 so the probability d of extinction is less than 1. If we run the program **Branch**, we can estimate that d is in fact only about .324. □

Distribution of Offspring

So far we have considered only the first of the two problems raised by Galton, namely the probability of extinction. We now consider the second problem, that is, the distribution of the number Z_n of offspring in the nth generation. The exact form of the distribution is not known except in very special cases. We shall see,

[6]N. Keyfitz, *Introduction to the Mathematics of Population,* rev. ed. (Reading, PA: Addison Wesley, 1977).

however, that we can describe the limiting behavior of Z_n as $n \to \infty$.

We first show that the generating function $h_n(z)$ of the distribution of Z_n can be obtained from $h(z)$ for any branching process.

We recall that the value of the generating function at the value z for any random variable X can be written as

$$h(z) = E(z^X) = p_0 + p_1 z + p_2 z^2 + \cdots \ .$$

That is, $h(z)$ is the expected value of an experiment which has outcome z^j with probability p_j.

Let $S_n = X_1 + X_2 + \cdots + X_n$ where each X_j has the same integer-valued distribution (p_j) with generating function $k(z) = p_0 + p_1 z + p_2 z^2 + \cdots$. Let $k_n(z)$ be the generating function of S_n. Then using one of the properties of ordinary generating functions discussed in Section 10.1, we have

$$k_n(z) = (k(z))^n \ ,$$

since the X_j's are independent and all have the same distribution.

Consider now the branching process Z_n. Let $h_n(z)$ be the generating function of Z_n. Then

$$\begin{aligned} h_{n+1}(z) &= E(z^{Z_{n+1}}) \\ &= \sum_k E(z^{Z_{n+1}} | Z_n = k) P(Z_n = k) \ . \end{aligned}$$

If $Z_n = k$, then $Z_{n+1} = X_1 + X_2 + \cdots + X_k$ where $X_1, X_2, \ldots, X_k$ are independent random variables with common generating function $h(z)$. Thus

$$E(z^{Z_{n+1}} | Z_n = k) = E(z^{X_1 + X_2 + \cdots + X_k}) = (h(z))^k \ ,$$

and

$$h_{n+1}(z) = \sum_k (h(z))^k P(Z_n = k) \ .$$

But

$$h_n(z) = \sum_k P(Z_n = k) z^k \ .$$

Thus,

$$h_{n+1}(z) = h_n(h(z)) \ . \tag{10.5}$$

If we differentiate Equation 10.5 and use the chain rule we have

$$h'_{n+1}(z) = h'_n(h(z)) h'(z).$$

Putting $z = 1$ and using the fact that $h(1) = 1$, $h'(1) = m$, and $h'_n(1) = m_n$ = the mean number of offspring in the n'th generation, we have

$$m_{n+1} = m_n \cdot m \ .$$

Thus, $m_2 = m \cdot m = m^2$, $m_3 = m^2 \cdot m = m^3$, and in general

$$m_n = m^n \ .$$

Thus, for a branching process with $m > 1$, the mean number of offspring grows exponentially at a rate m.

Examples

Example 10.10 For the branching process of Example 10.8 we have

$$\begin{aligned} h(z) &= 1/2 + (1/4)z + (1/4)z^2 \ , \\ h_2(z) &= h(h(z)) = 1/2 + (1/4)[1/2 + (1/4)z + (1/4)z^2] \\ &= +(1/4)[1/2 + (1/4)z + (1/4)z^2]^2 \\ &= 11/16 + (1/8)z + (9/64)z^2 + (1/32)z^3 + (1/64)z^4 \ . \end{aligned}$$

The probabilities for the number of offspring in the second generation agree with those obtained directly from the tree measure (see Figure 1). $\square$

It is clear that even in the simple case of at most two offspring, we cannot easily carry out the calculation of $h_n(z)$ by this method. However, there is one special case in which this can be done.

Example 10.11 Assume that the probabilities p_1, p_2, ... form a geometric series: $p_k = bc^{k-1}$, $k = 1, 2, \ldots$, with $0 < b \le 1 - c$ and $0 < c < 1$. Then we have

$$\begin{aligned} p_0 &= 1 - p_1 - p_2 - \cdots \\ &= 1 - b - bc - bc^2 - \cdots \\ &= 1 - \frac{b}{1-c} \ . \end{aligned}$$

The generating function $h(z)$ for this distribution is

$$\begin{aligned} h(z) &= p_0 + p_1 z + p_2 z^2 + \cdots \\ &= 1 - \frac{b}{1-c} + bz + bcz^2 + bc^2z^3 + \cdots \\ &= 1 - \frac{b}{1-c} + \frac{bz}{1-cz} \ . \end{aligned}$$

From this we find

$$h'(z) = \frac{bcz}{(1-cz)^2} + \frac{b}{1-cz} = \frac{b}{(1-cz)^2}$$

and

$$m = h'(1) = \frac{b}{(1-c)^2} \ .$$

We know that if $m \le 1$ the process will surely die out and $d = 1$. To find the probability d when $m > 1$ we must find a root $d < 1$ of the equation

$$z = h(z) \ ,$$

or

$$z = 1 - \frac{b}{1-c} + \frac{bz}{1-cz}.$$

This leads us to a quadratic equation. We know that $z = 1$ is one solution. The other is found to be

$$d = \frac{1 - b - c}{c(1 - c)}.$$

It is easy to verify that $d < 1$ just when $m > 1$.

It is possible in this case to find the distribution of Z_n. This is done by first finding the generating function $h_n(z)$.[7] The result for $m \neq 1$ is:

$$h_n(z) = 1 - m^n \left[\frac{1 - d}{m^n - d}\right] + \frac{m^n \left[\frac{1-d}{m^n - d}\right]^2 z}{1 - \left[\frac{m^n - 1}{m^n - d}\right] z}.$$

The coefficients of the powers of z give the distribution for Z_n:

$$P(Z_n = 0) = 1 - m^n \frac{1 - d}{m^n - d} = \frac{d(m^n - 1)}{m^n - d}$$

and

$$P(Z_n = j) = m^n \Big(\frac{1 - d}{m^n - d}\Big)^2 \cdot \Big(\frac{m^n - 1}{m^n - d}\Big)^{j-1},$$

for $j \geq 1$. □

Example 10.12 Let us re-examine the Keyfitz data to see if a distribution of the type considered in Example 10.11 could reasonably be used as a model for this population. We would have to estimate from the data the parameters b and c for the formula $p_k = bc^{k-1}$. Recall that

$$m = \frac{b}{(1 - c)^2} \tag{10.6}$$

and the probability d that the process dies out is

$$d = \frac{1 - b - c}{c(1 - c)}. \tag{10.7}$$

Solving Equation 10.6 and 10.7 for b and c gives

$$c = \frac{m - 1}{m - d}$$

and

$$b = m\Big(\frac{1 - d}{m - d}\Big)^2.$$

We shall use the value 1.837 for m and .324 for d that we found in the Keyfitz example. Using these values, we obtain $b = .3666$ and $c = .5533$. Note that $(1 - c)^2 < b < 1 - c$, as required. In Table 10.3 we give for comparison the probabilities p_0 through p_8 as calculated by the geometric distribution versus the empirical values.

[7] T. E. Harris, *The Theory of Branching Processes* (Berlin: Springer, 1963), p. 9.

p_j	Data	Geometric Model
0	.2092	.1816
1	.2584	.3666
2	.2360	.2028
3	.1593	.1122
4	.0828	.0621
5	.0357	.0344
6	.0133	.0190
7	.0042	.0105
8	.0011	.0058
9	.0002	.0032
10	.0000	.0018

Table 10.3: Comparison of observed and expected frequencies.

The geometric model tends to favor the larger numbers of offspring but is similar enough to show that this modified geometric distribution might be appropriate to use for studies of this kind.

Recall that if $S_n = X_1 + X_2 + \cdots + X_n$ is the sum of independent random variables with the same distribution then the Law of Large Numbers states that S_n/n converges to a constant, namely $E(X_1)$. It is natural to ask if there is a similar limiting theorem for branching processes.

Consider a branching process with Z_n representing the number of offspring after n generations. Then we have seen that the expected value of Z_n is m^n. Thus we can scale the random variable Z_n to have expected value 1 by considering the random variable

$$W_n = \frac{Z_n}{m^n} \ .$$

In the theory of branching processes it is proved that this random variable W_n will tend to a limit as n tends to infinity. However, unlike the case of the Law of Large Numbers where this limit is a constant, for a branching process the limiting value of the random variables W_n is itself a random variable.

Although we cannot prove this theorem here we can illustrate it by simulation. This requires a little care. When a branching process survives, the number of offspring is apt to get very large. If in a given generation there are 1000 offspring, the offspring of the next generation are the result of 1000 chance events, and it will take a while to simulate these 1000 experiments. However, since the final result is the sum of 1000 independent experiments we can use the Central Limit Theorem to replace these 1000 experiments by a single experiment with normal density having the appropriate mean and variance. The program **BranchingSimulation** carries out this process.

We have run this program for the Keyfitz example, carrying out 10 simulations and graphing the results in Figure 10.4.

The expected number of female offspring per female is 1.837, so that we are graphing the outcome for the random variables $W_n = Z_n/(1.837)^n$. For three of

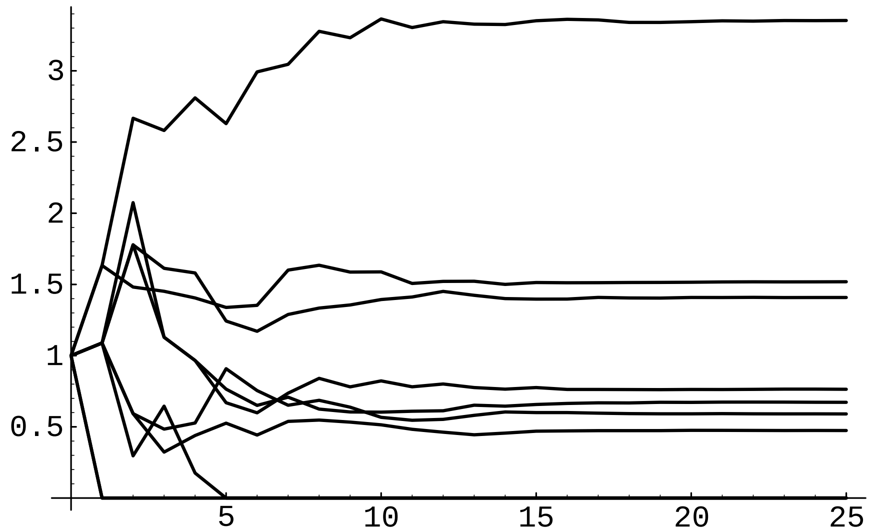

Figure 10.4: Simulation of Z_n/m^n for the Keyfitz example.

the simulations the process died out, which is consistent with the value $d = .3$ that we found for this example. For the other seven simulations the value of W_n tends to a limiting value which is different for each simulation. $\square$

Example 10.13 We now examine the random variable Z_n more closely for the case $m < 1$ (see Example 10.11). Fix a value $t > 0$; let $[tm^n]$ be the integer part of tm^n. Then

$$\begin{aligned} P(Z_n = [tm^n]) &= m^n (\frac{1-d}{m^n - d})^2 (\frac{m^n - 1}{m^n - d})^{[tm^n]-1} \\ &= \frac{1}{m^n} (\frac{1-d}{1-d/m^n})^2 (\frac{1-1/m^n}{1-d/m^n})^{tm^n+a} \ , \end{aligned}$$

where $|a| \leq 2$. Thus, as $n \to \infty$,

$$m^n P(Z_n = [tm^n]) \to (1-d)^2 \frac{e^{-t}}{e^{-td}} = (1-d)^2 e^{-t(1-d)} \ .$$

For $t = 0$,

$$P(Z_n = 0) \to d \ .$$

We can compare this result with the Central Limit Theorem for sums S_n of integer-valued independent random variables (see Theorem 9.3), which states that if t is an integer and $u = (t - n\mu)/\sqrt{\sigma^2 n}$, then as $n \to \infty$,

$$\sqrt{\sigma^2 n}\, P(S_n = u\sqrt{\sigma^2 n} + \mu n) \to \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \ .$$

We see that the form of these statements are quite similar. It is possible to prove a limit theorem for a general class of branching processes that states that under suitable hypotheses, as $n \to \infty$,

$$m^n P(Z_n = [tm^n]) \to k(t) ,$$

for $t > 0$, and

$$P(Z_n = 0) \to d .$$

However, unlike the Central Limit Theorem for sums of independent random variables, the function $k(t)$ will depend upon the basic distribution that determines the process. Its form is known for only a very few examples similar to the one we have considered here. □

Chain Letter Problem

Example 10.14 An interesting example of a branching process was suggested by Free Huizinga.[8] In 1978, a chain letter called the "Circle of Gold," believed to have started in California, found its way across the country to the theater district of New York. The chain required a participant to buy a letter containing a list of 12 names for 100 dollars. The buyer gives 50 dollars to the person from whom the letter was purchased and then sends 50 dollars to the person whose name is at the top of the list. The buyer then crosses off the name at the top of the list and adds her own name at the bottom in each letter before it is sold again.

Let us first assume that the buyer may sell the letter only to a single person. If you buy the letter you will want to compute your expected winnings. (We are ignoring here the fact that the passing on of chain letters through the mail is a federal offense with certain obvious resulting penalties.) Assume that each person involved has a probability p of selling the letter. Then you will receive 50 dollars with probability p and another 50 dollars if the letter is sold to 12 people, since then your name would have risen to the top of the list. This occurs with probability p^{12}, and so your expected winnings are $-100 + 50p + 50p^{12}$. Thus the chain in this situation is a highly unfavorable game.

It would be more reasonable to allow each person involved to make a copy of the list and try to sell the letter to at least 2 other people. Then you would have a chance of recovering your 100 dollars on these sales, and if any of the letters is sold 12 times you will receive a bonus of 50 dollars for each of these cases. We can consider this as a branching process with 12 generations. The members of the first generation are the letters you sell. The second generation consists of the letters sold by members of the first generation, and so forth.

Let us assume that the probabilities that each individual sells letters to 0, 1, or 2 others are p_0, p_1, and p_2, respectively. Let $Z_1, Z_2, \ldots, Z_{12}$ be the number of letters in the first 12 generations of this branching process. Then your expected winnings are

$$50(E(Z_1) + E(Z_{12})) = 50m + 50m^{12} ,$$

[8]Private communication.

where $m = p_1 + 2p_2$ is the expected number of letters you sold. Thus to be favorable we just have

$$50m + 50m^{12} > 100 \ ,$$

or

$$m + m^{12} > 2 \ .$$

But this will be true if and only if $m > 1$. We have seen that this will occur in the quadratic case if and only if $p_2 > p_0$. Let us assume for example that $p_0 = .2$, $p_1 = .5$, and $p_2 = .3$. Then $m = 1.1$ and the chain would be a favorable game. Your expected profit would be

$$50(1.1 + 1.1^{12}) - 100 \approx 112 \ .$$

The probability that you receive at least one payment from the 12th generation is $1 - d_{12}$. We find from our program **Branch** that $d_{12} = .599$. Thus, $1 - d_{12} = .401$ is the probability that you receive some bonus. The maximum that you could receive from the chain would be $50(2 + 2^{12}) = 204{,}900$ if everyone were to successfully sell two letters. Of course you can not always expect to be so lucky. (What is the probability of this happening?)

To simulate this game, we need only simulate a branching process for 12 generations. Using a slightly modified version of our program **BranchingSimulation** we carried out twenty such simulations, giving the results shown in Table 10.4.

Note that we were quite lucky on a few runs, but we came out ahead only a little less than half the time. The process died out by the twelfth generation in 12 out of the 20 experiments, in good agreement with the probability $d_{12} = .599$ that we calculated using the program **Branch**.

Let us modify the assumptions about our chain letter to let the buyer sell the letter to as many people as she can instead of to a maximum of two. We shall assume, in fact, that a person has a large number N of acquaintances and a small probability p of persuading any one of them to buy the letter. Then the distribution for the number of letters that she sells will be a binomial distribution with mean $m = Np$. Since N is large and p is small, we can assume that the probability p_j that an individual sells the letter to j people is given by the Poisson distribution

$$p_j = \frac{e^{-m} m^j}{j!} \ .$$

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Profit
1	0	0	0	0	0	0	0	0	0	0	0	-50
1	1	2	3	2	3	2	1	2	3	3	6	250
0	0	0	0	0	0	0	0	0	0	0	0	-100
2	4	4	2	3	4	4	3	2	2	1	1	50
1	2	3	5	4	3	3	3	5	8	6	6	250
0	0	0	0	0	0	0	0	0	0	0	0	-100
2	3	2	2	2	1	2	3	3	3	4	6	300
1	2	1	1	1	1	2	1	0	0	0	0	-50
0	0	0	0	0	0	0	0	0	0	0	0	-100
1	0	0	0	0	0	0	0	0	0	0	0	-50
2	3	2	3	3	3	5	9	12	12	13	15	750
1	1	1	0	0	0	0	0	0	0	0	0	-50
1	2	2	3	3	0	0	0	0	0	0	0	-50
1	1	1	1	2	2	3	4	4	6	4	5	200
1	1	0	0	0	0	0	0	0	0	0	0	-50
1	0	0	0	0	0	0	0	0	0	0	0	-50
1	0	0	0	0	0	0	0	0	0	0	0	-50
1	1	2	3	3	4	2	3	3	3	3	2	50
1	2	4	6	6	9	10	13	16	17	15	18	850
1	0	0	0	0	0	0	0	0	0	0	0	-50

Table 10.4: Simulation of chain letter (finite distribution case).

Z_1	Z_2	Z_3	Z_4	Z_5	Z_6	Z_7	Z_8	Z_9	Z_{10}	Z_{11}	Z_{12}	Profit
1	2	6	7	7	8	11	9	7	6	6	5	200
1	0	0	0	0	0	0	0	0	0	0	0	-50
1	0	0	0	0	0	0	0	0	0	0	0	-50
1	1	1	0	0	0	0	0	0	0	0	0	-50
0	0	0	0	0	0	0	0	0	0	0	0	-100
1	1	1	1	1	1	2	4	9	7	9	7	300
2	3	3	4	2	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	-50
2	1	0	0	0	0	0	0	0	0	0	0	0
3	3	4	7	11	17	14	11	11	10	16	25	1300
0	0	0	0	0	0	0	0	0	0	0	0	-100
1	2	2	1	1	3	1	0	0	0	0	0	-50
0	0	0	0	0	0	0	0	0	0	0	0	-100
2	3	1	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	50
1	0	0	0	0	0	0	0	0	0	0	0	-50
3	4	4	7	10	11	9	11	12	14	13	10	550
1	3	3	4	9	5	7	9	8	8	6	3	100
1	0	4	6	6	9	10	13	0	0	0	0	-50
1	0	0	0	0	0	0	0	0	0	0	0	-50

Table 10.5: Simulation of chain letter (Poisson case).

The generating function for the Poisson distribution is

$$\begin{aligned} h(z) &= \sum_{j=0}^{\infty} \frac{e^{-m} m^j z^j}{j!} \\ &= e^{-m} \sum_{j=0}^{\infty} \frac{m^j z^j}{j!} \\ &= e^{-m} e^{mz} = e^{m(z-1)} . \end{aligned}$$

The expected number of letters that an individual passes on is m, and again to be favorable we must have $m > 1$. Let us assume again that $m = 1.1$. Then we can find again the probability $1 - d_{12}$ of a bonus from **Branch**. The result is .232. Although the expected winnings are the same, the variance is larger in this case, and the buyer has a better chance for a reasonably large profit. We again carried out 20 simulations using the Poisson distribution with mean 1.1. The results are shown in Table 10.5.

We note that, as before, we came out ahead less than half the time, but we also had one large profit. In only 6 of the 20 cases did we receive any profit. This is again in reasonable agreement with our calculation of a probability .232 for this happening. □

Exercises

1 Let Z_1, Z_2, ..., Z_N describe a branching process in which each parent has j offspring with probability p_j. Find the probability d that the process eventually dies out if

(a) $p_0 = 1/2$, $p_1 = 1/4$, and $p_2 = 1/4$.

(b) $p_0 = 1/3$, $p_1 = 1/3$, and $p_2 = 1/3$.

(c) $p_0 = 1/3$, $p_1 = 0$, and $p_2 = 2/3$.

(d) $p_j = 1/2^{j+1}$, for $j = 0, 1, 2, \ldots$.

(e) $p_j = (1/3)(2/3)^j$, for $j = 0, 1, 2, \ldots$.

(f) $p_j = e^{-2}2^j/j!$, for $j = 0, 1, 2, \ldots$ (estimate d numerically).

2 Let Z_1, Z_2, ..., Z_N describe a branching process in which each parent has j offspring with probability p_j. Find the probability d that the process dies out if

(a) $p_0 = 1/2$, $p_1 = p_2 = 0$, and $p_3 = 1/2$.

(b) $p_0 = p_1 = p_2 = p_3 = 1/4$.

(c) $p_0 = t$, $p_1 = 1 - 2t$, $p_2 = 0$, and $p_3 = t$, where $t \leq 1/2$.

3 In the chain letter problem (see Example 10.14) find your expected profit if

(a) $p_0 = 1/2$, $p_1 = 0$, and $p_2 = 1/2$.

(b) $p_0 = 1/6$, $p_1 = 1/2$, and $p_2 = 1/3$.

Show that if $p_0 > 1/2$, you cannot expect to make a profit.

4 Let $S_N = X_1 + X_2 + \cdots + X_N$, where the X_i's are independent random variables with common distribution having generating function $f(z)$. Assume that N is an integer valued random variable independent of all of the X_j and having generating function $g(z)$. Show that the generating function for S_N is $h(z) = g(f(z))$. *Hint*: Use the fact that

$$h(z) = E(z^{S_N}) = \sum_k E(z^{S_N}|N = k)P(N = k) \ .$$

5 We have seen that if the generating function for the offspring of a single parent is $f(z)$, then the generating function for the number of offspring after two generations is given by $h(z) = f(f(z))$. Explain how this follows from the result of Exercise 4.

6 Consider a queueing process such that in each minute either 1 or 0 customers arrive with probabilities p or $q = 1 - p$, respectively. (The number p is called the *arrival rate*.) When a customer starts service she finishes in the next minute with probability r. The number r is called the *service rate*.) Thus when a customer begins being served she will finish being served in j minutes with probability $(1 - r)^{j-1}r$, for $j = 1, 2, 3, \ldots$.

(a) Find the generating function $f(z)$ for the number of customers who arrive in one minute and the generating function $g(z)$ for the length of time that a person spends in service once she begins service.

(b) Consider a *customer branching process* by considering the offspring of a customer to be the customers who arrive while she is being served. Using Exercise 4, show that the generating function for our customer branching process is $h(z) = g(f(z))$.

(c) If we start the branching process with the arrival of the first customer, then the length of time until the branching process dies out will be the *busy period* for the server. Find a condition in terms of the arrival rate and service rate that will assure that the server will ultimately have a time when he is not busy.

7 Let N be the expected total number of offspring in a branching process. Let m be the mean number of offspring of a single parent. Show that

$$N = 1 + \left(\sum p_k \cdot k\right) N = 1 + mN$$

and hence that N is finite if and only if $m < 1$ and in that case $N = 1/(1-m)$.

8 Consider a branching process such that the number of offspring of a parent is j with probability $1/2^{j+1}$ for $j = 0, 1, 2, \ldots$.

(a) Using the results of Example 10.11 show that the probability that there are j offspring in the nth generation is

$$p_j^{(n)} = \begin{cases} \frac{1}{n(n+1)}\left(\frac{n}{n+1}\right)^j, & \text{if } j \geq 1, \\ \frac{n}{n+1}, & \text{if } j = 0. \end{cases}$$

(b) Show that the probability that the process dies out exactly at the nth generation is $1/n(n+1)$.

(c) Show that the expected lifetime is infinite even though $d = 1$.

10.3 Generating Functions for Continuous Densities

In the previous section, we introduced the concepts of moments and moment generating functions for discrete random variables. These concepts have natural analogues for continuous random variables, provided some care is taken in arguments involving convergence.

Moments

If X is a continuous random variable defined on the probability space Ω, with density function f_X, then we define the nth moment of X by the formula

$$\mu_n = E(X^n) = \int_{-\infty}^{+\infty} x^n f_X(x)\, dx \ ,$$

provided the integral

$$\mu_n = E(X^n) = \int_{-\infty}^{+\infty} |x|^n f_X(x)\, dx\ ,$$

is finite. Then, just as in the discrete case, we see that $\mu_0 = 1$, $\mu_1 = \mu$, and $\mu_2 - \mu_1^2 = \sigma^2$.

Moment Generating Functions

Now we define the *moment generating function* $g(t)$ for X by the formula

$$\begin{aligned} g(t) &= \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} = \sum_{k=0}^{\infty} \frac{E(X^k) t^k}{k!} \\ &= E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx\ , \end{aligned}$$

provided this series converges. Then, as before, we have

$$\mu_n = g^{(n)}(0)\ .$$

Examples

Example 10.15 Let X be a continuous random variable with range $[0, 1]$ and density function $f_X(x) = 1$ for $0 \le x \le 1$ (uniform density). Then

$$\mu_n = \int_0^1 x^n\, dx = \frac{1}{n+1}\ ,$$

and

$$\begin{aligned} g(t) &= \sum_{k=0}^{\infty} \frac{t^k}{(k+1)!} \\ &= \frac{e^t - 1}{t}\ . \end{aligned}$$

Here the series converges for all t. Alternatively, we have

$$\begin{aligned} g(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x)\, dx \\ &= \int_0^1 e^{tx}\, dx = \frac{e^t - 1}{t}\ . \end{aligned}$$

Then (by L'Hôpital's rule)

$$\begin{aligned} \mu_0 &= g(0) = \lim_{t\to 0} \frac{e^t - 1}{t} = 1\ , \\ \mu_1 &= g'(0) = \lim_{t\to 0} \frac{te^t - e^t + 1}{t^2} = \frac{1}{2}\ , \\ \mu_2 &= g''(0) = \lim_{t\to 0} \frac{t^3 e^t - 2t^2 e^t + 2te^t - 2t}{t^4} = \frac{1}{3}\ . \end{aligned}$$

In particular, we verify that $\mu = g'(0) = 1/2$ and

$$\sigma^2 = g''(0) - (g'(0))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

as before (see Example 6.25). $\square$

Example 10.16 Let X have range $[0, \infty)$ and density function $f_X(x) = \lambda e^{-\lambda x}$ (exponential density with parameter λ). In this case

$$\begin{aligned} \mu_n &= \int_0^\infty x^n \lambda e^{-\lambda x}\, dx = \lambda(-1)^n \frac{d^n}{d\lambda^n} \int_0^\infty e^{-\lambda x}\, dx \\ &= \lambda(-1)^n \frac{d^n}{d\lambda^n}[\frac{1}{\lambda}] = \frac{n!}{\lambda^n}\ , \end{aligned}$$

and

$$\begin{aligned} g(t) &= \sum_{k=0}^\infty \frac{\mu_k t^k}{k!} \\ &= \sum_{k=0}^\infty [\frac{t}{\lambda}]^k = \frac{\lambda}{\lambda - t}\ . \end{aligned}$$

Here the series converges only for $|t| < \lambda$. Alternatively, we have

$$\begin{aligned} g(t) &= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx \\ &= \left.\frac{\lambda e^{(t-\lambda)x}}{t - \lambda}\right|_0^\infty = \frac{\lambda}{\lambda - t}\ . \end{aligned}$$

Now we can verify directly that

$$\mu_n = g^{(n)}(0) = \left.\frac{\lambda n!}{(\lambda - t)^{n+1}}\right|_{t=0} = \frac{n!}{\lambda^n}\ .$$

$\square$

Example 10.17 Let X have range $(-\infty, +\infty)$ and density function

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

(normal density). In this case we have

$$\begin{aligned} \mu_n &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2}\, dx \\ &= \begin{cases} \frac{(2m)!}{2^m m!}, & \text{if } n = 2m, \\ 0, & \text{if } n = 2m+1. \end{cases} \end{aligned}$$

(These moments are calculated by integrating once by parts to show that $\mu_n = (n-1)\mu_{n-2}$, and observing that $\mu_0 = 1$ and $\mu_1 = 0$.) Hence,

$$\begin{aligned} g(t) &= \sum_{n=0}^{\infty} \frac{\mu_n t^n}{n!} \\ &= \sum_{m=0}^{\infty} \frac{t^{2m}}{2^m m!} = e^{t^2/2} \ . \end{aligned}$$

This series converges for all values of t. Again we can verify that $g^{(n)}(0) = \mu_n$.

Let X be a normal random variable with parameters μ and σ. It is easy to show that the moment generating function of X is given by

$$e^{t\mu + (\sigma^2/2)t^2} \ .$$

Now suppose that X and Y are two independent normal random variables with parameters μ_1, σ_1, and μ_2, σ_2, respectively. Then, the product of the moment generating functions of X and Y is

$$e^{t(\mu_1+\mu_2) + ((\sigma_1^2+\sigma_2^2)/2)t^2} \ .$$

This is the moment generating function for a normal random variable with mean $\mu_1 + \mu_2$ and variance $\sigma_1^2 + \sigma_2^2$. Thus, the sum of two independent normal random variables is again normal. (This was proved for the special case that both summands are standard normal in Example 7.5.) □

In general, the series defining $g(t)$ will not converge for all t. But in the important special case where X is bounded (i.e., where the range of X is contained in a finite interval), we can show that the series does converge for all t.

Theorem 10.3 Suppose X is a continuous random variable with range contained in the interval $[-M, M]$. Then the series

$$g(t) = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!}$$

converges for all t to an infinitely differentiable function $g(t)$, and $g^{(n)}(0) = \mu_n$.

Proof. We have

$$\mu_k = \int_{-M}^{+M} x^k f_X(x)\, dx \ ,$$

so

$$\begin{aligned} |\mu_k| &\leq \int_{-M}^{+M} |x|^k f_X(x)\, dx \\ &\leq M^k \int_{-M}^{+M} f_X(x)\, dx = M^k \ . \end{aligned}$$

Hence, for all N we have

$$\sum_{k=0}^{N} \left| \frac{\mu_k t^k}{k!} \right| \leq \sum_{k=0}^{N} \frac{(M|t|)^k}{k!} \leq e^{M|t|} \ ,$$

which shows that the power series converges for all t. We know that the sum of a convergent power series is always differentiable. $\square$

Moment Problem

Theorem 10.4 If X is a bounded random variable, then the moment generating function $g_X(t)$ of x determines the density function $f_X(x)$ uniquely.

Sketch of the Proof. We know that

$$\begin{aligned} g_X(t) &= \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!} \\ &= \int_{-\infty}^{+\infty} e^{tx} f(x)\, dx \ . \end{aligned}$$

If we replace t by $i\tau$, where τ is real and $i = \sqrt{-1}$, then the series converges for all τ, and we can define the function

$$k_X(\tau) = g_X(i\tau) = \int_{-\infty}^{+\infty} e^{i\tau x} f_X(x)\, dx \ .$$

The function $k_X(\tau)$ is called the *characteristic function* of X, and is defined by the above equation even when the series for g_X does not converge. This equation says that k_X is the *Fourier transform* of f_X. It is known that the Fourier transform has an inverse, given by the formula

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-i\tau x} k_X(\tau)\, d\tau \ ,$$

suitably interpreted.[9] Here we see that the characteristic function k_X, and hence the moment generating function g_X, determines the density function f_X uniquely under our hypotheses. $\square$

Sketch of the Proof of the Central Limit Theorem

With the above result in mind, we can now sketch a proof of the Central Limit Theorem for bounded continuous random variables (see Theorem 9.6). To this end, let X be a continuous random variable with density function f_X, mean $\mu = 0$ and variance $\sigma^2 = 1$, and moment generating function $g(t)$ defined by its series for all t.

[9] H. Dym and H. P. McKean, *Fourier Series and Integrals* (New York: Academic Press, 1972).

Let X_1, X_2, …, X_n be an independent trials process with each X_i having density f_X, and let $S_n = X_1 + X_2 + \cdots + X_n$, and $S_n^* = (S_n - n\mu)/\sqrt{n\sigma^2} = S_n/\sqrt{n}$. Then each X_i has moment generating function $g(t)$, and since the X_i are independent, the sum S_n, just as in the discrete case (see Section 10.1), has moment generating function

$$g_n(t) = (g(t))^n \ ,$$

and the standardized sum S_n^* has moment generating function

$$g_n^*(t) = \left(g\left(\frac{t}{\sqrt{n}}\right)\right)^n \ .$$

We now show that, as $n \to \infty$, $g_n^*(t) \to e^{t^2/2}$, where $e^{t^2/2}$ is the moment generating function of the normal density $n(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ (see Example 10.17).

To show this, we set $u(t) = \log g(t)$, and

$$\begin{aligned} u_n^*(t) &= \log g_n^*(t) \\ &= n \log g\left(\frac{t}{\sqrt{n}}\right) = nu\left(\frac{t}{\sqrt{n}}\right) \ , \end{aligned}$$

and show that $u_n^*(t) \to t^2/2$ as $n \to \infty$. First we note that

$$\begin{aligned} u(0) &= \log g_n(0) = 0 \ , \\ u'(0) &= \frac{g'(0)}{g(0)} = \frac{\mu_1}{1} = 0 \ , \\ u''(0) &= \frac{g''(0)g(0) - (g'(0))^2}{(g(0))^2} \\ &= \frac{\mu_2 - \mu_1^2}{1} = \sigma^2 = 1 \ . \end{aligned}$$

Now by using L'Hôpital's rule twice, we get

$$\begin{aligned} \lim_{n\to\infty} u_n^*(t) &= \lim_{s\to\infty} \frac{u(t/\sqrt{s})}{s^{-1}} \\ &= \lim_{s\to\infty} \frac{u'(t/\sqrt{s})t}{2s^{-1/2}} \\ &= \lim_{s\to\infty} u''\left(\frac{t}{\sqrt{s}}\right)\frac{t^2}{2} = \sigma^2\frac{t^2}{2} = \frac{t^2}{2} \ . \end{aligned}$$

Hence, $g_n^*(t) \to e^{t^2/2}$ as $n \to \infty$. Now to complete the proof of the Central Limit Theorem, we must show that if $g_n^*(t) \to e^{t^2/2}$, then under our hypotheses the distribution functions $F_n^*(x)$ of the S_n^* must converge to the distribution function $F_N^*(x)$ of the normal variable N; that is, that

$$F_n^*(a) = P(S_n^* \le a) \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-x^2/2}\,dx \ ,$$

and furthermore, that the density functions $f_n^*(x)$ of the S_n^* must converge to the density function for N; that is, that

$$f_n^*(x) \to \frac{1}{\sqrt{2\pi}}e^{-x^2/2} \ ,$$

as $n \to \infty$.

Since the densities, and hence the distributions, of the S_n^* are uniquely determined by their moment generating functions under our hypotheses, these conclusions are certainly plausible, but their proofs involve a detailed examination of characteristic functions and Fourier transforms, and we shall not attempt them here.

In the same way, we can prove the Central Limit Theorem for bounded discrete random variables with integer values (see Theorem 9.4). Let X be a discrete random variable with density function $p(j)$, mean $\mu = 0$, variance $\sigma^2 = 1$, and moment generating function $g(t)$, and let $X_1, X_2, \ldots, X_n$ form an independent trials process with common density p. Let $S_n = X_1 + X_2 + \cdots + X_n$ and $S_n^* = S_n/\sqrt{n}$, with densities p_n and p_n^*, and moment generating functions $g_n(t)$ and $g_n^*(t) = \left(g(\frac{t}{\sqrt{n}})\right)^n$. Then we have

$$g_n^*(t) \to e^{t^2/2} \ ,$$

just as in the continuous case, and this implies in the same way that the distribution functions $F_n^*(x)$ converge to the normal distribution; that is, that

$$F_n^*(a) = P(S_n^* \le a) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-x^2/2}\, dx \ ,$$

as $n \to \infty$.

The corresponding statement about the distribution functions p_n^*, however, requires a little extra care (see Theorem 9.3). The trouble arises because the distribution $p(x)$ is not defined for all x, but only for integer x. It follows that the distribution $p_n^*(x)$ is defined only for x of the form $j/\sqrt{n}$, and these values change as n changes.

We can fix this, however, by introducing the function $\bar{p}(x)$, defined by the formula

$$\bar{p}(x) = \begin{cases} p(j), & \text{if } j - 1/2 \le x < j + 1/2, \\ 0 \ , & \text{otherwise.} \end{cases}$$

Then $\bar{p}(x)$ is defined for all x, $\bar{p}(j) = p(j)$, and the graph of $\bar{p}(x)$ is the step function for the distribution $p(j)$ (see Figure 3 of Section 9.1).

In the same way we introduce the step function $\bar{p}_n(x)$ and $\bar{p}_n^*(x)$ associated with the distributions p_n and p_n^*, and their moment generating functions $\bar{g}_n(t)$ and $\bar{g}_n^*(t)$. If we can show that $\bar{g}_n^*(t) \to e^{t^2/2}$, then we can conclude that

$$\bar{p}_n^*(x) \to \frac{1}{\sqrt{2\pi}} e^{t^2/2} \ ,$$

as $n \to \infty$, for all x, a conclusion strongly suggested by Figure 9.3.

Now $\bar{g}(t)$ is given by

$$\begin{aligned} \bar{g}(t) &= \int_{-\infty}^{+\infty} e^{tx} \bar{p}(x)\, dx \\ &= \sum_{j=-N}^{+N} \int_{j-1/2}^{j+1/2} e^{tx} p(j)\, dx \end{aligned}$$

$$\begin{aligned} &= \sum_{j=-N}^{+N} p(j)e^{tj}\frac{e^{t/2}-e^{-t/2}}{2t/2} \\ &= g(t)\frac{\sinh(t/2)}{t/2} , \end{aligned}$$

where we have put

$$\sinh(t/2) = \frac{e^{t/2}-e^{-t/2}}{2} .$$

In the same way, we find that

$$\begin{aligned} \bar{g}_n(t) &= g_n(t)\frac{\sinh(t/2)}{t/2} , \\ \bar{g}_n^*(t) &= g_n^*(t)\frac{\sinh(t/2\sqrt{n})}{t/2\sqrt{n}} . \end{aligned}$$

Now, as $n \to \infty$, we know that $g_n^*(t) \to e^{t^2/2}$, and, by L'Hôpital's rule,

$$\lim_{n\to\infty} \frac{\sinh(t/2\sqrt{n})}{t/2\sqrt{n}} = 1 .$$

It follows that

$$\bar{g}_n^*(t) \to e^{t^2/2} ,$$

and hence that

$$\bar{p}_n^*(x) \to \frac{1}{\sqrt{2\pi}}e^{-x^2/2} ,$$

as $n \to \infty$. The astute reader will note that in this sketch of the proof of Theorem 9.3, we never made use of the hypothesis that the greatest common divisor of the differences of all the values that the X_i can take on is 1. This is a technical point that we choose to ignore. A complete proof may be found in Gnedenko and Kolmogorov.[10]

Cauchy Density

The characteristic function of a continuous density is a useful tool even in cases when the moment series does not converge, or even in cases when the moments themselves are not finite. As an example, consider the Cauchy density with parameter $a = 1$ (see Example 5.10)

$$f(x) = \frac{1}{\pi(1+x^2)} .$$

If X and Y are independent random variables with Cauchy density $f(x)$, then the average $Z = (X+Y)/2$ also has Cauchy density $f(x)$, that is,

$$f_Z(x) = f(x) .$$

[10] B. V. Gnedenko and A. N. Kolomogorov, *Limit Distributions for Sums of Independent Random Variables* (Reading: Addison-Wesley, 1968), p. 233.

This is hard to check directly, but easy to check by using characteristic functions. Note first that

$$\mu_2 = E(X^2) = \int_{-\infty}^{+\infty} \frac{x^2}{\pi(1+x^2)}\, dx = \infty$$

so that μ_2 is infinite. Nevertheless, we can define the characteristic function $k_X(\tau)$ of x by the formula

$$k_X(\tau) = \int_{-\infty}^{+\infty} e^{i\tau x} \frac{1}{\pi(1+x^2)}\, dx\ .$$

This integral is easy to do by contour methods, and gives us

$$k_X(\tau) = k_Y(\tau) = e^{-|\tau|}\ .$$

Hence,

$$k_{X+Y}(\tau) = (e^{-|\tau|})^2 = e^{-2|\tau|}\ ,$$

and since

$$k_Z(\tau) = k_{X+Y}(\tau/2)\ ,$$

we have

$$k_Z(\tau) = e^{-2|\tau/2|} = e^{-|\tau|}\ .$$

This shows that $k_Z = k_X = k_Y$, and leads to the conclusions that $f_Z = f_X = f_Y$.

It follows from this that if X_1, X_2, ..., X_n is an independent trials process with common Cauchy density, and if

$$A_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is the average of the X_i, then A_n has the same density as do the X_i. This means that the Law of Large Numbers fails for this process; the distribution of the average A_n is exactly the same as for the individual terms. Our proof of the Law of Large Numbers fails in this case because the variance of X_i is not finite.

Exercises

1 Let X be a continuous random variable with values in $[0, 2]$ and density f_X. Find the moment generating function $g(t)$ for X if

(a) $f_X(x) = 1/2$.

(b) $f_X(x) = (1/2)x$.

(c) $f_X(x) = 1 - (1/2)x$.

(d) $f_X(x) = |1 - x|$.

(e) $f_X(x) = (3/8)x^2$.

Hint: Use the integral definition, as in Examples 10.15 and 10.16.

2 For each of the densities in Exercise 1 calculate the first and second moments, μ_1 and μ_2, directly from their definition and verify that $g(0) = 1$, $g'(0) = \mu_1$, and $g''(0) = \mu_2$.

3 Let X be a continuous random variable with values in $[0, \infty)$ and density f_X. Find the moment generating functions for X if

(a) $f_X(x) = 2e^{-2x}$.

(b) $f_X(x) = e^{-2x} + (1/2)e^{-x}$.

(c) $f_X(x) = 4xe^{-2x}$.

(d) $f_X(x) = \lambda(\lambda x)^{n-1}e^{-\lambda x}/(n-1)!$.

4 For each of the densities in Exercise 3, calculate the first and second moments, μ_1 and μ_2, directly from their definition and verify that $g(0) = 1$, $g'(0) = \mu_1$, and $g''(0) = \mu_2$.

5 Find the characteristic function $k_X(\tau)$ for each of the random variables X of Exercise 1.

6 Let X be a continuous random variable whose characteristic function $k_X(\tau)$ is

$$k_X(\tau) = e^{-|\tau|}, \qquad -\infty < \tau < +\infty \ .$$

Show directly that the density f_X of X is

$$f_X(x) = \frac{1}{\pi(1 + x^2)} \ .$$

7 Let X be a continuous random variable with values in $[0, 1]$, uniform density function $f_X(x) \equiv 1$ and moment generating function $g(t) = (e^t - 1)/t$. Find in terms of $g(t)$ the moment generating function for

(a) $-X$.

(b) $1 + X$.

(c) $3X$.

(d) $aX + b$.

8 Let X_1, X_2, ..., X_n be an independent trials process with uniform density. Find the moment generating function for

(a) X_1.

(b) $S_2 = X_1 + X_2$.

(c) $S_n = X_1 + X_2 + \cdots + X_n$.

(d) $A_n = S_n/n$.

(e) $S_n^* = (S_n - n\mu)/\sqrt{n\sigma^2}$.

9 Let X_1, X_2, ..., X_n be an independent trials process with normal density of mean 1 and variance 2. Find the moment generating function for

(a) X_1.

(b) $S_2 = X_1 + X_2$.

(c) $S_n = X_1 + X_2 + \cdots + X_n$.

(d) $A_n = S_n/n$.

(e) $S_n^* = (S_n - n\mu)/\sqrt{n\sigma^2}$.

10 Let X_1, X_2, ..., X_n be an independent trials process with density

$$f(x) = \frac{1}{2}e^{-|x|}, \qquad -\infty < x < +\infty\ .$$

(a) Find the mean and variance of $f(x)$.

(b) Find the moment generating function for X_1, S_n, A_n, and S_n^*.

(c) What can you say about the moment generating function of S_n^* as $n \to \infty$?

(d) What can you say about the moment generating function of A_n as $n \to \infty$?