11 Characteristic Functions and Their Applications

The moment generating function (m.g.f.) of a random variable X is defined as the average of the exponential function:

$$M_X(t) := E\left(e^{tX}\right) = \int_R e^{tx} F(dx) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n).$$

For example, if X is normally distributed with mean zero and variance 1, then

$$M_X(t) = (2\pi)^{-1/2} \int_R e^{tx} e^{-x^2/2}\, dx = (2\pi)^{-1/2} e^{t^2/2} \int_R e^{-(t-x)^2/2}\, dx = e^{t^2/2}.$$

From this, the moments are computed by expanding both sides in powers of t, which yields $E(X^{2n}) = (2n)!/2^n n!$ for even moments and zero for odd moments.

The m.g.f. is a useful computational device, which can be used to tabulate the moments of a large class of probability distributions, both discrete and continuous. However, the m.g.f. is not defined for all random variables, e.g., a Cauchy distribution, where $M_X(t) = (1/\pi)\int_R e^{tx}/\left(1+x^2\right) dx = +\infty$, for $t \neq 0$. Accordingly, we now introduce a universal label for an arbitrary distribution function, known as the *characteristic function* and defined as follows. We recall the complex exponential function, which satisfies $e^{it} = \cos t + i \sin t$, where t is a real number and $i = \sqrt{-1}$.

Exercise 11.1
If X is normally distributed with mean zero and variance 1, then $E(X^{2n}) = (2n)!/2^n n!$ for even moments and zero for odd moments.

11.1 Definition of the Characteristic Function

The characteristic function of a random variable X is defined as

$$\phi_X(t) := E\left(e^{itX}\right) = E(\cos tX) + iE(\sin tX), \qquad -\infty < t < \infty$$

The expectation is finite for all real values of t. In case X is a discrete random variable with discrete density f_X, we can write

$$\phi_X(t) = \sum_x e^{itx} f_X(x), \tag{11.1}$$

An Introduction to Stochastic Modeling. DOI: 10.1016/B978-0-12-381416-6.00011-3

whereas if X is continuous with density f_X, then

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\, dx. \tag{11.2}$$

If all the moments of X are finite, then the characteristic function can be computed by expanding the complex exponential function in its power series. We illustrate with the normal distribution $n\left(\mu, \sigma^2\right)$, for which we know the central moments. In this case,

$$\begin{aligned}
E\left[e^{itX}\right] &= e^{it\mu} E\left[e^{it(X-\mu)}\right] \\
&= e^{it\mu} \sum_{m=0}^{\infty} \frac{(it)^m}{m!} E(X-\mu)^m \\
&= e^{it\mu} \sum_{k=0}^{\infty} \frac{\left(-t^2\right)^k}{(2k)!} E(X-\mu)^{2k} \\
&= e^{it\mu} \sum_{k=0}^{\infty} \frac{\left(-t^2\right)^k}{(2k)!} \frac{\sigma^{2k}(2k)!}{2^k k!} \\
&= e^{it\mu} \sum_{k=0}^{\infty} \frac{\left(-\sigma^2 t^2/2\right)^k}{k!} \\
&= e^{it\mu} e^{-\sigma^2 t^2/2}.
\end{aligned}$$

which shows that the characteristic function has a normal-type dependence in the variable t.

If X is an arbitrary random variable, the characteristic function is a bounded continuous function of t with $|\phi_X(t)| \le 1, \phi_X(0) = 1$. This clearly holds for a normally distributed random variable and is easily proved in general. Additional smoothness properties of the characteristic function depend on the existence of higher moments, which is satisfied by the normal distribution, but not for an arbitrary random variable.

Exercise 11.2
Let X be a real-valued random variable. Show that the characteristic function is continuous in t.

Exercise 11.3
Let X be a real-valued random variable. Show that $|\phi_X(t)| \le 1$ and $\phi_X(0) = 1$.

11.1.1 Two Basic Properties of the Characteristic Function

In general, the characteristic function defines a *homomorphism*, converting sums of independent random variables into products. The precise statement is the following:

Theorem 11.1. *If X_1, X_2 are independent random variables, then*

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t)$$

The proof is a one-liner: if X_1, X_2 are independent, then

$$\phi_{X_1+X_2}(t) = E\left[\mathrm{e}^{it(X_1+X_2)}\right] = E\left[\mathrm{e}^{itX_1}\mathrm{e}^{itX_2}\right] = E\left[\mathrm{e}^{itX_1}\right]E\left[\mathrm{e}^{itX_2}\right] = \phi_{X_1}(t)\phi_{X_2}(t)$$

where we have first used the properties of the exponential function followed by independence in the last step.

The other important property of the characteristic function is that it serves as a label for the distribution function of the random variable. This is formalized as follows.

Theorem 11.2. *If X_1, X_2 are random variables with $\phi_{X_1}(t) = \phi_{X_2}(t)$ for all t, then $F_{X_1}(x) = F_{X_2}(x)$ for all x.*

A theorem of this type can be proved by first proving an inversion formula, where we explicitly display the density/distribution in terms of the characteristic function.

11.2 Inversion Formulas for Characteristic Functions

We first illustrate the proof of Theorem 11.2 in case of discrete random variables, where the characteristic function is written as an infinite series:

$$\phi_X(t) = \sum_{x\in R} \mathrm{e}^{itx} f_X(x).$$

We fix $y \in R$, multiply by the complex exponential e^{-ity}, and average on the interval $-L \le t \le L$ with the result

$$\begin{aligned}
\mathrm{e}^{-ity}\phi_X(t) &= \sum_{x\in R} \mathrm{e}^{it(x-y)} f_X(x) \\
\frac{1}{2L}\int_{-L}^{L} \mathrm{e}^{-ity}\phi_X(t)\,\mathrm{d}t &= \sum_{x\in R}\left(\frac{1}{2L}\int_{-L}^{L}\mathrm{e}^{it(x-y)}\,\mathrm{d}t\right) f_X(x) \\
&= \sum_{x\in R}\frac{\sin L(x-y)}{L(x-y)} f_X(x). \qquad (11.3)
\end{aligned}$$

(If $x = y$, then the integral on the right side has the value $2L$, which agrees with the limiting value of the indicated quotient.) In particular, if X takes only integer values $0, \pm 1, \pm 2, \ldots$ and y is an integer, then we can take $L = \pi$ and note that all of the terms on the right side are zero, except in the case that $x = y$, an integer. From this, we obtain the **inversion formula for integer-valued random variables**:

$$X \in \mathbf{Z} \Longrightarrow f_X(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ity} \phi_X(t)\, dt \qquad y = 0, \pm 1, \pm 2, \ldots \tag{11.4}$$

Example If X has a binomial distribution $B(n, p)$, then

$$\phi_X(t) = \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k q^{n-k} = \left(q + pe^{it}\right)^n \tag{11.5}$$

The inversion formula (11.4) takes the form

$$\binom{n}{y} p^y q^{n-y} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ity} \left(q + pe^{it}\right)^n dt, \qquad y = 0, 1, \ldots, n$$

Formula (11.4) will be used to prove the local limit theorem of de Moivre and Laplace.

Example If X has a Poisson distribution $\mathcal{P}(\lambda)$, then

$$\phi_X(t) = \sum_{k=0}^{\infty} e^{itk} \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(e^{it} - 1)}$$

and the inversion formula (11.4) takes the form

$$\frac{\lambda^y}{y!} e^{-\lambda} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ity} e^{\lambda(e^{it}-1)}\, dt, \qquad y = 0, 1, 2, \ldots \tag{11.6}$$

This will be used to do the proof of Stirling's formula.

In the case of a more general discrete random variable, we can take the limit $L \to \infty$ in (11.3). The terms on the right side are bounded by an absolutely convergent series and tend to zero, save for $x = y$, so that we obtain the **inversion formula for discrete random variables**:

$$X \in \mathbf{D} \Longrightarrow f_X(y) = \lim_{L \to \infty} \frac{1}{2L} \int_{-L}^{L} e^{-ity} \phi_X(t)\, dt, \qquad y \in \mathbf{R} \tag{11.7}$$

where $\mathbf{D}$ is the set of possible values of X, with $\sum_{x \in \mathbf{D}} f_X(x) = 1$. Formula (11.7) shows explicitly that ϕ_X determines f_X and thus F_X, since $F_X(x) = \sum_{z \le x} f_X(x)$. Hence, we have proved Theorem 11.2 in the case of general discrete random variables.

We can make a similar argument in the continuous case, when X has a density:

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx.$$

Again, we multiply by e^{-ity} and integrate over the real line tempered with the factor $e^{-\sigma^2 t^2/2}$— to ensure convergence of the improper integral; explicitly

$$\int_{-\infty}^{\infty} \phi_X(t) e^{-ity} e^{-\sigma^2 t^2/2}\,dt = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{it(x-y)} e^{-\sigma^2 t^2/2}\,dt \right) f_X(x)\,dx. \tag{11.8}$$

The inner integral is $\sqrt{2\pi}/\sigma \times$ the characteristic function of a normal density with mean zero and variance $1/\sigma^2$. In the special case, where ϕ_X is integrable over the real line, we can take the limit $\sigma \to 0$ to obtain the **Fourier inversion formula for integrable characteristic functions**:

$$\int_{-\infty}^{\infty} |\phi(t)|\,dt < \infty \Longrightarrow \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_X(t) e^{-ity}\,dt = f_X(y) \tag{11.9}$$

valid at all continuity points of f_X. More generally, if the limit $\bar{f}_X$ of f_X exists in some averaged sense at $x = y$, then we can take the limit in equation (11.8) to obtain the inversion formula

$$f_X(y+0) = f_X(y-0) \Longrightarrow \frac{1}{2\pi} \lim_{\sigma \to 0} \int_{-\infty}^{\infty} \phi_X(t) e^{-ity} e^{-\sigma^2 t^2/2}\,dt = \bar{f}_X(y) \tag{11.10}$$

If, e.g., f_X has a simple jump at y, then the right side of (11.10) needs to be interpreted as the average of the left and right limits at y.

Example In the case of the bilateral exponential density $f(x) = \frac{1}{2} e^{-|x|}$, the characteristic function is computed directly as $\phi(t) = 1/\left(1+t^2\right)$. This is an integrable function on the real line, so that the inversion formula (11.9) applies, to yield

$$\frac{1}{2} e^{-|y|} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{1+t^2} e^{-ity}\,dt, \qquad t \in \mathbf{R}$$

As a by-product, we can change the roles of y and t to obtain the characteristic function of the Cauchy density $f_X(x) = 1/\pi\left(1+x^2\right)$, namely $\phi_X(t) = e^{-|t|}$. Since this is integrable, we also have the inversion formula (11.9) without any limiting procedure.

Example In the case of the uniform density $f_X(x) = (1/(b-a))1_{[a,b]}(x)$, the characteristic function is $\phi_X(t) = \left(e^{itb} - e^{ita}\right)/(it(b-a))$, which is not integrable over the real line. Hence, we must use the general form (11.10) of the inversion formula.

Example In the case of a triangular density, e.g., $f(x) = 1 - |x|$ for $|x| \le 1$ and zero elsewhere, we may justify the inversion formula (11.9) by noting that f is the convolution of two uniform densities on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, for which $\phi(t) = O(1/t), t \to \infty$. Hence, by Theorem 11.1, the characteristic function of f is $O(1/t^2), t \to \infty$. Hence, we can apply (11.9) to obtain the Fourier inversion formula.

*11.2.1 Fourier Reciprocity/Local Non-Uniqueness**

The previous example can be re-written as a pair of Fourier integrals: Let $\phi(t) = 1 - |t|$ for $|t| < 1$ and zero elsewhere.

$$f(x) := \int_{-\infty}^{\infty} \phi(t) e^{itx}\, dt = \int_{-1}^{1} (1 - |t|) e^{itx}\, dt = 2\frac{1 - \cos x}{x^2} \qquad x \neq 0, f(0) = 1.$$

Applying the inversion formula for integrable characteristic functions, we have

$$\phi(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} 2\frac{1 - \cos x}{x^2}\, dx, \quad t \in \mathbf{R}$$

which shows that $\phi(t)$ is a characteristic function. Now, we *periodize* by defining

$$\Phi(t) := \sum_{k \in Z} \phi(t - 2k\pi)$$

which is a 2π periodic function on the line and which agrees with $\phi(t)$ for $|t| < 1$. Its Fourier coefficients are computed by

$$\begin{aligned}
\frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi(t) e^{-int}\, dt &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{k \in Z} \phi(t - 2k\pi) e^{-int}\, dt \\
&= \frac{1}{2\pi} \sum_{k \in Z} \int_{-(2k-1)\pi}^{(2k+1)\pi} \phi(y) e^{-in(y+2k\pi)}\, dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi(y) e^{-iny}\, dy \\
&= \frac{1}{\pi} \frac{1 - \cos k}{k^2}
\end{aligned}$$

* This section can be omitted without loss of continuity.

leading to the absolutely convergent Fourier series:

$$\Phi(t) = \frac{1}{\pi} \sum_{k \in Z} \frac{1 - \cos k}{k^2} e^{ikt}, \qquad t \in \mathbf{R}.$$

This allows one to define an integer-valued random variable by the distribution

$$p_k = \frac{1 - \cos k}{\pi k^2}, \qquad 0 \neq k \in Z, \quad p_0 = \frac{1}{2\pi}$$

which is clearly nonnegative and sums to $\Phi(0) = 1$. Since this distribution is concentrated on the integers, its characteristic function must be a periodic function, namely $\Phi(t)$. Clearly $\Phi(t) = \phi(t)$ for $|t| < 1$, but the equality fails outside of the interval $[-1, 1]$. In summary,

THERE EXIST TWO DISTINCT CHARACTERISTIC FUNCTIONS WHICH AGREE ON THE INTERVAL $[-1, 1]$.

11.2.2 Fourier Inversion and Parseval's Identity

The ideas used to prove the inversion formula (11.9) can be extended to treat the Fourier transform of an absolutely integrable function ψ, where we define

$$\hat{\psi}(t) = \int_{-\infty}^{\infty} \psi(x) e^{itx}\, dx \tag{11.11}$$

If ψ is the probability density of a random variable X, then $\hat{\psi} = \phi_X$, the characteristic function of X. In the more general case, we can apply the same transformations to $\hat{\psi}$ as above, namely multiply (11.11) by $e^{-ity} e^{-\sigma^2 t^2/2}$ and integrate, to obtain

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{\psi}(t) e^{-ity} e^{-\sigma^2 t^2/2}\, dt &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} e^{it(x-y)} e^{-\sigma^2 t^2/2}\, dt \right) \psi(x)\, dx \\ &= 2\pi \int_{-\infty}^{\infty} \frac{e^{-(x-y)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} \psi(x) dx\end{aligned}$$

If $\hat{\psi}$ is also absolutely integrable, then we can take the limit $\sigma \to 0$ and obtain the Fourier inversion formula

$$\psi(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{\psi}(t) e^{-ity}\, dt. \tag{11.12}$$

Applied to a random variable $y = X(\omega)$ and taking the expectation, we obtain a useful corollary.

Proposition 11.1 (Parseval's identity). Suppose that $\psi, \hat{\psi}$ are absolutely integrable. Then, we have the inversion formula (11.12) and for any random variable X, we have Parseval's identity:

$$E\psi(X) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{\psi}(t)\,\phi_X(-t)\,\mathrm{d}t \tag{11.13}$$

This will be used in the proof of the continuity theorem, below.

11.3 Inversion Formula for General Random Variables

In the general case, the random variable X is neither discrete nor continuous. It is still possible to obtain an inversion formula in this general case by inserting an additional integration. We begin with the symbolic formula

$$\phi_X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} F_X(\mathrm{d}x)$$

The distribution function F_X may be purely discrete, purely continuous, or a combination of both types. We use the above steps to write

$$\int_{-\infty}^{\infty} \phi_X(t)\mathrm{e}^{-ity}\mathrm{e}^{-\sigma^2 t^2/2}\,\mathrm{d}t = 2\pi\int_{-\infty}^{\infty} \frac{\mathrm{e}^{-(x-y)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} F_X(\mathrm{d}x)$$

Now, we integrate on the left side over the interval $a \le y \le b$ to obtain

$$\int_{-\infty}^{\infty} \phi_X(t)\left(\frac{\mathrm{e}^{-itb}-\mathrm{e}^{-ita}}{-it}\right)\mathrm{e}^{-\sigma^2 t^2/2}\,\mathrm{d}t = 2\pi\int_{-\infty}^{\infty}\left(\int_a^b \frac{\mathrm{e}^{-(x-y)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}\,\mathrm{d}y\right) F_X(\mathrm{d}x) \tag{11.14}$$

The integrand on the left side of (11.14) is defined by continuity at $t=0$. The integral inside the parentheses on the right side can be written in terms of the normal distribution function Φ as $\Phi((x-a)/\sigma) - \Phi((x-b)/\sigma)$. Using the properties that $\Phi(+\infty) = 1$, $\Phi(0) = \frac{1}{2}$, $\Phi(-\infty) = 0$, we see that

$$\lim_{\sigma\to 0}[\Phi((x-a)/\sigma) - \Phi((x-b)/\sigma)] = 1 \qquad a < x < b$$

$$\lim_{\sigma\to 0}[\Phi((x-a)/\sigma) - \Phi((x-b)/\sigma)] = \frac{1}{2} \qquad x = a \text{ or } x = b$$

$$\lim_{\sigma\to 0}[\Phi((x-a)/\sigma) - \Phi((x-b)/\sigma)] = 0 \qquad x < a \text{ or } x > b$$

Therefore, we obtain the **general form of the inversion formula**

$$\frac{1}{2\pi}\lim_{\sigma\to 0}\int_{-\infty}^{\infty}\phi_X(t)\left(\frac{\mathrm{e}^{-itb}-\mathrm{e}^{-ita}}{-it}\right)\mathrm{e}^{-\sigma^2t^2/2}\,\mathrm{d}t$$
$$= P[a<X<b]+\frac{1}{2}P[X=a]+\frac{1}{2}P[X=b], \tag{11.15}$$

which completes the proof of Theorem 11.2 in the most general case.

Corollary 11.1. If the distribution function of the random variable X is continuous at the points $x=a, x=b$, then

$$\frac{1}{2\pi}\lim_{\sigma\to 0}\int_{-\infty}^{\infty}\phi_X(t)\left(\frac{\mathrm{e}^{-itb}-\mathrm{e}^{-ita}}{-it}\right)\mathrm{e}^{-\sigma^2t^2/2}\,\mathrm{d}t = P[a<X<b] \tag{11.16}$$

Proof. Indeed, in this case $P[X=a]=0=P[X=b]$. ■

11.4 The Continuity Theorem

In order to use the characteristic function to prove limit theorems, we need to know that convergence of a sequence of characteristic functions implies convergence of the corresponding distribution functions, in an appropriate sense. The general result of this type is known as the *continuity theorem*, which is stated as follows.

Theorem 11.3. *Let $(X_n, n\geq 1)$ be a sequence of random variables with characteristic functions $\phi_n(t)$. If for each real number t, we have*

$$\lim_{n\to\infty}\phi_n(t)=\phi_X(t)$$

for some random variable X, then

$$\lim_{n\to\infty}P[a\leq X_n\leq b]=P[a\leq X\leq b]$$

provided that $P[X=a]=0=P[X=b]$; in particular, this occurs if X has a continuous distribution function.

Example Apply the continuity theorem to the (suitably normalized) binomial distribution with $p=\frac{1}{2}$.
Solution. If we have the binomial distribution with $p=\frac{1}{2}$, then the characteristic function of $X_n=(S_n-n/2)/\sqrt{n/4}$ is $\phi_n(t)=\cos(t/\sqrt{n})^n$. When $n\to\infty$, we have

$$\lim_n \phi_n(t)=\mathrm{e}^{-t^2/2}$$

which is the characteristic function of the standard normal distribution. Applying the continuity theorem shows that we have a limiting normal distribution.

If the limiting random variable X has a continuous distribution function, then $P[X=a]=0=P[X=b]$, so that we can assert that the probabilities of all intervals converge to the corresponding probabilities for the limiting random variable. In this case, we can assert that the probability of any interval, e.g., $[a,b)$, $(a,b]$ or (a,b) converges to the same limit.

The next example illustrates what can happen if the limiting distribution is not continuous.

Example Let the random variable $X_n := (-1)^n/n$, so that $X_n \to 0$ when $n \to \infty$. The distributions satisfy

$$P[0 \le X_n \le 1] = 1 \quad \text{if } n \text{ is even}, \qquad P[0 \le X_n \le 1] = 0 \quad \text{if } n \text{ is odd}$$

so that $\lim_n P[0 \le X_n \le 1]$ does not exist. This illustrates the possible limiting behavior when the limiting random variable has a discontinuous distribution function.

11.4.1 Proof of the Continuity Theorem*

The proof uses the notion of *upper limit* and *lower limit* of a sequence of real numbers. We begin with the *test functions* $\psi^{\epsilon}_{\pm}(x)$, depending on an additional parameter $\epsilon > 0$ and defined as follows: $\psi^{\epsilon}_{+}(x) = 1$ on the interval $[a,b]$ and $\psi(x) = 0$ if $x \le a-\epsilon$ or $x \ge b+\epsilon$. Otherwise, ψ^{ϵ}_{+} is a linear function, which interpolates between these values: $\psi_{+}(x) = (x-a+\epsilon)/\epsilon$ for $a-\epsilon \le x \le a$ and $\psi^{\epsilon}_{+}(x) = (b+\epsilon-x)/\epsilon$ for $b \le x \le b+\epsilon$. In the same manner, we define ψ^{ϵ}_{-}, which is piecewise linear, equal to 1 if $a+\epsilon \le x \le b-\epsilon$ and is zero for $x \le a$ and $x \ge b$. In particular, we have the double system of inequalities

$$\psi^{\epsilon}_{-}(x) \le 1_{[a,b]}(x) \le \psi^{\epsilon}_{+}(x) \tag{11.17}$$

On the other hand, both $\psi^{\epsilon}_{\pm}$ have trapezoidal profiles and can, thus, be expressed as the difference of two triangular profiles, both of which have integrable characteristic functions, from Equation (11.2). Hence, the Fourier inversion formula (11.12) and Parseval's identity (11.13) apply to both ψ^{ϵ}_{-} and ψ^{ϵ}_{+}. Applying both sides of (11.17) to X_n and taking the expectation, we have

$$E\left(\psi^{\epsilon}_{-}(X_n)\right) \le P[a \le X_n \le b] \le E\left(\psi^{\epsilon}_{+}(X_n)\right). \tag{11.18}$$

But for each n, we can use the Parseval's identity (11.13) to write

$$E\left(\psi^{\epsilon}_{\pm}(X_n)\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \phi_n(t)\hat{\psi}^{\epsilon}_{\pm}(-t)\,\mathrm{d}t.$$

* This section may be skipped on the first reading.

Taking the limit $n \to \infty$, we see that the right side converges, hence we have

$$\lim_n E\left(\psi_{\pm}^{\epsilon}(X_n)\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(t)\hat{\psi}_{\pm}^{\epsilon}(-t)\,\mathrm{d}t = E\left(\psi_{\pm}^{\epsilon}(X)\right)$$

where we have used Parseval's identity again. Referring to (11.18), we have the double system of inequalities

$$\limsup_n P[a \le X_n \le b] \le E\left(\psi_{+}^{\epsilon}(X)\right) \le P[a-\epsilon \le X \le b+\epsilon],$$
$$\liminf_n P[a \le X_n \le b] \ge E\left(\psi_{-}^{\epsilon}(X)\right) \ge P[a+\epsilon \le X \le b-\epsilon].$$

But the upper and lower limits do not depend on ϵ. Taking $\epsilon \to 0$, we obtain

$$P[a < X < b] \le \liminf_n P[a \le X_n \le b] \le \limsup_n P[a \le X_n \le b] \le P[a \le X \le b]$$

If $P[X=a] = 0 = P[X=b]$, then the two extreme members are equal and we have proved the required result.

Exercise 11.4
Show that the above proof applies equally well to compute $\lim_n P[a < X_n < b]$ or $\lim_n P[a \le X_n < b]$ or $\lim_n P[a < X_n \le b]$.

11.5 Proof of the Central Limit Theorem

The main application of the continuity theorem is to prove the classical CLT:

Theorem 11.4. *Let $Y_n, n \ge 1$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2 with $0 < \sigma^2 < \infty$. Denoting $S_n := Y_1 + \cdots + Y_n$, then for every pair of reals $a < b$*

$$\lim_n P\left[a \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le b\right] = \int_a^b \frac{\mathrm{e}^{-u^2/2}}{\sqrt{2\pi}}\,\mathrm{d}u. \tag{11.19}$$

Proof. This is proved by reducing to the case $\mu = 0, \sigma = 1$ as follows. Letting $Y_i' := (Y_i - \mu)/\sigma, S_n' = Y_1' + \cdots + Y_n'$, it is immediate that Y_i' has mean zero and variance 1. Furthermore $(S_n - n\mu)/\sigma\sqrt{n} = S_n'/\sqrt{n}$.

Assuming that $\mu = 0, \sigma = 1$, we have the characteristic function $\phi(t) = \phi_{Y_1}(t)$, a twice differentiable function with

$$|\phi(t)| \le 1, \quad \phi(0) = 1, \quad \phi'(0) = 0, \quad \phi''(0) = -1.$$

From Taylor's formula with remainder

$$\phi(s) = 1 - \frac{s^2}{2} + \epsilon_1(s), \qquad \lim_{s \to 0} \frac{\epsilon_1(s)}{s^2} = 0 \tag{11.20}$$

$$e^{-\frac{s^2}{2}} = 1 - \frac{s^2}{2} + \epsilon_2(s), \qquad \lim_{s \to 0} \frac{\epsilon_2(s)}{s^2} = 0. \tag{11.21}$$

The characteristic function of the normalized sum is

$$\phi_n(t) := E\left[e^{it\frac{S_n}{\sqrt{n}}}\right] = E\left[e^{it\frac{Y_1}{\sqrt{n}}}\right]^n = \phi\left(\frac{t}{\sqrt{n}}\right)^n$$

The characteristic function of the standard normal distribution is $e^{-t^2/2}$, so that the difference is written using the identity $A^n - B^n = (A - B)\left(A^{n-1} + \cdots + B^{n-1}\right)$:

$$\begin{aligned}\phi_n(t) - e^{-t^2/2} &= \phi\left(\frac{t}{\sqrt{n}}\right)^n - \left(e^{-t^2/2n}\right)^n \\ &= \left[\phi\left(\frac{t}{\sqrt{n}}\right) - e^{-t^2/2n}\right]\left[A^{n-1} + \cdots + B^{n-1}\right]\end{aligned}$$

where $A = \phi(t/\sqrt{n}), B = e^{-t^2/2n}$. Each of the n terms on the right is less than 1 in modulus, so that we can write

$$\left|\phi_n(t) - e^{-t^2/2}\right| \leq n\left|\phi\left(\frac{t}{\sqrt{n}}\right) - e^{-t^2/2n}\right|.$$

Setting $s = t/\sqrt{n}$ in (11.20) and (11.21) with t fixed, we have

$$\phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + \epsilon_1\left(\frac{t}{\sqrt{n}}\right), \qquad e^{-t^2/2n} = 1 - \frac{t^2}{2n} + \epsilon_2\left(\frac{t}{\sqrt{n}}\right).$$

Subtracting these two expressions, the first two terms cancel and we are left with terms of the form $n\epsilon(t/\sqrt{n})$, which tend to zero when $n \to \infty$ and t is fixed. We have proved that $\phi_n(t)$ converges to the standard normal characteristic function, which has a continuous distribution function. Hence, by the continuity theorem, the probabilities of all intervals converge, as required. ■

11.6 Stirling's Formula and Applications

Often, we encounter the factorial function of a large integer argument. The numerical evaluation of these expressions can be cumbersome, which leads one to search for an *asymptotic formula*, meaning a simpler formula, which provides a good approximation for large arguments.

Stirling's formula is the following limiting statement involving $n!$:

$$\lim_n \frac{n!}{n^{n+\frac{1}{2}}\mathrm{e}^{-n}} = \sqrt{2\pi} \tag{11.22}$$

where $e = 2.71828\cdots$ is the base of the natural logarithms. This is also written in the form

$$n! \sim n^{n+\frac{1}{2}}\mathrm{e}^{-n}\sqrt{2\pi}, \qquad n \to \infty \tag{11.23}$$

where the tilde sign means that the ratio of the two terms tends to 1 when $n \to \infty$. The Stirling's approximation (11.23) is already extremely accurate for small values of n; for example, if $n = 5$, then $n! = 120$, whereas the Stirling's approximation gives 118.019, an error of less than 2%. For $n = 10$, we have the exact value of 3,628,800, whereas Stirling's approximation is 3,598,690, an error of less than 1%.

We will prove Stirling's formula by representing the reciprocal of $n!$ in terms of a Poisson distribution, which we can estimate. No previous knowledge of the Poisson distribution is assumed.

11.6.1 Poisson Representation of n!

The Poisson distribution with parameter $\lambda > 0$ is defined by the sequence

$$p(k;\lambda) = \frac{\lambda^k}{k!}\mathrm{e}^{-\lambda}, \qquad k = 0, 1, 2, \ldots \tag{11.24}$$

It is immediate that $p(k;\lambda) > 0$ and $\sum_{k=0}^{\infty} p(k;\lambda) = 1$, so that we have a probability distribution on the nonnegative integers. The characteristic function is the following trigonometric series, which can be summed in closed form and which defines a 2π-periodic function:

$$\hat{p}(\theta;\lambda) = \sum_{k=0}^{\infty} p(k;\lambda)\mathrm{e}^{ik\theta} = \mathrm{e}^{\lambda(\mathrm{e}^{i\theta}-1)}, \qquad \theta \in \mathbf{R}. \tag{11.25}$$

For each $\lambda > 0$, the series (11.25) converges uniformly on $\mathbf{R}$, as well as the series obtained by multiplying by $\mathrm{e}^{-ik\theta}$. Hence, we can integrate term-by-term on any period interval to obtain

$$p(k;\lambda) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{p}(\theta,\lambda)\mathrm{e}^{-ik\theta}\mathrm{d}\theta, \qquad \lambda > 0, k = 0, 1, 2, \ldots \tag{11.26}$$

Now, we are free to take $\lambda = k$, to obtain the useful representation of the reciprocal factorial function:

$$p(k;k) = \frac{k^k}{k!}e^{-k} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{k\left(e^{i\theta}-1-i\theta\right)}\,d\theta, \qquad k = 0, 1, 2, \ldots. \tag{11.27}$$

11.6.2 Proof of Stirling's Formula

We, now, make the substitution $\psi = \theta\sqrt{k}$ to obtain the integral formula

$$\frac{k^{k+\frac{1}{2}}}{k!}e^{-k} = \frac{1}{2\pi}\int_{-\pi\sqrt{k}}^{\pi\sqrt{k}} e^{k\left(e^{i\psi/\sqrt{k}}-1-i\psi/\sqrt{k}\right)}\,d\psi, \qquad k = 1, 2, \ldots. \tag{11.28}$$

When $k \to \infty$, the integrand on the right side tends to $e^{-\psi^2/2}$ and is pointwise dominated by $e^{-\delta\psi^2}$, where $\delta := \inf_{0<|\theta|\le\pi}(1-\cos\theta)/\theta^2 = 2/\pi^2$, since the modulus of the exponential is the exponential of the real part, namely $k(\cos\theta - 1)$, which is bounded above by $-k\delta\theta^2$ for $|\theta| \le \pi$. Hence, by the dominated convergence theorem, we have

$$\lim_{k\to\infty}\frac{k^{k+\frac{1}{2}}}{k!}e^{-k} = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-\psi^2/2}\,d\psi = \frac{1}{\sqrt{2\pi}} \tag{11.29}$$

which is the statement of Stirling's formula, where we have used the normalization of the standard normal density: $\int_{-\infty}^{\infty} e^{-\psi^2/2}\,d\psi = \sqrt{2\pi}$.

Exercise 11.5
Prove the limiting relation $\lim_{z\to 0}\left(e^{iz} - 1 - iz\right)/z^2 = -1/2$.

Exercise 11.6
If $z = \alpha + i\beta$ is an arbitrary complex number, show that $\left|e^{\alpha+i\beta}\right| = e^{\alpha}$.

Exercise 11.7
Prove the upper and lower bounds $2\theta/\pi \le \sin\theta \le \theta$ for $0 \le \theta \le \pi/2$.
Hint: Look at the graphs of these three functions.

Exercise 11.8
Prove that $\int_R e^{-x^2/2}dx = \sqrt{2\pi}$.
Hint: Square both sides and use polar coordinates.

11.7 Local deMoivre–Laplace Theorem

We define the characteristic function and its normalized version by

$$F(\theta) := q + p\mathrm{e}^{i\theta}, \qquad F_0(\theta) := \mathrm{e}^{-ip\theta} F(\theta) = q\mathrm{e}^{-ip\theta} + p\mathrm{e}^{iq\theta} \tag{11.30}$$

The binomial probability is the Fourier coefficient of the characteristic function, which can be represented as a suitable integral on $(-\pi, \pi)$.

Lemma 11.1. *For $n = 0, 1, 2, \ldots$ and $k = 0, 1, \ldots, n$, let $x = x(k, n) = (k - np)/\sqrt{npq}$. Then,*

$$P_{kn} =: \binom{n}{k} p^k q^{n-k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\theta)^n \mathrm{e}^{-ik\theta}\, \mathrm{d}\theta \tag{11.31}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} F_0(\theta)^n \mathrm{e}^{-i\theta(k-np)}\, \mathrm{d}\theta \tag{11.32}$$

$$= \frac{1}{2\pi\sqrt{npq}} \int_{-\pi\sqrt{npq}}^{\pi\sqrt{npq}} F_0\left(\frac{\psi}{\sqrt{npq}}\right)^n \mathrm{e}^{-ix\psi}\, \mathrm{d}\psi \tag{11.33}$$

Proof. From the binomial theorem

$$F(\theta)^n = \left(q + p\mathrm{e}^{i\theta}\right)^n = \sum_{j=0}^{n} \binom{n}{k} p^j q^{n-j} \mathrm{e}^{ij\theta},$$

Multiply both sides by $\mathrm{e}^{-ik\theta}$ and integrate on $(-\pi, \pi)$, from which we conclude

$$P_{k,n} = \binom{n}{k} p^k q^{n-k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\theta)^n \mathrm{e}^{-ik\theta} \mathrm{d}\theta. \tag{11.34}$$

Formula (11.32) comes from the definition of $F_0(\theta)$. The final formula (11.33) comes from the definition of $x = x(k, n)$ and the substitution of $\psi = \theta\sqrt{npq}$, which completes the proof.

Now, we note

$$F_0(0) = 1, F_0'(0) = 0, F_0''(0) = -pq$$

$$|F_0(\theta)|^2 = q^2 + p^2 + 2pq\cos\theta = 1 - 2pq(1 - \cos\theta)$$

$$\leq \left(1 - 4pq\theta^2/\pi^2\right) \leq \mathrm{e}^{-4pq\theta^2/\pi^2}, \quad |\theta| < \pi$$

$$|F_0(\theta)| \leq \mathrm{e}^{-2pq\theta^2/\pi^2} \qquad |\theta| < \pi$$

$$\lim_{n\to\infty} F_0\left(\frac{\psi}{\sqrt{npq}}\right)^n = \mathrm{e}^{-\psi^2/2} \qquad \psi \in \mathbf{R}$$

For any real number x, let $k \to \infty, n \to \infty$ so that $(k - np)/\sqrt{npq} \to x$. Then, by the dominated convergence theorem,

$$\lim_{n\to\infty} \sqrt{npq}\, P_{kn} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{e}^{-\psi^2/2} \mathrm{e}^{-ix\psi}\, \mathrm{d}\psi = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

which is the statement of the *local central limit theorem of de Moivre and Laplace*.

This can be rewritten in a more intuitive form:

$$P(k,n) = p^k q^{n-k} \binom{n}{k} \sim \frac{\mathrm{e}^{-x^2/2}}{\sqrt{2\pi npq}}, \quad (n \to \infty)$$

■