not use the notation

$$\int_a^b h(x)\,\mathrm{d}F$$

unless there is no danger of ambiguity.

5.7 Characteristic functions

Probability generating functions proved to be very useful in handling non-negative integral random variables. For more general variables X it is natural to make the substitution $s = e^t$ in the quantity $G_X(s) = \mathbf{E}(s^X)$.

(1) **Definition.** The **moment generating function** of a variable X is a function $M: \mathbb{R} \to [0, \infty)$ given by

$$M(t) = \mathbf{E}(e^{tX}).$$

Moment generating functions are related to Laplace transforms, since

$$M(t) = \int e^{tx}\,\mathrm{d}F(x) = \int e^{tx} f(x)\,\mathrm{d}x$$

if X is continuous with density function f. They have properties similar to those of probability generating functions. For example, if $M(t) < \infty$ on some open interval containing the origin then

(a) $\mathbf{E}X = M'(0)$, $\mathbf{E}(X^k) = M^{(k)}(0)$,
(b) (*Taylor's theorem*)

$$M(t) = \sum_k \frac{\mathbf{E}(X^k)}{k!} t^k;$$

that is, M is the 'exponential generating function' of the sequence of moments of X.
(c) (*Laplace convolution theorem*) If X and Y are independent then

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

This is essentially the assertion that the Laplace transform of a convolution (see (4.8.2)) is the product of the Laplace transforms.

Moment generating functions provide a very useful technique but suffer the disadvantage that the integrals which define them may not always be finite. Rather than explore their properties in detail we move on immediately to another class of functions which are equally useful and whose finiteness is guaranteed.

(2) **Definition.** The **characteristic function** of X is the function $\phi: \mathbb{R} \to \mathbb{C}$ defined by

$$\phi(t) = \mathbf{E}(e^{itX}) \quad \text{where} \quad i = \sqrt{(-1)}.$$

Characteristic functions are related to Fourier transforms, since

$$\phi(t) = \int e^{itx}\,\mathrm{d}F(x).$$

In the notation of Section 5.6, ϕ is the abstract integral of a complex-valued random variable. It is well defined in the terms of Section 5.6 by

$$\phi(t) = \mathbf{E}(\cos tX) + i\mathbf{E}(\sin tX).$$

Furthermore, $\phi(t)$ is better behaved than $M(t)$.

(3) **Theorem.** *The characteristic function ϕ satisfies*

(a) $\phi(0) = 1$, $|\phi(t)| \leq 1$ *for all* t
(b) ϕ *is uniformly continuous on* $\mathbb{R}$
(c) ϕ *is non-negative definite, which is to say that*

$$\sum_{j,k} \phi(t_j - t_k) z_j \bar{z}_k \geq 0$$

for all real $t_1, \ldots, t_n$ *and complex* $z_1, \ldots, z_n$.

Proof.

(a) $\phi(0) = \mathbf{E}(1) = 1$.

$$|\phi(t)| \leq \int |e^{itx}|\,\mathrm{d}F = \int \mathrm{d}F = 1.$$

(b)
$$|\phi(t+h) - \phi(t)| = |\mathbf{E}(e^{i(t+h)X} - e^{itX})| \leq \mathbf{E}\,|e^{itX}(e^{ihX} - 1)| = \mathbf{E}(Y(h))$$

where $Y(h) = |e^{ihX} - 1|$. However, $|Y(h)| \leq 2$ and $Y(h) \to 0$ as $h \to 0$ and so $\mathbf{E}(Y(h)) \to 0$ (by bounded convergence (5.6.12)).

(c)
$$\sum_{j,k} \phi(t_j - t_k) z_j \bar{z}_k = \sum_{j,k} \int \{z_j \exp(it_j x)\}\{\bar{z}_k \exp(-it_k x)\}\,\mathrm{d}F = \mathbf{E}\left\{\left|\sum_j z_j \exp(it_j X)\right|^2\right\} \geq 0.$$

■

Actually, (3) characterizes characteristic functions in the sense that ϕ is a characteristic function if and only if it satisfies (a), (b), and (c). This is Bochner's theorem, for which we offer no proof. Many of the properties of characteristic functions rely for their proofs on a knowledge of complex analysis. This is a textbook on probability theory, and will not include such proofs unless they indicate some essential technique. We have asserted that the method of characteristic functions is very useful; however, we warn the reader that we shall not make use of them until Section 5.10. In the meantime we shall establish some of their properties.

First and foremost, from a knowledge of ϕ_X we can recapture the distribution of X. The full power of this statement is deferred until the next section; here we concern ourselves only with the moments of X.

Many of the interesting characteristic functions are not very well behaved, and we must move carefully.

(4) **Theorem.**

(a) *If* $\phi^{(k)}(0)$ *exists then* $\begin{cases} \mathbf{E}\,|X^k|<\infty & \text{if } k \text{ is even} \\ \mathbf{E}\,|X^{k-1}|<\infty & \text{if } k \text{ is odd.} \end{cases}$

(b) *If* $\mathbf{E}\,|X^k|<\infty$ *then*†

$$\phi(t)=\sum_{j=0}^{k}\frac{\mathbf{E}(X^j)}{j!}(it)^j+o(t^k),$$

and so $\phi^{(k)}(0)=i^k\mathbf{E}(X^k)$.

Proof. This is essentially Taylor's theorem for a function of a complex variable. For the proof, see Moran 1968 and Kingman and Taylor 1966. ■

One of the useful properties of characteristic functions is that they enable us to handle sums of independent variables with the minimum of fuss.

(5) **Theorem.** *If X and Y are independent then*

$$\phi_{X+Y}(t)=\phi_X(t)\phi_Y(t).$$

Proof.

$$\phi_{X+Y}(t)=\mathbf{E}(e^{it(X+Y)})=\mathbf{E}(e^{itX}e^{itY}).$$

Expand each exponential term into cosines and sines, multiply out, use independence, and put back together to obtain the result. ■

(6) **Theorem.** *If* $a, b\in\mathbb{R}$ *and* $Y=aX+b$ *then*

$$\phi_Y(t)=e^{itb}\phi_X(at).$$

Proof.

$$\phi_Y(t)=\mathbf{E}(e^{it(aX+b)})=\mathbf{E}(e^{itb}e^{i(at)X})$$
$$=e^{itb}\mathbf{E}(e^{i(at)X})=e^{itb}\phi_X(at).$$ ■

We shall make repeated use of these last two theorems. We sometimes need to study collections of variables which may be dependent.

(7) **Definition.** The **joint characteristic function** of X and Y is the function $\phi:\mathbb{R}^2\to\mathbb{R}$ given by

$$\phi(s,t)=\mathbf{E}(e^{isX}e^{itY}).$$

† See Subsection (10) of Appendix I for a reminder about Landau's O–o notation

Notice that $\phi(s,t)=\phi_{sX+tY}(1)$. As usual we shall be interested mostly in independent variables.

(8) **Theorem.** *X and Y are independent if and only if*

$$\phi_{X,Y}(s,t)=\phi_X(s)\phi_Y(t)\quad \text{for all } s \text{ and } t.$$

Proof. If X and Y are independent then the result follows by the argument of (5). The converse is proved by extending the inversion theorem of the next section to deal with joint distributions and showing that the joint distribution function factorizes. ■

Note particularly that for X and Y to be independent it is not sufficient that

(9) $$\phi_{X,Y}(t,t)=\phi_X(t)\phi_Y(t),\qquad \text{for all } t.$$

Exercise. Can you find an example of dependent variables which satisfy (9)?

We have seen (4) that it is an easy calculation to find the moments of X by differentiating its characteristic function $\phi_X(t)$ at $t=0$. A similar calculation gives the 'joint moments' $\mathbf{E}(X^jY^k)$ of two variables from a knowledge of their joint characteristic function $\phi_{X,Y}(s,t)$ (see Problem (5.11.23) for details).

The properties of moment generating functions are closely related to those of characteristic functions. In the rest of the text we shall use the latter whenever possible, but it will be appropriate to use the former for any topic whose analysis employs Laplace transforms; for example this is the case for the queueing theory of Chapter 11.

5.8 Examples of characteristic functions

Those who feel daunted by $\sqrt{(-1)}$ should find it a useful exercise to work through this section using $M(t)=\mathbf{E}(e^{tX})$ in place of $\phi(t)=\mathbf{E}(e^{itX})$. Many calculations here are left as *exercises*.

(1) **Example. Bernoulli distribution.** If X is Bernoulli parameter p then

$$\phi(t)=\mathbf{E}(e^{itX})=e^{it0}\,.\,q+e^{it1}\,.\,p=q+pe^{it}.$$ ●

(2) **Example. Binomial distribution.** If X is $B(n,p)$ then X has the same distribution as the sum of n independent Bernoulli variables $Y_1, Y_2, \ldots, Y_n$. Thus

$$\phi_X(t)=\phi_{Y_1}(t)\ldots\phi_{Y_n}(t)=(q+pe^{it})^n.$$ ●

(3) **Example. Exponential distribution.** If $f(x)=\lambda e^{-\lambda x}$ for $x\geqslant 0$ then

$$\phi(t)=\int_0^\infty e^{itx}\lambda e^{-\lambda x}\,dx.$$

This is a complex integral and its solution relies on a knowledge of how to integrate around contours in $\mathbb{R}^2$ (the appropriate contour is a sector). If you know about this then do it. Do not fall into the trap of treating i as if it were a real number, even though this malpractice yields the correct answer in this case:

$$\phi(t) = \frac{\lambda}{\lambda - it}.$$ ●

(4) **Example. Cauchy distribution.** If $f(x) = \{\pi(1+x^2)\}^{-1}$ then

$$\phi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{e^{itx}}{1+x^2}\,dx.$$

Treating i as a real number will not help you to avoid the contour integral this time. The answer is

$$\phi(t) = e^{-|t|}.$$

Those who are interested should try integrating around a semicircle with diameter $[-R, R]$ on the real axis. ●

(5) **Example. Normal distribution.** If X is $N(0, 1)$ then

$$\phi(t) = \mathbf{E}(e^{itX}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)}} \exp(itx - \tfrac{1}{2}x^2)\,dx.$$

Again, do not treat i as a real number. Consider instead the moment generating function of X

$$M(s) = \mathbf{E}(e^{sX}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{(2\pi)}} \exp(sx - \tfrac{1}{2}x^2)\,dx.$$

Complete the square in the integrand and use the hint at the end of Example (4.5.9) to obtain

$$M(s) = e^{s^2/2}.$$

We may not substitute $s = it$ without justification. In this particular instance the theory of analytic continuation of functions of a complex variable provides this justification, and we deduce that

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

By (5.7.6), the characteristic function of the $N(\mu, \sigma^2)$ variable $Y = \sigma X + \mu$ is

$$\phi_Y(t) = e^{it\mu}\phi_X(\sigma t) = \exp(i\mu t - \tfrac{1}{2}\sigma^2 t^2).$$ ●

(6) **Example. Multivariate normal distribution.** If $X_1, \ldots, X_n$ has the multivariate normal distribution $N(\mathbf{0}, \mathbf{V})$ then its joint density function is

$$f(\boldsymbol{x}) = \{(2\pi)^n\,|\mathbf{V}|\}^{-1/2} \exp(-\tfrac{1}{2}\boldsymbol{x}\mathbf{V}^{-1}\boldsymbol{x}').$$

The joint characteristic function of $X_1, \ldots, X_n$ is

$$\phi(\boldsymbol{t}) = \mathbf{E}(e^{i\boldsymbol{t}\mathbf{X}'})$$

where $\boldsymbol{t} = (t_1, \ldots, t_n)$ and $\mathbf{X} = (X_1, \ldots, X_n)$. Thus

(7) $$\phi(\boldsymbol{t}) = \int_{\mathbb{R}^n} \{(2\pi)^n\,|\mathbf{V}|\}^{-1/2} \exp(i\boldsymbol{t}\boldsymbol{x}' - \tfrac{1}{2}\boldsymbol{x}\mathbf{V}^{-1}\boldsymbol{x}')\,d\boldsymbol{x}.$$

As in the discussion of (4.7.9), there is a linear transformation $\boldsymbol{y} = \boldsymbol{x}\boldsymbol{B}$ such that

$$\boldsymbol{x}\mathbf{V}^{-1}\boldsymbol{x}' = \sum_j \lambda_j y_j^2$$

just as in (4.7.12). Make this transformation in (7) to see that the integrand factorizes into the product of functions of the single variables $y_1, y_2, \ldots, y_n$. Then use (5) to obtain

$$\phi(\boldsymbol{t}) = \exp(-\tfrac{1}{2}\boldsymbol{t}\mathbf{V}\boldsymbol{t}').$$

It is now an easy *exercise* to prove Theorem (4.7.14), that $\mathbf{V}$ is the covariance matrix of $\mathbf{X}$, by using the result of Problem (5.11.23). ●

(8) **Example. Gamma distribution.** If X is $\Gamma(\lambda, s)$ then

$$\phi(t) = \int_0^{\infty} \frac{1}{\Gamma(s)} \lambda^s x^{s-1} \exp(itx - \lambda x)\,dx.$$

As for the exponential distribution (3), routine methods of complex analysis give

$$\phi(t) = \left(\frac{\lambda}{\lambda - it}\right)^s.$$

Why is this similar to the result of (3)? This example includes the chi-squared distribution because a $\chi^2(d)$ variable is $\Gamma(\frac{1}{2}, \frac{1}{2}d)$ and thus has characteristic function

$$\phi(t) = (1 - 2it)^{-d/2}.$$

You may try to prove this from the result of Problem (4.10.10). ●

5.9 Inversion and continuity theorems

This section contains two major reasons why characteristic functions are useful. The first of these states that the distribution of a random variable is specified by its characteristic function. That is to say, if X and Y have the same characteristic function then they have the same distribution. Furthermore, there is a formula which tells us how to recapture the distribution function F corresponding to the characteristic function ϕ. Here is a special case first.

(1) **Theorem.** *If X is continuous with density function f and characteristic function ϕ then*

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\,dt$$

at every point x at which f is differentiable.

Proof. This is the Fourier inversion theorem and can be found in any introduction to Fourier transforms. If the integral fails to converge absolutely then we interpret it as its principal value (see Apostol 1957, p. 487). ■

A sufficient, but not necessary, condition that a characteristic function ϕ be the characteristic function of a continuous variable is that

$$\int_{-\infty}^{\infty} |\phi(t)|\,dt<\infty.$$

The general case is more complicated, and is contained in the next theorem.

(2) **Inversion Theorem.** *Let X have distribution function F and characteristic function ϕ. Define $\bar{F}:\mathbb{R}\to[0,1]$ by*

$$\bar{F}(x)=\tfrac{1}{2}\{F(x)+\lim_{y\uparrow x}F(y)\}.$$

Then

$$\bar{F}(b)-\bar{F}(a)=\lim_{N\to\infty}\int_{-N}^{N}\frac{e^{-iat}-e^{-ibt}}{2\pi it}\phi(t)\,dt.$$

Proof. See Kingman and Taylor 1966. ■

(3) **Corollary.** *X and Y have the same characteristic function if and only if they have the same distribution function.*

Proof. If $\phi_X=\phi_Y$ then, by (2),

$$\bar{F}_X(b)-\bar{F}_X(a)=\bar{F}_Y(b)-\bar{F}_Y(a).$$

Let $a\to-\infty$ to obtain

$$\bar{F}_X(b)=\bar{F}_Y(b);$$

now, for any fixed $x\in\mathbb{R}$, let $b\downarrow x$ and use right-continuity (2.1.6c) to obtain

$$F_X(x)=F_Y(x).$$ ■

Exactly similar results hold for jointly distributed random variables.

For example, if X and Y have joint density function f and joint characteristic function ϕ then whenever f is differentiable at (x, y)

$$f(x,y)=\frac{1}{4\pi^2}\iint_{\mathbb{R}^2} e^{-isx}e^{-ity}\phi(s,t)\,ds\,dt$$

and (5.7.8) follows straight away for this special case.

The second result of this section deals with a sequence $X_1, X_2, \dots$ of random variables. Roughly speaking it asserts that if the distribution functions $F_1, F_2, \dots$ of the sequence approach some limit F then the characteristic functions $\phi_1, \phi_2, \dots$ of the sequence approach the characteristic function of the distribution function F.

(4) **Definition.** We say that the sequence $F_1, F_2, \dots$ of distribution functions **converges** to the distribution function F, written $F_n\to F$, if $F(x)=\lim_{n\to\infty}F_n(x)$ at each point x where F is continuous.

The reason for the condition of continuity of F at x is indicated by the following. Define the distribution functions F_n and G_n by

$$F_n(x)=\begin{cases}0 & \text{if } x<\frac{1}{n}\\ 1 & \text{if } x\geq\frac{1}{n}\end{cases}\qquad G_n(x)=\begin{cases}0 & \text{if } x<-\frac{1}{n}\\ 1 & \text{if } x\geq-\frac{1}{n}.\end{cases}$$

As $n\to\infty$

$$F_n(x)\to F(x)\quad\text{if } x\neq0,\ F_n(0)\to0$$

$$G_n(x)\to F(x)\quad\text{for all } x$$

where $F(x)$ is the distribution function of a random variable which is constantly zero. Indeed $\lim_{n\to\infty}F_n(x)$ is not even a distribution function since it is not right-continuous at zero. It is intuitively reasonable to demand that the sequences $\{F_n\}$ and $\{G_n\}$ have the same limit, and so we drop the requirement that $F_n(x)\to F(x)$ at the point of discontinuity of F.

(5) **Continuity Theorem.** *Suppose that $F_1, F_2, \dots$ is a sequence of distribution functions with corresponding characteristic functions $\phi_1, \phi_2, \dots$.*

(a) *If $F_n\to F$ for some distribution function F with characteristic function ϕ, then $\phi_n(t)\to\phi(t)$ for all t.*
(b) *Conversely, if $\phi(t)=\lim_{n\to\infty}\phi_n(t)$ exists and is continuous at $t=0$, then ϕ is the characteristic function of some distribution function F and $F_n\to F$.*

Proof. As for (2). But see (5.11.30). ■

5.10 Two limit theorems

We are now in a position to prove two very celebrated theorems in probability theory, the 'Law of Large Numbers' and the 'Central Limit Theorem'. The first of these explains the remarks of Sections 1.1 and 1.3, where we discussed a heuristic foundation of probability theory. Part of our intuition about chance is that if we perform many repetitions of an experiment which has numerical outcomes then the average of all the outcomes settles down to some fixed number. This observation deals once again in the convergence of sequences of random variables, the general theory of which is dealt with later. Here it suffices to introduce only one new definition.

(1) **Definition.** If $X_1, X_2, \ldots, X$ is a collection of random variables with distribution functions $F_1, F_2, \ldots, F$, then we say that X_n **converges in distribution** to X, written $X_n \xrightarrow{D} X$, if $F_n \to F$.

This is just (5.9.4) rewritten in terms of random variables.

(2) **Theorem. Law of Large Numbers.** *Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite means μ. Their partial sums*

$$S_n = X_1 + X_2 + \ldots + X_n$$

satisfy

$$\frac{1}{n} S_n \xrightarrow{D} \mu \quad as \quad n \to \infty.$$

Proof. The theorem asserts that

$$\mathbf{P}(S_n \leq x) \to \begin{cases} 0 \text{ if } x < \mu \\ 1 \text{ if } x > \mu \end{cases} \quad \text{as} \quad n \to \infty.$$

The method of proof is clear. By the Continuity Theorem (5.9.5) we need to show that the characteristic function of S_n approaches the characteristic function of the constant random variable μ. This is easy. Let ϕ_X be the common characteristic function of the X's, and let ϕ_n be the characteristic function of $(1/n)S_n$. By (5.7.5) and (5.7.6),

(3) $$\phi_n(t) = \{\phi_X(t/n)\}^n.$$

The behaviour of $\phi_X(t/n)$ for large n is given by (5.7.4):

$$\phi_X(t) = 1 + it\mu + o(t).$$

Substitute into (3) to obtain

$$\phi_n(t) = \left(1 + \frac{i\mu t}{n} + o\left(\frac{t}{n}\right)\right)^n \to e^{i\mu t} \quad \text{as} \quad n \to \infty.$$

However, this limit is the characteristic function of the constant μ and the result is proved. ∎

So, for large n, S_n is about as big as $n\mu$. What can we say about the difference $S_n - n\mu$? There is an extraordinary answer to this question so long as the X's have finite variance:

(a) $S_n - n\mu$ is about as big as $n^{1/2}$,
(b) the distribution of $n^{-1/2}(S_n - n\mu)$ approaches the normal distribution as $n \to \infty$ *irrespective* of the distribution of the X's.

(4) **Central Limit Theorem.** *Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite means μ and finite non-zero variances σ^2, and let*

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

Then

$$\frac{S_n - n\mu}{\sqrt{(n\sigma^2)}} \xrightarrow{D} N(0, 1) \quad as \quad n \to \infty.$$

Note that the assertion of the theorem is an abuse of notation, since $N(0, 1)$ is a distribution and not a random variable; it is admissible because convergence in distribution involves only the corresponding distribution functions. The method of proof is the same as for the Law of Large Numbers.

Proof. First, write $Y_i = (X_i - \mu)/\sigma$, and let ϕ_Y be the characteristic function of the Y's. By (5.7.4)

$$\phi_Y(t) = 1 - \tfrac{1}{2}t^2 + o(t^2).$$

Also, the characteristic function ψ_n of

$$U_n = \frac{S_n - n\mu}{\sqrt{(n\sigma^2)}} = \frac{1}{\sqrt{n}} \sum_1^n Y_i$$

satisfies, by (5.7.5) and (5.7.6),

$$\begin{aligned} \psi_n(t) &= \{\phi_Y(tn^{-1/2})\}^n \\ &= \left\{1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right)\right\}^n \\ &\to e^{-t^2/2} \quad \text{as} \quad n \to \infty. \end{aligned}$$

However, this is the characteristic function of the $N(0, 1)$ distribution and the Continuity Theorem (5.9.5) completes the proof. ∎

Numerous generalizations of the Law of Large Numbers and the Central Limit Theorem are available. For example, in Chapter 7 we shall meet two stronger versions of (2), involving weaker assumptions on the