Chapter 7
Characteristic Functions

Abstract Section 7.1 begins with formal definitions and contains an extensive discussion of the basic properties of characteristic functions, including those related to the nature of the underlying distributions. Section 7.2 presents the proofs of the inversion formulas for both densities and distribution functions, and also in the space of square integrable functions. Then the fundamental continuity theorem relating pointwise convergence of characteristic functions to weak convergence of the respective distributions is proved in Sect. 7.3. The result is illustrated by proving the Poisson theorem, with a bound for the convergence rate, in Sect. 7.4. After that, the previously presented theory is extended in Sect. 7.5 to the multivariate case. Some applications of characteristic functions are discussed in Sect. 7.6, including the stability properties of the normal and Cauchy distributions and an in-depth discussion of the gamma distribution and its properties. Section 7.7 introduces the concept of generating functions and uses it to analyse the asymptotic behaviour of a simple Markov discrete time branching process. The obtained results include the formula for the eventual extinction probability, the asymptotic behaviour of the non-extinction probabilities in the critical case, and convergence in that case of the conditional distributions of the scaled population size given non-extinction to the exponential law.

7.1 Definition and Properties of Characteristic Functions

As a preliminary remark, note that together with real-valued random variables $\xi(\omega)$ we could also consider complex-valued random variables, by which we mean functions of the form $\xi_1(\omega) + i\xi_2(\omega)$, (ξ_1, ξ_2) being a random vector. It is natural to put $\mathbf{E}(\xi_1 + i\xi_2) = \mathbf{E}\xi_1 + i\mathbf{E}\xi_2$. Complex-valued random variables $\xi = \xi_1 + i\xi_2$ and $\eta = \eta_1 + i\eta_2$ are *independent* if the σ-algebras $\sigma(\xi_1, \xi_2)$ and $\sigma(\eta_1, \eta_2)$ generated by the vectors (ξ_1, ξ_2) and (η_1, η_2), respectively, are independent. It is not hard to verify that, for such random variables,

$$\mathbf{E}\xi\eta = \mathbf{E}\xi\mathbf{E}\eta.$$

Definition 7.1.1 The *characteristic function* (ch.f.) of a real-valued random variable ξ is the complex-valued function

A.A. Borovkov, *Probability Theory*, Universitext,
DOI 10.1007/978-1-4471-5201-9_7,

$$\varphi_\xi(t) := \mathbf{E}e^{it\xi} = \int e^{itx}\, dF(x),$$

where t is real.

If the distribution function $F(x)$ has a density $f(x)$ then the ch.f. is equal to

$$\varphi_\xi(t) = \int e^{itx} f(x)\, dx$$

and is just the Fourier transform of the function $f(x)$.[1] In the general case, the ch.f. is the *Fourier–Stieltjes transform* of the function $F(x)$.

The ch.f. exists for any random variable ξ. This follows immediately from the relation

$$\left|\varphi_\xi(t)\right| \le \int \left|e^{itx}\right| dF(x) \le \int 1\, dF(x) = 1.$$

Ch.f.s are a powerful tool for studying properties of the sums of independent random variables.

7.1.1 Properties of Characteristic Functions

1. *For any random variable* ξ,

$$\varphi_\xi(0) = 1 \quad \text{and} \quad \left|\varphi_\xi(t)\right| \le 1 \quad \text{for all } t.$$

This property is obvious.

2. *For any random variable* ξ,

$$\varphi_{a\xi+b}(t) = e^{itb}\varphi_\xi(ta).$$

Indeed,

$$\varphi_{a\xi+b}(t) = \mathbf{E}e^{it(a\xi+b)} = e^{itb}\mathbf{E}e^{iat\xi} = e^{itb}\varphi_\xi(ta). \qquad \square$$

[1] More precisely, in classical mathematical analysis, the Fourier transform $\varphi(t)$ of a function $f(t)$ from the space L_1 of integrable functions is defined by the equation

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx} f(t)\, dt$$

(the difference from ch.f. consists in the factor $1/\sqrt{2\pi}$). Under this definition the inversion formula has a "symmetric" form: if $\varphi \in L_1$ then

$$f(x) = \frac{1}{\sqrt{2\pi}} \int e^{-itx}\varphi(t)\, dt.$$

This representation is more symmetric than the inversion formula for ch.f. (7.2.1) in Sect. 7.2 below.

3. *If* $\xi_1, \dots, \xi_n$ *are independent random variables then the ch.f. of the sum* $S_n = \xi_1 + \dots + \xi_n$ *is equal to*

$$\varphi_{S_n}(t) = \varphi_{\xi_1}(t) \cdots \varphi_{\xi_n}(t).$$

Proof This follows from the properties of the expectation of the product of independent random variables. Indeed,

$$\begin{aligned} \varphi_{S_n}(t) = \mathbf{E}e^{it(\xi_1+\dots+\xi_n)} &= \mathbf{E}e^{it\xi_1}e^{it\xi_2}\dots e^{it\xi_n} \\ &= \mathbf{E}e^{it\xi_1}\mathbf{E}e^{it\xi_2}\dots\mathbf{E}e^{it\xi_n} = \varphi_{\xi_1}(t)\varphi_{\xi_2}(t)\cdots\varphi_{\xi_n}(t). \end{aligned}$$ □

Thus to the convolution $F_{\xi_1} * F_{\xi_2}$ there corresponds the product $\varphi_{\xi_1}\varphi_{\xi_2}$.

4. *The ch.f.* $\varphi_\xi(t)$ *is a uniformly continuous function.*
Indeed, as $h \to 0$,

$$\bigl|\varphi(t+h) - \varphi(t)\bigr| = \bigl|\mathbf{E}\bigl(e^{i(t+h)\xi} - e^{it\xi}\bigr)\bigr| \le \mathbf{E}\bigl|e^{ih\xi} - 1\bigr| \to 0$$

by the dominated convergence theorem (see Corollary 6.1.2) since $|e^{ih\xi} - 1| \xrightarrow{p} 0$ as $h \to 0$, and $|e^{ih\xi} - 1| \le 2$. □

5. *If the k-th moment exists*: $\mathbf{E}|\xi|^k < \infty$, $k \ge 1$, *then there exists a continuous k-th derivative of the function* $\varphi_\xi(t)$, *and* $\varphi^{(k)}(0) = i^k\mathbf{E}\xi^k$.

Proof Indeed, since

$$\left|\int ixe^{itx}\,dF(x)\right| \le \int |x|\,dF(x) = \mathbf{E}|\xi| < \infty,$$

the integral $\int ixe^{itx}\,dF(x)$ converges uniformly in t. Therefore one can differentiate under the integral sign:

$$\varphi'(t) = i\int xe^{itx}\,dF(x), \qquad \varphi'(0) = i\mathbf{E}\xi.$$

Further, one can argue by induction. If, for $l < k$,

$$\varphi^{(l)}(t) = i^l\int x^l e^{itx}\,dF(x),$$

then

$$\varphi^{(l+1)}(t) = i^{l+1}\int x^{l+1}e^{itx}\,dF(x)$$

by the uniform convergence of the integral on the right-hand side. Therefore

$$\varphi^{(l+1)}(0) = i^{l+1}\mathbf{E}\xi^{l+1}.$$ □

Property 5 implies that if $\mathbf{E}|\xi|^k < \infty$ then, in a neighbourhood of the point $t = 0$, one has the expansion

$$\varphi(t) = 1 + \sum_{j=1}^{k} \frac{(it)^j}{j!}\mathbf{E}\xi^j + o\big(|t^k|\big). \tag{7.1.1}$$

The converse assertion is only partially true:

If a derivative of an even order $\varphi^{(2k)}$ exists then

$$\mathbf{E}|\xi|^{2k} < \infty, \qquad \varphi^{(2k)}(0) = (-1)^k \mathbf{E}\xi^{2k}.$$

We will prove the property for $k = 1$ (for $k > 1$ one can employ induction). It suffices to verify that $\mathbf{E}|\xi|^2$ is finite. One has

$$-\frac{2\varphi(0) - \varphi(2h) - \varphi(-2h)}{4h^2} = \mathbf{E}\left(\frac{e^{ih\xi} - e^{-ih\xi}}{2h}\right)^2 = \mathbf{E}\frac{\sin^2 h\xi}{h^2}.$$

Since $h^{-2}\sin^2 h\xi \to \xi^2$ as $h \to 0$, by Fatou's lemma

$$\begin{aligned} -\varphi''(0) &= \lim_{h\to 0}\left(\frac{2\varphi(0) - \varphi(2h) - \varphi(-2h)}{4h^2}\right) = \lim_{h\to 0}\mathbf{E}\frac{\sin^2 h\xi}{h^2} \\ &\ge \mathbf{E}\lim_{h\to 0}\frac{\sin^2 h\xi}{h^2} = \mathbf{E}\xi^2. \end{aligned}$$

□

6. *If $\xi \ge 0$ then $\varphi_\xi(\lambda)$ is defined in the complex plane for* $\operatorname{Im}\lambda \ge 0$. *Moreover, $|\varphi_\xi(\lambda)| \le 1$ for such λ, and in the domain* $\operatorname{Im}\lambda > 0$, *$\varphi_\xi(\lambda)$ is analytic and continuous including on the boundary* $\operatorname{Im}\lambda = 0$.

Proof That $\varphi(\lambda)$ is analytic follows from the fact that, for $\operatorname{Im}\lambda > 0$, one can differentiate under the integral sign the right-hand side of

$$\varphi_\xi(\lambda) = \int_0^\infty e^{i\lambda x}\,dF(x).$$

(For $\operatorname{Im}\lambda > 0$ the integrand decreases exponentially fast as $x \to \infty$.) □

Continuity is proved in the same way as in property 4. This means that for non-negative ξ the ch.f. $\varphi_\xi(\lambda)$ uniquely determines the function

$$\psi(s) = \varphi_\xi(is) = \mathbf{E}e^{-s\xi}$$

of real variable $s \ge 0$, which is called the *Laplace* (or *Laplace–Stieltjes*) *transform* of the distribution of ξ.

The converse assertion also follows from properties of analytic functions: the *Laplace transform $\psi(s)$ on the half-line $s \ge 0$ uniquely determines the ch.f. $\varphi_\xi(\lambda)$.*

7. $\overline{\varphi}_\xi(t) = \varphi_\xi(-t) = \varphi_{-\xi}(t)$, *where the bar denotes the complex conjugate.*

Proof The relations follow from the equalities

$$\overline{\varphi}_\xi(t) = \overline{\mathbf{E}e^{it\xi}} = \mathbf{E}\overline{e^{it\xi}} = \mathbf{E}e^{-it\xi}. \qquad \square$$

This implies the following property.

7A. *If ξ is symmetric (has the same distribution as $-\xi$) then its ch.f. is real ($\varphi_\xi(t) = \varphi_\xi(-t)$).*

One can show that the converse is also true; to this end one has to make use of the uniqueness theorem to be discussed below.

Now we will find the ch.f.s of the basic probability laws.

Example 7.1.1 If $\xi = a$ with probability 1, i.e. $\xi \in \mathbf{I}_a$, then $\varphi_\xi(t) = e^{ita}$.

Example 7.1.2 If $\xi \in \mathbf{B}_p$ then $\varphi_\xi(t) = pe^{it} + (1-p) = 1 + p(e^{it} - 1)$.

Example 7.1.3 If $\xi \in \mathbf{\Phi}_{0,1}$ then $\varphi_\xi(t) = e^{-t^2/2}$.

Indeed,

$$\varphi(t) = \varphi_\xi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - x^2/2}\, dx.$$

Differentiating with respect to t and integrating by parts ($xe^{-x^2/2}\, dx = -de^{-x^2/2}$), we get

$$\varphi'(t) = \frac{1}{\sqrt{2\pi}} \int ixe^{itx-x^2/2}\, dx = -\frac{1}{\sqrt{2\pi}} \int te^{itx-x^2/2}\, dx = -t\varphi(t),$$

$$\big(\ln\varphi(t)\big)' = -t, \qquad \ln\varphi(t) = -\frac{t^2}{2} + c.$$

Since $\varphi(0) = 1$, one has $c = 0$ and $\varphi(t) = e^{-t^2/2}$. $\square$

Now let η be a normal random variable with parameters (a, σ). Then it can be represented as $\eta = \sigma\xi + a$, where ξ is normally distributed with parameters (0, 1). The ch.f. of η can be found using Property 2:

$$\varphi_\eta(t) = e^{ita} e^{-(t\sigma)^2/2} = e^{ita - t^2\sigma^2/2}.$$

Differentiating $\varphi_\eta(t)$ for $\eta \in \mathbf{\Phi}_{0,\sigma^2}$, we will obtain that $\mathbf{E}\eta^k = 0$ for odd k, and $\mathbf{E}\eta^k = \sigma^k(k-1)(k-3)\cdots 1$ for $k = 2, 4, \ldots$.

Example 7.1.4 If $\xi \in \mathbf{\Pi}_\mu$ then

$$\varphi_\xi(t) = \mathbf{E}e^{it\xi} = \sum_k e^{itk}\frac{\mu^k}{k!}e^{-\mu} = e^{-\mu}\sum_k \frac{(\mu e^{it})^k}{k!} = e^{-\mu}e^{\mu e^{it}} = \exp\big[\mu\big(e^{it} - 1\big)\big].$$

Example 7.1.5 If ξ has the exponential distribution $\mathbf{\Gamma}_\alpha$ with density $\alpha e^{-\alpha x}$ for $x \geq 0$, then

$$\varphi_\xi(t) = \alpha \int_0^\infty e^{itx-\alpha x}\, dx = \frac{\alpha}{\alpha - it}.$$

Therefore, if ξ has the "double" exponential distribution with density $\frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$, then

$$\varphi_\xi(t) = \frac{1}{2}\left(\frac{1}{1-it} + \frac{1}{1+it}\right) = \frac{1}{1+t^2}.$$

If ξ has the geometric distribution $\mathbf{P}(\xi = k) = (1-p)p^k$, $k = 0, 1, \ldots,$ then

$$\varphi_\xi(t) = \frac{1-p}{1-pe^{it}}.$$

Example 7.1.6 If $\xi \in \mathbf{K}_{0,1}$ (has the density $[\pi(1+x^2)]^{-1}$) then $\varphi_\xi(t) = e^{-|t|}$. The reader will easily be able to prove this somewhat later, using the inversion formula and Example 7.1.5.

Example 7.1.7 If $\xi \in \mathbf{U}_{0,1}$, then

$$\varphi_\xi(t) = \int_0^1 e^{itx}\, dx = \frac{e^{it}-1}{it}.$$

By virtue of Property 3, the ch.f.s of the sums $\xi_1 + \xi_2$, $\xi_1 + \xi_2 + \xi_3, \ldots$ that we considered in Example 3.6.1 will be equal to

$$\varphi_{\xi_1+\xi_2}(t) = -\frac{(e^{it}-1)^2}{t^2}, \quad \varphi_{\xi_1+\xi_2+\xi_3}(t) = -\frac{(e^{it}-1)^3}{t^3}, \quad \ldots.$$

We return to the general case. How can one verify whether one or another function φ is characteristic or not? Sometimes one can do this using the above properties. We suggest the reader to determine whether the functions $(1+t)^{-1}$, $1+t$, $\sin t$, $\cos t$ are characteristic, and if so, to which distributions they correspond.

In the general case the posed question is a difficult one. We state without proof one of the known results.

Bochner–Khinchin's Theorem *A necessary and sufficient condition for a continuous function $\varphi(t)$ with $\varphi(0) = 1$ to be characteristic is that it is nonnegatively defined, i.e., for any real $t_1, \ldots, t_n$ and complex $\lambda_1, \ldots, \lambda_n$, one has*

$$\sum_{k,j=1}^n \varphi(t_k - t_j)\lambda_k \overline{\lambda}_j \geq 0$$

($\overline{\lambda}$ *is the complex conjugate of* λ).

Note that the necessity of this condition is almost obvious, for if $\varphi(t) = \mathbf{E}e^{it\xi}$ then

$$\sum_{k,j=1}^{n} \varphi(t_k - t_j)\lambda_k\overline{\lambda}_j = \mathbf{E}\sum_{k,j=1}^{n} e^{i(t_k-t_j)\xi}\lambda_k\overline{\lambda}_j = \mathbf{E}\left|\sum_{k=1}^{n}\lambda_k e^{it_k\xi}\right|^2 \ge 0.$$

7.1.2 The Properties of Ch.F.s Related to the Structure of the Distribution of ξ

8. *If the distribution of ξ has a density then $\varphi_\xi(t) \to 0$ as $|t| \to \infty$.*

This is a direct consequence of the Lebesgue theorem on Fourier transforms. The converse assertion is false.

In general, the smoother $F(x)$ is the faster $\varphi_\xi(t)$ vanishes as $|t| \to \infty$. The formulas in Example 7.1.7 are typical in this respect. If the density $f(x)$ has an integrable k-th derivative then, by integrating by parts, we get

$$\varphi_\xi(t) = \int e^{itx} f(x)\,dx = \frac{1}{it}\int e^{itx} f'(x)\,dx = \cdots = \frac{1}{(it)^k}\int e^{itx} f^{(k)}(x)\,dx,$$

which implies that

$$\varphi_\xi(t) \le \frac{c}{|t|^k}.$$

8A. *If the distribution of ξ has a density of bounded variation then*

$$\left|\varphi_\xi(t)\right| \le \frac{c}{|t|}.$$

This property is also validated by integration by parts:

$$\left|\varphi_\xi(t)\right| = \left|\frac{1}{it}\int e^{itx}\,df(x)\right| \le \frac{1}{|t|}\int \left|df(x)\right|.$$

9. *A random variable ξ has a lattice distribution with span $h > 0$* (see Definition 3.2.3) *if and only if*

$$\left|\varphi_\xi\left(\frac{2\pi}{h}\right)\right| = 1, \qquad \left|\varphi_\xi\left(\frac{v}{h}\right)\right| < 1 \tag{7.1.2}$$

if v is not a multiple of 2π.

Clearly, without loss of generality we can assume $h = 1$. Moreover, since

$$\left|\varphi_{\xi-a}(t)\right| = \left|e^{-ita}\varphi_\xi(t)\right| = \left|\varphi_\xi(t)\right|,$$

the properties (7.1.2) are invariant with respect to the shift by a. Thus we can assume the shift a is equal to zero and thus change the lattice distribution condition in Property 9 to the arithmeticity condition (see Definition 3.2.3). Since $\varphi_\xi(t)$ is a periodic function, Property 9 can be rewritten in the following equivalent form:

The distribution of a random variable ξ *is arithmetic if and only if*

$$\varphi_\xi(2\pi) = 1, \qquad \bigl|\varphi_\xi(t)\bigr| < 1 \quad \text{for all } t \in (0, 2\pi). \tag{7.1.3}$$

Proof If ξ has an arithmetic distribution then

$$\varphi_\xi(t) = \sum_k \mathbf{P}(\xi = k)e^{itk} = 1$$

for $t = 2\pi$. Now let us prove the second relation in (7.1.3). Assume the contrary: for some $v \in (0, 2\pi)$, we have $|\varphi_\xi(v)| = 1$ or, which is the same,

$$\varphi_\xi(v) = e^{ibv}$$

for some real b. The last relation implies that

$$\varphi_{\xi-b}(v) = 1 = \mathbf{E}\cos v(\xi - b) + i\mathbf{E}\sin v(\xi - b), \qquad \mathbf{E}\bigl[1 - \cos v(\xi - b)\bigr] = 0.$$

Hence, by Property E4 in Sect. 4.1, $\cos v(\xi - b) = 1$ and $v(\xi - b) = 2\pi k(\omega)$ with probability 1, where $k(\omega)$ is an integer. Thus $\xi - b$ is a multiple of $2\pi/v > 1$. This contradicts the assumption that the span of the lattice equals 1, and hence proves (7.1.3).

Conversely, let (7.1.3) hold. As we saw, the first relation in (7.1.3) implies that ξ takes only integer values. If we assume that the lattice span equals $h > 1$ then, by the first part of the proof and the first relation in (7.1.2), we get $|\varphi(2\pi/h)| = 1$, which contradicts the first relation in (7.1.3). Property 9 is proved. □

The next definition looks like a tautology.

Definition 7.1.2 The distribution of ξ is called *non-lattice* if it is not a lattice distribution.

10. *If the distribution of* ξ *is non-lattice then*

$$\bigl|\varphi_\xi(t)\bigr| < 1 \quad \text{for all } t \neq 0.$$

Proof Indeed, if we assume the contrary, i.e. that $|\varphi(u)| = 1$ for some $u \neq 0$, then, by Property 9, we conclude that the distribution of ξ is a lattice with span $h = 2\pi/u$ or with a lesser span. □

11. *If the distribution of* ξ *has an absolutely continuous component of a positive mass* $p > 0$, *then it is clearly non-lattice and, moreover*,

$$\limsup_{|t|\to\infty} \bigl|\varphi_\xi(t)\bigr| \le 1 - p.$$

This assertion follows from Property 8.

Arithmetic distributions occupy an important place in the class of lattice distributions.

For arithmetic distributions, the ch.f. $\varphi_\xi(t)$ is a function of the variable $z = e^{it}$ and is periodic in t with period 2π. Hence, in this case it is sufficient to know the behaviour of the ch.f. on the interval $[-\pi, \pi]$ or, which is the same, to know the behaviour of the function

$$p_\xi(z) := \mathbf{E}z^\xi = \sum z^k \mathbf{P}(\xi = k)$$

on the unit circle $|z| = 1$.

Definition 7.1.3 The function $p_\xi(z)$ is called the *generating function of the random variable* ξ (or of the distribution of ξ).

Since $p_\xi(e^{it}) = \varphi_\xi(t)$ is a ch.f., all the properties of ch.f.s remain valid for generating functions, with the only changes corresponding to the change of variable. For more on applications of generating functions, see Sect. 7.7.

7.2 Inversion Formulas

Thus for any random variable there exists a corresponding ch.f. We will now show that the set $\mathcal{L}$ of functions e^{itx} is a distribution determining class, i.e. that the distribution can be uniquely reconstructed from its ch.f. This is proved using inversion formulas.

7.2.1 The Inversion Formula for Densities

Theorem 7.2.1 *If the ch.f. $\varphi(t)$ of a random variable ξ is integrable then the distribution of ξ has the bounded density*

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \varphi(t)\, dt. \tag{7.2.1}$$

This fact is known from classical Fourier analysis, but we shall give a proof of a probabilistic character.

Proof First we will establish the following (Parseval's) identity: for any fixed $\varepsilon > 0$,

$$\begin{aligned} p_\varepsilon(t) &:= \frac{1}{2\pi} \int e^{-itu} \varphi(u) e^{-\varepsilon^2 u^2/2}\, du \\ &\equiv \frac{1}{\sqrt{2\pi}\,\varepsilon} \int \exp\left\{-\frac{(u-t)^2}{2\varepsilon^2}\right\} \mathbf{F}(du), \end{aligned} \tag{7.2.2}$$

where $\mathbf{F}$ is the distribution of ξ. We begin with the equality

$$\frac{1}{\sqrt{2\pi}} \int \exp\left\{ix\frac{\xi - t}{\varepsilon} - \frac{x^2}{2}\right\} dx = \exp\left\{-\frac{(\xi - t)^2}{2\varepsilon^2}\right\}, \tag{7.2.3}$$

both sides of which being the value of the ch.f. of the normal distribution with parameters (0, 1) at the point $(\xi - t)/\varepsilon$. After changing the variable $x = \varepsilon u$, the left-hand side of this equality can be rewritten as

$$\frac{\varepsilon}{\sqrt{2\pi}} \int \exp\left\{iu(\xi - t) - \frac{\varepsilon^2 u^2}{2}\right\} du.$$

If we take expectations of both sides of (7.2.3), we obtain

$$\frac{\varepsilon}{\sqrt{2\pi}} \int e^{-iut} \varphi(u) e^{-\frac{\varepsilon^2 u^2}{2}}\, du = \int \exp\left\{-\frac{(u-t)^2}{2\varepsilon^2}\right\} \mathbf{F}(du).$$

This proves (7.2.2).

To prove the theorem first consider the left-hand side of the equality (7.2.2). Since $e^{-\varepsilon^2 u^2/2} \to 1$ as $\varepsilon \to 0$, $|e^{-\frac{\varepsilon^2 u^2}{2}}| \le 1$ and $\varphi(u)$ is integrable, as $\varepsilon \to 0$ one has

$$p_\varepsilon(t) \to \frac{1}{2\pi} \int e^{-itu} \varphi(u)\, du = p_0(t) \tag{7.2.4}$$

uniformly in t, because the integral on the left-hand side of (7.2.2) is uniformly continuous in t. This implies, in particular, that

$$\int_a^b p_\varepsilon(t)\, dt \to \int_a^b p_0(t). \tag{7.2.5}$$

Now consider the right-hand side of (7.2.2). It represents the density of the sum $\xi + \varepsilon\eta$, where ξ and η are independent and $\eta \Subset \mathbf{\Phi}_{0,1}$. Therefore

$$\int_a^b p_\varepsilon(t)\, dt = \mathbf{P}(a < \xi + \varepsilon\eta \le b). \tag{7.2.6}$$

Since $\xi + \varepsilon\eta \xrightarrow{p} \xi$ as $\varepsilon \to 0$ and the limit $\int_a^b p_\varepsilon(t)\, dt$ exists for any fixed a and b by virtue of (7.2.5), this limit (see (7.2.6)) cannot be anything other than $\mathbf{F}([a, b))$.

Thus, from (7.2.5) and (7.2.6) we get

$$\int_a^b p_0(t)\, dt = \mathbf{F}\bigl([a, b)\bigr).$$

This means that the distribution $\mathbf{F}$ has the density $p_0(t)$, which is defined by relation (7.2.4). The boundedness of $p_0(t)$ evidently follows from the integrability of φ:

$$p_0(t) \le \frac{1}{2\pi} \int \bigl|\varphi(t)\bigr|\, dt < \infty.$$

The theorem is proved. □

7.2.2 The Inversion Formula for Distributions

Theorem 7.2.2 *If $F(x)$ is the distribution function of a random variable ξ and $\varphi(t)$ is its ch.f., then, for any points of continuity x and y of the function $F(x)$,*[2]

$$F(y)-F(x)=\frac{1}{2\pi}\lim_{\sigma\to 0}\int\frac{e^{-itx}-e^{-ity}}{it}\varphi(t)e^{-t^2\sigma^2}\,dt. \tag{7.2.7}$$

If the function $\varphi(t)/t$ is integrable at infinity then the passage to the limit under the integral sign is justified and one can write

$$F(y)-F(x)=\frac{1}{2\pi}\int\frac{e^{-itx}-e^{-ity}}{it}\varphi(t)\,dt. \tag{7.2.8}$$

Proof Suppose first that the ch.f. $\varphi(t)$ is integrable. Then $F(x)$ has a density $f(x)$ and the assertion of the theorem in the form (7.2.8) follows if we integrate both sides of Eq. (7.2.1) over the interval with the end points x and y and change the order of integration (which is valid because of the absolute convergence).[3]

Now let $\varphi(t)$ be the characteristic function of a random variable ξ with an arbitrary distribution $\mathbf{F}$. On a common probability space with ξ, consider a random variable η which is independent of ξ and has the normal distribution with parameters $(0, 2\sigma^2)$. As we have already pointed out, the ch.f. of η is $e^{-t^2\sigma^2}$.

This means that the ch.f. of $\xi+\eta$, being equal to $\varphi(t)e^{-t^2\sigma^2}$, is integrable. Therefore by (7.2.8) one will have

$$F_{\xi+\eta}(y)-F_{\xi+\eta}(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-itx}-e^{-ity}}{it}\varphi(t)e^{-t^2\sigma^2}\,dt. \tag{7.2.9}$$

Since $\eta \xrightarrow{p} 0$ as $\sigma\to 0$, we have $\mathbf{F}_{\xi+\eta}\Rightarrow\mathbf{F}$ (see Chap. 6). Therefore, if x and y are points of continuity of $\mathbf{F}$, then $F(y)-F(x)=\lim_{\sigma\to 0}(F_{\xi+\eta}(y)-F_{\xi+\eta}(x))$. This, together with (7.2.9), proves the assertion of the theorem. □

In the proof of Theorem 7.2.2 we used a method which might be called the "smoothing" of distributions. It is often employed to overcome technical difficulties related to the inversion formula.

Corollary 7.2.1 (Uniqueness Theorem) *The ch.f. of a random variable uniquely determines its distribution function.*

---

[2]In the literature, the inversion formula is often given in the form

$$F(y)-F(x)=\frac{1}{2\pi}\lim_{A\to\infty}\int_{-A}^{A}\frac{e^{-itx}-e^{-ity}}{it}\varphi(t)\,dt$$

which is equivalent to (7.2.7).

[3]Formula (7.2.8) can also be obtained from (7.2.1) without integration by noting that $(F(x)-F(y))/(y-x)$ is the value at zero of the convolution of two densities: $f(x)$ and the uniform density over the interval $[-y,-x]$ (see also the remark at the end of Sect. 3.6). The ch.f. of the convolution is equal to $\frac{e^{-itx}-e^{-ity}}{(y-x)it}\varphi(t)$.

The proof follows from the inversion formula and the fact that $\mathbf{F}$ is uniquely determined by the differences $F(y) - F(x)$.

For *lattice* random variables the inversion formula becomes simpler. Let, for the sake of simplicity, ξ be an integer-valued random variable.

Theorem 7.2.3 *If $p_\xi(z) := \mathbf{E}z^\xi$ is the generating function of an arithmetic random variable then*

$$\mathbf{P}(\xi = k) = \frac{1}{2\pi i}\int_{|z|=1} p_\xi(z)z^{-k-1}\,dz. \tag{7.2.10}$$

Proof Turning to the ch.f. $\varphi_\xi(t) = \sum_j e^{itj}\mathbf{P}(\xi = j)$ and changing the variables $z = it$ in (7.2.10) we see that the right-hand side of (7.2.10) equals

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\varphi_\xi(t)\,dt = \frac{1}{2\pi}\sum_j \mathbf{P}(\xi = j)\int_{-\pi}^{\pi} e^{it(j-k)}\,dt.$$

Here all the integrals on the right-hand side are equal to zero, except for the integral with $j = k$ which is equal to 2π. Thus the right-hand side itself equals $\mathbf{P}(\xi = k)$. The theorem is proved. $\square$

Formula (7.2.10) is nothing else but the formula for Fourier coefficients and has a simple geometric interpretation. The functions $\{\mathrm{e}_k = e^{itk}\}$ form an orthonormal basis in the Hilbert space $L_2(-\pi, \pi)$ of square integrable complex-valued functions with the inner product

$$(f, g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\overline{g}(t)\,dt$$

($\overline{g}$ is the complex conjugate of g). If $\varphi_\xi = \sum \mathrm{e}_k\mathbf{P}(\xi = k)$ then it immediately follows from the equality $\varphi_\xi = \sum \mathrm{e}_k(_\xi, \mathrm{e}_k)$ that

$$\mathbf{P}(\xi = k) = (\varphi_\xi, \mathrm{e}_k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\varphi_\xi(t)\,dt.$$

7.2.3 *The Inversion Formula in L_2. The Class of Functions that Are Both Densities and Ch.F.s*

First consider some properties of ch.f.s related to the inversion formula. As a preliminary, note that, in classical Fourier analysis, one also considers the Fourier transforms of functions f from the space L_2 of square-integrable functions. Since in this case a function f is not necessarily integrable, the Fourier transform is defined as the integral in the principal value sense:[4]

$$\varphi(t) := \lim_{N\to\infty} \varphi_{(N)}(t), \qquad \varphi_{(N)}(t) := \int_{-N}^{N} e^{itx} f(x)\,dx, \tag{7.2.11}$$

where the limit is taken in the sense of convergence in L_2:

$$\int \left|\varphi(t) - \varphi_{(N)}(t)\right|^2 dx \to 0 \quad \text{as } N \to \infty.$$

Since by Parseval's equality

$$\|f\|_{L_2} = \frac{1}{2\pi}\|\varphi\|_{L_2}, \quad \text{where } \|g\|_{L_2} = \left[\int |g|^2(t)\,dt\right]^{1/2},$$

the Fourier transform maps the space L_2 into itself (there is no such isometricity for Fourier transforms in L_1). *Here the inversion formula* (7.2.1) *holds true but the integral in* (7.2.1) *is understood in the principal value sense.*

Denote by $\mathcal{F}$ and $\mathcal{H}$ the class of all densities and the class of all ch.f.s, respectively, and by $\mathcal{H}_{1,+} \subset L_1$ the class of *nonnegative real-valued integrable* ch.f.s, so that the elements of $\mathcal{H}_{1,+}$ are in $\mathcal{F}$ up to the normalising factors. Further, let $(\mathcal{H}_{1,+})^{(-1)}$ be the inverse image of the class $\mathcal{H}_{1,+}$ in $\mathcal{F}$ for the mapping $f \to \varphi$, i.e. the class of densities whose ch.f.s lie in $\mathcal{H}_{1,+}$. It is clear that functions f from $(\mathcal{H}_{1,+})^{(-1)}$ and φ from $\mathcal{H}_{1,+}$ are necessarily symmetric (see Property 7A in Sect. 7.1) and that $f(0) \in (0, \infty)$. The last relation follows from the fact that, by the inversion formula for $\varphi \in \mathcal{H}_{1,+}$, we have

$$\|\varphi\| := \|\varphi\|_{L_1} = \int \varphi(t)\,dt = 2\pi f(0).$$

Further, denote by $(\mathcal{H}_{1,+})_{\|\cdot\|}$ the class of *normalised* functions $\frac{\varphi}{\|\varphi\|}$, $\varphi \in \mathcal{H}_{1,+}$, so that $(\mathcal{H}_{1,+})_{\|\cdot\|} \subset \mathcal{F}$, and denote by $\mathcal{F}^{(2,*)}$ the class of *convolutions of symmetric densities* from L_2:

$$\mathcal{F}^{(2,*)} := \left\{ f^{(2)*}(x) : f \in L_2,\ f \text{ is symmetric}\right\},$$

where

$$f^{(2)*}(x) = \int f(t) f(x-t)\,dt.$$

Theorem 7.2.4 *The following relations hold true*:

$$(\mathcal{H}_{1,+})^{(-1)} = (\mathcal{H}_{1,+})_{\|\cdot\|}, \qquad \mathcal{F}^{(2,*)} \subset (\mathcal{H}_{1,+})_{\|\cdot\|}.$$

The class $(\mathcal{H}_{1,+})_{\|\cdot\|}$ may be called the class of *densities conjugate to* $f \in (\mathcal{H}_{1,+})^{(-1)}$. It turns out that this class coincides with the inverse image $(\mathcal{H}_{1,+})^{(-1)}$. The second statement of the theorem shows that this inverse image is a very rich

[4]Here we again omit the factor $\frac{1}{\sqrt{2\pi}}$ (cf. the footnote on page 154).

class and provides sufficient conditions for the density f to have a conjugate. We will need these conditions in Sect. 8.7.

Proof of Theorem 7.2.4 Let $f \in (\mathcal{H}_{1,+})^{(-1)}$. Then the corresponding ch.f. φ is in $\mathcal{H}_{1+}$ and the inversion formula (7.2.1) is applicable. Multiplying its right-hand side by $\frac{2\pi}{\|\varphi\|}$, we obtain an expression for the ch.f. (at the point $-t$) of the density $\frac{\varphi}{\|\varphi\|}$ (recall that $\varphi \ge 0$ is symmetric if $\varphi \in \mathcal{H}_{1,+}$). This means that $\frac{2\pi f}{\|\varphi\|}$ is a ch.f. and, moreover, that $f \in (\mathcal{H}_{1,+})_{\|\cdot\|}$.

Conversely, suppose that $f^* := \frac{\varphi}{\|\varphi\|} \in (\mathcal{H}_{1,+})_{\|\cdot\|}$. Then $f^* \in \mathcal{F}$ is symmetric, and the inversion formula can be applied to φ:

$$f(x) = \frac{1}{2\pi}\int e^{-itx}\varphi(t)\,dt = \frac{1}{2\pi}\int e^{itx}\varphi(t)\,dt, \qquad \frac{2\pi f(t)}{\|\varphi\|} = \int e^{itx} f^*(x)\,dx.$$

Since the ch.f. $\varphi^*(t) := \frac{2\pi f(t)}{\|\varphi\|}$ belongs to $\mathcal{H}_{1,+}$, one has $f^* \in (\mathcal{H}_{1,+})^{(-1)}$.

We now prove the second assertion. Suppose that $f \in L_2$. Then $\varphi \in L_2$ and $\varphi^2 \in L_1$. Moreover, by virtue of the symmetry of f and Property 7A in Sect. 7.1, the function φ is real-valued, so $\varphi^2 \ge 0$. This implies that $\varphi^2 \in \mathcal{H}_{1,+}$. Since φ^2 is the ch.f. of the density $f^{(2)*}$, we have $f^{(2)*} \in (\mathcal{H}_{1,+})^{(-1)}$. The theorem is proved. □

Note that *any bounded density f belongs to L_2*. Indeed, since the Lebesgue measure of $\{x : f(x) \ge 1\}$ is always less than 1, for $f(\cdot) \le N$ we have

$$\|f\|_{L_2}^2 = \int f^2(x)\,dx \le \int_{f(x)<1} f(x)\,dx + N^2 \int_{f(x)\ge 1} dx \le 1 + N^2. \qquad \square$$

Thus we have obtained the following result.

Corollary 7.2.2 *For any bounded symmetric density f, the convolution $f^{(2)*}$ is, up to a constant factor, the ch.f. of a random variable.*

Example 7.2.1 The "triangle" density

$$g(x) = \begin{cases} 1 - |x| & \text{if } |x| \le 1, \\ 0 & \text{if } |x| > 1, \end{cases}$$

being the convolution of the two uniform distributions on $[-1/2, 1/2]$ (cf. Example 3.6.1) is also a ch.f. We suggest the reader to verify that the preimage of this ch.f. is the density

$$f(x) = \frac{1}{2\pi}\frac{\sin^2 x/2}{x^2}$$

(the density conjugate to g). Conversely, the density g is conjugate to f, and the functions $8\pi f(t)$ and $g(t)$ will be ch.f.s for g and f, respectively.

These assertions will be useful in Sect. 8.7.

7.3 The Continuity (Convergence) Theorem

Let $\{\varphi_n(t)\}_{n=1}^{\infty}$ be a sequence of ch.f.s and $\{F_n\}_{n=1}^{\infty}$ the sequence of the respective distribution functions. Recall that the symbol $\Rightarrow$ denotes the weak convergence of distributions introduced in Chap. 6.

Theorem 7.3.1 (The Continuity Theorem) *A necessary and sufficient condition for the convergence $F_n \Rightarrow F$ as $n \to \infty$ is that $\varphi_n(t) \to \varphi(t)$ for any t, $\varphi(t)$ being the ch.f. corresponding to F.*

The theorem follows in an obvious way from Corollary 6.3.2 (here two of the three sufficient conditions from Corollary 6.3.2 are satisfied: conditions (2) and (3)). The proof of the theorem can be obtained in a simpler way as well. This way is presented in Sect. 7.4 of the previous editions of this book.

In Sect. 7.1, for nonnegative random variables ξ we introduced the notion of the Laplace transform $\psi(s) := \mathbf{E}e^{-s\xi}$. Let $\psi_n(s)$ and $\psi(s)$ be Laplace transforms corresponding to F_n and F. The following analogue of Theorem 7.3.1 holds for Laplace transforms:

In order that $F_n \Rightarrow F$ as $n \to \infty$ it is necessary and sufficient that $\psi_n(s) \to \psi(s)$ for each $s \ge 0$.

Just as in Theorem 7.3.1, this assertion follows from Corollary 6.3.2, since the class $\{f(x) = e^{-sx},\ s \ge 0\}$ is (like $\{e^{itx}\}$) a distribution determining class (see Property 6 in Sect. 7.1) and, moreover, the sufficient conditions (2) and (3) of Corollary 6.3.2 are satisfied.

Theorem 7.3.1 has a deficiency: one needs to know in advance that the function $\varphi(t)$ to which the ch.f.s converge is a ch.f. itself. However, one could have no such prior information (see e.g. Sect. 8.8). In this connection there arises a natural question under what conditions the limiting function $\varphi(t)$ will be characteristic.

The answer to this question is given by the following theorem.

Theorem 7.3.2 *Let*

$$\varphi_n(t) = \int e^{itx}\, dF_n(x)$$

be a sequence of ch.f.s and $\varphi_n(t) \to \varphi(t)$ as $n \to \infty$ for any t.

Then the following three conditions are equivalent:

(a) *$\varphi(t)$ is a ch.f.*;
(b) *$\varphi(t)$ is continuous at $t = 0$*;
(c) *the sequence $\{F_n\}$ is tight.*

Thus if we establish that $\varphi_n(t) \to \varphi(t)$ and one of the above three conditions is met, then we can assert that there exists a distribution F such that φ is the ch.f. of F and $F_n \Rightarrow F$.

Proof The equivalence of conditions (a) and (c) follows from Theorem 6.3.2. That (a) implies (b) is known. It remains to establish that (c) follows from (b). First we will show that the following lemma is true. □

Lemma 7.3.1 *If φ is the ch.f. of ξ then, for any $u > 0$,*

$$\mathbf{P}\left(|\xi| > \frac{2}{u}\right) \le \frac{1}{u}\int_{-u}^{u}\big[1-\varphi(t)\big]\,dt.$$

Proof The right-hand side of this inequality is equal to

$$\frac{1}{u}\int_{-u}^{u}\int_{-\infty}^{\infty}\big(1-e^{-itx}\big)\,dF(x)\,dt,$$

where F is the distribution function of ξ. Changing the order of integration and noting that

$$\int_{-u}^{u}\big(1-e^{-itx}\big)\,dt = \left(t+\frac{e^{-itx}}{ix}\right)\bigg|_{-u}^{u} = 2u\left(1-\frac{\sin ux}{ux}\right),$$

we obtain that

$$\begin{aligned}\frac{1}{u}\int_{-u}^{u}\big[1-\varphi(t)\big]\,dt &= 2\int_{-\infty}^{\infty}\left(1-\frac{\sin ux}{ux}\right)dF(x)\\ &\ge 2\int_{|x|>2/u}\left(1-\left|\frac{\sin ux}{ux}\right|\right)dF(x)\\ &\ge 2\int_{|x|>2/u}\left(1-\frac{1}{|ux|}\right)dF(x) \ge \int_{|x|>2/u}dF(x).\end{aligned}$$

The lemma is proved. □

Now suppose that condition (b) is met. By Lemma 7.3.1

$$\limsup_{n\to\infty}\int_{|x|>2/u}dF_n(x) \le \limsup_{n\to\infty}\frac{1}{u}\int_{-u}^{u}\big[1-\varphi_n(t)\big]\,dt = \frac{1}{u}\int_{-u}^{u}\big[1-\varphi(t)\big]\,dt.$$

Since $\varphi(t)$ is continuous at 0 and $\varphi(0)=1$, the mean value on the right-hand side can clearly be made arbitrarily small by choosing sufficiently small u. This obviously means that condition (c) is satisfied. The theorem is proved. □

Using ch.f.s one can not only establish convergence of distribution functions but also estimate the rate of this convergence in the cases when one can estimate how fast $\varphi_n - \varphi$ vanishes. We will encounter respective examples in Sect. 7.5.

We will mostly use the machinery of ch.f.s in Chaps. 8, 12 and 17. In the present chapter we will also touch upon some applications of ch.f.s, but they will only serve as illustrations.

7.4 The Application of Characteristic Functions in the Proof of the Poisson Theorem

Let $\xi_1, \dots, \xi_n$ be independent integer-valued random variables,

$$S_n = \sum_1^k \xi_k, \qquad \mathbf{P}(\xi_k = 1) = p_k, \qquad \mathbf{P}(\xi_k = 0) = 1 - p_k - q_k.$$

The theorem below is a generalisation of the theorems established in Sect. 5.4.[5]

Theorem 7.4.1 *One has*

$$\bigl|\mathbf{P}(S_n = k) - \boldsymbol{\Pi}_\mu\bigl(\{k\}\bigr)\bigr| \le \sum_{k=1}^n p_k^2 + 2\sum_{k=1}^n q_k, \quad \textit{where } \mu = \sum_{k=1}^n p_k.$$

Thus, if one is given a triangle array $\xi_{1n}, \xi_{2n}, \dots, \xi_{nn}$, $n = 1, 2, \dots$, of independent integer-valued random variables,

$$S_n = \sum_{k=1}^n \xi_{kn}, \qquad \mathbf{P}(\xi_{kn} = 1) = p_{kn}, \qquad \mathbf{P}(\xi_{kn} = 0) = 1 - p_{kn} - q_{kn},$$

$$\mu = \sum_{k=1}^n p_{kn},$$

then a sufficient condition for convergence of the difference $\mathbf{P}(S_n = k) - \boldsymbol{\Pi}_\mu(\{k\})$ to zero is that

$$\sum_{k=1}^n q_{kn} \to 0, \qquad \sum_{k=1}^n p_{kn}^2 \to 0.$$

Since

$$\sum_{k=1}^n p_{kn}^2 \le \mu \max_{k\le n} p_{kn},$$

the last condition is always met if

$$\max_{k\le n} p_{kn} \to 0, \qquad \mu \le \mu_0 = \text{const}.$$

[5] This extension is not really substantial since close results could be established using Theorem 5.2.2 in which ξ_k can only take the values 0 and 1. It suffices to observe that the probability of the event $A = \bigcup_k \{\xi_k \ne 0,\ \xi_k \ne 1\}$ is bounded by the sum $\sum q_k$ and therefore

$$\mathbf{P}(S_n = k) = \theta_1 \sum q_k + \Bigl(1 - \theta_2 \sum q_k\Bigr)\mathbf{P}(S_n = k|\overline{A}), \quad \theta_i \le 1,\ i = 1, 2,$$

where $\mathbf{P}(S_n = k|\overline{A}) = \mathbf{P}(S_n^* = k)$ and S_n^* are sums of independent random variables ξ_k^* with

$$\mathbf{P}\bigl(\xi_k^* = 1\bigr) = p_k^* = \frac{p_k}{1 - q_k}, \qquad \mathbf{P}\bigl(\xi_k^* = 0\bigr) = 1 - p_k^*.$$

To prove the theorem we will need two auxiliary assertions.

Lemma 7.4.1 *If* $\operatorname{Re}\beta \le 0$ *then*

$$\left|e^{\beta}-1\right| \le |\beta|, \qquad \left|e^{\beta}-1-\beta\right| \le |\beta|^2/2, \qquad \left|e^{\beta}-1-\beta-\beta^2/2\right| \le |\beta|^3/6.$$

Proof The first two inequalities follow from the relations (we use here the change of variables $t=\beta v$ and the fact that $|e^s| \le 1$ for $\operatorname{Re} s \le 0$)

$$\left|e^{\beta}-1\right| = \left|\int_0^{\beta} e^t\,dt\right| = \left|\beta\int_0^1 e^{\beta v}\,dv\right| \le |\beta|,$$
$$\left|e^{\beta}-1-\beta\right| = \left|\int_0^{\beta}\left(e^t-1\right)dt\right| = \left|\beta\int_0^1\left(e^{\beta v}-1\right)dv\right| \le |\beta|^2\int_0^1 v\,dv = \left|\beta^2\right|/2.$$

The last inequality is proved in the same way. □

Lemma 7.4.2 *If* $|a_k| \le 1$, $|b_k| \le 1$, $k=1,\dots,n$, *then*

$$\left|\prod_{k=1}^n a_k - \prod_{k=1}^n b_k\right| \le \sum_{k=1}^n |a_k-b_k|.$$

Thus if $\varphi_k(t)$ *and* $\theta_k(t)$ *are ch.f.s then, for any* t,

$$\left|\prod_{k=1}^n \varphi_k(t) - \prod_{k=1}^n \theta_k(t)\right| \le \sum_{k=1}^n \left|\varphi_k(t)-\theta_k(t)\right|.$$

Proof Put $A_n=\prod_{k=1}^n a_k$ and $B_n=\prod_{k=1}^n b_k$. Then $|A_n| \le 1$, $|B_n| \le 1$, and

$$\begin{aligned}|A_n-B_n| &= |A_{n-1}a_n - B_{n-1}b_n| \\ &= \left|(A_{n-1}-B_{n-1})a_n + (a_n-b_n)B_{n-1}\right| \le |A_{n-1}-B_{n-1}| + |a_n-b_n|.\end{aligned}$$

Applying this inequality n times, we obtain the required relation. □

Proof of Theorem 7.4.1 One has

$$\varphi_k(t) := \mathbf{E}e^{it\xi_k} = 1 + p_k\left(e^{it}-1\right) + q_k\left(\gamma_k(t)-1\right),$$

where $\gamma_k(t)$ is the ch.f. of some integer-valued random variable. By independence of the random variables ξ_k,

$$\varphi_{S_n}(t) = \prod_{k=1}^n \varphi_k(t).$$

Let further $\zeta \Subset \mathbf{\Pi}_\mu$. Then

$$\varphi_\zeta(t) = \mathbf{E}e^{it\zeta} = e^{\mu(e^{it}-1)} = \prod_{k=1}^n \theta_k(t),$$

where $\theta_k(t) = e^{p_k(e^{it}-1)}$. Therefore the difference between the ch.f.s φ_{S_n} and φ_ζ can be bounded by Lemma 7.4.2 as follows:

$$\bigl|\varphi_{S_n}(t) - \varphi_\zeta(t)\bigr| = \left|\prod_{k=1}^{n} \varphi_k - \prod_{k=1}^{n} \theta_k\right| \le \sum_{k=1}^{n} |\varphi_k - \theta_k|,$$

where by Lemma 7.4.1 (note that $\mathrm{Re}(e^{it} - 1) \le 0$)

$$\begin{aligned}\bigl|\theta_k(t) - 1 - p_k\bigl(e^{it} - 1\bigr)\bigr| &\le \frac{p_k^2|e^{it} - 1|^2}{2} = \frac{p_k^2}{2}\bigl(\sin^2 t + (1 - \cos t)^2\bigr) \\ &= p_k^2\left(\frac{\sin^2 t}{2} + 2\sin^4\frac{t}{2}\right),\end{aligned} \tag{7.4.1}$$

$$\sum_{k=1}^{n} |\varphi_k - \theta_k| \le 2\sum_{k=1}^{n} q_k + \sum_{k=1}^{n} p_k^2\left(\frac{\sin^2 t}{2} + 2\sin^4\frac{t}{2}\right).$$

It remains to make use of the inversion formula (7.2.10):

$$\begin{aligned}\bigl|\mathbf{P}(S_n = k) - \boldsymbol{\Pi}_\mu\bigl(\{k\}\bigr)\bigr| &\le \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\bigl(\varphi_{S_n}(t) - \varphi_\zeta(t)\bigr)\,dt\right| \\ &\le \frac{1}{\pi}\int_0^{\pi}\left[2\sum_{k=1}^{n} q_k + \sum_{k=1}^{n} p_k^2\left(\frac{\sin^2 t}{2} + 2\sin^4\frac{t}{2}\right)\right]dt \\ &= 2\sum_{k=1}^{n} q_k + \sum_{k=1}^{n} p_k^2,\end{aligned}$$

for

$$\frac{1}{2\pi}\int_0^{\pi} \sin^2 t\,dt = \frac{1}{4}, \qquad \frac{2}{\pi}\int_0^{\pi} \sin^4\frac{t}{2}\,dt = \frac{3}{4}.$$

The theorem is proved. □

If one makes use of the inequality $|e^{it} - 1| \le 2$ in (7.4.1), the computations will be simplified, there will be no need to calculate the last two integrals, but the bounds will be somewhat worse:

$$\sum |\varphi_k - \theta_k| \le 2\Bigl(\sum q_k + \sum p_k^2\Bigr),$$
$$\bigl|\mathbf{P}(S_n = k) - \boldsymbol{\Pi}_\mu\bigl(\{k\}\bigr)\bigr| \le 2\Bigl(\sum q_k + \sum p_k^2\Bigr).$$

7.5 Characteristic Functions of Multivariate Distributions. The Multivariate Normal Distribution

Definition 7.5.1 Given a random vector $\xi = (\xi_1, \xi_2, \ldots, \xi_d)$, its ch.f. (the ch.f. of its distribution) is defined as the function of the vector variable $t = (t_1, \ldots, t_d)$ equal to

$$\varphi_\xi(t) := \mathbf{E}e^{it\xi^T} = \mathbf{E}e^{i(t,\xi)} = \mathbf{E}\exp\left\{i\sum_{k=1}^{d} t_k\xi_k\right\}$$
$$= \int \exp\left\{i\sum_{k=1}^{d} t_k x_k\right\}\mathbf{F}_{\xi_1,\dots,\xi_d}(dx_1,\dots,dx_d),$$

where ξ^T is the transpose of ξ (a column vector), and (t,ξ) is the inner product.

The ch.f.s of multivariate distributions possess all the properties (with obvious amendments of their statements) listed in Sects. 7.1–7.3.

It is clear that $\varphi_\xi(0) = 1$ and that $|\varphi_\xi(t)| \le 1$ and $\varphi_\xi(-t) = \overline{\varphi_\xi(t)}$ always hold. Further, $\varphi_\xi(t)$ is everywhere continuous. If there exists a mixed moment $\mathbf{E}\xi_1^{k_1}\cdots\xi_d^{k_d}$ then φ_ξ has the respective derivative of order $k_1 + \cdots + k_d$:

$$\left.\frac{\partial\varphi_\xi^{k_1+\cdots+k_d}(t)}{\partial t_1^{k_1}\dots\partial t_d^{k_d}}\right|_{t=0} = i^{k_1+\cdots+k_d}\mathbf{E}\xi_1^{k_1}\cdots\xi_d^{k_d}.$$

If all the moments of some order exist, then an expansion of the function $\varphi_\xi(t)$ similar to (7.1.1) is valid in a neighbourhood of the point $t = 0$.

If $\varphi_\xi(t)$ is known, then the ch.f. of any subcollection of the random variables $(\xi_{k_1},\dots,\xi_{k_j})$ can obviously be obtained by setting all t_k except $t_{k_1},\dots,t_{k_j}$ to be equal to 0.

The following theorems are simple extensions of their univariate analogues.

Theorem 7.5.1 (The Inversion Formula) *If Δ is a parallelepiped defined by the inequalities $a_k < x < b_k$, $k = 1,\dots,d$, and the probability $\mathbf{P}(\xi \in \Delta)$ is continuous on the faces of the parallelepiped, then*

$$\mathbf{P}(\xi \in \Delta) = \lim_{\sigma\to 0}\frac{1}{(2\pi)^d}\int\cdots\int\left(\prod_{k=1}^{d}\frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k}e^{-t_k^2\sigma^2}\right)\varphi_\xi(t)\,dt_1\cdots dt_d.$$

If the random vector ξ has a density $f(x)$ and its ch.f. $\varphi_\xi(t)$ is integrable, then the inversion formula can be written in the form

$$f(x) = \frac{1}{(2\pi)^d}\int e^{-i(t,x)}\varphi_\xi(t)\,dt.$$

If a function $g(x)$ is such that its Fourier transform

$$\widetilde{g}(t) = \int e^{i(t,x)}g(x)\,dx$$

is integrable (and this is always the case for sufficiently smooth $g(x)$) then the Parseval equality holds:

$$\mathbf{E}g(\xi) = \mathbf{E}\frac{1}{(2\pi)^d}\int e^{-i(t,\xi)}\widetilde{g}(t)\,dt = \frac{1}{(2\pi)^d}\int\varphi_\xi(-t)\widetilde{g}(t)\,dt.$$

As before, the inversion formula implies the theorem on one-to-one correspondence between ch.f.s and distribution functions and together with it the fact that $\{e^{i(t,x)}\}$ is a distribution determining class (cf. Definition 6.3.2).

The weak convergence of distributions $\mathbf{F}_n(B)$ in the d-dimensional space to a distribution $\mathbf{F}(B)$ is defined in the same way as in the univariate case: $\mathbf{F}_{(n)} \Rightarrow \mathbf{F}$ if

$$\int f(x)\, d\mathbf{F}_{(n)}(dx) \to \int f(x)\, d\mathbf{F}(dx)$$

for any continuous and bounded function $f(x)$.

Denote by $\varphi_n(t)$ and $\varphi(t)$ the ch.f.s of distributions $\mathbf{F}_n$ and $\mathbf{F}$, respectively.

Theorem 7.5.2 (Continuity Theorem) *A necessary and sufficient condition for the weak convergence* $\mathbf{F}_{(n)} \Rightarrow \mathbf{F}$ *is that, for any* t, $\varphi_n(t) \to \varphi(t)$ *as* $n \to \infty$.

In the case where one can establish convergence of $\varphi_n(t)$ to some function $\varphi(t)$, there arises the question of whether $\varphi(t)$ will be the ch.f. of some distribution, or, which is the same, whether the sequence $\mathbf{F}_{(n)}$ will converge weakly to some distribution $\mathbf{F}$. Answers to these questions are given by the following assertion. Let Δ_N be the cube defined by the inequality $\max_k |x_k| < N$.

Theorem 7.5.3 (Continuity Theorem) *Suppose a sequence* $\varphi_n(t)$ *of ch.f.s converges as* $n \to \infty$ *to a function* $\varphi(t)$ *for each* t. *Then the following three conditions are equivalent*:

(a) $\varphi(t)$ *is a ch.f.*;
(b) $\varphi(t)$ *is continuous at the point* $t = 0$;
(c) $\limsup_{n\to\infty} \int_{x\notin\Delta_N} \mathbf{F}_{(n)}(dx) \to 0$ *as* $N \to \infty$.

All three theorems from this section can be proved in the same way as in the univariate case.

Example 7.5.1 The multivariate normal distribution is defined as a distribution with density (see Sect. 3.3)

$$f_\xi(x) = \frac{|A|^{1/2}}{(2\pi)^{d/2}} e^{-Q(x)/2},$$

where

$$Q(x) = xAx^T = \sum_{i,j=1}^{d} a_{ij} x_i x_j,$$

and $|A|$ is the determinant of a positive definite matrix $A = \|a_{ij}\|$.

This is a *centred* normal distribution for which $\mathbf{E}\xi = 0$. The distribution of the vector $\xi + a$ for any constant vector a is also called normal.

Find the ch.f. of ξ. Show that

$$\varphi_\xi(t) = \exp\left\{-\frac{t\sigma^2 t^T}{2}\right\}, \tag{7.5.1}$$

where $\sigma^2 = A^{-1}$ is the matrix inverse to A and coinciding with the covariance matrix $\|\sigma_{ij}\|$ of ξ:

$$\sigma_{ij} = \mathbf{E}\xi_i \xi_j .$$

Indeed,

$$\varphi_\xi(t) = \frac{\sqrt{|A|}}{(2\pi)^{d/2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{ itx^T - \frac{xAx^T}{2} \right\} dx_1 \cdots dx_d. \qquad (7.5.2)$$

Choose an orthogonal matrix C such that $CAC^T = D$ is a diagonal matrix, and denote by $\mu_1, \dots, \mu_n$ the values of its diagonal elements. Change the variables by putting $x = yC$ and $t = vC$. Then

$$|A| = |D| = \prod_{k=1}^{d} \mu_k ,$$

$$itx^T - \frac{1}{2} xAx^T = ivy^T - \frac{1}{2} yDy^T = i \sum_{k=1}^{d} v_k y_k - \frac{1}{2} \sum_{k=1}^{n} \mu_k y_k^2 ,$$

and, by Property 2 of ch.f.s of the univariate normal distributions,

$$\begin{aligned} \varphi_\xi(t) &= \frac{\sqrt{|A|}}{(2\pi)^{d/2}} \prod_{k=1}^{d} \int_{-\infty}^{\infty} \exp\left\{ iv_k y_k - \frac{\mu_k y_k^2}{2} \right\} dy_k = \sqrt{|A|} \prod_{k=1}^{d} \frac{1}{\sqrt{\mu_k}} \exp\left\{ -\frac{v_k^2}{2\mu_k} \right\} \\ &= \exp\left\{ -\frac{vD^{-1}v^T}{2} \right\} = \exp\left\{ -\frac{tC^T D^{-1} C t^T}{2} \right\} = \exp\left\{ -\frac{tA^{-1}t^T}{2} \right\}. \end{aligned}$$

On the other hand, since all the moments of ξ exist, in a neighbourhood of the point $t = 0$ one has

$$\varphi_\xi(t) = 1 - \frac{1}{2} tA^{-1}t^T + o\Big(\sum t_k^2 \Big) = 1 + it\mathbf{E}\xi^T + \frac{1}{2} t\sigma^2 t^T + o\Big(\sum t_k^2 \Big).$$

From this it follows that $\mathbf{E}\xi = 0$, $A^{-1} = \sigma^2$.

Formula (7.5.1) that we have just proved implies the following property of normal distributions: *the components of the vector* $(\xi_1, \dots, \xi_d)$ *are independent if and only if the correlation coefficients* $\rho(\xi_i, \xi_j)$ *are zero for all* $i \neq j$. Indeed, if σ^2 is a diagonal matrix, then $A = \sigma^{-2}$ is also diagonal and $f_\xi(x)$ is equal to the product of densities. Conversely, if $(\xi_1, \dots, \xi_d)$ are independent, then A is a diagonal matrix, and hence σ^2 is also diagonal.

7.6 Other Applications of Characteristic Functions. The Properties of the Gamma Distribution

7.6.1 Stability of the Distributions $\mathbf{\Phi}_{\alpha,\sigma^2}$ and $\mathbf{K}_{\alpha,\sigma}$

The stability property means, roughly speaking, that the distribution type is preserved under summation of random variables (this description of stability is not exact, for more detail see Sect. 8.8).

The sum of independent normally distributed random variables is also normally distributed. Indeed, let ξ_1 and ξ_2 be independent and normally distributed with parameters (a_1, σ_1^2) and (a_2, σ_2^2), respectively. Then the ch.f. of $\xi_1 + \xi_2$ is equal to

$$\begin{aligned}\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t)\varphi_{\xi_2}(t) &= \exp\left\{ita_1 - \frac{t^2\sigma_1^2}{2}\right\}\exp\left\{ita_2 - \frac{t^2\sigma_2^2}{2}\right\} \\ &= \exp\left\{it(a_1+a_2) - \frac{t^2}{2}\left(\sigma_1^2+\sigma_2^2\right)\right\}.\end{aligned}$$

Thus the sum $\xi_1 + \xi_2$ is again a normal random variable, with parameters $(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

Normality is also preserved when taking sums of *dependent* random variables (components of an arbitrary normally distributed random vector). This immediately follows from the form of the ch.f. of the multivariate normal law found in Sect. 7.5. One just has to note that to get the ch.f. of the sum $\xi_1 + \cdots + \xi_n$ it suffices to put $t_1 = \cdots = t_n = t$ in the expression

$$\varphi_{(\xi_1,\ldots,\xi_n)}(t_1,\ldots,t_n) = \mathbf{E}\exp\{it_1\xi_1 + \cdots + it_n\xi_n\}.$$

The sum of independent random variables distributed according to the Poisson law also has a Poisson distribution. Indeed, consider two independent random variables $\xi_1 \in \mathbf{\Pi}_{\lambda_1}$ and $\xi_2 \in \mathbf{\Pi}_{\lambda_2}$. The ch.f. of their sum is equal to

$$\varphi_{\xi_1+\xi_2}(t) = \exp\left\{\lambda_1\left(e^{it}-1\right)\right\}\exp\left\{\lambda_2\left(e^{it}-1\right)\right\} = \exp\left\{(\lambda_1+\lambda_2)\left(e^{it}-1\right)\right\}.$$

Therefore $\xi_1 + \xi_2 \in \mathbf{\Pi}_{\lambda_1+\lambda_2}$.

The sum of independent random variables distributed according to the Cauchy law also has a Cauchy distribution. Indeed, if $\xi_1 \in \mathbf{K}_{\alpha_1,\sigma_1}$ and $\xi_2 \in \mathbf{K}_{\alpha_2,\sigma_2}$, then

$$\begin{aligned}\varphi_{\xi_1+\xi_2}(t) &= \exp\left\{i\alpha_1 t - \sigma_1|t|\right\}\exp\left\{i\alpha_2 t - \sigma_2|t|\right\} \\ &= \exp\left\{i(\alpha_1+\alpha_2)t - (\sigma_1+\sigma_2)|t|\right\}; \\ &\xi_1 + \xi_2 \in \mathbf{K}_{\alpha_1+\alpha_2,\sigma_1+\sigma_2}.\end{aligned}$$

The above assertions are closely related to the fact that the normal and Poisson laws are, as we saw, limiting laws for sums of independent random variables (the Cauchy distribution has the same property, see Sect. 8.8). Indeed, if $S_{2n}/\sqrt{2n}$ converges in distribution to a normal law (where $S_k = \sum_{j=1}^k \xi_j$, ξ_j are independent and identically distributed) then it is clear that $S_n/\sqrt{n}$ and $(S_{2n} - S_n)/\sqrt{n}$ will also converge to a normal law so that the sum of two asymptotically normal random variables also has to be asymptotically normal.

Note, however, that due to its arithmetic structure the random variable $\xi \in \mathbf{\Pi}_\lambda$ (as opposed to $\xi \in \mathbf{\Phi}_{a,\sigma^2}$ or $\xi \in \mathbf{K}_{\alpha,\sigma}$) cannot be transformed by any normalisation (linear transformation) into a random variable again having the Poisson distribution but with another parameter. For this reason the Poisson distribution cannot be stable in the sense of Definition 8.8.2.

It is not hard to see that the other distributions we have met do not possess the above-mentioned property of preservation of the distribution type under summation of random variables. If, for instance, ξ_1 and ξ_2 are uniformly distributed over [0, 1] and independent then F_{ξ_1} and $F_{\xi_1+\xi_2}$ are substantially different functions (see Example 3.6.1).

7.6.2 The $\mathbf{\Gamma}$-distribution and its properties

In this subsection we will consider one more rather wide-spread type of distribution closely related to the normal distribution and frequently used in applications. This is the so-called *Pearson gamma distribution* $\mathbf{\Gamma}_{\alpha,\lambda}$. We will write $\xi \in \mathbf{\Gamma}_{\alpha,\lambda}$ if ξ has density

$$f(x;\alpha,\lambda) = \begin{cases} \frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}, & x \ge 0, \\ 0, & x < 0, \end{cases}$$

depending on two parameters $\alpha > 0$ and $\lambda > 0$, where $\Gamma(\lambda)$ is the gamma function

$$\Gamma(\lambda) = \int_0^\infty x^{\lambda-1} e^{-x}\,dx, \qquad \lambda > 0.$$

It follows from this equality that $\int f(x;\alpha,\lambda)\,dx = 1$ (one needs to make the variable change $\alpha x = y$). If one differentiates the ch.f.

$$\varphi(t) = \varphi(t;\alpha,\lambda) = \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty x^{\lambda-1} e^{itx-\alpha x}\,dx$$

with respect to t and then integrates by parts, the result will be

$$\begin{aligned}
\varphi'(t) &= \frac{\alpha^\lambda}{\Gamma(\lambda)} \int_0^\infty i x^\lambda e^{itx-\alpha x}\,dx = \frac{\alpha^\lambda}{\Gamma(\lambda)} \frac{i\lambda}{\alpha - it} \int_0^\infty x^{\lambda-1} e^{itx-\alpha x}\,dx \\
&= \frac{i\lambda}{\alpha - it}\varphi(t); \\
&\big(\ln\varphi(t)\big)' = \big(-\lambda \ln(\alpha - it)\big)', \qquad \varphi(t) = c(\alpha - it)^{-\lambda}.
\end{aligned}$$

Since $\varphi(0) = 1$ one has $c = \alpha^\lambda$ and $\varphi(t) = (1 - it/\alpha)^{-\lambda}$.

It follows from the form of the ch.f. that the subfamily of distributions $\mathbf{\Gamma}_{\alpha,\lambda}$ for a fixed α also has a certain stability property: if $\xi_1 \in \mathbf{\Gamma}_{\alpha,\lambda_1}$ and $\xi_2 \in \mathbf{\Gamma}_{\alpha,\lambda_2}$ are independent, then $\xi_1 + \xi_2 \in \mathbf{\Gamma}_{\alpha,\lambda_1+\lambda_2}$.

An example of a particular gamma distribution is given, for instance, by the distribution of the random variable

$$\chi_n^2 = \sum_{i=1}^{n} \xi_i^2,$$

where ξ_i are independent and normally distributed with parameters $(0, 1)$. This is the so-called *chi-squared distribution with n degrees of freedom* playing an important role in statistics.

To find the distribution of χ_n^2 it suffices to note that, by virtue of the equality

$$\mathbf{P}\big(\chi_1^2 < x\big) = \mathbf{P}\big(|\xi_1| < \sqrt{x}\big) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-u^2/2}\, du,$$

the density of χ_1^2 is equal to

$$\frac{1}{\sqrt{2\pi}} e^{-x/2} x^{-1/2} = f(x; 1/2, 1/2), \qquad \chi_1^2 \in \mathbf{\Gamma}_{1/2,1/2}.$$

This means that the ch.f. of χ_n^2 is

$$\varphi^n(t; 1/2, 1/2) = (1 - 2it)^{-n/2} = \varphi(t; 1/2, n/2)$$

and corresponds to the density $f(t; 1/2, n/2)$.

Another special case of the gamma distribution is the *exponential distribution* $\mathbf{\Gamma}_\alpha = \mathbf{\Gamma}_{\alpha,1}$ with density

$$f(x; \alpha, 1) = \alpha e^{-\alpha x}, \quad x \ge 0,$$

and characteristic function

$$\varphi(x; \alpha, 1) = \left(1 - \frac{it}{\alpha}\right)^{-1}.$$

We leave it to the reader to verify with the help of ch.f.s that if $\xi_j \in \mathbf{\Gamma}_{\alpha_j}$ and are independent, $\alpha_j \ne \alpha_l$ for $j \ne l$, then

$$\mathbf{P}\left(\sum_{j=1}^{n} \xi_j > x\right) = \sum_{j=1}^{n} e^{-\alpha_j x} \prod_{\substack{l=1 \\ l \ne j}}^{n} \left(1 - \frac{\alpha_j}{\alpha_l}\right)^{-1}.$$

In various applications (in particular, in queueing theory, cf. Sect. 12.4), the so-called *Erlang distribution* is also of importance. This is a distribution with density $f(x; \alpha, \lambda)$ for integer λ. The Erlang distribution is clearly a λ-fold convolution of the exponential distribution with itself.

We find the expectation and variance of a random variable ξ that has the gamma distribution with parameters (α, λ):

$$\mathbf{E}\xi = -i\varphi'(0; \alpha, \lambda) = \frac{\lambda}{\alpha}, \qquad \mathbf{E}\xi^2 = -i\varphi''(0; \alpha, \lambda) = \frac{\lambda(\lambda+1)}{\alpha^2},$$

$$\mathrm{Var}(\xi) = \frac{\lambda(\lambda+1)}{\alpha^2} - \left(\frac{\lambda}{\alpha}\right)^2 = \frac{\lambda}{\alpha^2}.$$

Distributions from the gamma family, and especially the exponential ones, are often (and justifiably) used to approximate distributions in various applied problems. We will present three relevant examples.

Example 7.6.1 Consider a complex device. The failure of at least one of n parts comprising the device means the breakdown of the whole device. The lifetime distribution of any of the parts is usually well described by the exponential law. (The reasons for this could be understood with the help of the Poisson theorem on rare events. See also Example 2.4.1 and Chap. 19.)

Thus if the lifetimes ξ_j of the parts are independent, and for the part number j one has

$$\mathbf{P}(\xi_j > x) = e^{-\alpha_j x}, \quad x > 0,$$

then the lifetime of the whole device will be equal to $\eta_n = \min(\xi_1, \dots, \xi_n)$ and we will get

$$\mathbf{P}(\eta_n > x) = \mathbf{P}\left(\bigcap_{j=1}^{n} \{\xi_j > x\}\right) = \prod_{j=1}^{n} \mathbf{P}(\xi_j > x) = \exp\left\{-x \sum_{i=1}^{n} \alpha_i\right\}.$$

This means that η_n will also have the exponential distribution, and since

$$\mathbf{E}\xi_j = 1/\alpha_j,$$

the mean failure-free operation time of the device will be equal to

$$\mathbf{E}\eta_n = \left(\sum_{i=1}^{n} \frac{1}{\mathbf{E}\xi_i}\right)^{-1}.$$

Example 7.6.2 Now turn to the distribution of $\zeta_n = \max(\xi_1, \dots, \xi_n)$, where ξ_i are independent and all have the $\boldsymbol{\Gamma}$-distribution with parameters (α, λ). We could consider, for instance, a queueing system with n channels. (That could be, say, a multiprocessor computer solving a problem using the complete enumeration algorithm, each of the processors of the machine checking a separate variant.) Channel number i is busy for a random time ξ_i. After what time will the whole system be free? This random time will clearly have the same distribution as ζ_n.

Since the ξ_i are independent, we have

$$\mathbf{P}(\zeta_n < x) = \mathbf{P}\left(\bigcap_{j=1}^{n} \{\xi_j < x\}\right) = \left[\mathbf{P}(\xi_1 < x)\right]^n. \tag{7.6.1}$$

If n is large, then for approximate calculations we could find the limiting distribution of ζ_n as $n \to \infty$. Note that, for any fixed x, $\mathbf{P}(\zeta_n < x) \to 0$ as $n \to \infty$.

Assuming for simplicity that $\alpha = 1$ (the general case can be reduced to this one by changing the scale), we apply L'Hospital's rule to see that, as $x \to \infty$,

$$\mathbf{P}(\xi_j < x) = \int_x^\infty \frac{1}{\Gamma(\lambda)} y^{\lambda-1} e^{-y}\, dy \sim \frac{x^{\lambda-1}}{\Gamma(\lambda)} e^{-x}.$$

Letting $n \to \infty$ and

$$x = x(n) = \ln\bigl[n(\ln n)^{\lambda-1}/\Gamma(\lambda)\bigr] + u, \quad u = \text{const},$$

we get

$$\mathbf{P}(\xi_j > x) \sim \frac{(\ln n)^{\lambda-1}}{\Gamma(\lambda)} \frac{\Gamma(\lambda)}{n(\ln n)^{\lambda-1}} e^{-u} = \frac{e^{-u}}{n}.$$

Therefore for such x and $n \to \infty$ we obtain by (7.6.1) that

$$\mathbf{P}(\zeta_n < x) = \left(1 - \frac{e^{-u}}{n}\bigl(1 + o(1)\bigr)\right)^n \to e^{-e^{-u}}.$$

Thus we have established the existence of the limit

$$\lim_{n\to\infty} \mathbf{P}\left(\zeta_n - \ln\left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)}\right] < u\right) = e^{-e^{-u}},$$

or, which is the same, that

$$\zeta_n - \ln\left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)}\right] \Subset\!\Rightarrow F_0, \qquad F_0(u) = e^{-e^{-u}}.$$

In other words, for large n the variable ζ_n admits the representation

$$\zeta_n \approx \ln\left[\frac{n(\ln n)^{\lambda-1}}{\Gamma(\lambda)}\right] + \zeta^0, \quad \text{where } \zeta^0 \Subset F_0.$$

Example 7.6.3 Let ξ_1 and ξ_2 be independent with $\xi_1 \Subset \mathbf{\Gamma}_{\alpha,\lambda_1}$ and $\xi_2 \Subset \mathbf{\Gamma}_{\alpha,\lambda_2}$. What is the distribution of $\xi_1/(\xi_1 + \xi_2)$? We will make use of Theorem 4.9.2. Since the joint density $f(x, y)$ of ξ_1 and $\eta = \xi_1 + \xi_2$ is equal to

$$f(x, y) = f(x; \alpha, \lambda_1) f(y - x; \alpha, \lambda_2),$$

the density of η is

$$q(y) = f(y; \alpha, \lambda_1 + \lambda_2),$$

and the conditional density $f(x \mid y)$ of ξ_1 given $\eta = y$ is equal to

$$f(x \mid y) = \frac{f(x, y)}{q(y)} = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} \frac{x^{\lambda_1-1}(y - x)^{\lambda_2-1}}{y^{\lambda_1+\lambda_2-1}}, \quad x \in [0, y].$$

By the formulas from Sect. 3.2 the conditional density of $\xi_1/y = \xi_1/(\xi_1 + \xi_2)$ (given the same condition $\xi_1 + \xi_2 = y$) is equal to

$$yf(yx \mid y) = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} x^{\lambda_1-1}(1 - x)^{\lambda_2-1}, \quad x \in [0, 1].$$

This distribution does not depend on y (nor on α). Hence the conditional density of $\xi_1/(\xi_1 + \xi_2)$ will have the same property, too. We obtain the so-called *beta distribution* $\mathbf{B}_{\lambda_1,\lambda_2}$ with parameters λ_1 and λ_2 defined on the interval [0, 1]. In particular, for $\lambda_1 = \lambda_2 = 1$, the distribution is uniform: $\mathbf{B}_{1,1} = \mathbf{U}_{0,1}$.

7.7 Generating Functions. Application to Branching Processes. A Problem on Extinction

7.7.1 Generating Functions

We already know that if a random variable ξ is integer-valued, i.e.

$$\mathbf{P}\biggl(\bigcup_k \{\xi = k\}\biggr) = 1,$$

then the ch.f. $\varphi_\xi(t)$ will actually be a function of $z = e^{it}$, and, along with its ch.f., the distribution of ξ can be specified by its generating function

$$p_\xi(z) := \mathbf{E}z^\xi = \sum_k z^k \mathbf{P}(\xi = k).$$

The inversion formula can be written here as

$$\mathbf{P}(\xi = k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\varphi_\xi(t)\,dt = \frac{1}{2\pi i}\int_{|z|=1} z^{-k-1} p_\xi(z)\,dz. \qquad (7.7.1)$$

As was already noted (see Sect. 7.2), relation (7.7.1) is simply the formula for Fourier coefficients (since $e^{itk} = \cos tk + i\sin tk$).

If ξ and η are independent random variables, then the distribution of $\xi + \eta$ will be given by the convolution of the sequences $\mathbf{P}(\xi = k)$ and $\mathbf{P}(\eta = k)$:

$$\mathbf{P}(\xi + \eta = n) = \sum_{k=-\infty}^{\infty} \mathbf{P}(\xi = k)\mathbf{P}(\eta = n - k)$$

(the total probability formula). To this convolution there corresponds the product of the generating functions:

$$p_{\xi+\eta}(z) = p_\xi(z)p_\eta(z).$$

It is clear from the examples considered in Sect. 7.1 that the generating functions of random variables distributed according to the Bernoulli and Poisson laws are

$$p_\xi(z) = 1 + p(z-1), \qquad p_\xi(z) = \exp\bigl\{\mu(z-1)\bigr\},$$

respectively.

One can see from the definition of the generating function that, for a nonnegative random variable $\xi \ge 0$, the function $p_\xi(z)$ is defined for $|z| \le 1$ and is analytic in the domain $|z| < 1$.

7.7.2 The Simplest Branching Processes

Now we turn to sequences of random variables which describe the so-called *branching processes*. We have already encountered a simple example of such a process when describing a chain reaction scheme in Example 4.4.4. Consider a more general scheme of a branching process. Imagine particles that can produce other particles of the same type; these could be neutrons in chain reactions, bacteria reproducing according to certain laws etc. Assume that initially there is a single particle (the "null generation") that, as a result of a "division" act, transforms with probabilities f_k, $k = 0, 1, 2, \dots$, into k particles of the same type,

$$\sum_{k=0}^{\infty} f_k = 1.$$

The new particles form the "first generation". Each of the particles from that generation behaves itself in the same way as the initial particle, independently of what happened before and of the other particles from that generation. Thus we obtain the "second generation", and so on. Denote by ζ_n the number of particles in the n-th generation. To describe the sequence ζ_n, introduce, as we did in Example 4.4.4, independent sequences of independent identically distributed random variables

$$\bigl\{\xi_j^{(1)}\bigr\}_{j=1}^{\infty}, \quad \bigl\{\xi_j^{(2)}\bigr\}_{j=1}^{\infty}, \quad \dots,$$

where $\xi_j^{(n)}$ have the distribution

$$\mathbf{P}\bigl(\xi_j^{(n)} = k\bigr) = f_k, \quad k = 0, 1, \dots.$$

Then the sequence ζ_n can be represented as

$$\begin{aligned}
\zeta_0 &= 1, \\
\zeta_1 &= \xi_1^{(1)}, \\
\zeta_2 &= \xi_1^{(2)} + \cdots + \xi_{\zeta_1}^{(2)}, \\
&\dots\dots\dots\dots\dots\dots \\
\zeta_n &= \xi_1^{(n)} + \cdots + \xi_{\zeta_{n-1}}^{(n)}.
\end{aligned}$$

These are sums of random numbers of random variables. Since $\xi_1^{(n)}, \xi_2^{(n)}, \dots$ do not depend on ζ_{n-1}, for the generating function $f_{(n)}(z) = \mathbf{E}z^{\zeta_n}$ we obtain by the total probability formula that

$$\begin{aligned}
f_{(n)}(z) &= \sum_{k=0}^{\infty} \mathbf{P}(\zeta_{n-1} = k)\mathbf{E}z^{\xi_1^{(n)} + \cdots + \xi_k^{(n)}} \\
&= \sum_{k=0}^{\infty} \mathbf{P}(\zeta_{n-1} = k) f^k(z) = f_{(n-1)}\bigl(f(z)\bigr),
\end{aligned} \tag{7.7.2}$$

where

$$f(z) := f_{(1)}(z) = \mathbf{E}z^{\xi_1^{(n)}} = \sum_{k=0}^{\infty} f_k z^k.$$

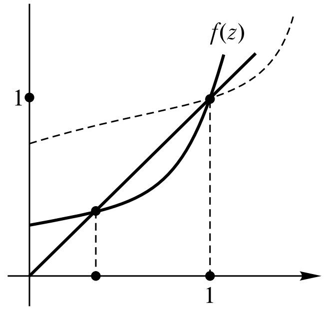

Fig. 7.1 Finding the extinction probability of a branching process: it is given by the smaller of the two solutions to the equation $z = f(z)$

Denote by $f_n(z)$ the n-th iterate of the function $f(z)$, i.e. $f_1(z) = f(z)$, $f_2(z) = f(f(z))$, $f_3(z) = f(f_2(z))$ and so on. Then we conclude from (7.7.2) by induction that the generating function of ζ_n is equal to the n-th iterate of $f(z)$:

$$\mathbf{E}z^{\zeta_n} = f_{(n)}(z).$$

From this one can easily obtain, by differentiating at the point $z = 1$, recursive relations for the moments of ζ_n.

How can one find the extinction probability of the process? By extinction we will understand the event that all ζ_n starting from some n will be equal to 0. (If $\zeta_n = 0$ then clearly $\zeta_{n+1} = \zeta_{n+2} = \cdots = 0$, because $\mathbf{P}(\zeta_{n+1} = 0 | \zeta_n = 0) = 1$.) Set $A_k = \{\zeta_k = 0\}$. Then extinction is the event $\bigcup_{k=1}^{\infty} A_k$. Since $A_n \subset A_{n+1}$, the extinction probability q is equal to $q = \lim_{n\to\infty} \mathbf{P}(A_n)$.

Theorem 7.7.1 *The extinction probability q is equal to the smallest nonnegative solution of the equation $q = f(q)$.*

Proof One has $\mathbf{P}(A_n) = f_n(0) \le 1$, and this sequence is non-increasing. Passing in the equality

$$f_{n+1}(0) = f\big(f_n(0)\big) \tag{7.7.3}$$

to the limit as $n \to \infty$, we obtain

$$q = f(q), \quad q \le 1.$$

This is an equation for the extinction probability. Let us analyse its solutions. The function $f(z)$ is convex (as $f''(z) \ge 0$) and non-decreasing in the domain $z \ge 0$ and $f'(1) = m$ is the mean number of offspring of a single particle. First assume that $\mathbf{P}(\xi_1^{(1)} = 1) < 1$. If $m \le 1$ then $f(z) > z$ for $z < 1$ and hence $q = 1$. If $m > 1$ then by convexity of f the equation $q = f(q)$ has exactly two solutions on the interval [0, 1]: $q_1 < 1$ and $q_2 = 1$ (see Fig. 7.1). Assume that $q = q_2 = 1$. Then the sequence $\delta_n = 1 - f_n(0)$ will monotonically converge to 0, and $f(1 - \delta_n) < 1 - \delta_n$ for sufficiently large n. Therefore, for such n,

$$\delta_{n+1} = 1 - f(1 - \delta_n) > \delta_n,$$

which is a contradiction as δ_n is a decreasing sequence. This means that $q = q_1 < 1$. Finally, in the case $\mathbf{P}(\xi_1^{(1)} = 1) = f_1 = 1$ one clearly has $f(z) \equiv z$ and $q = 0$. The theorem is proved. □

Now consider in more detail the case $m = 1$, which is called *critical*. We know that in this case the extinction probability q equals 1. Let $q_n = \mathbf{P}(A_n) = f_n(0)$ be the probability of extinction by time n. How fast does q_n converge to 1? By (7.7.3) one has $q_{n+1} = f(q_n)$. Therefore the probability $p_n = 1 - q_n$ of non-extinction of the process by time n satisfies the relation

$$p_{n+1} = g(p_n), \qquad g(x) = 1 - f(1 - x).$$

It is also clear that $\gamma_n = p_n - p_{n+1}$ is the probability that extinction will occur on step n.

Theorem 7.7.2 *If $m = f'(1) = 1$ and $0 < b := f''(1) < \infty$ then $\gamma_n \sim \frac{2}{bn^2}$ and $p_n \sim \frac{2}{bn}$ as $n \to \infty$.*

Proof If the second moment of the number of offspring of a single particle is finite ($b < \infty$) then the derivative $g''(0) = -b$ exists and therefore, since $g(0) = 0$ and $g'(0) = f'(1) = 1$, one has

$$g(x) = x - \frac{b}{2}x^2 + o\big(x^2\big), \quad x \to \infty.$$

Putting $x = p_n \to 0$, we find for the sequence $a_n = 1/p_n$ that

$$a_{n+1} - a_n = \frac{p_n - p_{n+1}}{p_n p_{n+1}} = \frac{bp_n^2(1 + o(1))}{2p_n^2(1 - bp_n/2 + o(p_n))} \to \frac{b}{2},$$

$$a_n = a_1 + \sum_{k=1}^{n-1}(a_{k+1} - a_k) \sim \frac{bn}{2}, \qquad p_n \sim \frac{2}{bn}.$$

The theorem is proved. □

Now consider the problem on the distribution of the number ζ_n of particles given $\zeta_n > 0$.

Theorem 7.7.3 *Under the assumptions of Theorem 7.7.2, the conditional distribution of $p_n\xi_n$ (or $2\zeta_n/(bn)$) given $\zeta_n > 0$ converges as $n \to \infty$ to the exponential distribution*:

$$\mathbf{P}(p_n\zeta_n > x | \zeta_n > 0) \to e^{-x}, \quad x > 0.$$

The above statement means, in particular, that given $\zeta_n > 0$, the number of particles ζ_n is of order n as $n \to \infty$.

Proof Consider the Laplace transform (see Property 6 in Sect. 7.1) of the conditional distribution of $p_n\zeta_n$ (given $\zeta_n > 0$):

$$\mathbf{E}\big(e^{-sp_n\zeta_n}|\zeta_n > 0\big) = \frac{1}{p_n}\sum_{k=1}^{\infty} e^{-skp_n}\mathbf{P}(\zeta_n = k). \tag{7.7.4}$$

We will make use of the fact that, if we could find an N such that $e^{-sp_n} = 1 - p_N$, which is the probability of extinction by time N, then the right-hand side of (7.7.4) will give, by the total probability formula, the conditional probability of the extinction of the process by time $n+N$ given its non-extinction at time n. We can evaluate this probability using Theorem 7.7.2.

Since $p_n \to 0$, for any fixed $s > 0$ one has

$$e^{-sp_n} - 1 \sim -sp_n \sim -\frac{2s}{bn}.$$

Clearly, one can always choose $N \sim n/s$, $s_n \sim s$, $s_n \downarrow s$ such that $e^{-s_n p_n} - 1 = -p_N$. Therefore $e^{-s_n p_n k} = (1-p_N)^k$ and the right-hand side of (7.7.4) can be rewritten for $s = s_n$ as

$$\begin{aligned}\frac{1}{p_n}\sum_{k=1}^{\infty}\mathbf{P}(\zeta_n = k)(1-p_N)^k &= \frac{1}{p_n}\mathbf{P}(\zeta_n > 0,\ \zeta_{n+N} = 0)\\ &= \frac{p_n - p_{n+N}}{p_n}\\ &= 1 - \frac{p_{n+N}}{p_n} \sim 1 - \frac{n}{n+N} = \frac{N}{n+N} \to \frac{1}{1+s}.\end{aligned}$$

Now note that

$$\mathbf{E}\big(e^{-sp_n\zeta_n}\big|\zeta_n > 0\big) - \mathbf{E}\big(e^{-s_n p_n\zeta_n}\big|\zeta_n > 0\big) = \mathbf{E}\big[e^{-sp_n\zeta_n}\big(1 - e^{-(s_n-s)p_n\zeta_n}\big|\zeta_n > 0\big)\big].$$

Since $e^{-\alpha} \le 1$ and $1 - e^{-\alpha} \le \alpha$ for $\alpha \ge 0$, and $\mathbf{E}\zeta_n = 1$, $\mathbf{E}(\zeta_n|\zeta_n > 0) = 1/p_n$, it is easily seen that the positive (since $s_n > s$) difference of the expectations in the last formula does not exceed

$$(s_n - s)p_n\mathbf{E}(\zeta_n|\zeta_n > 0) = s_n - s \to 0.$$

Therefore the Laplace transform (7.7.4) converges, as $n \to \infty$, to $1/(1+s)$. Since $1/(1+s)$ is the Laplace transform of the exponential distribution:

$$\int_0^{\infty} e^{-sx-x}\,dx = \frac{1}{1+s},$$

we conclude by the continuity theorem (see the remark after Theorem 7.3.1 in Sect. 7.3) that the conditional distribution of interest converges to the exponential law.[6]

In Sect. 15.4 (Example 15.4.1) we will obtain, as consequences of martingale convergence theorems, assertions about the behaviour of ζ_n as $n \to \infty$ for branching processes in the case $\mu > 1$ (the so-called *supercritical* processes). □

[6]The simple proof of Theorem 7.7.3 that we presented here is due to K.A. Borovkov.