Department of Probability and Statistics, School of Mathematics and Statistics, Xidian University

Alexander Iksanov

Characteristic functions and their relatives in probability theory

Lecture notes

Xi'an-2019

CONTENTS

INTRODUCTION

These lecture notes use material from various books and articles of other authors as well as lecture notes of the present author on the theory of characteristic functions published in 2004 in Ukrainian. It is not easy to list all the sources that were helpful in compiling this document. I only mention the sources Chapter 13 in [3], [7], Chapter 1 in [11] and Section 3.3 in [2] which are also highly recommended as further reading. Additional information about completely monotone functions and Laplace transforms can be found in the books [14] and [18]. More advanced material concerning characteristic functions is contained in the books [6, 9, 17].

The subject of these lecture notes is the use of generating functions, Laplace-Stieltjes transforms and characteristic functions (Fourier-Stieltjes tranforms) in probability theory. The general idea is that dealing with transforms is usually easier than with the underlying distributions. Thus, if we want to know some property of a probability distribution it is often helpful to reformulate the problem in terms of a transform, to solve a problem in the setting of transforms and than get back. The possibility of such an approach is justified by uniqueness theorems saying that a transform uniquely determines a distribution.

Here is an incomplete list of problems which can be effectively solved with the aid of transforms:

- If the underlying distribution is not known, it can be recovered provided that its transform is known.
- Transforms are useful in finding the distribution of a sum of independent random variables.
- Transforms are useful in calculating moments of the underlying distributions.
- Using continuity theorems, transforms are useful in finding limit distributions.
- Continuity properties of distributions are usually investigated with the help of characteristic functions.

One may wonder what is the reason to use several transforms rather than just one? This is justified by the range of applicability of these transforms. While generating functions are defined for random variables taking nonnegative integer values, Laplace transforms are defined for nonnegative, not necessarily discrete random variables. Finally, characteristic functions are defined for arbitrary random variables. Generating functions and Laplace transforms are real-valued infinitely differentiable functions. Furthermore, they are monotone, convex and have several other appealing features. Therefore, it is recommended to use these transforms whenever this is possible. Characteristic functions are complex-valued functions which are not necessarily differentiable and in general may exhibit an erratic behavior. These *have to* be used when underlying random variables take values of both signs. However, in some problems characteristic functions are also used for nonnegative random variables, both discrete and continuous. The reason is that characteristic functions contain more information than the other transforms. For instance, characteristic functions can be used to decide whether underlying distributions are absolutely continuous or continuous singular.

Chapter 1

Generating functions

1.1. Definition

Let ξ be a random variable taking nonnegative integer values. Set $p_k := \mathbb{P}\{\xi = k\}$ for $k \in \mathbb{N}_0$. These satisfy $p_k \geq 0$ and $\sum_{k\geq 0} p_k = 1$. The sequence $(p_k)_{k\in\mathbb{N}_0}$ is called *distribution* of the random variable ξ.

Definition 1. The generating function of the random variable ξ (or its distribution) is given by

$$f(s) := \mathbb{E}s^\xi = \sum_{k\geq 0} s^k p_k, \quad s \geq 0.$$

The series defining $f(s)$ trivially converges for each $s \in [0, 1)$ and furthermore, for such s, $f(s) \leq \sum_{k\geq 0} p_k = 1$. The function $s \mapsto \mathbb{E}s^\xi$ for $s > 1$ is not necessarily finite. Actually, the finiteness of $\mathbb{E}s^\xi$ for $s > 1$ is equivalent to the finiteness of the exponential moment $\mathbb{E}e^{\alpha\xi}$ for $\alpha = \log s$. If, for instance, $\lim_{k\to\infty} k^a p_k = b$ for some $a > 1$ and $b > 0$, then $\mathbb{E}s^\xi = \infty$ for all $s > 1$.

Assume that ξ takes nonnegative integer values and additionally allowance is made for $\mathbb{P}\{\xi = \infty\} > 0$, so that $\sum_{k\geq 1} p_k < 1$. Then the generating function $f(s) = \mathbb{E}s^\xi$ is still well-defined for $s \in (0, 1)$. For instance, if $\mathbb{P}\{\xi = 0\} = \mathbb{P}\{\xi = \infty\} = 1/2$, then $f(s) = 1/2$ for $s \in [0, 1)$.

1.2. Examples of generating functions

1. Let ξ have a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, that is,

$$\mathbb{P}\{\xi = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Then

$$f(s) = (ps + 1 - p)^n, \quad s \geq 0.$$

Proof. The binomial theorem tells us that, for $a, b \in \mathbb{R}$ and $n \in \mathbb{N}$,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

Hence, for $s \geq 0$, $f(s) = \sum_{k=0}^{n} \binom{n}{k} (ps)^k (1-p)^{n-k} = (ps + 1 - p)^n$. □

2. Let ξ have a Poisson distribution with parameter $\lambda > 0$, that is,

$$\mathbb{P}\{\xi = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

Then

$$f(s) = e^{-\lambda(1-s)}, \quad s \geq 0.$$

Proof. For $s \geq 0$, $f(s) = e^{-\lambda} \sum_{k \geq 0} \frac{(\lambda s)^k}{k!} = e^{-\lambda} e^{\lambda s} = e^{-\lambda(1-s)}$. □

3. Let ξ have a geometric distribution (starting at zero) with success probability $p \in (0,1)$, that is,

$$\mathbb{P}\{\xi = k\} = p(1-p)^k, \quad k \in \mathbb{N}_0.$$

Then

$$f(s) = \frac{p}{1-(1-p)s}, \quad s \in [0, (1-p)^{-1})$$

and $f(s) = +\infty$ for $s \geq (1-p)^{-1}$.

Proof. For $s \in [0, (1-p)^{-1})$, $f(s) = p \sum_{k \geq 0} ((1-p)s)^k = \frac{p}{1-(1-p)s}$. The condition $s < (1-p)^{-1}$ guarantees that the last series is the sum of infinite geometric progression with common ratio $(1-p)s < 1$ and, as such, is convergent. For $s \geq (1-p)^{-1}$, the common ratio $(1-p)s \geq 1$ which ensures that the series diverges. □

4. Let ξ have a negative binomial distribution with parameters $r > 0$ and $p \in (0,1)$, that is,

$$\mathbb{P}\{\xi = k\} = \frac{r(r+1) \cdot \ldots \cdot (r+k-1)}{k!} p^r (1-p)^k, \quad k \in \mathbb{N}_0.$$

Then

$$f(s) = \Big(\frac{p}{1-(1-p)s}\Big)^r, \quad s \in [0, (1-p)^{-1}).$$

Proof. According to a known generalization of the binomial theorem: for $c \in \mathbb{R}$ and $t \in (-1,1)$,

$$(1+t)^c = \sum_{k \geq 0} \frac{c(c-1) \cdot \ldots (c-k+1)}{k!} t^k = 1 + ct + \frac{c(c-1)}{2!} t^2 + \ldots \qquad (1.1)$$

Setting $c = -r$ and $t = -s$ we obtain

$$\begin{aligned}(1-s)^{-r} &= \sum_{k \geq 0} \frac{-r(-r-1) \cdot \ldots (-r-(k-1))}{k!} (-s)^k = \sum_{k \geq 0} \frac{r(r+1) \cdot \ldots (r+k-1)}{k!} s^k \\ &= 1 + rs + \frac{r(r+1)}{2!} s^2 + \ldots\end{aligned}$$

Using this formula yields

$$f(s) = p^r \sum_{k\geq 0} \frac{r(r+1)\cdot\ldots(r+k-1)}{k!}((1-p)s)^k = \Big(\frac{p}{1-(1-p)s}\Big)^r, \quad s \in [0, 1/(1-p)).$$

□

5. Let ξ have a logarithmic distribution with parameter $p \in (0,1)$, that is,

$$\mathbb{P}\{\xi = k\} = \frac{(1-p)^k}{\log(1/p)k}, \quad k \in \mathbb{N}.$$

Then

$$f(s) = \frac{\log(1-(1-p)s)}{\log p}, \quad s \in [0, (1-p)^{-1}).$$

Proof. Recalling the Maclaurin series expansion

$$-\log(1-x) = \sum_{n\geq 1} \frac{x^n}{n}, \quad x \in [0,1)$$

we obtain, for $s \in [0, 1/(1-p))$,

$$f(s) = \frac{1}{\log(1/p)} \sum_{k\geq 1} k^{-1}((1-p)s)^k = \frac{\log(1-(1-p)s)}{\log p}.$$

□

1.3. Properties of generating functions

1.3.1. Uniqueness theorem

Theorem 2. *The generating function uniquely determines the distribution. In other words, the generating functions of different distributions are different.*

Proof. Differentiating formally the power series defining $f(s)$ we obtain

$$f'(s) = \sum_{k\geq 1} s^{k-1} k p_k, \quad f''(s) = \sum_{k\geq 2} s^{k-2} k(k-1) p_k, \quad s \in [0,1)$$

and more generally, for $n \in \mathbb{N}$, (with $f^{(n)}$ denoting the nth derivative)

$$f^{(n)}(s) = \sum_{k\geq n} s^{k-n} k(k-1)\cdot\ldots\cdot(k-n+1) p_k, \quad s \in [0,1). \tag{1.2}$$

Of course, at $s=0$ the derivatives above are the right derivatives. Substituting now $s=0$ we infer $f^{(n)}(0) = n!p_n$, whence

$$p_n = \frac{f^{(n)}(0)}{n!}, \quad n \in \mathbb{N}.$$

This taken together with the trivial fact that $f(0) = p_0$ justifies the claim.

We are left with verifying that our formal differentiation was correct. Recall that in general interchanging the orders of differentiation (passage to the limit) and summation calls for justification. To explain the idea we only do it for the derivative of the first order. Write

$$f'(s) = \lim_{h\to 0} \frac{f(s+h)-f(s)}{h} = \lim_{h\to 0} \sum_{k\geq 1} \frac{(s+h)^k - s^k}{h} p_k.$$

Further, we want to interchange the passage to the limit and the summation. There are several sufficient conditions justifying the possibility of such an interchange. We are going to use the Lebesgue dominated convergence theorem which gives one collection of sufficient conditions. According to this theorem it suffices to find a sequence $(a_k)_{k\in\mathbb{N}}$ called *integrable or summable majorant* which does not depend on h and satisfies

$$\Big|\frac{(s+h)^k - s^k}{h} p_k\Big| \leq a_k \quad \text{for all} \quad k \in \mathbb{N}_0 \tag{1.3}$$

and

$$\sum_{k\geq 0} a_k < \infty. \tag{1.4}$$

Fix any $s \in [0,1)$. Assume first that $h > 0$. Since we shall send $h \to 0$ there is no loss of generality in assuming that $s + h \leq \gamma$ for some $\gamma \in (s,1)$. Write now

$$\frac{(s+h)^k - s^k}{h} = \frac{k}{h}\int_s^{s+h} y^{k-1}\mathrm{d}y \leq k(s+h)^{k-1} \leq k\gamma^{k-1}, \quad k \in \mathbb{N}.$$

Since $\sum_{k\geq 1} k\gamma^{k-1}p_k \leq \sum_{k\geq 1} k\gamma^{k-1} = (1-\gamma)^{-2} < \infty$, we conclude that we may take $a_k = k\gamma^{k-1}p_k$ for $k \in \mathbb{N}$ as a summable majorant. Fix now $s \in (0,1)$ and assume that $h < 0$, yet $s + h > 0$. We have

$$\Big|\frac{(s+h)^k - s^k}{h} p_k\Big| = \frac{s^k - (s+h)^k}{-h} p_k = \frac{k}{-h}\int_{s+h}^{s} y^{k-1}\mathrm{d}y \leq k(s+h)^{k-1} \leq ks^{k-1}, \quad k \in \mathbb{N}.$$

Arguing as above we infer that $a_k = ks^{k-1}p_k$ for $k \in \mathbb{N}$ is a summable majorant. Thus,

$$f'(s) = \lim_{h\to 0}\sum_{k\geq 1}\frac{(s+h)^k - s^k}{h}p_k = \sum_{k\geq 1}\lim_{h\to 0}\frac{(s+h)^k - s^k}{h}p_k = \sum_{k\geq 1}(s^k)'p_k = \sum_{k\geq 1} s^{k-1}kp_k.$$

The proof is complete. □

1.3.2. Generating functions and moments

Proposition 3. *Both sides of the equality that follows are either finite or infinite: for* $n \in \mathbb{N}$,

$$\mathbb{E}\xi(\xi-1)\cdot\ldots\cdot(\xi-n+1) = f^{(n)}(1-), \tag{1.5}$$

where $f^{(n)}(1-) = \lim_{s\to 1-0} f^{(n)}(s)$ *(the left limit at one).*

Proof. Equality (1.2) entails

$$\begin{aligned} f^{(n)}(1-) &= \lim_{s\to 1-0}\sum_{k\geq n} s^{k-n}k(k-1)\cdot\ldots\cdot(k-n+1)p_k \\ &= \sum_{k\geq n}\lim_{s\to 1-0} s^{k-n}k(k-1)\cdot\ldots\cdot(k-n+1)p_k \\ &= \sum_{k\geq n} k(k-1)\cdot\ldots\cdot(k-n+1)p_k. \end{aligned}$$

The interchange of the limit and the summation is justified by the Lévy monotone convergence theorem. It states that

$$\lim_{s\to s_0}\sum_n f_n(s)=\sum_n \lim_{s\to s_0} f_n(s),$$

whenever the functions f_n are nonnegative and monotone, and s_0 may be finite or infinite. Note that the series on the right-hand side may be infinite. In our situation the cited theorem applies because the functions $s\mapsto s^{k-n}$ are increasing on $[0,\infty)$. □

As an immediate consequence we obtain the following.

Corollary 4. $\mathbb{E}\xi=f'(1-)$, $\operatorname{Var}\xi=f''(1-)+f'(1-)-(f'(1-))^2$.

Proof. The formula for $\mathbb{E}\xi$ follows from (1.5) with $n=1$. Another appeal to (1.5) with $n=2$ shows that $f''(1-)=\mathbb{E}\xi(\xi-1)$. Hence,

$$f''(1-)+f'(1-)-(f'(1-))^2=\mathbb{E}\xi(\xi-1)+\mathbb{E}\xi-(\mathbb{E}\xi)^2=\mathbb{E}\xi^2-(\mathbb{E}\xi)^2=\operatorname{Var}\xi.$$

□

Example 5. Let ξ have a Poisson distribution with parameter $\lambda>0$. Then $f(s)=e^{-\lambda(1-s)}$ for $s\geq 0$, whence $f^{(n)}(s)=\lambda^n e^{-\lambda(1-s)}$ for $s\geq 0$ and $n\in\mathbb{N}$. In particular,

$$\mathbb{E}\xi(\xi-1)\cdot\ldots\cdot(\xi-n+1)=f^{(n)}(1)=\lambda^n,\quad n\in\mathbb{N},$$

so that $\mathbb{E}\xi=\operatorname{Var}\xi=\lambda$.

Sometimes the formula given below can be effectively used.

Proposition 6. *Let f be the generating function of a random variable ξ. Then*

$$\sum_{k\geq 0}s^k\mathbb{P}\{\xi>k\}=\frac{1-f(s)}{1-s},\quad s\in[0,1).$$

Proof. In what follows, for an event A we write $\mathbb{1}_A$ for a random variable which takes value 1 with probability $\mathbb{P}(A)$ and value 0 with probability $1-\mathbb{P}(A)$. In particular,

$$\mathbb{E}\mathbb{1}_A=\mathbb{P}(A).$$

Also, recall that, for any $u\in[0,1)$ and any $n\in\mathbb{N}$,

$$1+u+u^2+\ldots+u^{n-1}=\frac{1-u^n}{1-u}.$$

For $s\in[0,1)$, we have

$$\sum_{k\geq 0}s^k\mathbb{P}\{\xi>k\}=\sum_{k\geq 0}s^k\mathbb{E}\mathbb{1}_{\{\xi\geq k+1\}}=\mathbb{E}\sum_{k=0}^{\xi-1}s^k=\mathbb{E}\frac{1-s^\xi}{1-s}=\frac{1-f(s)}{1-s}.$$

Here, the sum $\sum_{k=0}^{\xi-1}$ is interpreted as 0 when $\xi=0$. For the second equality we have used Fubini's theorem which essentially states that to calculate a repeated integral of a positive function the corresponding single integrals can be dealt with in an arbitrary order. □

As a corollary, we can give an alternative derivation of the formula for the expectation. We first prove that

$$\mathbb{E}\xi = \sum_{k\geq 1} \mathbb{P}\{\xi \geq k\} = \sum_{k\geq 0} \mathbb{P}\{\xi > k\}, \tag{1.6}$$

where all parts of the equality may be infinite. Indeed,

$$\sum_{k\geq 1} \mathbb{P}\{\xi \geq k\} = \sum_{k\geq 1} \mathbb{E}\mathbb{1}_{\{\xi\geq k\}} = \mathbb{E}\sum_{k=1}^{\xi} 1 = \mathbb{E}\xi.$$

By Lévy's monotone convergence theorem,

$$\lim_{s\to 1-} \sum_{k\geq 0} s^k \mathbb{P}\{\xi > k\} = \sum_{k\geq 0} \lim_{s\to 1-} s^k \mathbb{P}\{\xi > k\} = \sum_{k\geq 0} \mathbb{P}\{\xi > k\} = \mathbb{E}\xi$$

having utilized (1.6) for the last equality. On the other hand,

$$\lim_{s\to 1-} \frac{1-f(s)}{1-s} = \lim_{h\to 0+} \frac{f(1)-f(1-h)}{h} = f'(1-).$$

Using Proposition 6 we finally infer

$$\mathbb{E}\xi = f'(1-)$$

which is the first part of Corollary 4.

1.3.3. Generating functions and convolutions Let ξ_1 and ξ_2 be independent random variables with distributions $(p_k)_{k\in\mathbb{N}_0}$ and $(q_k)_{k\in\mathbb{N}_0}$. Recall that this means that, for $k \in \mathbb{N}_0$, $p_k = \mathbb{P}\{\xi_1 = k\}$, $q_k = \mathbb{P}\{\xi_2 = k\}$, and $p_k, q_k \geq 0$, $\sum_{k\geq 0} p_k = \sum_{k\geq 0} q_k = 1$.

The distribution of the sum $\xi_1 + \xi_2$ is given by

$$\mathbb{P}\{\xi_1 + \xi_2 = n\} = \sum_{k=0}^{n} p_k q_{n-k} =: r_n, \quad n \in \mathbb{N}_0. \tag{1.7}$$

Definition 7. The sequence $(r_n)_{n\in\mathbb{N}_0}$ is called *convolution* of the distributions $(p_k)_{k\in\mathbb{N}_0}$ and $(q_k)_{k\in\mathbb{N}_0}$.

Proof of (1.7). We use the total probability formula: for $n \in \mathbb{N}_0$,

$$\begin{aligned}
\mathbb{P}\{\xi_1 + \xi_2 = n\} &= \sum_{k=0}^{n} \mathbb{P}\{\xi_1 + \xi_2 = n | \xi_1 = k\}\mathbb{P}\{\xi_1 = k\} \\
&= \sum_{k=0}^{n} \mathbb{P}\{\xi_2 = n-k | \xi_1 = k\}\mathbb{P}\{\xi_1 = k\} = \sum_{k=0}^{n} \mathbb{P}\{\xi_2 = n-k\}\mathbb{P}\{\xi_1 = k\} \\
&= \sum_{k=0}^{n} q_{n-k} p_k = \sum_{k=0}^{n} p_k q_{n-k}.
\end{aligned}$$

The third equality follows from the independence of ξ_1 and ξ_2. Recall that $\mathbb{P}(A|B) = \mathbb{P}(A)$ whenever the events A and B are independent, and $\mathbb{P}(B) > 0$. □

The convolution is a non-trivial operation. It is rarely possible to find the convolution explicitly. Using generating functions instead of distributions gives a great advantage as the following proposition demonstrates.

Proposition 8. *Let ξ_1 and ξ_2 be independent random variables with the generating functions f_1 and f_2. Then the generating function of $\xi_1 + \xi_2$ is $f_1 f_2$.*

Proof. For $s \geq 0$, $\mathbb{E}s^{\xi_1+\xi_2} = \mathbb{E}s^{\xi_1}s^{\xi_2} = \mathbb{E}s^{\xi_1}\mathbb{E}s^{\xi_2} = f_1(s)f_2(s)$. For the second equality we have used the fact that the random variables s^{ξ_1} and s^{ξ_2} are independent, and that the expectation of the product of independent random variables is equal to the product of the expectations. □

Example 9. Let ξ_1 and ξ_2 be independent random variables having Poisson distribution with parameters $\lambda > 0$ and $\mu > 0$, respectively. Then $\xi_1 + \xi_2$ has a Poisson distribution with parameter $\lambda + \mu$.

Proof. We give two proofs.
Proof via generating functions. We already know that, for $s \geq 0$, $f_1(s) = \mathbb{E}s^{\xi_1} = e^{-\lambda(1-s)}$ and $f_2(s) = \mathbb{E}s^{\xi_2} = e^{-\mu(1-s)}$. By Proposition 8, $\mathbb{E}s^{\xi_1+\xi_2} = f_1(s)f_2(s) = e^{-(\lambda+\mu)(1-s)}$. This is nothing else but the generating function of a Poisson distribution with parameter $\lambda + \mu$.
Direct proof via distributions. Using (1.7) yields, for $n \in \mathbb{N}_0$,

$$\mathbb{P}\{\xi_1 + \xi_2 = n\} = \sum_{k=0}^{n} e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!} = \frac{e^{-\lambda-\mu}}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda^k\mu^{n-k} = e^{-\lambda-\mu}\frac{(\lambda+\mu)^n}{n!}$$

having utilized the binomial theorem at the last step. Thus, $\xi_1 + \xi_2$ has a Poisson distribution with parameter $\lambda + \mu$. □

With the help of Proposition 8 we can give an interpretation of the binomial distribution and the negative binomial distribution as the distributions of sums of independent identically distributed random variables.

Example 10. Fix any $n \in \mathbb{N}$. Let $\xi_1, \ldots, \xi_n$ be independent identically distributed random variables with the distribution $\mathbb{P}\{\xi_1 = 1\} = p$ and $\mathbb{P}\{\xi_1 = 0\} = 1 - p$ for $p \in (0, 1)$. Then $\xi_1 + \ldots + \xi_n$ has a binomial distribution with parameters n and p.

Proof. It is clear that $f(s) = \mathbb{E}s^{\xi_1} = ps + 1 - p$ for $s \geq 0$. By Proposition 8, $\mathbb{E}s^{\xi_1+\ldots+\xi_n} = f^n(s) = (ps + 1 - p)^n$. This is the generating function of a binomial distribution with parameters n and p. □

Example 11. Fix any $r \in \mathbb{N}$. Let $\xi_1, \ldots, \xi_r$ be independent identically distributed random variables with a geometric distribution with success probability $p \in (0, 1)$, that is, $\mathbb{P}\{\xi_1 = k\} = p(1-p)^k$, $k \in \mathbb{N}_0$. Then $\xi_1 + \ldots + \xi_r$ has a negative binomial distribution with parameters r and p.

Proof. Since $\mathbb{E}s^{\xi_1} = p/(1 - (1-p)s)$ for $s \in [0, 1/(1-p))$ an application of Proposition 8 leads to the conclusion

$$\mathbb{E}s^{\xi_1+\ldots+\xi_r} = \Big(\frac{p}{1-(1-p)s}\Big)^r, \quad s \in [0, 1/(1-p)).$$

This is the generating function of a negative binomial distribution with parameters r and p, see Example 4 in Section 1.2.. $\square$

Recall that a common interpretation of a random variable ξ having a geometric distribution (starting at zero) with success probability p is that it represents the number of failures preceding the first success in the sequence of independent trials with success probability p. From this it is clear that $\xi_1 + \ldots + \xi_r$ defined in the last example can be thought of as the number of failures preceding the rth success in the sequence of independent trials with success probability p. This gives an interpretation of a negative binomial distribution with positive integer parameter r.

1.3.4. Generating functions of compound distributions

Proposition 12. *Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables taking nonnegative integer values which are independent of a random variable N also taking nonnegative integer values. If the generating functions of ξ_1 and N are $f(s)$ and $g(s)$, respectively, then the generating function of*[1] *$\sum_{k=1}^N \xi_k$ is $g(f(s))$.*

Proof. Using the total expectation formula we obtain, for $s \geq 0$,

$$\mathbb{E}s^{\sum_{k=1}^N \xi_k} = \sum_{n\geq 0} \mathbb{E}\Big(s^{\sum_{k=1}^N \xi_k}\mathbb{1}_{\{N=n\}}\Big) = \sum_{n\geq 0} \mathbb{E}\Big(s^{\sum_{k=1}^n \xi_k}\mathbb{1}_{\{N=n\}}\Big) = \sum_{n\geq 0} \mathbb{E}s^{\sum_{k=1}^n \xi_k}\mathbb{P}\{N=n\}$$
$$= \sum_{n\geq 0} f^n(s)\mathbb{P}\{N=n\} = g(f(s)).$$

Here, the third equality is guaranteed by the independence, and the fourth is a consequence of Proposition 8. $\square$

Definition 13. The distribution of the random sum $\sum_{k=1}^N \xi_k$ as in Proposition 12 is called *compound* distribution. If the distribution of N is Poisson, then the distribution of $\sum_{k=1}^N \xi_k$ is called *compound Poisson.* If the distribution of N is geometric, then the distribution of $\sum_{k=1}^N \xi_k$ is called *compound geometric.*

From Proposition 12 it follows that the generating function of a compound Poisson distribution is $e^{-\lambda(1-f(s))}$, where $\lambda > 0$ is the parameter of the underlying Poisson distribution, and that the generating function of a compound geometric distribution is $p/(1-(1-p)f(s))$, where $p \in (0,1)$ is the success probability of the underlying geometric distribution.

Example 14. The number of orders that a delivery service for pizzas receives on the given day has a Poisson distribution with parameter $\lambda > 0$. The operator got sick and there is only probability $p \in (0,1)$ that she notes the address of a caller correctly. What is the distribution of the number of pizzas successfully delivered on the given day?

Proof. Let $\xi_k = 1$ if the address of the kth caller is correctly noted, and $\xi_k = 0$, otherwise. Then $\mathbb{P}\{\xi_k = 1\} = p$, $\mathbb{P}\{\xi_k = 0\} = 1-p$, so that $f(s) = \mathbb{E}s^{\xi_k} = ps+1-p$ for $s \geq 0$. Denote by

[1] Here and hereafter, the sum $\sum_{k=1}^0$ is interpreted as 0.

N the number of orders. By assumption, N has a Poisson distribution with parameter λ. It is natural to assume that $\xi_1, \xi_2, \dots$ and N are independent. We are looking for the distribution of $\sum_{k=1}^N \xi_k$. By Proposition 12 (or just by the previous remark), its generating function is $e^{-\lambda(1-f(s))} = e^{-\lambda p(1-s)}$. Hence, the distribution in question is Poisson with parameter λp. □

Example 15. The negative binomial distribution is a compound Poisson distribution.

Proof. The claim is equivalent to saying that the generating functions of these distributions are the same, that is, for $p \in (0,1)$, $r, \lambda > 0$,

$$\left(\frac{p}{1-(1-p)s}\right)^r = e^{-\lambda(1-f(s))}, \quad s \in [0,1)$$

Thus, we have to point out the value of $\lambda > 0$ and prove that the function $f(s)$ defined by this equality is a generating function of some distribution. Setting $\lambda = r\log(1/p)$ and passing to logarithms we obtain

$$-r\log(1/p)\left(1 - \frac{\log(1-(1-p)s)}{\log p}\right) = -r\log(1/p)(1-f(s)).$$

Thus,

$$f(s) = \frac{\log(1-(1-p)s)}{\log p}, \quad s \in [0, 1/(1-p)).$$

This is the generating function of a logarithmic distribution with parameter p, see Example 5 in Section 1.2..

□

1.4. The Galton-Watson branching process

1.4.1. Formula for generating function Let ξ be a random variable taking nonnegative integer values with distribution $p_k = \mathbb{P}\{\xi = k\}$, $k \in \mathbb{N}_0$ and generating function f.

To give an informal description of the Galton-Watson branching process imagine the evolution of the following population. The 0th generation of the population is given by the initial ancestor. She gives birth to offspring which form the 1st generation of the population, their number being a copy of ξ. Each of the 1st generation individuals, independently of each other and the initial ancestor, produces offspring, the number of these being again a copy of ξ. All the children of the 1st generation individuals form the 2nd generation. The subsequent evolution of the population is analogous. The most important feature is that all individuals reproduce independently of each other.

Formally, denote by $(\xi_{1,j})_{j\in\mathbb{N}}$, $(\xi_{2,j})_{j\in\mathbb{N}}, \dots$ independent copies of the random variable ξ. For $n \in \mathbb{N}_0$, denote by Z_n the number of individuals in the nth generation of the population. Then

$$Z_0 = 1,$$

$$Z_1 = \xi_{1,Z_0} = \xi_{1,1},$$

$$Z_2 = \xi_{2,1} + \xi_{2,2} + \ldots + \xi_{2,Z_1},$$

$$\ldots$$

$$Z_n = \xi_{n,1} + \xi_{n,2} + \ldots + \xi_{n,Z_{n-1}},$$

$$\ldots$$

Enumerating the individuals of the $(i-1)$st generation in some way the random variable $\xi_{i,j}$ can be thought of as the number of offspring of the jth individual in the $(i-1)$st generation.

Definition 16. The random sequence $(Z_n)_{n\in\mathbb{N}_0}$ is called *Galton-Watson branching process.*

For $n \in \mathbb{N}_0$, set $f_n(s) := \mathbb{E}s^{Z_n}$, $s \geq 0$ the generating function of Z_n. Obviously, $f_0(s) = s$ and $f_1(s) = f(s)$.

Proposition 17. *For $n \in \mathbb{N}\backslash\{1\}$,*

$$f_n(s) = \underbrace{f \circ \ldots \circ f}_{n \text{ times}}(s), \tag{1.8}$$

where $\circ$ denotes composition. In particular, $f_2(s) = f \circ f(s) = f(f(s))$, $f_3(s) = f \circ f \circ f(s) = f(f(f(s)))$ and so on.

Proof. The distribution of $Z_n = \sum_{k=1}^{Z_{n-1}} \xi_{n,k}$ is compound. Therefore, by Proposition 12,

$$f_n(s) = f_{n-1}(f(s)) = f_{n-1} \circ f(s), \quad n \geq 2 \tag{1.9}$$

because the generating function of Z_{n-1} is $f_{n-1}(s)$.

To prove (1.8) we use the mathematical induction. If $n = 2$, then $f_2(s) = f \circ f(s)$ by (1.9). Assume that (1.8) holds with $n = k$, that is,

$$f_k(s) = \underbrace{f \circ \ldots \circ f}_{k \text{ times}}(s).$$

Then another appeal to (1.9) yields

$$f_{k+1}(s) = f_k \circ f(s) = \underbrace{f \circ \ldots \circ f}_{k \text{ times}} \circ f(s) = \underbrace{f \circ \ldots \circ f}_{(k+1) \text{ times}}(s),$$

thereby completing the proof. □

Example 18. In the situation that every individual gives birth to either one child with probability $p \in (0,1)$ or does not have offspring at all with probability $1-p$, the generating function of Z_n can be calculated explicitly. Indeed,

$$f_1(s) = f(s) = ps + 1 - p,$$

$$f_2(s) = p(ps + 1 - p) + 1 - p = p^2 s + (p+1)(1-p),$$

$$\ldots$$

$$f_n(s) = p^n s + (p^{n-1} + p^{n-2} + \ldots + 1)(1-p)$$

for $n \in \mathbb{N}$.

1.4.2. Moments

Lemma 19. *Assume that* $\mathbb{E}\xi = \mu \in (0,\infty)$. *Then* $\mathbb{E}Z_n = \mu^n$ *for* $n \in \mathbb{N}$.

Proof. We know that $\mathbb{E}s^{Z_n} = f_n(s) = f_{n-1}(f(s))$ for $n \in \mathbb{N}$ and $s \in [0,1)$. Differentiating we obtain $f_n'(s) = f_{n-1}'(f(s))f'(s)$ for $s \in (0,1)$. Letting $s \to 1-$ and using Corollary 1.5 we infer

$$\mathbb{E}Z_n = f_n'(1-) = f_{n-1}'(1)f'(1-) = \mathbb{E}Z_{n-1}\mathbb{E}\xi = \mathbb{E}Z_{n-1}\mu$$

for each $n \in \mathbb{N}$. Iterating this yields

$$\mathbb{E}Z_n = \mathbb{E}Z_{n-1}\mu = \mathbb{E}Z_{n-2}\mu^2 = \ldots = \mathbb{E}Z_1\mu^{n-1} = \mu^n.$$

□

In a similar but more tedious way one can check that when $\sigma^2 = \operatorname{Var}\xi \in (0,\infty)$ we have

$$\operatorname{Var} Z_n = \begin{cases} \sigma^2\mu^{n-1}\frac{1-\mu^n}{1-\mu}, & \text{if } \mu \neq 1, \\ \sigma^2 n, & \text{if } \mu = 1. \end{cases}$$

This is left as an exercise.

1.4.3. Extinction probability Let us find the extinction probability of the Galton-Watson process, that is, $\rho := \mathbb{P}(A)$, where $A = \{\text{the population dies out}\}$. Plainly, the population dies out if, and only if, some generation has no individuals. Thus, setting, for $n \in \mathbb{N}_0$, $A_n = \{\text{the } n\text{th generation has no individuals}\}$ we infer $A = \cup_{n\geq 0}A_n$. Observe that $A_n \subseteq A_{n+1}$ because if the nth generation has no individuals, so does the $(n+1)$st generation. Hence,

$$\mathbb{P}(A) = \mathbb{P}(\cup_{n\geq 0}A_n) = \lim_{n\to\infty}\mathbb{P}(A_n). \tag{1.10}$$

To check this, observe that $A_0 = \oslash$ and $\cup_{j\geq 0}A_n = \cup_{j\geq 0}(A_{j+1}\backslash A_j)$. Since the events $A_1\backslash A_0$, $A_2\backslash A_1, \ldots$ are disjoint, we infer with the help of σ-additivity of $\mathbb{P}$:

$$\mathbb{P}(\cup_{j\geq 0}A_j) = \mathbb{P}(\cup_{j\geq 0}(A_{j+1}\backslash A_j)) = \sum_{j\geq 0}\mathbb{P}(A_{j+1}\backslash A_j) = \lim_{n\to\infty}\sum_{j=0}^{n-1}\mathbb{P}(A_{j+1}\backslash A_j) = \lim_{n\to\infty}\mathbb{P}(A_n).$$

If $\mathbb{P}\{\xi = 0\} = 0$, then each individual has a positive number of offspring, hence the extinction does not occur, that is, $\rho = 0$. If $\mathbb{P}\{\xi = 1\} = 1$, then there is exactly one individual in each generation, hence the population survives forever. These two cases are somewhat degenerate. If these conditions do not hold we have the following.

Theorem 20. *Assume that* $\mathbb{P}\{\xi = 0\} > 0$, $\mathbb{P}\{\xi = 1\} < 1$ *and* $\mu = \mathbb{E}\xi \in (0,\infty)$. *If* $\mu \leq 1$, *then* $\rho = 1$. *If* $\mu > 1$, *then* $\rho \in (0,1)$ *is the smallest solution to the equation* $f(s) = s$ *on* $[0,1]$.

Proof. We first show that

$$f(\rho) = \rho. \tag{1.11}$$

For $n \in \mathbb{N}_0$, set $\rho_n = \mathbb{P}(A_n) = \mathbb{P}(Z_n = 0) = f_n(0)$, where $f_n(s)$ is the generating function of Z_n. According to (1.10), $\lim_{n\to\infty} \rho_n = \rho$ and thereupon

$$\lim_{n\to\infty} f(\rho_n) = f(\rho) \tag{1.12}$$

by continuity of f. On the other hand, $f(\rho_n) = f(f_n(0)) = f_{n+1}(0) = \rho_{n+1}$, whence

$$\lim_{n\to\infty} f(\rho_n) = \rho. \tag{1.13}$$

Comparing (1.12) and (1.13) we arrive at (1.11).

Next, we show that ρ is the smallest solution to $f(s) = s$ on $[0, 1]$. Let x be any solution to the equation $f(s) = s$ on $[0, 1]$. Then $x \geq 0$ entails $x = f(x) \geq f(0) = \rho_1$ by monotonicity of f. Similarly, $x = f(x) \geq f(\rho_1) = \rho_2$ and, more generally, $x \geq \rho_n$ for each $n \in \mathbb{N}$. Therefore, $x \geq \rho$, and the claim follows.

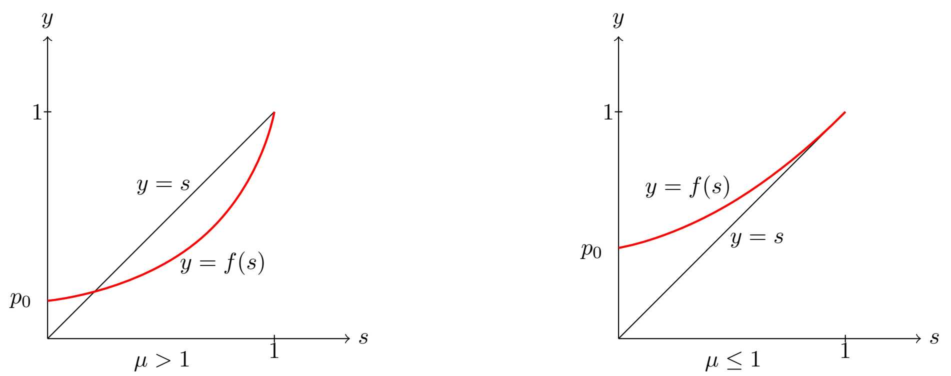

Figure 1.1: Graphs of $y = f(s)$

To proceed, we note that since $f''(s) = \sum_{k\geq 2} k(k-1)p_k s^{k-2} \geq 0$ for $s \in [0, 1)$, $f(s)$ is a convex function on $[0, 1)$. In view of this and the fact that $f(0) = p_0 > 0$, the graphs $y = f(s)$ and $y = s$ have at most two points in common on $[0, 1]$. One of these is $s = 1$. If $f'(1-) = \mu \leq 1$, then in a left neighborhood of 1 the graph of $y = f(s)$ cannot be below that of $y = s$. Hence, by convexity of $f(s)$ the only intersection is $s = 1$. If $f'(1-) = \mu > 1$, then in a left neighborhood of 1 the graph of $y = f(s)$ is below that of $y = s$. Hence, there must be an additional intersection to the left of 1 of the two graphs. See Figure 1.1. □

Example 21. Let ξ have distribution $\mathbb{P}\{\xi = 1\} = p \in (0, 1)$ and $\mathbb{P}\{\xi = 0\} = 1 - p$. Then the corresponding Galton-Watson process dies out with probability 1. Indeed, the equation $f(s) = s$ reads $ps + 1 - p = s$. This has the only solution $s = 1$ which agrees with the fact that $\mathbb{E}\xi = p < 1$.

Example 22. The secretary works with documents. Each document requires 3 minutes of her attention. While a document is being treated two, one or no new documents arrive with probabilities 3/5, 1/5 and 1/5, respectively. The secretary cannot take a coffee break until no documents remain. If present conditions persist, what is the probability that the secretary ever takes a coffee break?

Proof. We have to find the extinction probability ρ of the Galton-Watson process with $\mathbb{P}\{\xi = 2\} = 3/5$, $\mathbb{P}\{\xi = 1\} = \mathbb{P}\{\xi = 0\} = 1/5$. In this case $\mathbb{E}\xi = 7/5 > 1$. Therefore, we are looking for the smallest solution to the equation

$$3s^2/5 + s/5 + 1/5 = s$$

on $[0, 1]$. The equation has solutions $s = 1/3$ and $s = 1$. Hence, $\rho = 1/3$. □

1.5. Continuity theorem for generating functions

Let $\xi_1, \xi_2, \ldots$ be some nonnegative integer-valued random variables with distributions $\mathbb{P}\{\xi_n = k\} = p_k^{(n)}$ for $n \in \mathbb{N}$ and $k \in \mathbb{N}_0$. For $n \in \mathbb{N}$, denote by $F_n(x)$ the distribution function of ξ_n. Then $F_n(x) = \sum_{k=0}^{\lfloor x \rfloor} p_k^{(n)}$ for $x \geq 0$ and $F_n(x) = 0$ for $x < 0$.

Definition 23. The random variables ξ_n are said to *converge in distribution* to a random variable ξ (notation $\xi_n \overset{\mathrm{d}}{\to} \xi$) if

$$\lim_{n\to\infty} p_k^{(n)} = p_k =: \mathbb{P}\{\xi = k\} \quad \text{for each } k \in \mathbb{N}_0 \tag{1.14}$$

or equivalently if the distribution functions F_n *weakly converge* to F, that is, $\lim_{n\to\infty} F_n(x) = F(x) =: \mathbb{P}\{\xi \leq x\}$ at every x which is a continuity point of F.

Since we are considering nonnegative integer-valued random variables the set of discontinuity points of F is a subset of $\mathbb{N}_0$.

Checking relation (1.14) may be problematic. Is it sufficient instead to prove convergence of the corresponding generating functions? The answer is 'yes', as is stated in the result given next which is called *continuity theorem for generating functions*. The existence of such a result is one confirmation of the usefulness of generating functions in probability theory.

For $n \in \mathbb{N}$, set $f^{(n)}(s) := \mathbb{E}s^{\xi_n}$ for $s \geq 0$.

Theorem 24. *A sequence $(p_k)_{k\in\mathbb{N}_0}$ satisfying $\lim_{n\to\infty} p_k^{(n)} = p_k$ for each $k \in \mathbb{N}_0$ exists if, and only if, there exists a function f such that $\lim_{n\to\infty} f^{(n)}(s) = f(s)$ for $s \in [0, 1)$. Furthermore, $(p_k)_{k\in\mathbb{N}_0}$ is the distribution of a proper random variable and $f(s)$ is the generating function of this distribution if, and only if, $\lim_{s\to 1-} f(s) = 1$.*

Proof. Assume that $\lim_{n\to\infty} p_k^{(n)} = p_k$ for each $k \in \mathbb{N}_0$. Set $f(s) = \sum_{k\geq 0} s^k p_k$ for $s \in [0, 1)$. It is possible that $\sum_{k\geq 0} p_k < 1$.

Write, for $s \in [0, 1)$ and each fixed $m \in \mathbb{N}$,

$$|f^{(n)}(s) - f(s)| = \Big|\sum_{k\geq 0} s^k (p_k^{(n)} - p_k)\Big| \leq \sum_{k=0}^{m} s^k |p_k^{(n)} - p_k| + 2\sum_{k\geq m+1} s^k$$

having utilized the crude estimate $|p_k^{(n)} - p_k| \leq 2$ for the second summand on the right-hand side. By assumption,

$$\limsup_{n\to\infty} |f^{(n)}(s) - f(s)| \leq 2s^{m+1}(1 - s)^{-1}.$$

Letting now $m \to \infty$ completes the proof of $\lim_{n\to\infty} f^{(n)}(s) = f(s)$ for $s \in [0,1)$. Noting that $\lim_{s\to 1-} f(s) = \sum_{k\geq 0} p_k$ justifies the last claim of the theorem.

The proof of the other implication is more involved and will not be given. We refer to p. 29 in [11]. □

1.6. The law of rare events

Theorem 25. *Assume that, for each* $n \in \mathbb{N}$, $(X_{n,k})_{k\in\mathbb{N}}$ *are independent (not necessarily identically distributed) random variables with distribution* $p_k^{(n)} = \mathbb{P}\{X_{n,k} = 1\} = 1 - \mathbb{P}\{X_{n,k} = 0\}$. *If*

$$\lim_{n\to\infty} \sum_{k=1}^{n} p_k^{(n)} = \lambda \in (0,\infty) \tag{1.15}$$

and

$$\lim_{n\to\infty} \max_{1\leq k\leq n} p_k^{(n)} = 0, \tag{1.16}$$

then $\sum_{k=1}^{n} X_{n,k} \overset{\mathrm{d}}{\to} X$ *as* $n \to \infty$, *where* X *is a random variable having a Poisson distribution with parameter* λ.

What does this theorem tell us? The effect of (1.16) is that each $X_{n,k}$, $k = 1, 2, \ldots, n$ has a uniformly small probability of being 1. We can think of $X_{n,k}$ as the indicator of some event $A_{n,k}$ with $\mathbb{P}(A_{n,k}) = p_k^{(n)}$. Then $\sum_{k=1}^{n} X_{n,k}$ is the number of $A_{n,1}, \ldots, A_{n,n}$ which occur. So when each of a large number of independent events has a small probability of occurring (that is, the events are rare), the number of events which occur is approximately Poisson distributed.

Proof of Theorem 25. By the continuity theorem for generating functions (Theorem 24) it is sufficient to prove the convergence of generating functions, that is,

$$\lim_{n\to\infty} \mathbb{E} s^{\sum_{k=1}^{n} X_{n,k}} = e^{-\lambda(1-s)}, \quad s \in [0,1). \tag{1.17}$$

Recall that $s \mapsto e^{-\lambda(1-s)}$ is the generating function of a Poisson distribution with parameter $\lambda > 0$. Since $\mathbb{E} s^{X_{n,k}} = s p_k^{(n)} + 1 - p_k^{(n)}$ for $n \in \mathbb{N}$, $k \in \mathbb{N}$ and $s \geq 0$, and $X_{n,1}, X_{n,2}, \ldots$ are independent by assumption we infer

$$\mathbb{E} s^{\sum_{k=1}^{n} X_{n,k}} = \prod_{k=1}^{n} (s p_k^{(n)} + 1 - p_k^{(n)}) = \prod_{k=1}^{n} (1 - (1-s) p_k^{(n)}).$$

Passing to the logarithms we see that (1.17) is equivalent to

$$\lim_{n\to\infty} \sum_{k=1}^{n} -\log(1 - (1-s) p_k^{(n)}) = \lambda(1-s), \quad s \in [0,1).$$

We need some preparatory work. Let us check that

$$-\log(1-x) = x + R(x), \quad x \in [0,1) \tag{1.18}$$

with

$$R(x) \le 2x^2, \quad x \in [0, 1/2]. \tag{1.19}$$

To this end, we use the Maclaurin series representation

$$-\log(1-x) = \sum_{n\ge 1} \frac{x^n}{n} = x + \sum_{n\ge 2} \frac{x^n}{n}, \quad x \in [0,1)$$

and set $R(x) = \sum_{n\ge 2} n^{-1}x^n$. Since $n^{-1}x^n \le x^n$ we have

$$R(x) \le \sum_{n\ge 2} x^n = \frac{x^2}{1-x}, \quad x \in [0,1).$$

If $x \in [0, 1/2]$, then $(1-x)^{-1} \le 2$, giving (1.19).

Using representation (1.18) we can write, for $s \in [0,1)$,

$$\sum_{k=1}^{n} -\log(1-(1-s)p_k^{(n)}) = (1-s)\sum_{k=1}^{n} p_k^{(n)} + \sum_{k=1}^{n} R((1-s)p_k^{(n)}).$$

In view of (1.15) the first term on the right-hand side converges to $\lambda(1-s)$ as $n \to \infty$. Thus, we are left with showing that the second term on the right-hand side converges to 0 as $n \to \infty$. Assumption (1.16) ensures that $(1-s)p_k^{(n)} \le p_k^{(n)} \le 1/2$ for $s \in [0,1)$, $k = 1, 2, \ldots, n$ and large enough n. Therefore, for such n we can apply (1.19) to obtain

$$\sum_{k=1}^{n} R((1-s)p_k^{(n)}) \le 2(1-s)^2 \sum_{k=1}^{n} (p_k^{(n)})^2 \le 2(1-s)^2 \max_{1\le k\le n} p_k^{(n)} \sum_{k=1}^{n} p_k^{(n)}.$$

The right-hand side converges to 0 as $n \to \infty$ by (1.15) and (1.16). The proof of Theorem 25 is complete. □

1.7. Simple random walk

Let ξ_1, $\xi_2, \ldots$ be independent identically distributed random variables with distribution $\mathbb{P}\{\xi_k = 1\} = p \in (0,1)$ and $\mathbb{P}\{\xi_k = -1\} = 1 - p =: q$. Set $S_0 := 0$ and $S_n := \xi_1 + \ldots + \xi_n$ for $n \in \mathbb{N}$.

Definition 26. The random sequence $(S_n)_{n\in\mathbb{N}_0}$ is called *simple random walk*.

The S_n can be thought of as the position of a particle which starts at 0 at time 0 and jumps randomly by ± 1 at every positive integer time.

We are interested in the distribution of

$$N := \inf\{n \in \mathbb{N} : S_n = 1\}.$$

Set $r_n := \mathbb{P}\{N = n\}$ for $n \in \mathbb{N}_0$. It is clear that $r_0 = 0$, $r_1 = p$, $r_2 = 0$ and more generally $r_n = 0$ for even n. The latter follows from the fact that at even times the walk is located at even states. To find r_n for $n \geq 3$ we are going to write a difference equation and solve it using generating functions. The claim is that

$$r_n = q \sum_{j=1}^{n-2} r_j r_{n-j-1}, \quad n \geq 3. \tag{1.20}$$

What is the intuitive meaning of this equality? Starting at 0 we want to get into 1 in more than two steps. Thus, the first step should be to -1 (the probability of this is q), then we must make our way back up to 0 in j steps (the probability of this is r_j because it is the same as the probability to reach 1 for the first time when starting at 0). Finally, from 0 we still must get up to 1 in $n-j-1$ steps (the probability of this is r_{n-j-1}). Thus, formula (1.20) is plausible. To give a formal proof, define the events

$$\begin{aligned} A_j &:= \{\text{the random walk makes a first return from } -1 \text{ to } 0 \text{ in } j \text{ steps}\} \\ &= \{\inf\{n \in \mathbb{N} : \xi_2 + \ldots + \xi_{n+1} = 1\} = j\} \end{aligned}$$

and

$$\begin{aligned} B_{n-j-1} &:= \{\text{the random walk makes a first passage from 0 to 1 in } n-j-1 \text{ steps}\} \\ &= \{\inf\{n \in \mathbb{N} : \xi_{j+2} + \ldots + \xi_{j+1+n} = 1\} = n-j-1\}. \end{aligned}$$

Then

$$\{N = n\} = \{\xi_1 = -1\} \cap \bigcup_{j=1}^{n-2} (A_j \cap B_{n-j+1}), \quad n \geq 3. \tag{1.21}$$

Since the event A_j is determined by $\xi_2, \xi_3, \ldots, \xi_{j+1}$ and similarly B_{n-j-1} is determined by $\xi_{j+2}, \ldots, \xi_{j+1+n}$, the three events

$$\{\xi_1 = 1\}, \quad A_j, \quad B_{n-j-1}$$

are independent, whence

$$\mathbb{P}\{N = n\} = q \sum_{j=1}^{n-2} \mathbb{P}(A_j)\mathbb{P}(B_{n-j-1}).$$

Here, we have used the fact that the union in (1.21) is a union of disjoint events. Therefore, the probability of the union is equal to the sum of probabilities. Further, $(\xi_2, \xi_3, \ldots) \stackrel{\text{d}}{=} (\xi_1, \xi_2, \ldots)$ which means that, for any $m \in \mathbb{N}$ and any $k_1, \ldots, k_m \in \{-1, 1\}$, $\mathbb{P}\{\xi_2 = k_1, \ldots, \xi_{m+1} = k_m\} = \mathbb{P}\{\xi_1 = k_1, \ldots, \xi_m = k_m\}$. As a consequence,

$$\mathbb{P}(A_j) = \mathbb{P}\{\inf\{n \in \mathbb{N} : \xi_1 + \ldots + \xi_n = 1\} = j\} = r_j$$

and similarly

$$\mathbb{P}(B_{n-j-1}) = r_{n-j-1}.$$

We have proved that (1.20) does indeed hold.

Let us now solve (1.20) with the help of generating functions. Set $f(s) := \mathbb{E}s^N = \sum_{n\geq 1} s^n r_n$ for $s \geq 0$ (recall that $r_0 = 0$). We already know that $f(s) = ps + \sum_{n\geq 3} s^n r_n$. Furthermore, in view of (1.20)

$$\sum_{n\geq 3} s^n r_n = q \sum_{n\geq 3} s^n \sum_{j=1}^{n-2} r_j r_{n-j-1} = qs \sum_{j\geq 1} s^j r_j \sum_{n\geq j+2} s^{n-j-1} r_j r_{n-j-1} = qsf^2(s).$$

Hence,

$$f(s) = ps + qsf^2(s).$$

Solving this quadratic equation for the unknown $f(s)$ we obtain

$$f(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2qs}.$$

The solution with the + sign cannot be a generating function because $\lim_{s\to 0+} \frac{1+\sqrt{1-4pqs^2}}{2qs} = \infty$ and we know that a generating function at 0 does not exceed 1. Hence,

$$f(s) = \mathbb{E}s^N = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}, \quad s \in (0, 1/2\sqrt{pq}). \tag{1.22}$$

To find r_n for odd n we use formula (1.1). This yields

$$(1 - 4pqs^2)^{1/2} = \sum_{k\geq 0} \frac{(1/2)(1/2-1)\cdot\ldots\cdot(1/2-k+1)}{k!}(-1)^k(4pq)^k s^{2k}$$
$$= 1 - \sum_{k\geq 1} \frac{(2k-3)!!}{2^k k!}(4pq)^k s^{2k},$$

where $(2k-3)!! = 1\cdot 3\cdot 5\cdot\ldots\cdot(2k-3)$ for integer $k \geq 2$ and $(2k-3)!! = 1$ for $k = 1$. Summarizing,

$$f(s) = \frac{1 - (1 - 4pqs^2)^{1/2}}{2qs} = \sum_{k\geq 1} \frac{(2k-3)!!}{k!} 2^{k-1} p^k q^{k-1} s^{2k-1},$$

whence

$$\mathbb{P}\{N = 2k-1\} = \frac{(2k-3)!!}{k!} 2^{k-1} p^k q^{k-1}, \quad k \in \mathbb{N}.$$

Although this formula is explicit it is not easy to work with. For instance, it is more practical to use generating functions in order to find $\mathbb{P}\{N < \infty\}$, moments of N etc.

Recall that $\mathbb{P}\{N < \infty\} = \sum_{k\geq 0} \mathbb{P}\{N = k\} = f(1-)$. Hence,

$$\mathbb{P}\{N < \infty\} = f(1-) = \frac{1 - \sqrt{1 - 4p(1-p)}}{2(1-p)} = \frac{1 - |2p-1|}{2(1-p)} = \begin{cases} 1, & \text{if } p \in [1/2, 1), \\ \frac{p}{1-p}, & \text{if } p \in (0, 1/2). \end{cases}$$

Thus, if $p < 1/2$, that is, the pressure pushing the random walk in the positive direction is weak we have $\mathbb{P}\{N = \infty\} > 0$.

Assuming that $p \in [1/2, 1)$ let us find $\mathbb{E}N$. For other values of p $\mathbb{P}\{N = \infty\} > 0$ whence $\mathbb{E}N = \infty$. To this end, we first differentiate $f(s)$ to obtain

$$f'(s) = \frac{2qs(1/2)(1 - 4pqs^2)^{-1/2} 8pqs - (1 - (1 - 4pqs^2)^{1/2})2q}{4q^2 s^2}.$$

This in combination with Corollary 4 gives

$$\mathbb{E}N = f'(1-) = \frac{(1-4pq)^{-1/2}8pq^2 - (1-(1-4pq)^{1/2})2q}{4q^2} = \frac{2p}{2p-1} - \frac{1-(2p-1)}{2(1-p)}$$
$$= \begin{cases} \infty, & \text{if } p = 1/2, \\ \frac{1}{2p-1}, & \text{if } p \in (1/2, 1). \end{cases}$$

having utilized the formula $(1-4pq)^{1/2} = (4p^2-4p+1)^{1/2} = |2p-1| = 2p-1$ for $p \in [1/2, 1)$.

Set

$$N_0 := \inf\{n \in \mathbb{N} : S_n = 0\}.$$

An extension of this technique can be used to analyze the distribution of N_0 the first return time to 0. Set $f_n := \mathbb{P}\{N = n\}$ for $n \in \mathbb{N}_0$. It is clear that $f_0 = 0$ and that $f_n = 0$ for odd n which follows from the fact that at odd times the walk is located at odd states. We shall use the following decomposition

$$N_0 = \begin{cases} 1 + \inf\{n \in \mathbb{N} : \xi_2 + \ldots + \xi_{n+1} = 1\}, & \text{on } \xi_1 = -1, \\ 1 + \inf\{n \in \mathbb{N} : \xi_2 + \ldots + \xi_{n+1} = -1\}, & \text{on } \xi_1 = 1. \end{cases}$$

To explain the first line observe that the first jump contributes 1 to N_0 and if the first jump is to -1, then we wait until the random walk with jumps ξ_2, $\xi_3, \ldots$ makes the first passage from 0 to 1 (the original random walk then pass from -1 to 0). Introduce the notation

$$N^+ := \inf\{n \in \mathbb{N} : \xi_2 + \ldots + \xi_{n+1} = 1\} \quad \text{and} \quad N^- := \inf\{n \in \mathbb{N} : \xi_2 + \ldots + \xi_{n+1} = -1\}.$$

As was discussed above $(\xi_2, \xi_3, \ldots) \overset{\mathrm{d}}{=} (\xi_1, \xi_2, \ldots)$ which ensures that $N^+ \overset{\mathrm{d}}{=} N = \inf\{n \in \mathbb{N} : S_n = 1\}$. Since N^+ and N^- are determined by ξ_2, $\xi_3, \ldots$, they both are independent of ξ_1. Further,

$$N^- = \inf\{n \in \mathbb{N} : \xi_2 + \ldots + \xi_{n+1} = -1\} \overset{\mathrm{d}}{=} \inf\{n \in \mathbb{N} : \xi_1 + \ldots + \xi_n = -1\}$$
$$= \inf\{n \in \mathbb{N} : (-\xi_1) + \ldots + (-\xi_n) = 1\}.$$

Thus, N^- has the same distribution as N for a simple random walk which moves one unit to the right with probability q and one unit to the left with probability p. Equivalently, the generating function of N^- is the same as the generating function in (1.22) with p and q reversed, that is,

$$\mathbb{E}s^{N^-} = \frac{1-\sqrt{1-4pqs^2}}{2ps}, \quad s \in (0, 1/2\sqrt{pq}).$$

Set

$$f_0(s) := \mathbb{E}s^{N_0} = \sum_{n \geq 0} f_{2n}s^{2n}, \quad s \in [0, 1).$$

Using the decomposition above yields

$$\begin{aligned} f_0(s) &= \mathbb{E}s^{N_0} = \mathbb{E}s^{N_0}\mathbb{1}_{\{\xi_1=-1\}} + \mathbb{E}s^{N_0}\mathbb{1}_{\{\xi_1=1\}} = \mathbb{E}s^{1+N^+}\mathbb{1}_{\xi_1=-1} + \mathbb{E}s^{1+N^-}\mathbb{1}_{\{\xi_1=1\}} \\ &= s\mathbb{E}s^{N^+}q + s\mathbb{E}s^{N^-}p = sq\frac{1-\sqrt{1-4pqs^2}}{2qs} + sp\frac{1-\sqrt{1-4pqs^2}}{2ps} = 1 - \sqrt{1-4pqs^2}, \end{aligned}$$

where the fourth equality is implied by independence of N^+ and ξ_1, N^- and ξ_1. Using this we infer

$$\mathbb{P}\{N_0 < \infty\} = f_0(1-) = 1 - \sqrt{1-4pq} = 1 - |2p-1| = \begin{cases} 1, & \text{if } p = 1/2, \\ 2(1-p), & \text{if } p \in (1/2, 1), \\ 2p, & \text{if } p \in (0, 1/2) \end{cases}$$

which shows that only in the balanced case $p = 1/2$ does the random walk return to 0 with probability 1. In this case $f_0(s) = 1 - \sqrt{1-s^2}$ and $f_0'(s) = s(1-s^2)^{-1/2}$. Hence, $\mathbb{E}N_0 = f'(1-) = \infty$, that is, a return to the origin is certain but only after a random amount of time whose expectation is infinite.

1.8. Left-continuous random walk

Let ξ_1, $\xi_2, \ldots$ be independent identically distributed random variables with distribution $\mathbb{P}\{\xi_i = k-1\} = p_k$ for $k \in \mathbb{N}_0$, where $p_0 \in (0,1)$. Set $S_0 := 1$ and $S_n := 1 + \xi_1 + \ldots + \xi_n$ for $n \in \mathbb{N}$.

Definition 27. The random sequence $(S_n)_{\mathbb{N}_0}$ is called *left-continuous random walk* or *skip-free to the left random walk.*

When the walk moves to the left, it does so only by jumps of the size -1. Thus, the walk cannot jump over states when moving to the left, hence the term.

Set

$$N := \inf\{n \in \mathbb{N} : S_n = 0\}.$$

This is the time needed for the first passage from 1 to 0.

Proposition 28. *Let $f(s) := \mathbb{E}s^N$, $s \in [0,1)$ be the generating function of N. Then*

$$f(s) = sg(f(s)), \quad s \in [0,1), \tag{1.23}$$

where $g(s) = \mathbb{E}s^{\xi_1+1} = \sum_{k\geq 0} s^k p_k$.

Proof. We claim that

$$N \overset{\mathrm{d}}{=} 1 + N_1 + \ldots + N_{\xi_1+1}, \tag{1.24}$$

where N_1, $N_2, \ldots$ are independent copies of N which are also independent of ξ_1.

To check this, observe that if $\xi_1 = -1$, then $N = 1$. If $\xi_1 = 0$, then $N = 1 + N_1^{(1,0)}$, where $N_1^{(1,0)} = \inf\{k \geq 2 : S_k - \xi_1 = 0\}$ is the time needed for the first passage from 1 to 0 by the random walk $(1 + \xi_2 + \xi_3 + \ldots + \xi_k)$ (this is the original random walk in which the jump ξ_1 is omitted). Since $(\xi_2, \xi_3, \ldots) \overset{\mathrm{d}}{=} (\xi_1, \xi_2, \ldots)$, and $N_1^{(1,0)}$ is the same functional of $(\xi_2, \xi_3, \ldots)$ as N of $(\xi_1, \xi_2, \ldots)$, we conclude that $N_1^{(1,0)} \overset{\mathrm{d}}{=} N$ and that $N_1^{(1,0)}$ is independent of ξ_1. If $\xi_1 = 1$, then $N = 1 + N_1^{(2,1)} + N_2^{(1,0)}$, where $N_1^{(2,1)} = \inf\{k \geq 2 : S_k - \xi_1 = 1\}$ is the time needed for the first passage from 2 to 1 and $N_2^{(1,0)} = \inf\{k \geq N_1^{(2,1)} + 2 : S_k - S_{N_1^{(2,1)}+1} = 0\}$ is the time needed for the first passage from 1 to 0. Let us show that $N_1^{(2,1)}$ and $N_2^{(1,0)}$ are independent copies of N which are also independent of ξ_1. To this end, first note that, for $k \in \mathbb{N}$,

$$\{N = k\} = \{(\xi_1, \ldots, \xi_{k-1}) \in A_k, \xi_k = -1\},$$

where the A_k are some sets which can be specified precisely but are of no importance for us. Write

$$\begin{aligned}
& \mathbb{P}\{N_1^{(2,1)} = i, N_2^{(1,0)} = j, \xi_1 = 1\} \\
= \; & \mathbb{P}\{(\xi_2, \ldots, \xi_i) \in A_i, \xi_{i+1} = -1, (\xi_{i+2}, \ldots, \xi_{i+j}) \in A_j, \xi_{i+j+1} = -1, \xi_1 = 1\} \\
= \; & \mathbb{P}\{(\xi_2, \ldots, \xi_i) \in A_i, \xi_{i+1} = -1\}\mathbb{P}\{(\xi_{i+2}, \ldots, \xi_{i+j}) \in A_j, \xi_{i+j+1} = -1\}\mathbb{P}\{\xi_1 = 1\} \\
= \; & \mathbb{P}\{(\xi_1, \ldots, \xi_{i-1}) \in A_i, \xi_i = -1\}\mathbb{P}\{(\xi_1, \ldots, \xi_{j-1}) \in A_j, \xi_j = -1\}\mathbb{P}\{\xi_1 = 1\} \\
= \; & \mathbb{P}\{N = i\}\mathbb{P}\{N = j\}\mathbb{P}\{\xi_1 = 1\}.
\end{aligned}$$

This establishes the desired independence and distributional equalities. In the general situation, the claim follows from the representation

$$N = 1 + N_1^{(\xi_1+1,\xi_1)} + N_2^{(\xi_1,\xi_1-1)} \ldots + N_{\xi_1+1}^{(1,0)}$$

which tells us that in order to pass from 1 to 0 for the first time the walk first gets to the position ξ_1+1 and then make subsequent passages from ξ_1+1 to ξ_1, from ξ_1 to $\xi_1-1, \ldots$, from 1 to 0. Of course, during every particular passage from i to $i-1$ arbitrarily long excursions to the right are possible.

Using (1.24) in combination with Proposition 12 we immediately conclude that

$$f(s) = \mathbb{E}s^N = s\mathbb{E}s^{\sum_{i=1}^{\xi_1+1} N_i} = sg(f(s)).$$

□

Corollary 29. $\mathbb{P}\{N < \infty\} = 1$ if $\mathbb{E}\xi \le 0$ (and then necessarily $\mathbb{E}\xi \in (-1, 0]$). Further, $\mathbb{E}N = \mu^{-1}$ if $\mu := -\mathbb{E}\xi > 0$, and $\mathbb{E}N = \infty$ if $\mu = 0$; and $\operatorname{Var} N = \sigma^2\mu^{-3}$ provided that $\mu > 0$ and $\sigma^2 := \operatorname{Var}\xi < \infty$.

Proof. Set $\mathbb{P}\{N < \infty\} = f(1-) = \gamma$ and note that $\gamma \in (0, 1]$. Then using (1.23) we obtain

$$\gamma = \lim_{s\to 1-} f(s) = \lim_{s\to 1-} sg(f(s)) = g(\gamma).$$

According to the proof of Theorem 20 we know that the equation $s = g(s)$ has a unique solution $s = 1$ on $[0, 1]$ provided that $\mathbb{E}(\xi + 1) \le 1$. This entails $\gamma = 1$.

Assume now that $\mu > 0$. According to the previous part of the proof we know that $f(1-) = 1$. Differentiating (1.23) yields

$$f'(s) = g(f(s)) + sg'(f(s))f'(s), \quad s \in [0, 1)$$

or equivalently

$$f'(s) = \frac{g(f(s))}{1 - sg'(f(s))}.$$

Thus,

$$\mathbb{E}N = \lim_{s\to 1-} f'(s) = \lim_{s\to 1-} \frac{g(f(s))}{1 - sg'(f(s))} = \frac{g(1-)}{1-g'(1-)} = \frac{1}{1-\mathbb{E}(\xi+1)} = \frac{1}{\mu}.$$

This formula is also valid when $\mu = 0$ in which case the right-hand side is interpreted as ∞.

The formula for the variance is left as an exercise. □

Now we are going to exhibit an interesting connection between left-continuous random walks and a Galton-Watson branching process.

Theorem 30. *Let $(Z_n)_{n\in\mathbb{N}_0}$ be a Galton-Watson process for which $Z_1 \overset{\mathrm{d}}{=} \xi_1 + 1$, where $\mathbb{P}\{\xi_1 = k-1\} = p_k$ for $k \in \mathbb{N}_0$, and $p_0 \in (0,1)$ (that is, the distribution of ξ_1 is the same as at the beginning of this section). Denote by*

$$Z = \sum_{n\geq 0} Z_n = 1 + \sum_{n\geq 1} Z_n$$

the total population size, that is, the total number of individuals ever born (the case that $Z = \infty$ with positive probability is not excluded). Then $Z \overset{\mathrm{d}}{=} N$, where N is as defined at the beginning of this section.

We do not know yet what is the value of $\mathbb{P}\{N < \infty\}$ when $\mathbb{E}\xi_1 > 0$. This case remained untouched in Corollary 29.

Corollary 31. If $\mathbb{E}\xi_1 > 0$, then $\mathbb{P}\{N < \infty\} = \rho \in (0,1)$, where ρ is the extinction probability of the Galton-Watson process in Theorem 30.

Proof. In view of Theorem 30 it suffices to prove that $\mathbb{P}\{Z < \infty\} = \rho \in (0,1)$. This is a consequence of $\{Z < \infty\} = \{T < \infty\}$, where

$$T := \inf\{n \in \mathbb{N} : Z_n = 0\},$$

and $\mathbb{P}\{T < \infty\} = \rho$. Finally, the condition $\mathbb{E}Z_1 > 1$ ensures that $\rho \in (0,1)$ by Theorem 20. □

Before giving a proof of Theorem 30 we consider an example in which explicit calculations are possible.

Example 32. Let $\mathbb{P}\{\xi = 0\} = p \in (0,1)$ and $\mathbb{P}\{\xi = 1\} = 1-p$. Denote by

$$T := \inf\{n \in \mathbb{N} : Z_n = 0\}$$

the extinction time of the Galton-Watson process. For $k \in \mathbb{N}$, the event $\{T = k\}$ is equivalent to the following: the generations 1 through $k-1$ contain exactly one individual, whereas the generation k contains no individuals, the probability of this being $(1-p)^{k-1}p$. Thus, we have proved that T has a geometric distribution (starting at 1) with success probability p. The corresponding generating function is

$$\mathbb{E}s^T = \frac{ps}{1-(1-p)s}, \quad s \in [0, 1/(1-p)).$$

From the representation $Z = 1 + Z_1 + \ldots + Z_{T-1}$ and the fact that $Z_k = 1$ for $k \le T-1$ we infer $Z = T$ with probability one. Hence, Z too has a geometric distribution (starting at 1) with success probability p.

On the other hand, using (1.23) with $g(s) = (1-p)s + p$ we infer $f(s) = \mathbb{E}s^N = s((1-p)f(s)+p)$, whence

$$\mathbb{E}s^N = \frac{ps}{1-(1-p)s} = \mathbb{E}s^Z$$

and thereupon $N \overset{\mathrm{d}}{=} Z$ by the uniqueness theorem for generating functions (Theorem 2).

Proof of Theorem 30. The total population size Z is equal to 1 (for the initial ancestor) plus the total sizes of the subpopulations initiated by the individuals of the first generation, the number of these being Z_1. The total sizes of the subpopulations are independent of each other and of Z_1 because the individuals of the first generation reproduce independently of each other and their number. Hence,

$$Z = 1 + Z^{(1)} + \ldots + Z^{(Z_1)},$$

where, given $Z_1 = k$, $Z^{(1)}, \ldots, Z^{(k)}$ are independent copies of Z. Using Proposition 12 yields

$$f(s) := \mathbb{E}s^Z = s\mathbb{E}s^{\sum_{i=1}^{Z_1} Z^{(i)}} = sg(f(s)) = \mathbb{E}s^N,$$

where $g(s) = \mathbb{E}s^{Z_1} = \mathbb{E}s^{\xi_1+1}$, and the last equality follows from Theorem 28. □

From the proof of Corollary 31 we know that when $\mathbb{E}\xi_1 > 0$ the random variable Z is infinite with positive probability and a fortiori $\mathbb{E}Z = \infty$. In the last result of this section we shall show that the mean value of Z conditionally on the extinction is finite and find its exact value.

Proposition 33. *If $\mathbb{E}\xi_1 > 0$, then*

$$\mathbb{E}(Z|A) = \frac{1}{1-f'(\rho)} < \infty,$$

where $A = \{T < \infty\}$ is the extinction event for the Galton-Watson process, and $\rho = \mathbb{P}(A)$ is the extinction probability.

Proof. By definition, $\mathbb{E}(Z|A) = \mathbb{E}(Z\mathbb{1}_A)/\mathbb{P}(A) = \rho^{-1}\mathbb{E}Z\mathbb{1}_{\{T<\infty\}}$. So, the problem boils down to finding the last expectation. Write

$$\mathbb{E}Z\mathbb{1}_{\{T<\infty\}} = \mathbb{E}\sum_{n=0}^{T-1} Z_n\mathbb{1}_{\{T<\infty\}} = \mathbb{E}\sum_{n\ge 0} Z_n\mathbb{1}_{\{n<T<\infty\}} = \sum_{n\ge 0}\mathbb{E}Z_n\mathbb{1}_{\{n<T<\infty\}},$$

where the last equality is justified by Fubini's theorem in combination with the fact that the random variables involved are nonnegative. Further,

$$\mathbb{E}Z_n\mathbb{1}_{\{n<T<\infty\}} = \sum_{k\ge 1} k\mathbb{P}\{Z_n = k, n < T < \infty\} = \sum_{k\ge 1} k\mathbb{P}\{Z_n = k\}\rho^k = \rho f_n'(\rho) = \rho(f'(\rho))^n, \tag{1.25}$$

where, as before, $f_n(s) = \mathbb{E}s^{Z_n}$ and $f(s) = f_1(s)$.

To explain the equality

$$\mathbb{P}\{Z_n = k, n < T < \infty\} = \mathbb{P}\{Z_n = k\}\rho^k, \quad k \in \mathbb{N}, \tag{1.26}$$

observe that $\{Z_n = k\} \subseteq \{T > n\} = \{Z_n > 0\}$, whence $\mathbb{P}\{Z_n = k, n < T < \infty\} = \mathbb{P}\{Z_n = k, T < \infty\}$. The event $\{Z_n = k, T < \infty\}$ coincides with the following: $Z_n = k$ and each of the k subpopulations begot by the individuals of the nth generation dies out. Since the subpopulations evolve independently of each other and Z_n, and the probability that a particular subpopulation dies out is equal to ρ we infer (1.26). To prove the last equality in (1.25) we use the mathematical induction. When $n = 1$ the equality holds trivially. Assuming it holds for $n = m$, write, for $s \in (0, 1)$, $f'_{m+1}(s) = (f_m(f(s)))' = f'_m(f(s))f'(s)$. Substituting now $s = \rho$ and recalling that $f(\rho) = \rho$ by Theorem 20 we infer

$$f'_{m+1}(\rho) = f'_m(\rho)f'(\rho) = (f'(\rho))^m f'(\rho) = (f'(\rho))^{m+1}$$

having utilized the induction assumption for the penultimate equality. To proceed we need to know that $f'(\rho) < 1$. This is easily seen at Figure 1.1. Indeed, $f'(\rho)$ is the slope of f at ρ which obviously has to be smaller than 1.

Combining pieces together we arrive at

$$\mathbb{E}Z\mathbb{1}_{\{T<\infty\}} = \rho\sum_{n\geq 0}(f'(\rho))^n = \frac{\rho}{1 - f'(\rho)}.$$

□

1.9. Problems

Problem 34. Find the distribution of a random variable ξ taking nonnegative integer values if its generating function is

(a)

$$f(s) = \frac{e^{1+s} + e^{1-s}}{e^2 + 1}, \quad s \geq 0;$$

(b)

$$f(s) = 1 + \frac{1 - s}{s}\log(1 - s), \quad s \in (0, 1);$$

(c) for $\alpha \in (0, 1)$,

$$f(s) = 1 - (1 - s)^\alpha, \quad s \in [0, 1)$$

(the corresponding distribution is called *Sibuya distribution*).

Problem 35. Let ξ_1, $\xi_2, \ldots$ be independent identically distributed random variables with distribution $\mathbb{P}\{\xi_1 = 1\} = p \in (0, 1)$ and $\mathbb{P}\{\xi_1 = 0\} = 1-p$. Set $S_0 := 0$ and $S_n := \xi_1 + \ldots + \xi_n$ for $n \in \mathbb{N}$. Prove that, for each $k \in \mathbb{N}_0$, $\sum_{n\geq 0}\mathbb{P}\{S_n = k\} = 1/p$.

Problem 36. Let g be defined on $[0,1)$ and satisfy one of the following conditions:
(a) $g(2^{-n}) = 2^{-n}$ for all $n \in \mathbb{N}$;
(b) $g(2^{-n}) = n(n+1)^{-1}2^{-n}$ for all $n \in \mathbb{N}$.

Is g a generating function of a probability distribution? If yes, then find all distributions with generating function g.

Problem 37. Prove that the function g defined on $[0,1]$ by

$$g(s) = \begin{cases} 3s, & \text{for } s \in [0,1/4], \\ (s+2)/3, & \text{for } s \in (1/4,1]. \end{cases}$$

is not the generating function of a random variable ξ taking nonnegative integer values.

Problem 38. Let $\eta_1, \eta_2, \ldots$ be (not necessarily integer-valued) independent random variables with the distribution function $F(x)$. Assume that $\eta_1, \eta_2, \ldots$ are independent of a random variable N taking nonnegative integer values with the generating function $f(s)$. Prove that $\mathbb{P}\{\max(\eta_1, \eta_2, \ldots, \eta_N) \le x\} = f(F(x))$ for $x \in \mathbb{R}$.

Problem 39. Let $\xi_1, \xi_2, \ldots$ be independent random variables with distribution $\mathbb{P}\{\xi_j = 1\} = 1/j = 1 - \mathbb{P}\{\xi_j = 0\}$ for $j \in \mathbb{N}$. Use the continuity theorem for generating functions to prove that $\xi_{n+1} + \xi_{n+2} + \ldots + \xi_{\lfloor ne \rfloor} \overset{\mathrm{d}}{\to} X$ as $n \to \infty$, where X is a random variable with a Poisson distribution with parameter 1, and $\lfloor x \rfloor$ is the greatest integer $\le x$.

Problem 40. Let $(Z_n)_{n \in \mathbb{N}_0}$ be a Galton-Watson process with $\sigma^2 = \operatorname{Var}\xi \in (0,\infty)$ and $\mu = \mathbb{E}Z_1 < \infty$. Prove that

$$\operatorname{Var} Z_n = \begin{cases} \sigma^2 \mu^{n-1} \frac{1-\mu^n}{1-\mu}, & \text{if } \mu \neq 1, \\ \sigma^2 n, & \text{if } \mu = 1. \end{cases}$$

Problem 41. Let $(Z_n)_{n \in \mathbb{N}_0}$ be a Galton-Watson process with $\mathbb{E}s^{Z_1} = p(1-(1-p)s)^{-1}$ for $p \in (0,1)$ and $s \in [0, 1/(1-p))$. Prove the following explicit formulae for the survival probability in the first n generations:

$$\mathbb{P}\{Z_n > 0\} = \frac{((1-p)/p)^n(1-(1-p)/p)}{1-((1-p)/p)^{n+1}}$$

if $p \neq 1/2$, and

$$\mathbb{P}\{Z_n > 0\} = \frac{1}{n+1},$$

if $p = 1/2$.

Problem 42. Let $(Z_n)_{n \in \mathbb{N}_0}$ be a Galton-Watson process with $\mathbb{E}s^{Z_1} = (2-s)^{-1}$, $s \in [0,2)$ (thus, the distribution of Z_1 is geometric (starting at zero) with success probability 1/2). Find the generating function of $Z = \sum_{n \ge 0} Z_n$ and use it to find the distribution of Z.

Problem 43. Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables with distribution $\mathbb{P}\{\xi_i = k-1\} = p_k$ for $k \in \mathbb{N}_0$, where $p_0 \in (0,1)$. Set $S_0 := 1$, $S_n := 1+\xi_1+\ldots+\xi_n$ for $n \in \mathbb{N}$ and then .

$$N := \inf\{n \in \mathbb{N} : S_n = 0\}.$$

Prove that $\operatorname{Var} N = \sigma^2 \mu^{-3}$ provided that $\mu = -\mathbb{E}\xi > 0$ and $\sigma^2 = \operatorname{Var}\xi < \infty$.

Chapter 2

Laplace-Stieltjes transforms

The first section of this chapter does not treat Laplace-Stieltjes transforms and any other transforms. Its subject is a moment problem. We shall need this material later while proving the uniqueness theorem for Laplace-Stieltjes transforms.

2.1. Moment problem

The moment problem consists in the following. Let F be a distribution function with all moments finite, that is,

$$m_n := \int_{\mathbb{R}} x^n \mathrm{d}F(x), \quad n \in \mathbb{N}$$

satisfy $|m_n| < \infty$. The question is whether there exists a different distribution function G having the same moments, that is,

$$\int_{\mathbb{R}} x^n \mathrm{d}G(x) = m_n, \quad n \in \mathbb{N}?$$

The answer may be positive in which case the moment problem is called *indeterminate* and may be negative in which case the moment problem is called *determinate*.

We now give an example of an indeterminate moment problem.

Example 44. Let ξ be a random variable with a standard normal distribution, that is, the distribution of ξ is given by the density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.$$

The distribution of the random variable $\eta = e^{\xi}$ is called *log-normal distribution.* The log-normal distribution is given by its density

$$g(x) = x^{-1} f(\log x) = \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}, \quad x > 0.$$

Let us check that

$$m_n = \mathbb{E}\eta^n = e^{n^2/2}, \quad n \in \mathbb{N}.$$

Indeed, for $n \in \mathbb{N}$,

$$\mathbb{E}\eta^n = \mathbb{E}e^{n\xi} = \int_{\mathbb{R}} e^{nx} f(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{nx} e^{-x^2/2}\mathrm{d}x = \frac{1}{\sqrt{2\pi}} e^{n^2/2} \int_{\mathbb{R}} e^{-(x^2-2nx+n^2)/2}\mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi}} e^{n^2/2} \int_{\mathbb{R}} e^{-(x-n)^2/2}\mathrm{d}x = \frac{1}{\sqrt{2\pi}} e^{n^2/2} \int_{\mathbb{R}} e^{-x^2/2}\mathrm{d}x = e^{n^2/2} \int_{\mathbb{R}} f(x)\mathrm{d}x = e^{n^2/2}.$$

We intend to show that the moment problem for the log-normal distribution is indeterminate. To this end, we have to point out a distribution different from the log-normal but having the same moments as the log-normal distribution.

Set $h(x) = g(x)(1 + \sin(2\pi \log x))$ for $x > 0$ and let us check that $h(x)$ is a density of some distribution. To ensure this two properties have to be verified: 1) $h(x) \geq 0$ for $x > 0$ and 2) $\int_0^\infty h(x)\mathrm{d}x = 1$. The nonnegativity of h is a consequence of nonnegativity of g and the fact that $\sin y \geq -1$. Since $\int_0^\infty g(x)\mathrm{d}x = 1$ (because g is a density), to check the second property it suffices to prove that

$$\int_0^\infty g(x) \sin(2\pi \log x)\mathrm{d}x = 0.$$

We first note that the integral is convergent in view of the estimate $|g(x)\sin(2\pi \log x)| \leq g(x)$ and the fact that $\int_0^\infty g(x)\mathrm{d}x < \infty$. Changing the variable $y = \log x$ yields

$$\int_0^\infty g(x) \sin(2\pi \log x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-y^2/2} \sin(2\pi y)\mathrm{d}y = 0. \tag{2.1}$$

The last integral is equal to 0 because the function $y \mapsto e^{-y^2/2}\sin(2\pi y)$ is odd and any convergent integral of an odd function over an interval symmetric around 0 is equal to zero. It remains to show that

$$\int_0^\infty x^n h(x)\mathrm{d}x = e^{n^2/2}, \quad n \in \mathbb{N}$$

or equivalently that

$$\int_0^\infty x^n g(x) \sin(2\pi \log x)\mathrm{d}x = 0, \quad n \in \mathbb{N}.$$

This can be done as follows. Changing the variable $\log x = y + n$ we infer

$$\int_0^\infty x^n g(x) \sin(2\pi \log x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_0^\infty x^{n-1} e^{-(\log x)^2/2} \sin(2\pi \log x)\mathrm{d}x$$
$$= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{n(y+n)} e^{-(y+n)^2/2} \sin(2\pi(y+n))\mathrm{d}y = \frac{1}{\sqrt{2\pi}} e^{n^2/2} \int_{\mathbb{R}} e^{-y^2/2} \sin(2\pi y)\mathrm{d}y = 0,$$

where the penultimate equality is implied by periodicity of $y \mapsto \sin(2\pi y)$ with period n, and the last equality is justified by (2.1).

It is a remarkable fact that there exist discrete distributions having the same moments as the log-normal distribution, see Exercise 97.

Many other interesting examples in this vein can be found in Chapter 'The moment problem' in [15].

Let us now move to a discussion of determinate moment problems. The theorem given next provides sufficient conditions under which the moment problem is determinate. The essence of the theorem is that the moment sequence of a determinate moment problem cannot grow too fast. For instance, it can be easily checked that the moment sequence of the log-normal distribution does not satisfy the assumption of the theorem as it must be.

Theorem 45. *Let $(m_n)_{n\in\mathbb{N}}$ be a sequence of moments of a random variable and assume that*

$$\limsup_{n\to\infty} \frac{m_{2n}^{1/(2n)}}{2n} < \infty.$$

Then the sequence $(m_n)_{n\in\mathbb{N}}$ determines distribution uniquely.

The proof of Theorem 2.15 essentially uses characteristic functions. Therefore, it will be given on p. 74 in Chapter 3.

We proceed by giving two corollaries with even simpler sufficient conditions. The first of these asserts that the moment problem is determinate for any distribution with some finite exponential moment. By this corollary, the moment problems for an exponential and normal distributions are determinate.

Corollary 46. If $\mathbb{E}e^{s|\xi|} < \infty$ for some $s > 0$, then the moment sequence of ξ determines distribution uniquely.

Proof. Set $n_k := \mathbb{E}|\xi|^k$ for $k \in \mathbb{N}$. Since $|\xi|$ has a finite exponential moment, $n_k < \infty$ for all $k \in \mathbb{N}$. The function $t \mapsto \mathbb{E}e^{t|\xi|}$ is finite on $[0, s]$ and can be represented as a power series on this (and possibly larger) interval

$$\mathbb{E}e^{t|\xi|} = \mathbb{E}\sum_{k\ge 0} \frac{t^k}{k!}|\xi|^k = \sum_{k\ge 0} \frac{t^k}{k!} n_k.$$

Denote by r the radius of convergence of this power series. By assumption $r \ge s > 0$ and the Cauchy-Hadamard formula tells us that

$$\limsup_{k\to\infty} \left(\frac{n_k}{k!}\right)^{1/k} = \frac{1}{r} < \infty.$$

Since

$$(k!)^{1/k} = e^{(\log k!)/k} = e^{(\log 2+\ldots+\log k)/k} \le k,$$

so that

$$\frac{n_k^{1/k}}{k} \le \frac{n_k^{1/k}}{(k!)^{1/k}},$$

we conclude that

$$\limsup_{k\to\infty} \frac{n_k^{1/k}}{k} < \infty.$$

This ensures

$$\limsup_{k\to\infty} \frac{n_{2k}^{1/(2k)}}{2k} < \infty$$

and thereupon

$$\limsup_{k\to\infty} \frac{m_{2k}^{1/(2k)}}{2k} < \infty$$

because $\mathbb{E}\xi^{2k} = m_{2k} = n_{2k}$ for $k \in \mathbb{N}$. The proof of the corollary finishes by an appeal to Theorem 2.15. □

The second corollary states that the moment problem for any distribution with a compact support is determinate.

Corollary 47. Let ξ be a random variable taking values in an interval $[a,b]$ for $-\infty < a < b < \infty$. Then its moments $(\mathbb{E}\xi^n)_{n\in\mathbb{N}}$ determine distribution uniquely.

Proof. Actually, this is a consequence of Corollary 46. Since $|\xi| \le c := \max(|a|,|b|)$ a.s., so that $e^{s|\xi|} \le e^{sc}$ a.s. for all $s>0$ we infer $\mathbb{E}e^{s|\xi|} \le e^{sc} < \infty$ for all $s>0$. It remains to apply Corollary 46. □

2.2. Completely monotone functions

2.2.1. Definition and examples

Definition 48. A function $f:(0,\infty)\to[0,\infty)$ is called *completely monotone* if it is infinitely differentiable and $(-1)^n f^{(n)}(s) \ge 0$ for all $n\in\mathbb{N}$ and $s>0$, where $f^{(n)}$ is the derivative of the nth order.

Thus, a completely monotone function is nonnegative, nonincreasing and convex. Completely monotone functions are necessarily bounded outside zero but may be unbounded at zero (for instance, the function f_1 below with $a=0$).

Example 49. Simple examples of completely monotone functions are $f_1(s) = 1/(s+a)^\beta$ and $f_2(s) = e^{-\beta s}$ for $a \ge 0$ and $\beta > 0$. Indeed, these functions are completely monotone in view of $(-1)^n f_1^{(n)}(s) = \beta(\beta+1)\cdot\ldots\cdot(\beta+n-1)(s+a)^{-n-\beta} > 0$ and $(-1)^n f_2^{(n)}(s) = \beta^n e^{-\beta s} > 0$ for $n\in\mathbb{N}$ and $s>0$.

2.2.2. Bernstein criterion

Definition 50. Let $G:[0,\infty)\to[0,\infty)$ be a nondecreasing and right-continuous function. If the integral $\int_{[0,\infty)} e^{-ax}\mathrm{d}G(x) < \infty$ for some $a>0$, then the function

$$f^*(s) := \int_{[0,\infty)} e^{-sx}\mathrm{d}G(x) < \infty, \quad s \ge a$$

is called *Laplace-Stieltjes transform* of G.

Note that if a nondecreasing and right-continuous function G grows sufficiently fast, then its Laplace-Stieltjes transform may be infinite at all positive arguments. This is the case, for instance, if $\mathrm{d}G(x) = e^{x^2}\mathbb{1}_{[0,\infty)}(x)\mathrm{d}x$. Indeed, then

$$\int_{[0,\infty)} e^{-sx}\mathrm{d}G(x) = \int_0^\infty e^{-sx+x^2}\mathrm{d}x = \infty, \quad s \ge 0.$$

The most important result concerning completely monotone functions is the Bernstein criterion. It states that the class of completely monotone functions coincides with the class of Laplace-Stieltjes transforms of nondecreasing and right-continuous functions defined on $[0,\infty)$.

Theorem 51. *A function $f : (0,\infty) \to [0,\infty)$ is completely monotone if, and only if, there exists a function $G : [0,\infty) \to [0,\infty]$ which is nondecreasing and right-continuous and satisfies*

$$f(s) = \int_{[0,\infty)} e^{-sx} \mathrm{d}G(x), \quad s > 0.$$

Example 52. If $G(x) = x^\beta/\Gamma(\beta+1)$ for $x \geq 0$ and $\beta > 0$, then the corresponding completely monotone function is $f(s) = s^{-\beta}$. Here, $\Gamma(z) := \int_0^\infty e^{-y} y^{z-1} \mathrm{d}y$ for $z > 0$ is the Euler gamma function.

Proof. Indeed, for $s > 0$,

$$f(s) = \int_{[0,\infty)} e^{-sx} \mathrm{d}(x^\beta/\Gamma(\beta+1)) = \frac{\beta}{\Gamma(\beta+1)} \int_0^\infty e^{-sx} x^{\beta-1} \mathrm{d}x$$
$$= \frac{1}{s^\beta \Gamma(\beta)} \int_0^\infty e^{-y} y^{\beta-1} \mathrm{d}y = \frac{1}{s^\beta}.$$

For the penultimate equality we have used the change of variable $y = sx$ and the equality

$$\Gamma(x+1) = x\Gamma(x), \quad x > 0 \tag{2.2}$$

which can be checked with the help of integration by parts as follows:

$$\Gamma(x+1) = \int_0^\infty e^{-y} y^x \mathrm{d}y = \int_0^\infty y^x \mathrm{d}(-e^{-y}) = -y^x e^{-y}\Big|_0^\infty + x \int_0^\infty e^{-y} y^{x-1} \mathrm{d}y = x\Gamma(x).$$

□

To prove the Bernstein criterion we need several auxiliary notions and results. Under the additional requirement $\lim_{x\to\infty} G(x) = 1$ a nondecreasing and right-continuous function G appearing in Definition 2 is a distribution function of a nonnegative random variable. We need a generalization of the notion of weak convergence introduced in Definition 23 for distribution functions.

Definition 53. For each $n \in \mathbb{N}$, let $G, G_n : [0,\infty) \to [0,\infty]$ be nondecreasing and right-continuous functions. One says that the functions G_n *weakly converge* to G if $\lim_{n\to\infty} G_n(x) = G(x)$ at every $x \geq 0$ which is a continuity point of G.

Theorem 54 given next is called a continuity theorem for Laplace-Stieltjes transforms. The proof will not be given. We refer instead to Theorem 2a on p. 433 in [3].

Theorem 54. *For each $n \in \mathbb{N}$, let $G_n : [0,\infty) \to [0,\infty]$ be nondecreasing and right-continuous functions.*
Assume that, for some $a > 0$, $\lim_{n\to\infty} \int_{[0,\infty)} e^{-sx} \mathrm{d}G_n(x) = g(s)$ for $s > a$. Then there exists a nondecreasing and right-continuous function $G : [0,\infty) \to [0,\infty]$ such that $g(s) = \int_0^\infty e^{-sx} \mathrm{d}G(x)$ for $s > a$, and the functions G_n weakly converge to G as $n \to \infty$.

Conversely, assume that, as $n \to \infty$, the functions G_n weakly converge to a nondecreasing and right-continuous function G. If, for some $a > 0$, the sequence $\left(\int_{[0,\infty)} e^{-ax} \mathrm{d}G_n(x) \right)_{n \in \mathbb{N}}$ is bounded, then

$$\lim_{n \to \infty} \int_{[0,\infty)} e^{-sx} \mathrm{d}G_n(x) = \int_{[0,\infty)} e^{-sx} \mathrm{d}G(x), \quad s > a.$$

As a final preparation for the proof of the Bernstein criterion we establish a technical lemma.

Lemma 55. *Let $G : [0, \infty) \to [0, \infty)$ be a nondecreasing and right-continuous function such that*

$$g(s) = \int_{[0,\infty)} e^{-sx} \mathrm{d}G(x) < \infty \quad \text{for } s > 0.$$

Then, for $n \in \mathbb{N}_0$,

$$(-1)^n g^{(n)}(s) = \int_{[0,\infty)} e^{-sx} x^n \mathrm{d}G(x), \quad s > 0. \tag{2.3}$$

Proof. We use induction on n. When $n = 0$ we interpret $g^{(0)}(s)$ as $g(s)$. Then (2.3) is just the definition of $g(s)$. Assume that (2.3) holds with $n = k$ for some $k \in \mathbb{N}$. Then, for $s > 0$,

$$(-1)^{k+1} g^{(k+1)}(s) = (-1)^{k+1} \lim_{h \to 0} \frac{g^{(k)}(s+h) - g^{(k)}(s)}{h} = \lim_{h \to 0} \int_{[0,\infty)} e^{-sx} x^k \frac{1 - e^{-hx}}{h} \mathrm{d}G(x),$$

where for the last equality we have used the induction assumption. Let us show that we can interchange the limit and the integral in the expression on the right-hand side. According to Lebesgue's dominated convergence theorem it suffices to find a function $t(x)$ which does not depend on h such that

$$e^{-sx} x^k \frac{1 - e^{-hx}}{h} \le t(x) \quad \text{for all } x > 0 \tag{2.4}$$

and $\int_{[0,\infty)} t(x) \mathrm{d}G(x) < \infty$. Set $t(x) := (2(k+1)/s)^{k+1} e^{-k-1} e^{-sx/2}$ for $x \ge 0$. Since $(1 - e^{-hx})/h \le x$ for $x \ge 0$ we have

$$e^{-sx} x^k \frac{1 - e^{-hx}}{h} \le e^{-sx} x^{k+1} = e^{-sx/2} x^{k+1} e^{-sx/2}.$$

Taking the derivative of $x \mapsto e^{-sx/2} x^{k+1}$ we infer that the latter function attains its maximal value $(2(k+1)/s)^{k+1} e^{-k-1}$ on $[0, \infty)$ at point $x = 2(k+1)/s$, that is,

$$e^{-sx/2} x^{k+1} \le (2(k+1)/s)^{k+1} e^{-k-1} \quad \text{for all } x \ge 0.$$

Thus, we have checked (2.4). Further,

$$\begin{aligned} \int_{[0,\infty)} t(x) \mathrm{d}G(x) &= (2(k+1)/s)^{k+1} e^{-k-1} \int_{[0,\infty)} e^{-sx/2} \mathrm{d}G(x) \\ &= (2(k+1)/s)^{k+1} e^{-k-1} g(s/2) < \infty. \end{aligned}$$

Thus,

$$(-1)^{k+1}g^{(k+1)}(s) = \lim_{h\to 0}\int_{[0,\infty)} e^{-sx}x^k\frac{1-e^{-hx}}{h}\mathrm{d}G(x) = \int_{[0,\infty)} e^{-sx}x^k \lim_{h\to 0}\Big(\frac{1-e^{-hx}}{h}\Big)\mathrm{d}G(x)$$
$$= \int_{[0,\infty)} e^{-sx}x^{k+1}\mathrm{d}G(x)$$

which shows that (2.3) holds with $n = k+1$. The proof of the lemma is complete. □

Proof of Theorem 51. Assume that $f(s) = \int_{[0,\infty)} e^{-sx}\mathrm{d}G(x)$ for $s>0$. By Lemma 55,

$$(-1)^n f^{(n)}(s) = \int_{[0,\infty)} e^{-sx}x^n\mathrm{d}G(x), \quad s>0,\ n\in\mathbb{N}_0.$$

It remains to note that the integral on the right-hand side is nonnegative.

Conversely, assume that f is completely monotone and particularly infinitely differentiable. Pick any $0<x<y$. We intend to show that $f(x)$ can be represented as the Taylor series

$$f(x) = \sum_{k\ge 0}\frac{(-1)^k f^{(k)}(y)}{k!}(y-x)^k, \tag{2.5}$$

where we write $f^{(0)}$ for f. Recall that a real-valued function is called analytic if it can be represented by its Taylor series (in other words, the Taylor series of a function converges to the function). Of course, every analytic function is infinitely differentiable. On the other hand, an infinitely differentiable real-valued function is not necessarily analytic. For instance, it may happen that the corresponding Taylor series diverges or it converges yet does not coincide with the original function[1]. Summarizing, formula (2.5) does require justification.

By Taylor's formula we have, for any $n\in\mathbb{N}$,

$$f(x) = f(y) + f'(y)(x-y) + \ldots + \frac{f^{(n)}(y)}{n!}(x-y)^n - \int_x^y \frac{f^{(n+1)}(z)}{n!}(x-z)^n\mathrm{d}z$$
$$= \sum_{k=0}^n \frac{(-1)^k f^{(k)}(y)}{k!}(y-x)^k + \int_x^y \frac{(-1)^{n+1}f^{(n+1)}(z)}{n!}(z-x)^n\mathrm{d}z.$$

Thus, we are left with showing that

$$R_n(x) := \int_x^y \frac{(-1)^{n+1}f^{(n+1)}(z)}{n!}(z-x)^n\mathrm{d}z \to 0, \quad n\to\infty.$$

For later needs, observe that

$$R_n(x) = f(x) - \sum_{k=0}^n \frac{(-1)^k f^{(k)}(y)}{k!}(y-x)^k \le f(x) \tag{2.6}$$

[1]A standard example of the second situation is given by the function $g(x) = e^{-1/x^2}$ for $x>0$ and $g(0)=0$. Then g is infinitely differentiable with the derivative of each order being equal to 0 at 0. Thus, the Taylor series of g is identically zero and as such does not represent g. An example of the first situation is $g(x) = \sum_{n\ge 0}(n!(1+a^n x))^{-1}$, $x\ge 0$ for fixed $a>1$. One can check that $g^{(k)}(0+) = (-1)^k k!\exp(a^k)$ for $k\in\mathbb{N}$, so by the Cauchy-Hadamard formula the Maclaurin series of g has zero radius of convergence. See [1] for more on this.

because each summand under the sum is nonnegative. Changing the variable $z = (y-x)t+x$ we obtain

$$R_n(x) = \frac{(y-x)^{n+1}}{n!}\int_0^1 (-1)^{n+1} f^{(n+1)}((y-x)t+x)\mathrm{d}t$$
$$= \frac{(y-x)^{n+1}}{n!}\int_0^1 (-1)^{n+1} f^{(n+1)}(x(1-t)+yt)\mathrm{d}t.$$

Since x is positive, there exists $b > 0$ such that $x > b$. The function $(-1)^{n+1}f^{(n+1)}$ is nonincreasing because its derivative $(-1)^{n+1}f^{(n+2)} \le 0$. Thus,

$$R_n(x) \le \frac{(y-x)^{n+1}}{n!}\int_0^1 (-1)^{n+1} f^{(n+1)}(b(1-t)+yt)\mathrm{d}t = \frac{(y-x)^{n+1}}{(y-b)^{n+1}}R_n(b) \le \frac{(y-x)^{n+1}}{(y-b)^{n+1}}f(b)$$

having utilized inequality (2.6) with $x = b$ at the last step. The right-hand side converges to 0 as $n \to \infty$ by our choice of b. This completes the proof of (2.5).

Let $a, u > 0$ be arbitrary. A specialization of formula (2.5) with $x = a(1-e^{-u/a})$ and $y = a$ yields

$$f(a(1-e^{-u/a})) = \sum_{k\ge 0} e^{-uk/a}\frac{(-1)^k a^k f^{(k)}(a)}{k!}.$$

Thus, $u \mapsto f(a(1-e^{-u/a}))$ is the Laplace-Stieltjes transform of the piecewise constant function that we denote by G_a with jumps of size $(-1)^k a^k f^{(k)}(a)/k!$ at points k/a, $k \in \mathbb{N}_0$. Note that G_a is nondecreasing and right-continuous. Since $\lim_{a\to\infty} a(1-e^{-u/a}) = u$ for $u > 0$, and the function f is continuous we infer

$$\lim_{a\to\infty} f(a(1-e^{-u/a})) = f(u), \quad u > 0.$$

By the direct part of Theorem 54, as $a \to \infty$ the functions G_a weakly converge to a nondecreasing and right-continuous function G, and $f(u) = \int_{[0,\infty)} e^{-ux}\mathrm{d}G(x)$ for $u > 0$. □

2.2.3. Various properties In the lemmas stated below we discuss various properties of completely monotone functions.

Lemma 56. *A linear combination with positive coefficients of completely monotone functions is completely monotone, that is, if, $\alpha_1, \dots, \alpha_j$ are nonnegative numbers and $f_1(s), \dots, f_j(s)$ are completely monotone functions, then $\alpha_1 f_1(s) + \dots + \alpha_j f_j(s)$ is completely monotone.*

Proof. Let $n \in \mathbb{N}$. The claim is immediate from

$$(-1)^n(\alpha_1 f_1(s) + \dots + \alpha_j f_j(s))^{(n)} = \alpha_1(-1)^n f_1^{(n)}(s) + \dots + \alpha_j(-1)^n f_j^{(n)}(s) \ge 0, \quad s > 0.$$

□

Lemma 57. *The product of completely monotone functions is completely monotone.*

Proof. Let f_1 and f_2 be completely monotone. It is clear that $f_1 f_2 \geq 0$. Further, using Leibniz rule for differentiation we infer, for $n \in \mathbb{N}$ and $s > 0$,

$$(-1)^n (f_1(s) f_2(s))^{(n)} = (-1)^n \sum_{k=0}^{n} \binom{n}{k} f_1^{(k)}(s) f_2^{(n-k)}(s)$$
$$= \sum_{k=0}^{n} \binom{n}{k} ((-1)^k f_1^{(k)}(s))((-1)^{n-k} f_2^{(n-k)}(s)) \geq 0.$$

□

Example 58. Let us show that $f(s) = \log(1 + 1/s)$, $s > 0$ is completely monotone. It is clear that a function g is completely monotone if, and only if, it is nonnegative and $-g'$ is completely monotone. According to Example 49, $-f'(s) = (s(s+1))^{-1}$ is the product of two completely monotone functions, hence $-f'$ is completely monotone by Lemma 57. Since $f \geq 0$ we conclude that f is completely monotone.

Lemma 59. *Let $f(s)$ be a completely monotone function and $g(s)$ a positive function with completely monotone derivative. Then $f(g(s))$ is completely monotone.*

Proof. We start by noting that $f(g(s)) \geq 0$ because f is nonnegative on $(0, \infty)$ and $g(s) > 0$ by assumption. We shall use the Faà di Bruno formula for the nth derivative of the composition $f(g(s))$ (the formula holds true for arbitrary differentiable functions):

$$(f(g(s)))^{(n)} = \sum_{(m, i_1, \ldots, i_l)} \frac{n!}{i_1! \cdot \ldots \cdot i_l!} f^{(m)}(g(s)) \prod_{j=1}^{l} \Big(\frac{g^{(j)}(s)}{j!} \Big)^{i_j},$$

where $(m, i_1, \ldots, i_l)$ stands for the summation over all $l \in \mathbb{N}$ and all nonnegative integer $i_1, \ldots, i_l$ such that $\sum_{j=1}^{l} j i_j = n$ and $\sum_{j=1}^{l} i_j = m$. Observe that $n = m + \sum_{j=1}^{l} (j-1) i_j$ and that, for each $j \in \mathbb{N}$, $(-1)^{j-1} g^{(j)}(s) \geq 0$. This yields, for each $n \in \mathbb{N}$,

$$(-1)^n (f(g(s)))^{(n)} = \sum_{(m, i_1, \ldots, i_l)} \frac{n!}{i_1! \cdot \ldots \cdot i_l!} (-1)^m f^{(m)}(g(s)) \prod_{j=1}^{l} \Big(\frac{(-1)^{j-1} g^{(j)}(s)}{j!} \Big)^{i_j} \geq 0.$$

□

Example 60. The functions $g_1(s) = s^\alpha$ for $\alpha \in (0, 1]$ and $g_2(s) = \log(1 + s)$ are positive on $(0, \infty)$ and their derivatives are completely monotone. Indeed, while for $h_1(s) := g_1'(s) = \alpha s^{\alpha-1}$ we have $h_1(s) > 0$ for $s > 0$ and $(-1)^n h_1^{(n)}(s) = \alpha(1-\alpha) \cdot \ldots \cdot (n-\alpha) s^{\alpha-n-1}$ for $n \in \mathbb{N}$ and $s > 0$; complete monotonicity of $h_2(s) := g_2'(s) = (s+1)^{-1}$ is known from Example 49. Thus, whenever $f(s)$ is completely monotone, so are the functions $f(s^\alpha)$ for $\alpha \in (0, 1)$ and $f(\log(1 + s))$. In particular, the functions $s \mapsto e^{-\beta s^\alpha}$ and $s \mapsto \log^{-\beta}(1 + s)$ for $\alpha \in (0, 1)$ and $\beta > 0$ are completely monotone. The former is the composition of completely monotone $s \mapsto e^{-\beta s}$ and $g_1(s)$, and the latter is the composition of completely monotone $s \mapsto s^{-\beta}$ and $g_2(s)$.

Lemma 61. *Let $(f_n(s))_{n\in\mathbb{N}}$ be a sequence of completely monotone functions. If the limit $g(s) := \lim_{n\to\infty} f_n(s)$ exists, then $g(s)$ is completely monotone.*

Proof. By Theorem 51, there exist nondecreasing and right-continuous functions $G_n : [0,\infty) \to [0,\infty)$ such that

$$f_n(s) = \int_{[0,\infty)} e^{-sx} \mathrm{d}G_n(x), \quad s > 0.$$

Since $\lim_{n\to\infty} f_n(s) = g(s)$ we infer by the direct part of Theorem 54 that

$$g(s) = \int_{[0,\infty)} e^{-sx} \mathrm{d}G(x)$$

for a nondecreasing and right-continuous function $G : [0,\infty) \to [0,\infty)$. Thus, g is completely monotone by another appeal to Theorem 51. □

Lemma 62. *Let $(a_k)_{k\in\mathbb{N}_0}$ be a sequence of nonnegative numbers. Assume that the series $h(s) := \sum_{k\geq 0} a_k s^k$ converges for $s \in (0, s_0]$ for some $s_0 > 0$. If $f(s)$ is a completely monotone function satisfying $f(s) \leq s_0$ for $s > 0$, then $h(f(s))$ is completely monotone.*

Proof. For each $n \in \mathbb{N}$, set $h_n(s) := \sum_{k=0}^{n} a_k s^k$. The function $h_n(f(s)) = \sum_{k=0}^{n} a_k f^k(s)$ is completely monotone. Indeed, while $f^0(s) \equiv 1$ for $s > 0$ is trivially completely monotone (it is positive and all its derivatives are equal to zero), the functions $f^k(s)$ for $k \in \mathbb{N}$ are completely monotone by Lemma 57. Thus, $h_n(f(s))$ is completely monotone by Lemma 56 as the linear combination with nonnegative coefficients of completely monotone functions. Finally, $h(f(s)) = \lim_{n\to\infty} h_n(f(s))$ is completely monotone by Lemma 61. □

Example 63. Let us show that $g(s) = e^{1/s}$ is completely monotone. Take $a_k = 1/k!$ for $k \in \mathbb{N}_0$. Then $h(s) = \sum_{k\geq 0} s^k/k! = e^s$ for all $s \geq 0$. According to Example 49, $f(s) = 1/s$ is completely monotone. By Lemma 62 $h(f(s)) = e^{1/s}$ is completely monotone.

2.3. Definition and examples of Laplace transforms

Let ξ be a nonnegative random variable with distribution function $F(x) := \mathbb{P}\{\xi \leq x\}$ for $x \in \mathbb{R}$. Note that $F(x) = 0$ for $x < 0$.

Definition 64. The function $\ell : [0,\infty) \to [0,\infty)$ defined by

$$\ell(s) := \mathbb{E}e^{-s\xi} = \int_{[0,\,\infty)} e^{-sx} \mathrm{d}F(x), \quad s \geq 0$$

is called *Laplace transform* of ξ or *Laplace-Stieltjes transform* of F.

If the distribution of ξ has density $h(x)$, so that $\mathrm{d}F(x) = h(x)\mathrm{d}x$, the integral defining $\ell(s)$ simplifies to

$$\ell(s) = \int_0^\infty e^{-sx} h(x) \mathrm{d}x, \quad s \geq 0.$$

If the distribution of ξ is discrete, that is, $\mathbb{P}\{\xi = x_k\} = p_k$ for $k \in \mathbb{N}$, where $x_k, p_k \geq 0$ and $\sum_{k\geq 1} p_k = 1$, then

$$\ell(s) = \sum_{k\geq 1} e^{-sx_k} p_k, \quad s \geq 0.$$

Now we give several examples of Laplace transforms.

1. Let ξ have a uniform distribution on (a, b) for $0 \leq a < b < \infty$, that is, its density is given by

$$h(x) = (b-a)^{-1} \mathbb{1}_{(a,b)}(x).$$

Then

$$\ell(s) = \frac{e^{-as} - e^{-bs}}{(b-a)s}, \quad s \geq 0. \tag{2.7}$$

Proof.

$$\ell(s) = \int_0^\infty e^{-sx} h(x) \mathrm{d}x = \frac{1}{b-a} \int_a^b e^{-sx} \mathrm{d}x = \frac{e^{-as} - e^{-bs}}{(b-a)s}.$$

□

2. Let ξ have a gamma distribution with positive parameters α and β, that is, its density is given by

$$h(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0,$$

where $\Gamma(z) = \int_0^\infty e^{-y} y^{z-1} \mathrm{d}y$ for $z > 0$ is the Euler gamma function. Then

$$\ell(s) = \Big(\frac{\beta}{s+\beta}\Big)^\alpha, \quad s \geq 0.$$

Note that a gamma distribution with parameters 1 and β is an exponential distribution with parameter β. Hence, the Laplace-Stieltjes transform of this distribution is

$$\ell(s) = \frac{\beta}{s+\beta}, \quad s \geq 0. \tag{2.8}$$

Proof.

$$\ell(s) = \int_0^\infty e^{-sx} h(x) \mathrm{d}x = \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-(s+\beta)x} x^{\alpha-1} \mathrm{d}x.$$

Changing the variable $y = (s+\beta)x$ gives

$$\ell(s) = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(s+\beta)^\alpha} \int_0^\infty e^{-y} y^{\alpha-1} \mathrm{d}y = \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{1}{(s+\beta)^\alpha} \Gamma(\alpha) = \Big(\frac{\beta}{\beta+s}\Big)^\alpha.$$

□

It is clear that if ξ takes nonnegative integer values and has generating function f, then its Laplace transform $\ell(s) = f(e^{-s})$ for $s \geq 0$. Using the examples from Section 1.2. we conclude the following.

3. If ξ has a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0, 1)$, that is,

$$\mathbb{P}\{\xi = k\} = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

then

$$\ell(s) = (pe^{-s} + 1 - p)^n, \quad s \geq 0.$$

If ξ has a Poisson distribution with parameter $\lambda > 0$, that is,

$$\mathbb{P}\{\xi = k\} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0,$$

then

$$\ell(s) = e^{-\lambda(1-e^{-s})}, \quad s \geq 0.$$

If ξ has a geometric distribution (starting at zero) with success probability $p \in (0, 1)$, that is,

$$\mathbb{P}\{\xi = k\} = p(1-p)^k, \quad k \in \mathbb{N}_0,$$

then

$$\ell(s) = \frac{p}{1 - (1-p)e^{-s}}, \quad s \geq 0.$$

2.4. Properties of Laplace transforms

2.4.1. Uniqueness theorem

Theorem 65. *The Laplace-Stieltjes transform determines distribution uniquely, that is, if ξ_1 and ξ_2 are two nonnegative random variables and $\mathbb{E}e^{-s\xi_1} = \mathbb{E}e^{-s\xi_2}$ for all $s \geq 0$, then ξ_1 has the same distribution as ξ_2.*

Proof. Set $\eta_1 := e^{-\xi_1}$ and $\eta_2 := e^{-\xi_2}$. By assumption, $\mathbb{E}\eta_1^s = \mathbb{E}\eta_2^s$ for all $s \geq 0$ and a fortiori $\mathbb{E}\eta_1^n = \mathbb{E}\eta_2^n$ for all $n \in \mathbb{N}$. In other words, the random variables η_1 and η_2 share the same moment sequence. Since they take values in $[0, 1]$, Corollary 47 applies and enables us to conclude that $\eta_1 \overset{\mathrm{d}}{=} \eta_2$, whence $\xi_1 \overset{\mathrm{d}}{=} \xi_2$. □

2.4.2. Laplace transforms and complete monotonicity Theorem 66 which is a consequence of Theorem 51 and as such also called Bernstein's criterion sets the link between Laplace transforms and completely monotone functions.

Theorem 66. *A function ℓ is the Laplace transform of a nonnegative random variable if, and only if, it is completely monotone and $\ell(0+) = 1$.*

Proof. Assume that $\ell(s) = \mathbb{E}e^{-s\xi} = \int_{[0,\infty)} e^{-sx} \mathrm{d}\mathbb{P}\{\xi \leq x\}$ for $s \geq 0$. Then ℓ is completely monotone by Theorem 51. Since $\ell(0+) = \mathbb{P}\{\xi < \infty\}$, and $\mathbb{P}\{\xi < \infty\} = 1$ we conclude that $\ell(0+) = 1$.

Conversely, assume that ℓ is completely monotone and $\ell(0+) = 1$. By Theorem 51, $\ell(s) = \int_{[0,\infty)} e^{-sx} \mathrm{d}G(x)$ for $s > 0$, where G is nondecreasing and right-continuous function. Since $\ell(0+) = G(+\infty)$, and $\ell(0+) = 1$ we infer $G(+\infty) = 1$. Thus, G is the distribution function of a nonnegative random variable. □

Example 67. Let $\alpha \in (0,1)$ and $\beta > 0$. The function ℓ given by

$$\ell(s) = e^{-\beta s^\alpha}, \quad s \geq 0$$

is the Laplace transform of a nonnegative random variable.

Proof. The function ℓ is the composition of completely monotone function $s \mapsto e^{-\beta s}$ (see Example 49) and positive function $s \mapsto s^\alpha$ with completely monotone derivative (see Example 49). Thus, ℓ is completely monotone by Lemma 59. Since $\ell(0+) = 1$, according to Theorem 66, ℓ is the Laplace transform of a nonnegative random variable. □

Definition 68. The distribution with the Laplace-Stieltjes transform $\ell(s) = e^{-\beta s^\alpha}$, $s \geq 0$, where $\alpha \in (0,1)$ and $\beta > 0$, is called *positive α-stable.*

2.4.3. Laplace transforms and log-convexity

Definition 69. Let I be an interval (closed or open or semiopen) of the real line. A function $g : I \to \mathbb{R}$ is called convex on I if for any $\alpha, \beta \geq 0$ such that $\alpha + \beta = 1$ and any $x, y \in I$

$$g(\alpha x + \beta y) \leq \alpha g(x) + \beta g(y).$$

A function $g : I \to \mathbb{R}$ is called concave on I if for any $\alpha, \beta \geq 0$ such that $\alpha + \beta = 1$ and any $x, y \in I$

$$g(\alpha x + \beta y) \geq \alpha g(x) + \beta g(y).$$

Well-known sufficient conditions for convexity (concavity) are.
(1) If a function g is twice differentiable on I and $g''(x) \geq 0$ ($g''(x) \leq 0$) for $x \in I$, then g is convex (concave) on I.
(2) If a function g is differentiable on I and $g'(x)$ is nondecreasing (nonincreasing) on I, then g is convex (concave) on I.

Definition 70. Let I be an interval (closed or open or semiopen) of the real line. A function $g : I \to (0, \infty)$ is called *log-convex* on I if $\log g$ is a convex function, that is, for any $\alpha, \beta \geq 0$ such that $\alpha + \beta = 1$ and any $x, y \in I$

$$g(\alpha x + \beta y) \leq g^\alpha(x) g^\beta(y).$$

Lemma 71. *Every log-convex function is also convex.*

Proof. We shall use the Young inequality: for any $p, q > 0$ such that $1/p + 1/q = 1$ and any $a, b \geq 0$

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}. \tag{2.9}$$

By assumption, for any $\alpha, \beta \geq 0$ such that $\alpha + \beta = 1$ and any $x, y \in I$

$$g(\alpha x + \beta y) \leq g^\alpha(x) g^\beta(y).$$

We have to prove that

$$g(\alpha x + \beta y) \leq \alpha g(x) + \beta g(y).$$

Thus, it suffices to check that

$$g^\alpha(x) g^\beta(y) \leq \alpha g(x) + \beta g(y).$$

This is an immediate consequence of the Young inequality with $a = g^\alpha(x)$, $b = g^\beta(y)$, $p = 1/\alpha$ and $q = 1/\beta$. $\square$

On the other hand, a positive convex function is not necessarily log-convex. For instance, $g(x) = x^2$ is convex on $[0, \infty)$, however, $\log g(x) = 2 \log x$ is concave on $[0, \infty)$.

Let ℓ be the Laplace transform of a nonnegative random variable. By Theorem 66, the function ℓ is completely monotone. In particular, $\ell''(s) \geq 0$ which shows that ℓ is convex on $[0, \infty)$. Is ℓ log-convex on I? From the discussion above it follows that log-convexity is not implied by convexity.

Proposition 72. *The Laplace transform of a nonnegative random variable is log-convex on* $[0, \infty)$.

For the proof we need a classical inequality.

Hölder inequality. Let p and q be positive numbers satisfying $1/p + 1/q = 1$. For nonnegative random variables X and Y

$$\mathbb{E}XY \leq (\mathbb{E}X^p)^{1/p} (\mathbb{E}Y^q)^{1/q} \tag{2.10}$$

provided that the right-hand side is finite.

Proof. The inequality is trivial if one of the random variables takes value 0 with probability one. Hence, we assume in what follows that $\mathbb{E}X^p > 0$ and $\mathbb{E}Y^q > 0$. Use the Young inequality (2.9) with $a = X/(\mathbb{E}X^p)^{1/p}$ and $b = Y/(\mathbb{E}Y^q)^{1/q}$ to obtain

$$\frac{X}{(\mathbb{E}X^p)^{1/p}} \frac{Y}{(\mathbb{E}Y^q)^{1/q}} \leq \frac{X^p}{p\mathbb{E}X^p} + \frac{Y^q}{q\mathbb{E}Y^q} \quad \text{a.s.}$$

Passing to expectations gives

$$\frac{\mathbb{E}XY}{(\mathbb{E}X^p)^{1/p}(\mathbb{E}Y^q)^{1/q}} \leq \frac{\mathbb{E}X^p}{p\mathbb{E}X^p} + \frac{\mathbb{E}Y^q}{q\mathbb{E}Y^q} = \frac{1}{p} + \frac{1}{q} = 1,$$

whence (2.10). $\square$

Proof of Proposition 72. Let $s, t \geq 0$. Use Holder's inequality with $X = e^{-\alpha s \xi}$, $Y = e^{-\beta t \xi}$ for nonnegative α and β such that $\alpha + \beta = 1$ and with $p = 1/\alpha$ and $q = 1/\beta$. This gives

$$\ell(\alpha s + \beta t) = \mathbb{E}e^{-\alpha s \xi} e^{-\beta t \xi} \leq (\mathbb{E}e^{-s\xi})^\alpha (\mathbb{E}e^{-t\xi})^\beta = \ell^\alpha(s) \ell^\beta(t).$$

Thus, ℓ is indeed log-convex. $\square$

2.4.4. Laplace transforms and arithmetic operations Let ξ_1 and ξ_2 be nonnegative independent random variables with distribution functions $F(x)$ and $G(x)$ for $x \in \mathbb{R}$. Obviously, $F(x) = G(x) = 0$ for $x < 0$.

The distribution of the sum $\xi_1 + \xi_2$ is given by

$$(F \star G)(x) := \mathbb{P}\{\xi_1 + \xi_2 \le x\} = \int_{[0,\, x]} F(x-y)\mathrm{d}G(y) = \int_{[0,\, x]} G(x-y)\mathrm{d}F(y), \quad x \ge 0 \tag{2.11}$$

and $\mathbb{P}\{\xi_1 + \xi_2 \le x\} = 0$ for $x < 0$.

Definition 73. The function $F \star G$ is called *convolution* of the distribution functions F and G.

The operation $\star$ is also called convolution. From (2.11) it follows that convolution is a commutative operation, that is, $F \star G = G \star F$.

Proof of (2.11). We use the total probability formula in an integral form: for $x \ge 0$,

$$\begin{aligned}
\mathbb{P}\{\xi_1 + \xi_2 \le x\} &= \int_{[0,\infty)} \mathbb{P}\{\xi_1 + \xi_2 \le x | \xi_2 = y\}\mathrm{d}\mathbb{P}\{\xi_2 \le y\} \\
&= \int_{[0,\,\infty)} \mathbb{P}\{\xi_1 \le x - y | \xi_2 = y\}\mathrm{d}\mathbb{P}\{\xi_2 \le y\} = \int_{[0,\,\infty)} \mathbb{P}\{\xi_1 \le x - y\}\mathrm{d}\mathbb{P}\{\xi_2 \le y\} \\
&= \int_{[0,\,x]} F(x-y)\mathrm{d}G(y)
\end{aligned}$$

The third equality is implied by the independence of ξ_1 and ξ_2. In the last equality we narrowed the interval of integration to $[0,\, x]$ instead of $[0,\, \infty)$ because $F(x-y) = 0$ for $y > x$. The second equality in (2.11) follows similarly: just condition on ξ_1 rather than ξ_2. □

Proposition 74 is a counterpart of Proposition 8 for generating functions. It tells us that although convolution is a complex operation, the corresponding Laplace transform can be easily calculated.

Proposition 74. *Let ξ_1 and ξ_2 be nonnegative independent random variables with Laplace transforms ℓ_1 and ℓ_2. Then the Laplace transform of $\xi_1 + \xi_2$ is $\ell_1 \ell_2$.*

Proof. For $s \ge 0$, $\mathbb{E}e^{-s(\xi_1+\xi_2)} = \mathbb{E}e^{-s\xi_1}e^{-s\xi_2} = \mathbb{E}e^{-s\xi_1}\mathbb{E}e^{-s\xi_2} = \ell_1(s)\ell_2(s)$. For the second equality we have used the fact that the random variables $e^{-s\xi_1}$ and $e^{-s\xi_2}$ are independent, and that the expectation of the product of independent random variables is equal to the product of the expectations. □

Example 75. Let $n \in \mathbb{N}$ and denote by $\xi_1, \xi_2, \ldots, \xi_n$ independent random variables with a uniform distribution on $(0, 1)$. Set $S_n := \xi_1 + \ldots + \xi_n$. Let us show that the distribution function of S_n is given by

$$\mathbb{P}\{S_n \le x\} = \frac{1}{n!} \sum_{k=0}^{n} (-1)^k \binom{n}{k} (x-k)_+^n, \quad x \ge 0, \tag{2.12}$$

where $y_+ = y$ if $y \geq 0$ and $= 0$ if $y < 0$. It is clear that $\mathbb{P}\{S_n \leq x\} = 0$ for $x < 0$ and $= 1$ for $x > n$ and $\mathbb{P}\{S_n \leq x\} = \frac{1}{n!}\sum_{k=0}^{n-1}(-1)^k\binom{n}{k}(x-k)_+^n$ for $x \in [0, n]$.

Proof. From (2.7) we know that $\mathbb{E}e^{-s\xi_1} = s^{-1}(1-e^{-s})$ for $s > 0$. Therefore, by Proposition 74,

$$\mathbb{E}e^{-sS_n} = \Big(\frac{1-e^{-s}}{s}\Big)^n.$$

Using the binomial theorem yields

$$\Big(\frac{1-e^{-s}}{s}\Big)^n = \sum_{k=0}^n (-1)^k \binom{n}{k} e^{-ks}s^{-n}, \quad s \geq 0.$$

Set $T_a(x) = ((x-a)_+^n + c)/n!$, where $a, c \geq 0$ are arbitrary constants. Then

$$\begin{aligned}\int_{[0,\infty)} e^{-sx}\mathrm{d}T_a(x) &= \frac{e^{-as}}{(n-1)!}\int_a^\infty e^{-s(x-a)}(x-a)^{n-1}\mathrm{d}x = \frac{e^{-as}}{(n-1)!}\int_0^\infty e^{-sy}y^{n-1}\mathrm{d}y \\ &= \frac{e^{-as}}{s^n(n-1)!}\int_0^\infty e^{-z}z^{n-1}\mathrm{d}z = \frac{e^{-as}}{s^n(n-1)!}\Gamma(n) = e^{-as}s^{-n}.\end{aligned}$$

Here the second and third equalities are obtained by the change of variables $y = x - a$ and $z = sy$, respectively. In the last equality we have used the fact that $\Gamma(n) = (n-1)!$. This follows from $\Gamma(1) = \int_0^\infty e^{-y}\mathrm{d}y = 1$ and $\Gamma(x+1) = x\Gamma(x)$ for $x > 0$ which can be seen with the help of integration by parts. Thus,

$$\sum_{k=0}^n (-1)^k\binom{n}{k}e^{-ks}s^{-n} = \sum_{k=0}^n (-1)^k\binom{n}{k}\int_0^\infty e^{-sx}\mathrm{d}T_k(x) = \int_0^\infty e^{-sx}\mathrm{d}\Big(\sum_{k=0}^n(-1)^k\binom{n}{k}T_k(x)\Big).$$

We have to check that

$$\sum_{k=0}^n (-1)^k\binom{n}{k}T_k(x) = \frac{1}{n!}\sum_{k=0}^n(-1)^k\binom{n}{k}((x-k)_+^n + c)$$

is the distribution function of a positive random variable. The function is obviously continuous on $[0, \infty)$. Differentiating it we see it is nondecreasing. To ensure that this function is 0 at 0 we have to put $c = 0$. Finally, since

$$\Big(\frac{1-e^{-s}}{s}\Big)^n = \int_0^\infty e^{-sx}\mathrm{d}\Big(\frac{1}{n!}\sum_{k=0}^n(-1)^k\binom{n}{k}(x-k)_+^n\Big)$$

and $\lim_{s\to 0+}((1-e^{-s})/s)^n = 1$ we infer that the function in focus converges to 1 as $x \to \infty$. Hence, it is indeed a distribution function. By the uniqueness theorem for Laplace transforms (Theorem 65) formula (2.12) follows. □

According to (2.12), for $x \in [0, 2]$,

$$\mathbb{P}\{S_2 \leq x\} = (x^2 - 2(x-1)_+^2)/2 = \begin{cases} x^2/2, & \text{if } x \in [0,1], \\ -x^2/2 + 2x - 1, & \text{if } x \in (1, 2]. \end{cases}$$

As a check, let us show this directly, not resorting to Laplace transforms. Recall that the distribution function of ξ_1 is $F(x) = x$ for $x \in [0, 1]$ and $= 1$ for $x > 1$ and that the density of ξ_2 is $h(x) = \mathbb{1}_{(0,1)}(x)$. Using now (2.11) in combination with $\mathrm{d}G(x) = h(x)\mathrm{d}x$ we obtain

$$\mathbb{P}\{S_2 \le x\} = \mathbb{P}\{\xi_1 + \xi_2 \le x\} = \int_0^x F(x-y)h(y)\mathrm{d}y.$$

We have to consider two cases separately.
Case $x \in [0, 1]$. Then

$$\int_0^x F(x-y)h(y)\mathrm{d}y = \int_0^x F(x-y)\mathrm{d}y = \int_0^x F(y)\mathrm{d}y = \int_0^x y\mathrm{d}y = x^2/2.$$

Case $x \in (1, 2]$. Then

$$\int_0^x F(x-y)h(y)\mathrm{d}y = \int_0^1 F(x-y)\mathrm{d}y = \int_{x-1}^x F(y)\mathrm{d}y = \int_{x-1}^1 F(y)\mathrm{d}y + \int_1^x F(y)\mathrm{d}y$$
$$= \int_{x-1}^1 y\mathrm{d}y + \int_1^x \mathrm{d}y = (1-(x-1)^2)/2 + x - 1 = -x^2/2 + 2x - 1.$$

Example 76. Let $\theta_1, \theta_2, \ldots$ be independent random variables with distributions $\mathbb{P}\{\theta_k = 0\} = \mathbb{P}\{\theta_k = 2^{-k}\} = 1/2$ for $k \in \mathbb{N}$. Then, as $n \to \infty$, $\theta_1 + \ldots + \theta_n$ converges in distribution to a random variable having a uniform distribution on $(0, 1)$.

Proof. By continuity theorem for Laplace-Stieltjes transforms (Theorem 54) it is sufficient to prove that, for $s > 0$,

$$\lim_{n\to\infty} \mathbb{E}e^{-s(\theta_1+\ldots+\theta_n)} = \frac{1-e^{-s}}{s}. \tag{2.13}$$

Recall that $s \mapsto s^{-1}(1-e^{-s})$ is the Laplace-Stieltjes transform of a uniform distribution on $(0, 1)$.

Since $\mathbb{E}e^{-s\theta_k} = (1+e^{-s2^{-k}})/2$ we obtain with the help of Proposition 74 that

$$\mathbb{E}e^{-s(\theta_1+\ldots+\theta_n)} = (1+e^{-s/2})/2 \cdot \ldots \cdot (1+e^{-s2^{-n}})/2.$$

Multiplying and then dividing the right-hand side by $1 - e^{-s2^{-n}}$ and using $(1-x)(1+x) = 1 - x^2$ we arrive at

$$\mathbb{E}e^{-s(\theta_1+\ldots+\theta_n)} = \frac{1-e^{-s}}{2^n(1-e^{-s2^{-n}})}.$$

Since $\lim_{x\to 0}(1-e^{-x})/x = 1$ we have $\lim_{n\to\infty} 2^n(1-e^{-s2^{-n}}) = s$ and (2.13) follows.

□

Lemma 77. *Let ξ_1 and ξ_2 be nonnegative independent random variables with distribution functions F_1 and F_2 and Laplace transforms ℓ_1 and ℓ_2. Then, for $s \ge 0$,*

$$\mathbb{E}e^{-s\xi_1\xi_2} = \int_0^\infty \ell_1(sy)\mathrm{d}F_2(y) = \int_0^\infty \ell_2(sy)\mathrm{d}F_1(y). \tag{2.14}$$

Proof. We use the total expectation formula in an integral form: for $s \geq 0$,

$$\mathbb{E}e^{-s\xi_1\xi_2} = \int_{[0,\infty)} \mathbb{E}(e^{-s\xi_1\xi_2}|\xi_2 = y)\mathrm{d}\mathbb{P}\{\xi_2 \leq y\} = \int_{[0,\infty)} \mathbb{E}(e^{-sy\xi_1}|\xi_2 = y)\mathrm{d}\mathbb{P}\{\xi_2 \leq y\}$$
$$= \int_{[0,\infty)} \mathbb{E}e^{-sy\xi_1}\mathrm{d}\mathbb{P}\{\xi_2 \leq y\} = \int_{[0,\infty)} \ell_1(sy)\mathrm{d}F_2(y).$$

The second equality is implied by the independence of ξ_1 and ξ_2. The second equality in (2.14) follows similarly: just condition on ξ_1 rather than ξ_2. □

Example 78. Let ξ_1 and ξ_2 be independent random variables such that ξ_1 has a gamma distribution with parameters 2 and 1 and ξ_2 has a uniform distribution on $(0,1)$. Then $\xi_1\xi_2$ has an exponential distribution with parameter 1.

Proof. The random variable ξ_2 has density $h(x) = \mathbb{1}_{(0,1)}(x)$. According to (2.8), $\mathbb{E}e^{-s\xi_1} = (s+1)^{-2}$ for $s \geq 0$. Hence, by Lemma 77

$$\mathbb{E}e^{-s\xi_1\xi_2} = \int_0^\infty \mathbb{E}e^{-sy\xi_1}h(y)\mathrm{d}y = \int_0^1 (sy+1)^{-2}\mathrm{d}y = -s^{-1}(sy+1)^{-1}|_0^1 = (s+1)^{-1}.$$

This is the Laplace-Stieltjes transform of an exponential distribution with parameter 1, see (2.8). □

Lemma 79 is a counterpart of Proposition 12. Its proof is left as an exercise.

Lemma 79. *Let ξ_1, $\xi_2, \ldots$ be nonnegative independent identically distributed random variables which are independent of a random variable N taking nonnegative integer values. If the Laplace transform of ξ_1 is $\ell(s)$ and the generating function of N is $f(s)$, then the Laplace transform of $\sum_{k=1}^N \xi_k$ is $f(\ell(s))$.*

Corollary 80. If $\ell(s)$ is a Laplace transform of a nonnegative random variable, then so are $e^{-\lambda(1-\ell(s))}$ and $p/(1-(1-p)\ell(s))$ for any $\lambda > 0$ and any $p \in (0,1)$.

Proof. Just observe that both functions are of the form $f(\ell(s))$, where in the former case $f(s) = e^{-\lambda(1-s)}$ is the generating function of a Poisson distribution with parameter λ and in the latter case $f(s) = p/(1-(1-p)s)$ is the generating function of a geometric distribution (starting at zero) with success probability p. □

2.4.5. Laplace transforms and moments

Proposition 81. *Let ξ be a nonnegative random variable with Laplace transform ℓ. Both sides of the equality that follows are either finite or infinite: for $n \in \mathbb{N}$,*

$$\mathbb{E}\xi^n = (-1)^n \ell^{(n)}(0+), \tag{2.15}$$

where $\ell^{(n)}$ is the derivative of the nth order.

Proof. Using (2.3) we conclude that, for $n \in \mathbb{N}$,

$$(-1)^n \ell^{(n)}(0+) = \lim_{s\to 0+} (-1)^n \ell^{(n)}(s) = \lim_{s\to 0+} \int_{(0,\infty)} e^{-sx} x^n \mathrm{d}F(x) = \int_{(0,\infty)} \lim_{s\to 0+} e^{-sx} x^n \mathrm{d}F(x)$$
$$= \int_{(0,\infty)} x^n \mathrm{d}F(x) = \mathbb{E}\xi^n.$$

To interchange the limit and the integral we have used Lévy monotone convergence theorem which is applicable because the function $s \mapsto e^{-sx}$ is decreasing for each $x > 0$. $\square$

Example 82. Let ξ be a random variable with an exponential distribution with parameter $\beta > 0$. Since $\ell(s) = \mathbb{E}e^{-s\xi} = \beta(s+\beta)^{-1}$ for $s \geq 0$ and, for $n \in \mathbb{N}$,

$$(-1)^n \ell^{(n)}(s) = \frac{\beta n!}{(s+\beta)^{n+1}},$$

we infer

$$\mathbb{E}\xi^n = (-1)^n \ell^{(n)}(0+) = \frac{n!}{\beta^n}, \quad n \in \mathbb{N}.$$

Example 83. Let ξ be a random variable with the Sibuya distribution with parameter $\alpha \in (0,1)$. According to Problem 34(c), its generating function is $f(s) = 1-(1-s)^\alpha$ for $s \in [0,1)$. Hence, its Laplace transform is $\ell(s) = f(e^{-s}) = 1-(1-e^{-s})^\alpha$. Since $-\ell'(s) = \alpha e^{-s}(1-e^{-s})^{\alpha-1}$, we conclude that $\mathbb{E}\xi = -\ell'(0+) = \infty$.

We already know how to calculate the moments of positive integer orders with the aid of Laplace transforms. What about moments of fractional order? May Laplace transforms be of any help? The answer is ‘yes’. Before giving the corresponding statements we formulate two auxiliary results.

Lemma 84. *Let ξ have an exponential distribution with parameter 1 and η be a nonnegative random variable with Laplace transform ℓ. Assume that ξ and η are independent. Then $\ell(s) = \mathbb{P}\{\xi/\eta > s\}$ for $s \geq 0$. In other words, the distribution function of the random variable ξ/η is $1-\ell$.*

Proof. We use the total probability formula in an integral form: for $s \geq 0$:

$$\mathbb{P}\{\xi/\eta > s\} = \int_{[0,\infty)} \mathbb{P}\{\xi/\eta > s | \eta = x\} \mathrm{d}\mathbb{P}\{\eta \leq x\} = \int_{[0,\infty)} \mathbb{P}\{\xi > sx | \eta = x\} \mathrm{d}\mathbb{P}\{\eta \leq x\}$$
$$= \int_{[0,\infty)} \mathbb{P}\{\xi > sx\} \mathrm{d}\mathbb{P}\{\eta \leq x\} = \int_{[0,\infty)} e^{-sx} \mathrm{d}\mathbb{P}\{\eta \leq x\} = \mathbb{E}e^{-s\eta} = \ell(s).$$

The third equality follows from the independence of ξ and η. $\square$

Lemma 85. *Let ξ be a positive random variable. Both sides of the equality that follows are either finite or infinite: for $a > 0$,*

$$\mathbb{E}\xi^{-a} = a \int_0^\infty x^{-a-1} \mathbb{P}\{\xi \leq x\} \mathrm{d}x. \tag{2.16}$$

Proof. Let A be an event. Then the random variable $\mathbb{1}_A$ takes two values: 1 with probability $\mathbb{P}(A)$ and 0 with probability $1-\mathbb{P}(A)$. In particular, $\mathbb{E}\mathbb{1}_A=\mathbb{P}(A)$.

Using the last equality with $A=\{\xi\leq x\}$ we obtain

$$a\int_0^\infty x^{-a-1}\mathbb{P}\{\xi\leq x\}\mathrm{d}x=a\int_0^\infty x^{-a-1}\mathbb{E}\mathbb{1}_{\{\xi\leq x\}}\mathrm{d}x=a\mathbb{E}\int_0^\infty x^{-a-1}\mathbb{1}_{\{\xi\leq x\}}\mathrm{d}x$$
$$=a\mathbb{E}\int_\xi^\infty x^{-a-1}\mathrm{d}x=\mathbb{E}\xi^{-a}$$

having utilized Fubini's theorem for the second equality. The theorem which is immediately applicable because the integrand is nonnegative enabled us to interchange the expectation and the integral. □

Proposition 86. *Let η be a nonnegative random variable with Laplace transform ℓ. Both sides of the equality that follows are either finite or infinite: for $\gamma\in(0,1)$,*

$$\mathbb{E}\eta^\gamma=\frac{\gamma}{\Gamma(1-\gamma)}\int_0^\infty s^{-\gamma-1}(1-\ell(s))\mathrm{d}s,$$

where $\Gamma(\cdot)$ is the Euler gamma function.

Proof. Let ξ have an exponential distribution with parameter 1 and be independent of η. We have

$$\mathbb{E}(\xi/\eta)^{-\gamma}=\gamma\int_0^\infty s^{-\gamma-1}\mathbb{P}\{\xi/\eta\leq s\}\mathrm{d}s=\gamma\int_0^\infty s^{-\gamma-1}(1-\ell(s))\mathrm{d}s,$$

where the first equality follows from (2.16), and the second equality is a consequence of Lemma 84. On the other hand,

$$\mathbb{E}(\xi/\eta)^{-\gamma}=\mathbb{E}\xi^{-\gamma}\mathbb{E}\eta^\gamma=\Gamma(1-\gamma)\mathbb{E}\eta^\gamma$$

because $\mathbb{E}\xi^{-\gamma}=\int_0^\infty x^{-\gamma}e^{-x}\mathrm{d}x=\Gamma(1-\gamma)$. □

Example 87. Let η be a random variable with Laplace transform $\ell(s)=e^{-s^\alpha}$ for some $\alpha\in(0,1)$ (thus, the distribution of η is positive α-stable, see Definition 68). Then $\mathbb{E}\eta^\gamma=\Gamma(1-\gamma/\alpha)/\Gamma(1-\gamma)<\infty$ for $\gamma\in(0,\alpha)$ and $\mathbb{E}\eta^\gamma=\infty$ for $\gamma\geq\alpha$.

Proof. By Proposition 86,

$$\mathbb{E}\eta^\gamma=\frac{\gamma}{\Gamma(1-\gamma)}\int_0^\infty s^{-\gamma-1}(1-e^{-s^\alpha})\mathrm{d}s=\frac{\gamma}{\alpha\Gamma(1-\gamma)}\int_0^\infty y^{-\gamma/\alpha-1}(1-e^{-y})\mathrm{d}y. \tag{2.17}$$

For the last equality we have used the change of variable $y=s^\alpha$ so that $\mathrm{d}y/y=\alpha\mathrm{d}s/s$.

Assume that $\gamma/\alpha\geq 1$. As $y\to 0+$, $y^{-\gamma/\alpha-1}(1-e^{-y})\sim y^{-\gamma/\alpha}$, where $g_1\sim g_2$ means that $\lim_{y\to0+}(g_1(y)/g_2(y)=1$. Since the function $y\mapsto y^{-\gamma/\alpha}$ is not integrable near zero, the integral on the right-hand side of (2.17) diverges, whence $\mathbb{E}\eta^\gamma=\infty$.

Assume now that $0<\gamma/\alpha<1$. Integration by parts yields

$$\int_0^\infty y^{-\gamma/\alpha-1}(1-e^{-y})\mathrm{d}y=(\alpha/\gamma)\int_0^\infty(1-e^{-y})\mathrm{d}(-y^{-\gamma/\alpha})=(\alpha/\gamma)\int_0^\infty y^{-\gamma/\alpha}e^{-y}\mathrm{d}y$$
$$=\alpha\Gamma(1-\gamma/\alpha)/\gamma<\infty.$$

□

The proof of the next proposition which is similar to that of Proposition 86 is left as an exercise.

Proposition 88. *Let η be a nonnegative random variable with Laplace transform ℓ. Both sides of the equality that follows are either finite or infinite: for $\gamma > 0$,*

$$\mathbb{E}\eta^{-\gamma} = \frac{\gamma}{\Gamma(1+\gamma)} \int_0^\infty s^{\gamma-1}\ell(s)\mathrm{d}s,$$

where $\Gamma(\cdot)$ is the Euler gamma function.

Example 89. Let η be a random variable with Laplace transform $\ell(s) = e^{-s^\alpha}$ for some $\alpha \in (0,1)$, that is, the distribution of η is positive α-stable, see Definition 68. Set $\theta_\alpha := \eta^{-\alpha}$. Then

$$\mathbb{E}\theta_\alpha^n = \frac{n!}{\Gamma(1+\alpha n)}, \quad n \in \mathbb{N}$$

and the moment generating function of θ_α is given by

$$\mathbb{E}e^{s\theta_\alpha} = \sum_{n\geq 0} \frac{s^n}{\Gamma(1+n\alpha)}, \quad s \geq 0. \tag{2.18}$$

Proof. According to Proposition 88,

$$\mathbb{E}\theta_\alpha^n = \mathbb{E}\eta^{-\alpha n} = \frac{\alpha n}{\Gamma(1+\alpha n)} \int_0^\infty s^{\alpha n-1} e^{-s^\alpha}\mathrm{d}s = \frac{n}{\Gamma(1+\alpha n)} \int_0^\infty y^{n-1}e^{-y}\mathrm{d}y = \frac{n\Gamma(n)}{\Gamma(1+\alpha n)}$$
$$= \frac{n!}{\Gamma(1+\alpha n)}.$$

For the second equality we have used the change of variable $y = s^\alpha$.

Equality (2.18) follows immediately:

$$\mathbb{E}e^{s\theta_\alpha} = \sum_{n\geq 0} \frac{s^n}{n!}\mathbb{E}\theta_\alpha^n = \sum_{n\geq 0} \frac{s^n}{\Gamma(1+\alpha n)}, \quad s \geq 0.$$

However, the series may diverge. Therefore, we have to check that it converges for all $s \geq 0$ (in other words, its radius of convergence $r = \infty$). The radius of convergence of any power series $\sum_{n\geq 0} s^n a_n$ is given by the Cauchy-Hadamard formula

$$1/r = \limsup_{n\to\infty} |a_n|^{1/n}.$$

Thus, we are left with showing that

$$\lim_{n\to\infty} (\Gamma(1+\alpha n))^{1/n} = \infty.$$

This is a consequence of

$$\Gamma(1+\alpha n) = \int_0^\infty x^{\alpha n}e^{-x}\mathrm{d}x \geq \int_n^\infty x^{\alpha n}e^{-x}\mathrm{d}x \geq n^{\alpha n}e^{-n}$$

which shows that $(\Gamma(1+\alpha n))^{1/n} \geq e^{-1}n^\alpha \to \infty$ as $n \to \infty$.

□

Definition 90. For $\alpha \in (0,1]$, the function $E_\alpha(s) := \sum_{n\geq 0}(s^n/\Gamma(1+n\alpha))$, $s \geq 0$ is called *Mittag-Leffler function* with parameter α.

Note that the Mittag-Leffler function with parameter 1 is just the exponential function:

$$E_1(s) = \sum_{n\geq 0}(s^n/\Gamma(1+n)) = e^s, \quad s \geq 0.$$

In view of the discussion above the term introduced in the next definition is quite natural.

Definition 91. The distribution of a random variable θ_α is called *Mittag-Leffler distribution* with parameter α.

2.4.6. Further properties

Lemma 92. *Let ξ be a nonnegative random variable with Laplace transform ℓ. Then $\mathbb{P}\{\xi = 0\} = \lim_{s\to\infty}\ell(s)$.*

Proof. Write

$$\begin{aligned}\ell(s) = \int_{[0,\infty)} e^{-sx}\mathrm{d}\mathbb{P}\{\xi \leq x\} &= \int_{\{0\}} e^{-sx}\mathrm{d}\mathbb{P}\{\xi \leq x\} + \int_{(0,\infty)} e^{-sx}\mathrm{d}\mathbb{P}\{\xi \leq x\} \\ &= \mathbb{P}\{\xi = 0\} + \int_{(0,\infty)} e^{-sx}\mathrm{d}\mathbb{P}\{\xi \leq x\}.\end{aligned}$$

When $x \in (0,\infty)$, $\lim_{s\to\infty} e^{-sx} = 0$. Since the function $s \mapsto e^{-sx}$ is decreasing for each $x > 0$ we can use Lévy's monotone convergence theorem to infer

$$\lim_{s\to\infty}\int_{(0,\infty)} e^{-sx}\mathbb{P}\{\xi \leq x\} = \int_{(0,\infty)}(\lim_{s\to\infty} e^{-sx})\mathbb{P}\{\xi \leq x\} = 0.$$

Alternatively, we can interchange the limit and the integral by an appeal to Lebesgue's dominated convergence theorem. Indeed, observe that $e^{-sx} \leq 1$ for each $s, x > 0$ and $\int_{(0,\infty)} 1 \cdot \mathrm{d}\mathbb{P}\{\xi \leq x\} = \mathbb{P}\{\xi > 0\} < \infty$. ☐

Example 93. If ξ has a Poisson distribution with parameter $\lambda > 0$, then $\mathbb{P}\{\xi = 0\} = e^{-\lambda}$. Let us now confirm this fact with the help of Lemma 92. Since the Laplace transform of ξ is $\ell(s) = e^{-\lambda(1-e^{-s})}$, we conclude that $\lim_{s\to\infty}\ell(s) = e^{-\lambda}$, and the right-hand side is indeed $\mathbb{P}\{\xi = 0\}$.

Example 94. If ξ has an exponential distribution with parameter β, then $\mathbb{P}\{\xi = 0\} = 0$ because the distribution is continuous. The Laplace transform of ξ is $\ell(s) = \beta(s+\beta)^{-1}$, and $\lim_{s\to\infty}\ell(s) = 0$ as it must be. Assume now that $F(x) = \mathbb{P}\{\xi \leq x\} = 1 - (1-\gamma)e^{-x}$ for $x \geq 0$ and some $\gamma \in (0,1)$. Then $\mathbb{P}\{\xi = 0\} = F(0+) = \gamma$. The Laplace transform of ξ is

$$\begin{aligned}\ell(s) = \int_{[0,\infty)} e^{-sx}\mathrm{d}F(x) &= \int_{\{0\}} e^{-sx}\mathrm{d}F(x) + \int_{(0,\infty)} e^{-sx}F'(x)\mathrm{d}x \\ &= F(0+) + (1-\gamma)\int_0^\infty e^{-(s+1)x}\mathrm{d}x = \gamma + \frac{1-\gamma}{s+1} = \frac{\gamma s+1}{s+1}.\end{aligned}$$

Therefore, $\lim_{s\to\infty}\ell(s) = \gamma$, and the right-hand side is indeed $\mathbb{P}\{\xi = 0\}$.

Lemma 95. *Let ξ be a nonnegative random variable with Laplace transform ℓ and density h. If $\lim_{x\to 0+} h(x) = c \in [0, \infty]$, then $\lim_{s\to+\infty} s\ell(s) = c$.*

Proof. Assume that $\lim_{x\to 0+} h(x) = c \in [0, \infty)$. Then, given $\varepsilon > 0$ there exists $x_0 > 0$ such that $c - \varepsilon \le h(x) \le c + \varepsilon$ (if $c = 0$, the left-hand inequality is not needed) whenever $x \in (0, x_0]$.

Write $s\ell(s) = s\int_0^{x_0} e^{-sx}h(x)\mathrm{d}x + s\int_{x_0}^\infty e^{-sx}h(x)\mathrm{d}x$. For the second integral we have $\lim_{s\to\infty} s\int_{x_0}^\infty e^{-sx}h(x)\mathrm{d}x = 0$ in view of the inequality

$$0 \le s\int_{x_0}^\infty e^{-sx}h(x)\mathrm{d}x \le se^{-sx_0}\int_{x_0}^\infty h(x)\mathrm{d}x \le se^{-sx_0}$$

and the fact that the function on the right-hand side converges to 0 as $s \to \infty$.

For the first integral we obtain

$$s\int_0^{x_0} e^{-sx}h(x)\mathrm{d}x \le (c+\varepsilon)s\int_0^{x_0} e^{-sx}\mathrm{d}x = (c+\varepsilon)(1 - e^{-sx_0}).$$

Thus, $\limsup_{s\to\infty} s\int_0^{x_0} e^{-sx}h(x)\mathrm{d}x \le c+\varepsilon$ and thereupon $\limsup_{s\to\infty} s\ell(s) \le c+\varepsilon$. Letting $\varepsilon \to 0+$ yields $\limsup_{s\to\infty} s\ell(s) \le c$. This completes the proof when $c = 0$. When $c \in (0, \infty)$ we argue similarly. From

$$s\int_0^{x_0} e^{-sx}h(x)\mathrm{d}x \ge (c-\varepsilon)s\int_0^{x_0} e^{-sx}\mathrm{d}x = (c-\varepsilon)(1 - e^{-sx_0})$$

we first derive $\liminf_{s\to\infty} s\ell(s) \ge c - \varepsilon$ and then $\liminf_{s\to\infty} s\ell(s) \ge c$. Combining the results for the limit superior and the limit inferior completes the proof in the case $c \in (0, \infty)$. The proof in the case $c = \infty$ is left as an exercise. □

Example 96. Let ξ be a random variable having an exponential distribution with parameter $\beta > 0$. Then its density $h(x) = \beta e^{-\beta x}\mathbb{1}_{(0,\infty)}(x)$ satisfies $\lim_{x\to 0+} h(x) = \beta$. This is in full agreement with Lemma 95 because $\ell(s) = \beta(s+\beta)^{-1}$ the Laplace transform of ξ satisfies $\lim_{s\to\infty} s\ell(s) = \beta$.

2.5. Problems

Problem 97. Let $a > 0$ and ξ_a be a random variable with distribution

$$\mathbb{P}\{\xi_a = ae^k\} = a^{-k}e^{-k^2/2}/c_a, \quad k \in \mathbb{Z},$$

where $c_a = \sum_{k\in\mathbb{Z}} a^{-k}e^{-k^2/2}$. Prove that $\mathbb{E}\xi_a^n = e^{n^2/2}$ for $n \in \mathbb{N}$, that is, the moment sequence of ξ_a is the same as that of the log-normal distribution.

Problem 98. Let ξ_1, $\xi_2, \ldots$ be nonnegative independent identically distributed random variables which are independent of a random variable N taking nonnegative integer values. Prove the following. If the Laplace transform of ξ_1 is $\ell(s)$ and the generating function of N is $f(s)$, then the Laplace transform of $\sum_{k=1}^N \xi_k$ is $f(\ell(s))$.

Problem 99. Let $\ell(s)$, $s \geq 0$ be the Laplace transform of a nonnegative random variable. Is

(a)

$$\psi(s) = 1 - (1 - \ell(s))^{\alpha}, \quad s \geq 0$$

for $\alpha \in (0, 1)$;

(b)

$$\psi(s) = \frac{1}{1 + \ell(s)}, \quad s \geq 0$$

(c)

$$\psi(s) = \ell(\log(1 + s)), \quad s \geq 0$$

the Laplace transform of a nonnegative random variable?

Problem 100. Let ℓ be the Laplace transform of a nonnegative random variable ξ. Prove that the function $s \mapsto (1 - \ell(s))/s$ is completely monotone. Show that it is the Laplace transform of a nonnegative random variable if, and only if, $\mathbb{E}\xi = 1$.

Problem 101. Let $\gamma \in (0, 1)$ and θ_1, $\theta_2, \ldots$ be independent identically distributed random variables with Laplace transform $\mathbb{E}e^{-s\theta_1} = (1 + \gamma s)/(1 + s)$ for $s \geq 0$. Show that $\theta_1 + \gamma\theta_2 + \ldots + \gamma^{k-1}\theta_k$ converges in distribution as $k \to \infty$ and identify the limit distribution.

Problem 102. Let $\gamma > 0$ and η be a nonnegative random variable with Laplace transform ℓ. Prove that

$$\mathbb{E}\eta^{-\gamma} = \frac{\gamma}{\Gamma(1 + \gamma)} \int_0^{\infty} s^{\gamma - 1} \ell(s) \mathrm{d}s,$$

where $\Gamma(\cdot)$ is the Euler gamma function. The case that both sides of the equality are infinite is not excluded.

Problem 103. Let η be a random variable with Laplace transform $\ell(s) = e^{-s^{\alpha}}$ for some $\alpha \in (0, 1)$. Prove that, for any $\gamma > 0$,

$$\mathbb{E}\eta^{-\gamma} = \frac{\Gamma(1 + \gamma/\alpha)}{\Gamma(1 + \gamma)} < \infty.$$

Problem 104. Let ξ_1, $\xi_2, \ldots$ be independent random variables having a gamma distribution with parameters 2 and $\lambda > 0$, that is, $\mathbb{P}\{\xi_1 \leq x\} = 1 - (1 + \lambda x)e^{-\lambda x}$ for $x \geq 0$. Set $S_0 := 0$ and $S_n := \xi_1 + \ldots + \xi_n$ for $n \in \mathbb{N}$. Use Laplace transforms to check that, for $x \geq 0$,

$$\sum_{n \geq 0} \mathbb{P}\{S_n \leq x\} = 1 + \frac{\lambda x}{2} - \frac{1 - e^{-2\lambda x}}{4}.$$

Problem 105. Let ξ_1, $\xi_2, \ldots$ be independent identically distributed nonnegative random variables with mean $\mu \in (0, \infty)$ which are independent of a random variable ξ_0 with distribution function

$$\mathbb{P}\{\xi_0 \leq x\} = \frac{1}{\mu} \int_0^x \mathbb{P}\{\xi > y\} \mathrm{d}y, \quad x > 0.$$

Set $S_n := \xi_0 + \xi_1 + \ldots + \xi_n$ for $n \in \mathbb{N}_0$. Use Laplace transforms to check that, for $x \geq 0$,

$$\sum_{n \geq 0} \mathbb{P}\{S_n \leq x\} = \frac{x}{\mu}.$$

Problem 106. Let ξ be a nonnegative random variable with Laplace transform ℓ and density h. Assuming that $\lim_{x \to 0+} h(x) = \infty$, prove that $\lim_{s \to +\infty} s\ell(s) = \infty$.

Chapter 3

Characteristic functions

3.1. Definition and examples

Let ξ be a random variable with distribution function $F(x) = \mathbb{P}\{\xi \leq x\}$ for $x \in \mathbb{R}$. The distribution function is characterized by the following properties:
(a) $F(-\infty) = 0$, $F(+\infty) = 1$;
(b) F is right-continuous;
(c) F is nondecreasing.

Throughout the chapter i denotes the imaginary unit, that is, $\mathrm{i} = \sqrt{-1}$.

Definition 107. The function $\varphi : \mathbb{R} \to \mathbb{C}$ defined by

$$\varphi(t) := \mathbb{E}e^{\mathrm{i}t\xi} = \int_{\mathbb{R}} e^{\mathrm{i}tx} \mathrm{d}F(x), \quad t \in \mathbb{R} \tag{3.1}$$

is called *characteristic function* of ξ or *Fourier-Stieltjes transform* of its distribution.

Euler's formula tells us that

$$e^{\mathrm{i}x} = \cos x + \mathrm{i} \sin x, \quad x \in \mathbb{R} \tag{3.2}$$

which particularly implies that $|e^{\mathrm{i}x}| = \sqrt{\cos^2 x + \sin^2 x} = 1$. From this we infer

$$|\mathbb{E}e^{\mathrm{i}t\xi}| \leq \mathbb{E}|e^{\mathrm{i}t\xi}| = 1, \quad t \in \mathbb{R} \tag{3.3}$$

which shows that the characteristic function is defined for any random variable. Further, since, for each $t \in \mathbb{R}$, the function $x \mapsto e^{\mathrm{i}tx}$ is continuous and bounded, the integral in (3.1) exists as a Riemann-Stieltjes integral.

Any complex number $a + \mathrm{i}b$ can be identified with the point (a, b) on the plane $\mathbb{R}^2$. Similarly, with the help of (3.2) we can represent the characteristic function φ as follows:

$$\varphi(t) = \int_{\mathbb{R}} \cos(tx) \mathrm{d}F(x) + \mathrm{i} \int_{\mathbb{R}} \sin(tx) \mathrm{d}F(x), \quad t \in \mathbb{R}.$$

Thus, rather than analyzing the complex-valued function φ one can investigate the two integrals above which are real-valued functions.

If the distribution of ξ is discrete, that is, ξ takes at most countable number of values $\ldots, x_{-1}, x_0, x_1, \ldots$, then the integral defining φ becomes the sum

$$\varphi(t) = \sum_{k\in\mathbb{Z}} e^{\mathrm{i}tx_k}\mathbb{P}\{\xi = x_k\}, \quad t \in \mathbb{R}.$$

Here are several examples of characteristic functions of discrete random variables.
1. Let $\mathbb{P}\{\xi = a\} = 1$ for some $a \in \mathbb{R}$. Then

$$\varphi(t) = e^{\mathrm{i}at}, \quad t \in \mathbb{R}.$$

2. Let $\mathbb{P}\{\xi = -1\} = \mathbb{P}\{\xi = 1\} = 1/2$. Then

$$\varphi(t) = \cos t, \quad t \in \mathbb{R}.$$

Proof. Indeed, using Euler's formula (3.2) we obtain

$$\mathbb{E}e^{\mathrm{i}t\xi} = \frac{e^{\mathrm{i}t} + e^{-\mathrm{i}t}}{2} = \frac{\cos t + \mathrm{i}\sin t + \cos t - \mathrm{i}\sin t}{2} = \cos t.$$

□

3. If ξ has a Poisson distribution with parameter $\lambda > 0$, that is,

$$\mathbb{P}\{\xi = k\} = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0,$$

then

$$\varphi(t) = e^{\lambda(e^{\mathrm{i}t}-1)}, \quad t \in \mathbb{R}.$$

Proof.

$$\mathbb{E}e^{\mathrm{i}t\xi} = e^{-\lambda}\sum_{k\geq 0}\frac{(e^{\mathrm{i}t}\lambda)^k}{k!} = e^{-\lambda}e^{\lambda e^{\mathrm{i}t}} = e^{\lambda(e^{\mathrm{i}t}-1)}.$$

□

4. If ξ has a binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$, that is,

$$\mathbb{P}\{\xi = k\} = \binom{n}{k}p^k(1-p)^{n-k}, \quad k = 0, 1, \ldots, n,$$

then

$$\varphi(t) = (pe^{\mathrm{i}t} + 1 - p)^n, \quad t \in \mathbb{R}.$$

Proof. The binomial theorem for complex numbers tells us that, for $a, b \in \mathbb{C}$ and $n \in \mathbb{N}$,

$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}.$$

Hence, $\mathbb{E}e^{\mathrm{i}t\xi} = \sum_{k=0}^{n}\binom{n}{k}(pe^{\mathrm{i}t})^k(1-p)^{n-k} = (pe^{\mathrm{i}t} + 1 - p)^n$ for $t \in \mathbb{R}$. □

Before giving a more exotic example we need a definition.

Definition 108. Let $a \in (0,1]$. The function $z \mapsto \zeta(z,a)$ defined by

$$\zeta(z,a) = \sum_{k\geq 0} \frac{1}{(k+a)^z}, \quad z \in \mathbb{C}, \operatorname{Re} z > 1$$

is called *Hurwitz zeta function.* The function $z \mapsto \zeta(z,1)$ is called *Riemann zeta function.*

5. Let ξ have a Hurwitz zeta distribution with parameters $a \in (0,1]$ and $\sigma > 1$, that is,

$$\mathbb{P}\{\xi = -\log(k+a)\} = \frac{1}{\zeta(\sigma,a)(k+a)^\sigma}, \quad k \in \mathbb{N}_0.$$

Then

$$\varphi(t) = \frac{\zeta(\sigma + \mathrm{i}t, a)}{\zeta(\sigma,a)}, \quad t \in \mathbb{R}.$$

Proof.

$$\mathbb{E}e^{\mathrm{i}t\xi} = \sum_{k\geq 0} \frac{e^{-\mathrm{i}t\log(k+a)}}{\zeta(\sigma,a)(k+a)^\sigma} = \frac{1}{\zeta(\sigma,a)} \sum_{k\geq 0} \frac{1}{(k+a)^{\sigma+\mathrm{i}t}} = \frac{\zeta(\sigma+\mathrm{i}t,a)}{\zeta(\sigma,a)}.$$

□

If the distribution of ξ has density h, so that $\mathrm{d}F(x) = h(x)\mathrm{d}x$, the integral defining φ simplifies to

$$\varphi(t) = \int_{\mathbb{R}} e^{\mathrm{i}tx} h(x)\mathrm{d}x, \quad t \in \mathbb{R}.$$

We proceed by giving several examples of the corresponding characteristic functions.

6. If ξ has a uniform distribution on (a,b) for $-\infty < a < b < \infty$, that is, $h(x) = (b-a)^{-1}\mathbb{1}_{(a,b)}(x)$, then

$$\varphi(t) = \frac{e^{\mathrm{i}bt} - e^{\mathrm{i}at}}{\mathrm{i}(b-a)t}, \quad t \in \mathbb{R}.$$

In particular, if $a = -b$, $b > 0$, then

$$\varphi(t) = \frac{\sin(bt)}{bt}, \quad t \in \mathbb{R}. \tag{3.4}$$

Proof. We have

$$\mathbb{E}e^{\mathrm{i}t\xi} = \frac{1}{b-a}\int_a^b e^{\mathrm{i}tx}\mathrm{d}x = \frac{e^{\mathrm{i}bt} - e^{\mathrm{i}at}}{\mathrm{i}(b-a)t}.$$

If $a = -b$, $b > 0$, then

$$\mathbb{E}e^{\mathrm{i}t\xi} = \frac{e^{\mathrm{i}bt} - e^{-\mathrm{i}bt}}{2\mathrm{i}bt} = \frac{\cos(bt) + \mathrm{i}\sin(bt) - \cos(-bt) - \mathrm{i}\sin(-bt)}{2\mathrm{i}bt} = \frac{\sin(bt)}{bt}$$

having utilized the fact that $x \mapsto \cos x$ is an even function, that is, $\cos(-x) = \cos x$, whereas $x \mapsto \sin x$ is an odd function, that is, $\sin(-x) = -\sin x$. □

7. If ξ has a standard normal distribution (with mean 0 and variance 1), that is,

$$h(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R},$$

then

$$\varphi(t) = e^{-t^2/2}, \quad t \in \mathbb{R}.$$

Proof. Write

$$\varphi(t)=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{\mathrm{i}tx}e^{-x^2/2}\mathrm{d}x=\frac{1}{\sqrt{2\pi}}\Big(\int_{\mathbb{R}}\cos(tx)e^{-x^2/2}\mathrm{d}x+\mathrm{i}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\sin(tx)e^{-x^2/2}\mathrm{d}x\Big)$$
$$=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}\cos(tx)e^{-x^2/2}\mathrm{d}x.$$

We have used Euler's formula for the second equality. The third equality follows from the fact that, for each $t\neq 0$, the function $x\mapsto \sin(tx)e^{-x^2/2}$ is integrable on $\mathbb{R}$ (the inequality $|\sin(tx)e^{-x^2/2}|\leq e^{-x^2/2}$ shows it is actually absolutely integrable) and odd (as the product of the odd function $x\mapsto\sin(tx)$ and the even function $x\mapsto e^{-x^2/2}$). The integral of any such a function over any interval symmetric with respect to zero, particularly, over $\mathbb{R}$ is equal to zero.

Let us now show that we can differentiate in t under the integral to obtain

$$\sqrt{2\pi}\varphi'(t)=\Big(\int_{\mathbb{R}}\cos(tx)e^{-x^2/2}\mathrm{d}x\Big)'=-\int_{\mathbb{R}}\sin(tx)xe^{-x^2/2}\mathrm{d}x. \tag{3.5}$$

We shall need a standard equality

$$\cos u-\cos v=-2\sin((u-v)/2)\sin((u+v)/2),\quad u,v\in\mathbb{R}.$$

Using it yields, for each $t\in\mathbb{R}$,

$$\lim_{h\to 0}\frac{\cos((t+h)x)-\cos(tx)}{h}=-x\sin(tx). \tag{3.6}$$

Indeed, $\cos((t+h)x)-\cos(tx)=-2\sin(hx/2)\sin((t+h/2)x)$. It is well-known that $\lim_{h\to 0}(\sin h/h)=1$, whence $\lim_{h\to\,0}(2\sin(hx/2)/h)=x$. Since the function $y\mapsto\sin y$ is continuous, we infer $\lim_{h\to 0}\sin((t+h/2)x)=\sin(tx)$. Thus, (3.6) has been proved. Since

$$\Big(\int_{\mathbb{R}}\cos(tx)e^{-x^2/2}\mathrm{d}x\Big)'=\lim_{h\to 0}\int_{\mathbb{R}}\frac{\cos((t+h)x)-\cos(tx)}{h}e^{-x^2/2}\mathrm{d}x,$$

and (3.6) holds, we just have to show that

$$\lim_{h\to 0}\int_{\mathbb{R}}\frac{\cos((t+h)x)-\cos(tx)}{h}e^{-x^2/2}\mathrm{d}x=\int_{\mathbb{R}}\lim_{h\to 0}\frac{\cos((t+h)x)-\cos(tx)}{h}e^{-x^2/2}\mathrm{d}x.$$

According to Lebesgue's dominated convergence theorem the latter holds true provided we can find a function $q(x)$ which does not depend on h such that

$$\Big|\frac{\cos((t+h)x)-\cos(tx)}{h}\Big|=\frac{2|\sin(hx/2)||\sin((t+h/2))x|}{|h|}\leq q(x)\quad\text{for all } x\in\mathbb{R} \tag{3.7}$$

and $\int_{\mathbb{R}}q(x)e^{-x^2/2}\mathrm{d}x<\infty$. Since $|\sin y|\leq\min(|y|\wedge 1)$ for $y\in\mathbb{R}$ we can take $q(x):=|x|$ for $x\in\mathbb{R}$. This choice is relevant, for

$$\int_{\mathbb{R}}q(x)e^{-x^2/2}\mathrm{d}x=\int_{\mathbb{R}}|x|e^{-x^2/2}\mathrm{d}x=2\int_0^\infty xe^{-x^2/2}\mathrm{d}x=2\int_0^\infty e^{-y}\mathrm{d}y=2.$$

The second equality follows from the observation: the integral of any integrable even function over $\mathbb{R}$ is equal to 2 times the integral over $(0,\infty)$. The third equality is obtained by the change of variable $y=x^2/2$, so that $\mathrm{d}y=x\mathrm{d}x$. Thus, equality (3.5) holds.

Integrating by parts yields

$$-\int_{\mathbb{R}} \sin(tx)xe^{-x^2/2}\mathrm{d}x = \int_{\mathbb{R}} \sin(tx)\mathrm{d}(e^{-x^2/2}) = \sin(tx)e^{-x^2/2}\Big|_{-\infty}^{\infty} - t\int_{\mathbb{R}} \cos(tx)e^{-x^2/2}\mathrm{d}x$$
$$= -t\int_{\mathbb{R}} \cos(tx)e^{-x^2/2}\mathrm{d}x = -\sqrt{2\pi}t\varphi(t).$$

Thus, $\varphi'(t) = -t\varphi(t)$ for $t \in \mathbb{R}$. Solving this differential equation we obtain $\varphi(t) = Ce^{-t^2/2}$ for some constant $C \in \mathbb{R}$. Since $\varphi(0) = 1$ we infer $C = 1$, whence the claim. □

3.2. Simple properties of characteristic functions

Here, we collect some standard simple properties of characteristic functions which can be found in most of textbooks. More sophisticated properties will be discussed in the next sections.

For $x = a+\mathrm{i}b$ ($a, b \in \mathbb{R}$) we shall write $\overline{x}$ for the complex conjugate of x, that is, $\overline{x} = a-\mathrm{i}b$.

Lemma 109. *Let φ be the characteristic function of a random variable ξ. Then*
(a) $\varphi(0) = 1$*;*
(b) $|\varphi(t)| \le 1$ *for all* $t \in \mathbb{R}$*;*
(c) $\varphi(-t) = \overline{\varphi(t)}$ *for all* $t \in \mathbb{R}$ *(this is called Hermitian property);*
(d) *the characteristic function ψ of the random variable $\eta = a\xi + b$, where $a, b \in \mathbb{R}$, is given by $\psi(t) = e^{\mathrm{i}tb}\varphi(at)$ for $t \in \mathbb{R}$;*
(e) *φ is uniformly continuous on* $\mathbb{R}$.

Proof. While (a) is obvious, part (b) has been proved in (3.3).
To prove (c), note that

$$\varphi(t) = \mathbb{E}e^{\mathrm{i}t\xi} = \mathbb{E}(\cos(t\xi) + \mathrm{i}\sin(t\xi)) = \mathbb{E}\cos(t\xi) + \mathrm{i}\mathbb{E}\sin(t\xi)$$

which implies that

$$\overline{\varphi(t)} = \mathbb{E}\cos(t\xi) - \mathrm{i}\mathbb{E}\sin(t\xi).$$

On the other hand,

$$\varphi(-t) = \mathbb{E}e^{-\mathrm{i}t\xi} = \mathbb{E}(\cos(-t\xi) + \mathrm{i}\sin(-t\xi)) = \mathbb{E}\cos(t\xi) - \mathrm{i}\mathbb{E}\sin(t\xi)) = \overline{\varphi(t)}.$$

As for (d), write

$$\psi(t) = \mathbb{E}e^{\mathrm{i}t(a\xi+b)} = e^{\mathrm{i}tb}\mathbb{E}e^{\mathrm{i}ta\xi} = e^{\mathrm{i}tb}\varphi(at).$$

Passing to (e) we first recall that uniform continuity of φ means that $\lim_{h\to 0}|\varphi(t+h)-\varphi(t)| = 0$ uniformly in $t \in \mathbb{R}$. To prove this, write

$$|\varphi(t+h) - \varphi(t)| = |\mathbb{E}(e^{\mathrm{i}(t+h)\xi} - e^{\mathrm{i}t\xi})| \le \mathbb{E}|e^{\mathrm{i}t\xi}||e^{\mathrm{i}h\xi} - 1| = \mathbb{E}|e^{\mathrm{i}h\xi} - 1|.$$

Observe that

$$e^{\mathrm{i}h\xi} - 1 = \cos(h\xi) - 1 + \mathrm{i}\sin(h\xi) = -2\sin^2(h\xi/2) + 2\mathrm{i}\sin(h\xi/2)\cos(h\xi/2)$$
$$= -2\sin(h\xi/2)(\sin(h\xi/2) - \mathrm{i}\cos(h\xi/2)).$$

Here, we have used the half-angle formulae

$$1-\cos(2x)=2\sin^2 x,\quad \sin(2x)=2\sin x\cos x,\quad x\in\mathbb{R}. \tag{3.8}$$

Hence, $|e^{\mathrm{i}h\xi}-1|=2|\sin(h\xi/2)|$ because

$$|\sin(h\xi/2)-\mathrm{i}\cos(h\xi/2)|=\sqrt{\sin^2(h\xi/2)+\cos^2(h\xi/2)}=1.$$

For any $A>0$, write

$$\mathbb{E}|\sin(h\xi/2)|=\mathbb{E}|\sin(h\xi/2)|\mathbb{1}_{\{|\xi|\le A\}}+\mathbb{E}|\sin(h\xi/2)|\mathbb{1}_{\{|\xi|>A\}}\le |h|A/2+\mathbb{P}\{|\xi|>A\}.$$

The first term on the right-hand side is obtained with the help of inequality $|\sin y|\le|y|$ for $y\in\mathbb{R}$ as follows:

$$|\sin(h\xi/2)|\mathbb{1}_{\{|\xi|\le A\}}\le(|h\xi|/2)\mathbb{1}_{\{|\xi|\le A\}}\le(A|h|/2)\mathbb{1}_{\{|\xi|\le A\}}\le A|h|/2.$$

For the second term we have used $|\sin y|\le 1$ for $y\in\mathbb{R}$ and $\mathbb{E}\mathbb{1}_B=\mathbb{P}(B)$ for any event B. Thus, for any $A>0$, $\limsup_{h\to0}\mathbb{E}|\sin(h\xi/2)|\le\mathbb{P}\{|\xi|>A\}$ and sending $A\to\infty$ we infer $\lim_{h\to0}\mathbb{E}|\sin(h\xi/2)|=0$ and thereupon $\lim_{h\to0}|\varphi(t+h)-\varphi(t)|=0$. The latter convergence is uniform in $t\in\mathbb{R}$ because the upper bound $\mathbb{E}|e^{\mathrm{i}h\xi}-1|$ does not depend on t. □

3.3. Inversion formula and uniqueness theorem

Theorem 110 given next provides *inversion formulae* which enable us to recover the distribution function provided that the corresponding characteristic function is known. It should be stressed that the formulae are of mainly theoretical interest because finding an explicit form of distribution functions with the help of these formulae is rarely possible. On the other hand, Theorem 110 is an important ingredient in proving the uniqueness theorem for characteristic functions to be discussed later in this section.

Theorem 110. *Let φ be a characteristic function of a random variable ξ with distribution function F.*
(a) If $x_1<x_2$, $x_1,x_2\in\mathbb{R}$, then

$$\frac{1}{2}\mathbb{P}\{\xi=x_1\}+\frac{1}{2}\mathbb{P}\{\xi=x_2\}+\mathbb{P}\{x_1<\xi<x_2\}=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-\mathrm{i}tx_1}-e^{-\mathrm{i}tx_2}}{\mathrm{i}t}\varphi(t)\mathrm{d}t. \tag{3.9}$$

(b) Let a and $a+h$ ($h>0$) be continuity points of F, that is, $\mathbb{P}\{\xi=a\}=\mathbb{P}\{\xi=a+h\}=0$. Then

$$F(a+h)-F(a)=\mathbb{P}\{a<\xi<a+h\}=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{1-e^{-\mathrm{i}th}}{\mathrm{i}t}e^{-\mathrm{i}ta}\varphi(t)\mathrm{d}t. \tag{3.10}$$

Proof. (a) We divide the proof into several steps.
Step 1. Here, we prove that

$$\lim_{t\to\infty}\int_0^t\frac{\sin(\alpha x)}{x}\mathrm{d}x=\frac{\pi}{2}\operatorname{sgn}\alpha, \tag{3.11}$$

where $\operatorname{sgn}\alpha = 1$ if $\alpha > 0$, $= 0$ if $\alpha = 0$ and -1 if $\alpha < 0$.

If $\alpha > 0$, then changing the variable yields

$$\int_0^t \frac{\sin(\alpha x)}{x}\mathrm{d}x = \int_0^{\alpha t} \frac{\sin x}{x}\mathrm{d}x.$$

If $\alpha < 0$, then since $\sin(\alpha x) = -\sin(-\alpha x)$ we obtain

$$\int_0^t \frac{\sin(\alpha x)}{x}\mathrm{d}x = -\int_0^{-\alpha t} \frac{\sin x}{x}\mathrm{d}x.$$

Thus, in any case it suffices to prove that

$$\lim_{t\to\infty}\int_0^t \frac{\sin x}{x}\mathrm{d}x = \frac{\pi}{2}. \tag{3.12}$$

Since $1/x = \int_0^\infty e^{-yx}\mathrm{d}y$ we conclude that, for all $t > 0$,

$$\int_0^t\int_0^\infty |\sin x e^{-yx}|\mathrm{d}y\mathrm{d}x = \int_0^t \frac{|\sin x|}{x}\mathrm{d}x \le t$$

having utilized the inequality $|\sin x| \le x$ for $x \ge 0$ for the last equality. Thus, we may apply Fubini's theorem to obtain

$$\int_0^t \frac{\sin x}{x}\mathrm{d}x = \int_0^t \sin x\int_0^\infty e^{-yx}\mathrm{d}y\mathrm{d}x = \int_0^\infty\int_0^t \sin x e^{-yx}\mathrm{d}x\mathrm{d}y.$$

In order to evaluate the inner integral we integrate by parts twice:

$$\begin{aligned} I_t(y) &:= \int_0^t \sin x e^{-yx}\mathrm{d}x = \int_0^t e^{-yx}\mathrm{d}(-\cos x) = (-(\cos x)e^{-yx})\Big|_0^t - \int_0^t(\cos x)ye^{-yx}\mathrm{d}x \\ &= 1-(\cos t)e^{-yt} - \int_0^t ye^{-yx}\mathrm{d}(\sin x) = 1-(\cos t)e^{-yt} - ((\sin x)ye^{-yx})\Big|_0^t - \int_0^t(\sin x)y^2e^{-yx}\mathrm{d}x \\ &= 1-(\cos t)e^{-yt} - (\sin t)ye^{-yt} - y^2 I_t(y), \end{aligned}$$

so that

$$I_t(y) = \frac{1-e^{-yt}(\cos t + y\sin t)}{1+y^2}, \quad y > 0.$$

Inserting this into the double integral and recalling that

$$\int_0^\infty \frac{1}{1+y^2}\mathrm{d}y = \arctan y\Big|_0^\infty = \frac{\pi}{2}$$

yields

$$\int_0^t \frac{\sin x}{x}\mathrm{d}x = \int_0^\infty I_t(y)\mathrm{d}y = \frac{\pi}{2} - \int_0^\infty \frac{e^{-yt}(\cos t + y\sin t)}{1+y^2}\mathrm{d}y.$$

Observe that, for each $y > 0$,

$$\lim_{t\to\infty}\frac{e^{-yt}(\cos t + y\sin t)}{1+y^2} = 0$$

and, for all $t \ge 1$,

$$\Big|\frac{e^{-yt}(\cos t + y\sin t)}{1+y^2}\Big| \le \frac{1+e^{-1}}{1+y^2}$$

(here, we use the estimate $ye^{-yt} \le (et)^{-1} \le e^{-1}$ for $t \ge 1$). Since the right-hand side is integrable, an application of Lebesgue's dominated convergence theorem leads to the conclusion

$$\lim_{t\to\infty} \int_0^\infty \frac{1}{1+y^2} e^{-yt}(\cos t + y \sin t)\mathrm{d}y = 0.$$

Thus, (3.12) has been proved.

STEP 2. The purpose of this step is to show that, for any $y > 0$ and $\alpha \in \mathbb{R}$

$$\Big|\int_0^y \frac{\sin(\alpha t)}{t}\mathrm{d}t\Big| \le \int_0^\pi \frac{\sin t}{t}\mathrm{d}t. \tag{3.13}$$

Arguing as at the beginning of Step 1 we conclude it is enough to check that, for any $y > 0$,

$$0 < \int_0^y \frac{\sin t}{t}\mathrm{d}t \le \int_0^\pi \frac{\sin t}{t}\mathrm{d}t. \tag{3.14}$$

For any $y > 0$ there exists $k \in \mathbb{N}_0$ such that $y \in (\pi k, \pi(k+1)]$. We shall use a decomposition

$$\int_0^y \frac{\sin t}{t}\mathrm{d}t = \int_0^{\pi k} \frac{\sin t}{t}\mathrm{d}t + \int_{\pi k}^y \frac{\sin t}{t}\mathrm{d}t.$$

If k is odd, then the second integral is negative and

$$\int_0^y \frac{\sin t}{t}\mathrm{d}t \le \int_0^{\pi k} \frac{\sin t}{t}\mathrm{d}t.$$

If k is even, then $0 < \int_{\pi k}^y (\sin t/t)\mathrm{d}t \le \int_{\pi k}^{\pi(k+1)} (\sin t/t)\mathrm{d}t$, and

$$\int_0^y \frac{\sin t}{t}\mathrm{d}t \le \int_0^{\pi(k+1)} \frac{\sin t}{t}\mathrm{d}t.$$

Write, for $j \in \mathbb{N}_0$,

$$\int_0^{(2j+1)\pi} \frac{\sin t}{t}\mathrm{d}t = \sum_{i=0}^{2j} \int_{\pi i}^{\pi(i+1)} \frac{\sin t}{t}\mathrm{d}t = \sum_{i=0}^{2j} \int_0^\pi \frac{\sin(t+\pi i)}{t+\pi i}\mathrm{d}t = \sum_{i=0}^{2j} (-1)^i \int_0^\pi \frac{\sin t}{t+\pi i}\mathrm{d}t$$
$$= a_0 - (a_1 - a_2) - \ldots - (a_{2j-1} - a_{2j}) \le a_0.$$

Here, we have set $a_i := \int_0^\pi \frac{\sin t}{t+\pi i}\mathrm{d}t$ for $i \in \mathbb{N}_0$. All a_i are positive and the sequence $(a_i)_{i\in\mathbb{N}_0}$ is decreasing which implies $a_i - a_{i+1} > 0$. The positivity of $\int_0^y (\sin t/t)\mathrm{d}t$ can be proved along similar lines.

STEP 3. We are ready to prove (3.9). Write

$$J(T) := \frac{1}{2\pi} \int_{-T}^T \frac{e^{-\mathrm{i}tx_1} - e^{-\mathrm{i}tx_2}}{\mathrm{i}t}\varphi(t)\mathrm{d}t = \frac{1}{2\pi} \int_{-T}^T \frac{e^{-\mathrm{i}tx_1} - e^{-\mathrm{i}tx_2}}{\mathrm{i}t} \int_\mathbb{R} e^{\mathrm{i}tx}\mathrm{d}F(x)\mathrm{d}t$$
$$= \frac{1}{\pi} \int_\mathbb{R} \mathrm{d}F(x) \int_{-T}^T \frac{e^{\mathrm{i}t(x-x_1)} - e^{\mathrm{i}t(x-x_2)}}{2\mathrm{i}t}\mathrm{d}t.$$

The possibility of interchanging the order of integration is justified by Fubini's theorem and the following inequalities:

$$\Big|\frac{e^{\mathrm{i}t(x-x_1)} - e^{\mathrm{i}t(x-x_2)}}{\mathrm{i}t}\Big| = \Big|\frac{e^{-\mathrm{i}tx_1} - e^{-\mathrm{i}tx_2)}}{\mathrm{i}t}\Big| = \Big|\int_{x_1}^{x_2} e^{-\mathrm{i}ty}\mathrm{d}y\Big| \le \int_{x_1}^{x_2} |e^{\mathrm{i}ty}|\mathrm{d}y = x_2 - x_1$$

which entail

$$\Big|\int_{-T}^{T}\frac{e^{-\mathrm{i}tx_1}-e^{-\mathrm{i}tx_2}}{\mathrm{i}t}\varphi(t)\mathrm{d}t\Big|\leq\int_{-T}^{T}\Big|\frac{e^{-\mathrm{i}tx_1}-e^{-\mathrm{i}tx_2}}{\mathrm{i}t}\Big|\mathrm{d}t\leq 2T(x_2-x_1).$$

Here, we have used the fact that $|\varphi(t)|\leq 1$ for $t\in\mathbb{R}$, see part (b) of Lemma 109.

We continue as follows:

$$\begin{aligned}
&J(T)\\
=\;&\frac{1}{\pi}\int_{\mathbb{R}}\mathrm{d}F(x)\int_{-T}^{T}\frac{\cos(t(x-x_1))+\mathrm{i}\sin(t(x-x_1))-\cos(t(x-x_2))-\mathrm{i}\sin(t(x-x_2))}{2\mathrm{i}t}\mathrm{d}t\\
=\;&\frac{1}{\pi}\int_{\mathbb{R}}\mathrm{d}F(x)\int_{0}^{T}\frac{\sin(t(x-x_1))-\sin(t(x-x_2))}{t}\mathrm{d}t.
\end{aligned}$$

The last equality follows from the fact that, for $i=1,2$, $t\mapsto\cos(t(x-x_i))/t$ is an odd function integrable on $[-T,T]$ and $t\mapsto\sin(t(x-x_i))/t$ is an even function integrable on $[-T,T]$. Thus, the integral over $[-T,T]$ of the first function is equal to 0, whereas the integral over $[-T,T]$ of the second function is equal 2 times the integral over $[0,T]$. According to (3.11),

$$\lim_{T\to\infty}\int_0^T\frac{\sin(t(x-x_1))-\sin(t(x-x_2))}{t}\mathrm{d}t=\begin{cases}-\pi/2+\pi/2=0, & \text{if } x<x_1,\\ \pi/2, & \text{if } x=x_1,\\ \pi/2-(-\pi/2)=\pi, & \text{if } x\in(x_1,x_2),\\ \pi/2, & \text{if } x=x_2,\\ \pi/2-\pi/2=0, & \text{if } x>x_2.\end{cases}$$

Further, in view of (3.13),

$$\Big|\int_0^T\frac{\sin(t(x-x_1))-\sin(t(x-x_2))}{t}\mathrm{d}t\Big|\leq 2\int_0^{\pi}\frac{\sin t}{t}\mathrm{d}t.$$

Hence, by Lebesgue's dominated convergence theorem

$$\begin{aligned}
\lim_{T\to\infty}J(T)&=\frac{1}{\pi}\int_{\mathbb{R}}\mathrm{d}F(x)\lim_{T\to\infty}\int_0^T\frac{\sin(t(x-x_1))-\sin(t(x-x_2))}{t}\mathrm{d}t\\
&=\frac{1}{2}\int_{\{x_1\}}\mathrm{d}F(x)+\int_{(x_1,x_2)}\mathrm{d}F(x)+\frac{1}{2}\int_{\{x_2\}}\mathrm{d}F(x)\\
&=\frac{1}{2}\mathbb{P}\{\xi=x_1\}+\mathbb{P}\{x_1<\xi<x_2\}+\frac{1}{2}\mathbb{P}\{\xi=x_2\}.
\end{aligned}$$

(b) Formula (3.10) is a particular case of (3.9). Just put $x_1=a$, $x_2=a+h$ and notice that $\mathbb{P}\{\xi=a\}=\mathbb{P}\{\xi=a+h\}=0$ because a and $a+h$ are continuity points of F by assumption. □

Under a minor additional assumption the inversion formula can be simplified.

Proposition 111. *Let φ be a characteristic function of a random variable ξ with distribution function F. If $\mathbb{E}\log(1+|\xi|)<\infty$, then*

$$F(x)-\frac{1}{2}\mathbb{P}\{\xi=x\}=\frac{1}{2}-\frac{1}{\pi}\int_0^{\infty}\frac{\mathrm{Im}\,(e^{-\mathrm{i}tx}\varphi(t))}{t}\mathrm{d}t,\quad x\in\mathbb{R}.\tag{3.15}$$

Remark 112. The condition $\mathbb{E}\log(1+|\xi|) < \infty$ ensures that the integral in (3.15) is absolutely convergent.

Formula (3.15) which was obtained in [4] is called *Gil-Palaez formula*. We refrain from giving a proof, referring instead to the proof of Theorem 2 in [12].

Characteristic functions were of little use in probability theory if these would not define the corresponding distributions uniquely. Theorem 113 given next is called the uniqueness theorem for characteristic functions.

Theorem 113. *Let F_1 and F_2 be distribution functions with the corresponding characteristic functions φ_1 and φ_2. Then $F_1 \equiv F_2$ if, and only if, $\varphi_1(t) = \varphi_2(t)$ for each t from some dense set in* $(0, \infty)$.

Proof. The one way implication is trivial, for $F_1 \equiv F_2$ implies

$$\varphi_1(t) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \mathrm{d}F_1(x) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \mathrm{d}F_2(x) = \varphi_2(t) \quad \text{for all } t \in \mathbb{R}.$$

Assume now that $\varphi_1(t) = \varphi_2(t)$ for all t from some dense set in $(0, \infty)$. According to part (c) of Lemma 109 the latter equality holds for all $t \in A$, where A is a dense set in $\mathbb{R}$. Thus, given $t \in \mathbb{R}$ there exists a sequence $(t_n) \in A$ such that $\lim_{n\to\infty} t_n = t$. Write

$$|\varphi_1(t) - \varphi_2(t)| \leq |\varphi_1(t) - \varphi_1(t_n)| + |\varphi_1(t_n) - \varphi_2(t_n)| + |\varphi_2(t_n) - \varphi_2(t)|.$$

The first and the third term converges to 0 as $n \to \infty$ because φ_1 and φ_2 are continuous, see Lemma 109 (e). The second term is equal to 0 by assumption. Thus, we have proved that $\varphi_1(t) = \varphi_2(t)$ for all $t \in \mathbb{R}$.

To complete the proof we shall use formula (3.10). Any distribution function F is nondecreasing. Hence, the number of its discontinuity points is at most countable. Let a be a continuity point of F. There exists a sequence of positive numbers $(h_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} h_n = \infty$ such that $a + h_1$, $a + h_2, \ldots$ are continuity points of F. According to (3.10),

$$1 - F(a) = \lim_{n\to\infty} (F(a + h_n) - F(a))$$

is uniquely determined by the corresponding characteristic function. Hence, F is uniquely determined by the characteristic function at all continuity points of F, hence, at all points because F is right-continuous. Indeed, let b be a discontinuity point of F. Let (a_n) be a sequence of continuity points of F approaching b from above. Then, by right-continuity, $\lim_{n\to\infty} F(a_n) = F(b)$, that is, F is uniquely determined by its values at continuity points. □

Remark 114. Let F_1 and F_2 be distribution functions with the corresponding characteristic functions φ_1 and φ_2. The equality $\varphi_1(t) = \varphi_2(t)$ for all t in a *finite interval* does not necessarily implies that $F_1 \equiv F_2$.

To give the corresponding example we shall use the result of Problem 189. There it is shown that $\varphi(t) = (1 - |t|)\mathbb{1}_{(-1,1)}(t)$ is the characteristic function of a random variable having a density, and that the function φ_1 which is a periodic continuation with period 2 of $t \mapsto 1-|t|$, $t \in [-1,1]$ is also the characteristic function of a discrete random variable. Thus, even though $\varphi(t) = \varphi_1(t)$ for $t \in [-1,1]$, the corresponding distributions are different. More examples of this flavor can be found in Remarks 185 and 186.

3.4. Lévy's continuity theorem for characteristic functions

Recall that one says that distribution functions F_n weakly converge as $n \to \infty$ to a distribution function F, if $\lim_{n\to\infty} F_n(x) = F(x)$ at every continuity point x of F.

Theorem 115 given next is called *Lévy continuity theorem for characteristic functions.*

Theorem 115. *(a) If distribution functions F_n weakly converge as $n \to \infty$ to a distribution function F, then*

$$\lim_{n\to\infty} \int_{\mathbb{R}} e^{itx} dF_n(x) = \int_{\mathbb{R}} e^{itx} dF(x)$$

locally uniformly in t (that is, uniformly on every finite interval).
(b) For $n \in \mathbb{N}$, let F_n be a distribution function with the characteristic function φ_n. Assume that there exists a function f continuous at 0 and such that $\lim_{n\to\infty} \varphi_n(t) = f(t)$ (pointwise). Then the distribution functions F_n weakly converge as $n \to \infty$ to a distribution function F, and f is the characteristic function of F.

The proof of Theorem 115 can be found on p. 113 in [2].

Part (b) of Theorem 115 together with part (e) of Lemma 109 imply that if the pointwise limit of the sequence of characteristic functions is continuous at 0, then it is continuous on $\mathbb{R}$. The example given next demonstrates that the requirement of continuity at 0 in Theorem 115 cannot be dispensed with.

Example 116. For $n \in \mathbb{N}$, let F_n be the distribution function of a uniform distribution on $(-n, n)$, that is,

$$F_n(x) = \begin{cases} 0, & \text{if } x < -n, \\ \frac{n+x}{2n}, & \text{if } -n \le x < n, \\ 1, & \text{if } x \ge n. \end{cases}$$

For each $x \in \mathbb{R}$, we have $\lim_{n\to\infty} F_n(x) = 1/2$, so that the limit function is not a distribution function. On the other hand, according to formula (3.4),

$$\varphi_n(t) = \int_{\mathbb{R}} e^{itx} dF_n(x) = \frac{\sin(nt)}{nt}, \quad n \in \mathbb{N}, t \in \mathbb{R}.$$

Thus,

$$\lim_{n\to\infty} \varphi_n(t) = \begin{cases} 1, & \text{if } t = 0, \\ 0, & \text{if } t \ne 0, \end{cases}$$

and we see that the limit function is not continuous at 0.

3.5. Characteristic functions and arithmetic operations

Let ξ_1 and ξ_2 be independent random variables with distribution functions $F(x)$ and $G(x)$ for $x \in \mathbb{R}$. The distribution of the sum $\xi_1 + \xi_2$ is given by

$$(F \star G)(x) := \mathbb{P}\{\xi_1 + \xi_2 \leq x\} = \int_{\mathbb{R}} F(x-y)\mathrm{d}G(y) = \int_{\mathbb{R}} G(x-y)\mathrm{d}F(y), \quad x \in \mathbb{R}. \quad (3.16)$$

Similarly to the definition given in Section 3.5. the function $F \star G$ is also called *convolution* of the distribution functions F and G. The proof of (3.16) is analogous to that of (2.11), hence omitted.

Proposition 117. *Let ξ_1 and ξ_2 be independent random variables with characteristic functions φ_1 and φ_2. Then the characteristic function of $\xi_1 + \xi_2$ is $\varphi_1\varphi_2$.*

Proof. If ξ_1 and ξ_2 are independent random variables, then so are $e^{\mathrm{i}t\xi_1}$ and $e^{\mathrm{i}t\xi_2}$. Hence, $\mathbb{E}e^{\mathrm{i}t(\xi_1+\xi_2)} = \mathbb{E}e^{\mathrm{i}t\xi_1}\mathbb{E}e^{t\xi_2}$. Here, we have used the fact that the expectation of the product of independent random variables is equal to the product of their expectations. □

Example 118. If η is a random variable with *triangular* density h given by

$$h(x) = (1 - |x|)\mathbb{1}_{(-1,1)}(x),$$

then its characteristic function is

$$\varphi(t) = \frac{4\sin^2(t/2)}{t^2} = \frac{2(1-\cos t)}{t^2}, \quad t \in \mathbb{R}.$$

Proof. Let ξ_1 and ξ_2 be independent random variables with a uniform distribution on $(-1/2, 1/2)$. In particular, their density h_1 is given by $h_1(x) = \mathbb{1}_{(-1/2,\,1/2)}(x)$. We intend to show that η has the same distribution as $\xi_1 + \xi_2$ (actually we shall show that their densities are the same). Then

$$\mathbb{E}e^{\mathrm{i}t\eta} = \mathbb{E}e^{\mathrm{i}t(\xi_1+\xi_2)} = (\mathbb{E}e^{\mathrm{i}t\xi_1})^2 = \frac{4\sin^2(t/2)}{t^2} = \frac{2(1-\cos t)}{t^2},$$

where for the second equality we have used Proposition 117, the penultimate equality is a consequence of (3.4) with $b = 1/2$, and the last equality is ensured by the half-angle formula (3.8).

The density g of $\xi_1 + \xi_2$ is given by

$$g(x) = \int_{\mathbb{R}} h_1(x-y)h_1(y)\mathrm{d}y = \int_{-1/2}^{1/2} h_1(x-y)\mathrm{d}y, \quad x \in \mathbb{R}.$$

The integrand is not zero if, and only if, $x - y \in (-1/2, 1/2)$ or equivalently $x - 1/2 < y < x + 1/2$. Consider the two cases.
Case $x \geq 0$. The intervals $(-1/2, 1/2)$ and $(x - 1/2, x + 1/2)$ intersect if, and only if, $x \in [0, 1)$ in which case the intersection is $(x - 1/2, 1/2)$. So

$$g(x) = \int_{-1/2}^{1/2} h_1(x-y)\mathrm{d}y = \int_{x-1/2}^{1/2} \mathrm{d}y = 1 - x, \quad x \in [0, 1).$$

Case $x < 0$. The intervals $(-1/2, 1/2)$ and $(x - 1/2, x + 1/2)$ intersect if, and only if, $x \in (-1, 0)$ in which case the intersection is $(-1/2, x + 1/2)$. So

$$g(x) = \int_{-1/2}^{1/2} h_1(x-y)\mathrm{d}y = \int_{-1/2}^{x+1/2} \mathrm{d}y = 1 + x, \quad x \in (-1, 0).$$

Thus, $g(x) = h(x)$ for $x \in (-1, 1)$, and it is clear that $g(x) = 0$ for other x because $\xi_1 + \xi_2$ takes values in $(-1, 1)$. □

Here is a useful consequence of part (d) of Lemma 109 and Proposition 117.

Corollary 119. Let φ be the characteristic function of a random variable ξ. Then $t \mapsto \varphi(-t)$ is the characteristic function of a random variable $(-\xi)$, and $t \mapsto |\varphi(t)|^2$ is the characteristic function of a random variable $\xi_1 - \xi_2$, where ξ_1 and ξ_2 are independent copies of ξ.

Proof. The first claim follows from part (d) of Lemma 109 with $a = -1$ and $b = 0$. For the second claim, write

$$\mathbb{E}e^{\mathrm{i}t(\xi_1 - \xi_2)} = \varphi(t)\varphi(-t) = \varphi(t)\overline{\varphi(t)} = |\varphi(t)|^2.$$

The first equality follows from Proposition 117 and the first claim of the present lemma. The second equality is a consequence of the Hermitian property (see part (c) of Lemma 109). Finally, the last equality is implied by $|x|^2 = x\overline{x}$ which holds for any $x \in \mathbb{C}$. □

Remark 120. $|\varphi|$ is not necessarily a characteristic function.

The corresponding example will be given in Example 132.

Lemma 121. *Let $\varphi_1, \ldots, \varphi_n$ be characteristic functions of random variables and $\alpha_1, \ldots, \alpha_n$ be nonnegative numbers satisfying $\alpha_1 + \ldots + \alpha_n = 1$. Then $\alpha_1\varphi_1 + \ldots + \alpha_n\varphi_n$ is a characteristic function of a random variable.*

Proof. Denote by $F_1, \ldots, F_n$ distribution functions of random variables with the characteristic functions $\varphi_1, \ldots, \varphi_n$. Then $\alpha_1 F_1 + \ldots + \alpha_n F_n$ is the distribution function called *finite mixture* of the distribution functions $F_1, \ldots, F_n$ or *convex linear combination* of these. Here, the three properties characterizing distribution functions can be checked easily:
(a) $\alpha_1 F_1(-\infty) + \ldots + \alpha_n F_n(-\infty) = 0$, $\alpha_1 F_1(\infty) + \ldots + \alpha_n F_n(\infty) = 1$;
(b) $\alpha_1 F_1 + \ldots + \alpha_n F_n$ is right-continuous;
(c) $\alpha_1 F_1 + \ldots + \alpha_n F_n$ is nondecreasing.

The condition $\alpha_1 + \ldots + \alpha_n = 1$ ensures the second equality in part (a). For instance, part (b) is justified by

$$\begin{aligned}\lim_{h \to 0+} (\alpha_1 F_1(x+h) + \ldots + \alpha_n F_n(x+h)) &= \alpha_1 \lim_{h \to 0+} F_1(x+h) + \ldots + \alpha_n \lim_{h \to 0+} F_n(x+h) \\ &= \alpha_1 F_1(x) + \ldots + \alpha_n F_n(x)\end{aligned}$$

for each $x \in \mathbb{R}$, where the last equality is implied by right-continuity of $F_1, \dots, F_n$. It is obvious that

$$\sum_{j=1}^{n} \alpha_j \varphi_j(t) = \sum_{j=1}^{n} \alpha_j \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \mathrm{d}F_j(x) = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}tx} \mathrm{d}\Big(\sum_{j=1}^{n} \alpha_j F_j(x)\Big),$$

that is, $\alpha_1\varphi_1 + \ldots + \alpha_n\varphi_n$ is the characteristic function of a random variable with the distribution function $\alpha_1 F_1 + \ldots + \alpha_n F_n$. In other words, the characteristic function of a finite mixture of distribution functions is the finite mixture of characteristic functions. □

Example 122. If ξ is a random variable having a bilateral exponential distribution with parameter $\beta > 0$, that is, its density h is given by

$$h(x) = \frac{\beta}{2} e^{-\beta|x|}, \quad x \in \mathbb{R},$$

then its characteristic function is

$$\varphi(t) = \frac{\beta^2}{\beta^2 + t^2}, \quad t \in \mathbb{R}.$$

Proof. Set

$$h_1(x) = \beta e^{-\beta x} \mathbb{1}_{(0,\infty)}(x) \quad \text{and} \quad h_2(x) = \beta e^{\beta x} \mathbb{1}_{(-\infty,0)}(x)$$

and note that h_1 is the density of a random variable η having an exponential distribution with parameter β, and h_2 is the density of $(-\eta)$. Arguing as in the case of Laplace transforms (see Section 2.3.) we obtain

$$\varphi_1(t) := \mathbb{E} e^{\mathrm{i}t\eta} = \int_{[0,\infty)} e^{\mathrm{i}tx} h_1(x) \mathrm{d}x = \frac{\beta}{\beta - \mathrm{i}t}, \quad t \in \mathbb{R}$$

and

$$\varphi_2(t) := \mathbb{E} e^{-\mathrm{i}t\eta} = \int_{(-\infty,0]} e^{\mathrm{i}tx} h_2(x) \mathrm{d}x = \frac{\beta}{\beta + \mathrm{i}t}, \quad t \in \mathbb{R}.$$

Since

$$h(x) = \frac{1}{2} h_1(x) + \frac{1}{2} h_2(x), \quad x \in \mathbb{R}$$

we conclude that the distribution of ξ is the finite mixture of distributions. Thus, according to Lemma 121 and its proof,

$$\varphi(t) = \frac{1}{2}\varphi_1(t) + \frac{1}{2}\varphi_2(t) = \frac{\beta^2}{\beta^2 + t^2}.$$

□

Recall that the *real part* and the *imaginary part* of a complex number x is denoted by $\operatorname{Re} x$ and $\operatorname{Im} x$, respectively. Thus, if $x = a + \mathrm{i}b$ for $a, b \in \mathbb{R}$, then $\operatorname{Re} x = a$ and $\operatorname{Im} x = b$. Note that

$$\operatorname{Re} x = (x + \overline{x})/2 \quad \text{and} \quad \operatorname{Im} x = (x - \overline{x})/(2i). \tag{3.17}$$

Corollary 123. Let φ be the characteristic function of a random variable ξ. Then $\operatorname{Re}\varphi$ is also the characteristic function of a random variable. Furthermore, the following equalities hold

$$\operatorname{Re}\varphi(t)=\frac{1}{2}(\varphi(t)+\varphi(-t))=\int_{\mathbb{R}}\cos(tx)\mathrm{d}\mathbb{P}\{\xi\leq x\}. \tag{3.18}$$

Proof. The first equality in (3.18) follows by a combination of the first equality in (3.17) in which we set $x=\varphi(t)$ and the Hermitian property (part (c) of Lemma 109) which states that $\overline{\varphi(t)}=\varphi(-t)$. The second equality is immediate from the representation

$$\varphi(t)=\int_{\mathbb{R}}\cos(tx)\mathrm{d}\mathbb{P}\{\xi\leq x\}+\mathrm{i}\int_{\mathbb{R}}\sin(tx)\mathrm{d}\mathbb{P}\{\xi\leq x\}.$$

According to Corollary 119, $t\mapsto\varphi(-t)$ is the characteristic function of $(-\xi)$. Hence, the first equality in (3.18) tells us that $\operatorname{Re}\varphi$ is the finite mixture of characteristic functions. By Lemma 121, $\operatorname{Re}\varphi$ is a characteristic function of a random variable. □

Remark 124. If φ is the characteristic function of a random variable, then $\operatorname{Im}\varphi$ is never the characteristic function of a random variable.

For the proof just observe that $\operatorname{Im}\varphi(0)=0$. If $\operatorname{Im}\varphi$ were a characteristic function, this value should have been equal to 1.

3.6. Characteristic functions and moments

3.6.1. Introductory comments

Definition 125. The absolute moment of order γ of a random variable ξ is defined by

$$n_\gamma:=\mathbb{E}|\xi|^\gamma=\int_{\mathbb{R}}|x|^\gamma\mathrm{d}F(x),$$

where F is the distribution function of ξ.

Definition 126. The moment of order $\gamma>0$ of a random variable ξ is defined by

$$m_\gamma:=\mathbb{E}\xi^\gamma=\int_{\mathbb{R}}x^\gamma\mathrm{d}F(x).$$

If a random variable ξ is nonnegative, then $n_\gamma=m_\gamma$, possibly infinite. If ξ takes values of both signs, then only moments of positive integer orders are well-defined, and $n_\gamma=m_\gamma$, possibly infinite, for even γ.

Let us show that if $n_\gamma<\infty$, then $n_\theta<\infty$ for all $\theta\in(0,\gamma)$. Indeed, $|x|^\gamma>|x|^\theta$ if $|x|>1$. Hence,

$$n_\theta=\int_{\mathbb{R}}|x|^\theta\mathrm{d}F(x)=\int_{[-1,\,1]}|x|^\theta\mathrm{d}F(x)+\int_{\mathbb{R}\setminus[-1,\,1]}|x|^\theta\mathrm{d}F(x)\leq\int_{[-1,\,1]}\mathrm{d}F(x)$$
$$+\int_{\mathbb{R}\setminus[-1,\,1]}|x|^\gamma\mathrm{d}F(x):=I_1+I_2<\infty,$$

because $I_1 = F(1) - F((-1)-) < \infty$ and, by assumption, $I_2 \le n_\gamma < \infty$. Similarly, one can check that if $n_\gamma = \infty$, then $n_\theta = \infty$ for all $\theta > \gamma$.

Let ξ be a random variable taking values in the finite interval $[a, b]$, where $-\infty < a < b < \infty$. Then $|\xi| \le \max(|a|, |b|)$ with probability 1, whence $n_\gamma = \mathbb{E}|\xi|^\gamma \le \max(|a|^\gamma, |b|^\gamma) < \infty$ for all $\gamma > 0$. On the other hand, there exist distributions of random variables ξ taking values in infinite intervals such that, still, $n_\gamma < \infty$ for all $\gamma > 0$. For instance, let ξ be a random variable with an exponential distribution with parameter $\beta > 0$, that is, its density h is given by $h(x) = \beta e^{-\beta x} \mathbb{1}_{(0,\infty)}(x)$. Then, extending Example 82 we have, for all $\gamma > 0$, $n_\gamma = m_\gamma = \beta \int_0^\infty x^\gamma e^{-\beta x} \mathrm{d}x = \beta^{-\gamma} \Gamma(\gamma + 1) < \infty$. Here, Γ is the gamma function defined by $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} \mathrm{d}x$ for $\alpha > 0$.

For a random variable ξ, put $\xi^+ := \max(\xi, 0) = \xi \mathbb{1}_{\{\xi \ge 0\}}$ and $\xi^- := \max(-\xi, 0) = -\xi \mathbb{1}_{\{\xi < 0\}}$. Note that ξ^+ and ξ^- are nonnegative random variables. Let us prove that the moments of these random variables can be calculated as follows: for $\gamma > 0$,

$$\mathbb{E}(\xi^+)^\gamma = \gamma \int_0^\infty x^{\gamma-1} \mathbb{P}\{\xi > x\} \mathrm{d}x \quad \text{and} \quad \mathbb{E}(\xi^-)^\gamma = \gamma \int_0^\infty x^{\gamma-1} \mathbb{P}\{-\xi > x\} \mathrm{d}x. \tag{3.19}$$

Here, both sides of the equalities may be infinite. Indeed,

$$\gamma \int_0^\infty x^{\gamma-1} \mathbb{P}\{\xi > x\} \mathrm{d}x = \gamma \int_0^\infty x^{\gamma-1} \mathbb{E} \mathbb{1}_{\{\xi > x\}} \mathrm{d}x = \gamma \mathbb{E} \int_0^\infty x^{\gamma-1} \mathbb{1}_{\{\xi > x\}} \mathrm{d}x$$
$$= \mathbb{E} \int_0^\xi \gamma x^{\gamma-1} \mathrm{d}x \mathbb{1}_{\{\xi \ge 0\}} = \mathbb{E} \xi^\gamma \mathbb{1}_{\{\xi \ge 0\}} = \mathbb{E}(\xi^+)^\gamma.$$

We have used Fubini's theorem for the second equality. The formula for $\mathbb{E}(\xi^-)^\gamma$ can be proved along similar lines.

Now we give the example of distribution which have infinite moments of some orders.

Example 127. Assume that

$$\mathbb{P}\{\xi > x\} := \begin{cases} 1, & \text{if } x \le 1, \\ x^{-\beta}, & \text{if } x > 1. \end{cases}$$

Using (3.20) we obtain $m_\gamma = \gamma \int_1^\infty x^{\gamma-1} x^{-\beta} \mathrm{d}x = \frac{\gamma}{\beta - \gamma}$. Thus, $m_\gamma < \infty$ for $\gamma \in (0, \beta)$ and $m_\gamma = \infty$ for $\gamma = \beta$. Hence, also $m_\gamma = \infty$ for $\gamma \ge \beta$.

In the last example for each $\gamma < \beta$ there exists $\varepsilon_\gamma > 0$ such that $m_{\gamma+\varepsilon_\gamma} < \infty$. The example given next is intended to demonstrate that such a behavior is not always the case, that is, it is well-possible that $m_\gamma < \infty$ and $m_{\gamma+\varepsilon} = \infty$ for all $\varepsilon > 0$.

Example 128. Let

$$\mathbb{P}\{\xi > x\} := \begin{cases} 1, & \text{if } x \le e, \\ \frac{e}{x \log^2 x}, & \text{if } x > e. \end{cases}$$

Given $\varepsilon > 0$ there exists $A_\varepsilon > 0$ such that $x^\varepsilon > \log^2 x$ for $x > A_\varepsilon$. Therefore, we obtain with the help of (3.20)

$$m_{1+\varepsilon} = (1 + \varepsilon) \int_e^\infty x^\varepsilon \frac{e}{x \log^2 x} \mathrm{d}x > (1 + \varepsilon) e \int_{A_\varepsilon}^\infty x^{-1} \mathrm{d}x = \infty.$$

On the other hand, $m_1 = \int_e^\infty \frac{e}{x\log^2 x}\mathrm{d}x = e\int_1^\infty y^{-2}\mathrm{d}y = e$.

Finally, we give the example of distribution with all moments of positive orders being infinite.

Example 129. Set

$$h(x) = \frac{1}{x\log^2 x}\mathbb{1}_{(e,\infty)}(x)$$

and note that $h(x) \geq 0$ and according to the previous example $\int_{\mathbb{R}} h(x)\mathrm{d}x = 1$. Therefore, h is a density of some distribution and, for all $\varepsilon > 0$, $m_\varepsilon = \int_0^\infty x^\varepsilon h(x)\mathrm{d}x = \int_e^\infty \frac{x^\varepsilon}{x\log^2 x}\mathrm{d}x = \infty$.

In the examples above the moments were either finite or infinite. However, there exist distributions of random variables taking values in the whole $\mathbb{R}$, some moments of which do not exist (are not defined) rather than finite or infinite. For instance, if $\mathbb{E}\xi^+ = \mathbb{E}\xi^- = \infty$, then $\mathbb{E}\xi$ does not exist. The explicit example of such a situation can be found in Example 136.

3.6.2. Moments of positive integer orders

Theorem 130. *Let φ be a characteristic function of a random variable ξ with distribution function F.*
(a) Let $k \in \mathbb{N}$. The finiteness of $n_k = \mathbb{E}|\xi|^k$ ensures that $\varphi^{(k)}$ the derivative of the kth order of characteristic function φ exists and is continuous. Furthermore, for $t \in \mathbb{R}$,

$$\varphi^{(p)}(t) = \mathrm{i}^p \int_{\mathbb{R}} x^p e^{\mathrm{i}tx}\mathrm{d}F(x), \quad p \in \mathbb{N}_0, p \leq k. \tag{3.20}$$

(b) Assume that $\varphi^{(k)}(0)$ the value at 0 of the kth order derivative of φ exists and is finite. Then $n_k < \infty$ if k is even, and $n_{k-1} < \infty$ if k is odd.

Proof. (a) The condition $n_k < \infty$ entails $n_p < \infty$ for all $p \in (0, k]$. We shall use the mathematical induction. Assume that for some integer $p \in (0, k)$ the derivative $\varphi^{(p)}$ exists and is finite, and that (3.20) holds. We intend to show that $n_{p+1} < \infty$ ensures the existence of $\varphi^{(p+1)}$ which satisfies (3.20) with $p + 1$ replacing p. Using (3.20) yields

$$\frac{\varphi^{(p)}(t+h) - \varphi^{(p)}(t)}{h} = \mathrm{i}^p \int_{\mathbb{R}} x^p e^{\mathrm{i}tx} \frac{e^{\mathrm{i}hx} - 1}{h}\mathrm{d}F(x).$$

Recall that

$$|e^{\mathrm{i}hx} - 1| = 2|\sin(hx/2)| \leq |hx|, \quad x \in \mathbb{R},$$

whence

$$\left|\frac{e^{\mathrm{i}hx} - 1}{h} x^p e^{\mathrm{i}tx}\right| \leq |x|^{p+1}.$$

Since $\int_{\mathbb{R}} |x|^{p+1}\mathrm{d}F(x) = n_{p+1} < \infty$ by assumption we infer

$$\mathrm{i}^p \lim_{h\to 0} \int_{\mathbb{R}} x^p e^{\mathrm{i}tx} x^p \frac{e^{\mathrm{i}hx} - 1}{h}\mathrm{d}F(x) = \mathrm{i}^p \int_{\mathbb{R}} x^p e^{\mathrm{i}tx} \lim_{h\to 0}\Big(\frac{e^{\mathrm{i}hx} - 1}{h}\Big)\mathrm{d}F(x) = \mathrm{i}^{p+1} \int_{\mathbb{R}} x^{p+1} e^{\mathrm{i}tx}\mathrm{d}F(x)$$

by Lebesque's dominated convergence theorem. This demonstrates that the derivative $\varphi^{(p+1)}$ exists and satisfies (3.20) with $p + 1$ replacing p.

Left with the proof of continuity of $\varphi^{(k)}$ we use (3.20) with $p = k$ to obtain, for $t, h \in \mathbb{R}$,

$$|\varphi^{(k)}(t+h) - \varphi^{(k)}(t)| = \Big|\mathrm{i}^k \int_{\mathbb{R}} x^k e^{\mathrm{i}tx}(e^{\mathrm{i}hx} - 1)\mathrm{d}F(x)\Big| \le 2\int_{\mathbb{R}} |x|^k |\sin(hx/2)| \mathrm{d}F(x).$$

Since $|x|^k|\sin(hx/2)| \le |x|^k$ and $n_k = \int_{\mathbb{R}} |x|^k \mathrm{d}F(x)$ is finite by assumption we conclude with the help of Lebesgue's dominated convergence theorem that

$$\lim_{h\to 0} \int_{\mathbb{R}} |x|^k |\sin(hx/2)| \mathrm{d}F(x) = \int_{\mathbb{R}} |x|^k (\lim_{h\to 0} |\sin(hx/2)|) \mathrm{d}F(x) = 0$$

because $\lim_{h\to 0} |\sin(hx/2)| = 0$ for any fixed $x \in \mathbb{R}$. Hence, $\lim_{h\to 0} |\varphi^{(k)}(t+h) - \varphi^{(k)}(t)| = 0$, and continuity of $\varphi^{(k)}$ follows.

(b) Denote by $2q_0$ the largest even number satisfying $2q_0 \le k$. It is sufficient to prove that $n_{2q} < \infty$ for each $q \in (0, q_0]$. Assume that $n_{2q} < \infty$ for some $q < q_0$. According to part (a) of the theorem the derivative $\varphi^{(2q)}$ exists and formula (3.20) holds with $p = 2q$. This entails

$$\begin{aligned} \frac{\varphi^{(2q)}(h) + \varphi^{(2q)}(-h) - 2\varphi^{(2q)}(0)}{h^2} &= (-1)^q \int_{\mathbb{R}} x^{2q} \frac{e^{\mathrm{i}hx} + e^{-\mathrm{i}hx} - 2}{h^2} \mathrm{d}F(x) \qquad (3.21) \\ &= 4(-1)^{q+1} \int_{\mathbb{R}} x^{2q} \frac{\sin^2(hx/2)}{h^2} \mathrm{d}F(x) \end{aligned}$$

for $h \in \mathbb{R}$, having utilized

$$e^{\mathrm{i}hx} + e^{-\mathrm{i}hx} - 2 = 2\cos(hx) - 2 = -4\sin^2(hx/2).$$

As $h \to 0$, the limit of the left hand-side in (3.21) is equal to $\varphi^{(2q+2)}(0)$ and thus finite by assumption. By Fatou's lemma,

$$\begin{aligned} \infty > (-1)^{q+1}\varphi^{(2q+2)}(0) &= 4 \lim_{h\to 0} \int_{\mathbb{R}} x^{2q} \frac{\sin^2(hx/2)}{h^2} \mathrm{d}F(x) \\ &\ge 4 \int_{\mathbb{R}} \liminf_{h\to 0} \Big(x^{2q} \frac{\sin^2(hx/2)}{h^2} \Big) \mathrm{d}F(x) = \int_{\mathbb{R}} x^{2q+2} \mathrm{d}F(x) = n_{2q+2}. \end{aligned}$$

Thus, we have proved that $n_{2q+2} < \infty$ which proves part (b) with the help of induction. The proof of the theorem is complete. □

There are several important corollaries to Theorem 130.

Corollary 131. Let $p \in \mathbb{N}$ be even. If the characteristic function φ is p-times differentiable at 0, then it is p-times continuously differentiable on $\mathbb{R}$.

This corollary allows us to give an example in which φ is a characteristic function, whereas $|\varphi|$ is not.

Example 132. Let ξ be a random variable with distribution $\mathbb{P}\{\xi = 1\} = \mathbb{P}\{\xi = -1\} = 1/2$. Then $\varphi(t) = \mathbb{E}e^{\mathrm{i}t\xi} = \cos t$. We claim that $\psi(t) = |\varphi(t)| = |\cos t|$ is not a characteristic function.

Assume on the contrary that $\psi(t)$ is a characteristic function. Since $\varphi''(0) = -1$ and $\varphi(t) = \psi(t)$ in a neighborhood of zero we conclude that $\psi''(0) = -1$. Therefore, according to Corollary 131 the function $\psi(t)$ has to be twice differentiable on $\mathbb{R}$. A contradiction is obtained by noting that $\psi(t)$ is not differentiable at points $t = \pi/2 + \pi k$ for $k \in \mathbb{Z}$.

Corollary 133. Let $p \in \mathbb{N}$ be even. The finiteness of $\varphi^{(p)}(0)$ is necessary and sufficient for the finiteness of $n_p = m_p$.

Let us note that there exist characteristic functions which are not differentiable at every point. One example is given by a Weierstrass function

$$\varphi(t) = \sum_{k \geq 0} 2^{-k-1} \cos(\pi a^k t), \quad t \in \mathbb{R}$$

where $a > 2 + 3\pi$ is an odd number. The proof of nondifferentiability can be found on pp. 352-353 in [16]. The corresponding distribution function F is given by

$$F(x) = \sum_{k \geq 0} 2^{-k-1} F_k(x), \quad x \in \mathbb{R},$$

where F_k is the distribution function of a random variable taking values $\pm \pi a^k$ with probability 1/2.

Corollary 134. Let $k \in \mathbb{N}$. If the moment m_k is finite, then $\varphi^{(n)}(0) = \mathrm{i}^n m_n$ for $n \in \mathbb{N}$, $n \leq k$, and $\operatorname{sign} \varphi^{(2p)}(0) = (-1)^p$ for $p \in \mathbb{N}$, $p \leq 2\lfloor k/2 \rfloor$.

Remark 135. The finiteness of $\varphi^{(k)}(0)$ for odd k does not in general entail $n_k < \infty$.

Here is a supporting example.

Example 136. Let us show that φ given by

$$\varphi(t) = \alpha \sum_{j \geq 2} \frac{\cos(jt)}{j^2 \log j}, \quad t \in \mathbb{R} \tag{3.22}$$

(for appropriate $\alpha > 0$) is a differentiable characteristic function, yet $\mathbb{E}\xi^+ = \mathbb{E}\xi^- = \infty$, whence $n_1 = \mathbb{E}|\xi| = \infty$. Here, ξ is the corresponding random variable, and $\xi^+ = \max(\xi, 0)$ and $\xi^- = \max(-\xi, 0)$.

Proof. For $j \in \mathbb{N}$, let ξ_j be a random variable with distribution $\mathbb{P}\{\xi_j = j\} = \mathbb{P}\{\xi_j = -j\} = 1/2$. Its characteristic function is

$$\varphi_j(t) = \mathbb{E}e^{\mathrm{i}t\xi_j} = \cos(jt), \quad t \in \mathbb{R}.$$

The series $\sum_{j \geq 2} \frac{1}{j^2 \log j}$ converges. Set $\alpha := 1/\sum_{j \geq 2} \frac{1}{j^2 \log j}$ and then $\alpha_j = \alpha/(j^2 \log j)$ for integer $j \geq 2$. According to Problem 192, the function φ defined by

$$\varphi(t) = \sum_{j \geq 2} \alpha_j \varphi_j(t) = \alpha \sum_{j \geq 2} \frac{\cos(jt)}{j^2 \log j}, \quad t \in \mathbb{R}$$

is a characteristic function. Furthermore, the corresponding distribution function F is given by

$$F(x) = \sum_{j \geq 2} \alpha_j F_j(x) = \alpha \sum_{j \geq 2} \frac{F_j(x)}{j^2 \log j}, \quad x \in \mathbb{R},$$

where F_j is the distribution function of ξ_j.

To prove that $n_1 = \infty$, we first observe that $\mathbb{P}\{\xi_j^\pm = j\} = \mathbb{P}\{\xi_j^\pm = 0\} = 1/2$, whence $\mathbb{E}\xi_j^\pm = j/2$ and then write

$$\mathbb{E}\xi^+ = \int_0^\infty (1-F(x))\mathrm{d}x = \int_0^\infty \sum_{j\geq 2} \alpha_j (1-F_j(x))\mathrm{d}x = \sum_{j\geq 2} \alpha_j \int_0^\infty (1-F_j(x))\mathrm{d}x = \sum_{j\geq 2} \alpha_j \mathbb{E}\xi_j^+$$
$$= (\alpha/2) \sum_{j\geq 2} \frac{1}{j\log j} = \infty.$$

We have used formula (3.20) for the first equality and Fubini's theorem for the third. Similarly

$$\mathbb{E}\xi^- = \int_{-\infty}^0 F(x)\mathrm{d}x = \int_{-\infty}^0 \sum_{j\geq 2} \alpha_j F_j(x)\mathrm{d}x = \sum_{j\geq 2} \alpha_j \int_{-\infty}^0 F_j(x)\mathrm{d}x = \sum_{j\geq 2} \alpha_j \mathbb{E}\xi_j^-$$
$$= (\alpha/2) \sum_{j\geq 2} \frac{1}{j\log j} = \infty.$$

Thus, $\mathbb{E}|\xi| = \mathbb{E}\xi^+ + +\mathbb{E}\xi^- = \infty$.

To proceed, we need two standard analytic facts. The first of these is called *Dirichlet's test for uniform convergence of functional series.*

Lemma 137. *The functional series $\sum_{k\geq 1} u_k(x) v_k(x)$ converges uniformly on some interval I if there exists a constant $A > 0$ such that, for all $n \in \mathbb{N}$ and all $x \in I$,*

$$\left|\sum_{k=1}^n u_k(x)\right| \leq A$$

and, for each $x \in I$, the sequence $(v_k(x))_{k\in\mathbb{N}}$ is nonincreasing and $\lim_{k\to\infty} v_k(x) = 0$ uniformly in $x \in I$.

Also, we need to know when the functional series can be differentiated termwise.

Lemma 138. *For $k \in \mathbb{N}$, let $u_k(x)$ be a continuously differentiable function. Assume that the series $\sum_{k\geq 1} u_k(x)$ converges, and the series $\sum_{k\geq 1} u_k'(x)$ converges uniformly. Then*

$$\left(\sum_{k\geq 1} u_k(x)\right)' = \sum_{k\geq 1} u_k'(x).$$

Now we are ready to show that φ given in (3.22) is a differentiable function. To this end, we are going to use Lemma 138. For each integer $j \geq 2$, the function $t \mapsto \alpha_j \cos(jt)$ is continuously differentiable. The series defining φ trivially convergence, for otherwise φ were not a characteristic function. Thus, it remains to check that the series

$$\sum_{j\geq 2} \frac{\sin(jt)}{j\log j}$$

obtained by formal differentiation of the series defining φ converges uniformly on $\mathbb{R}$. For $x \in \mathbb{R}$, set $u_j(x) = \sin(jx)/j$ and $v_j(x) = 1/\log j$. The proof of the fact that there exists $A > 0$ such that, for each $x \in \mathbb{R}$ and each integer $n \geq 2$,

$$\Big|\sum_{j=2}^n \frac{\sin(jx)}{j}\Big| \leq A$$

can be found on p. 6-7 in [16]. Further, v_j does not depend on x and monotonically decreases to 0 as $j \to \infty$. Thus, the series in the last centered formula does indeed converge uniformly. With this at hand we conclude that the function φ is differentiable. □

At this point we are ready to prove Theorem 2.15 from Chapter 2.

Proof of Theorem 2.15. By assumption given $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\frac{m_{2n}^{1/(2n)}}{2n} \le r + \varepsilon$$

or equivalently

$$m_{2n} \le (2n)^{2n}(r+\varepsilon)^{2n} \tag{3.23}$$

for some $r > 0$ whenever $n \ge n_0$.

Let ξ be a random variable with $m_n = \mathbb{E}\xi^n$ for $n \in \mathbb{N}$. As before, we use the notation $\nu_n = \mathbb{E}|\xi|^n$ for $n \in \mathbb{N}$. We intend to prove that

$$\nu_n \le 2n^n(r+\varepsilon)^n \quad \text{for all } n \ge n_0. \tag{3.24}$$

Since $\nu_{2k} = m_{2k}$ for $k \in \mathbb{N}$, the latter inequality is a consequence (3.23) for even indices n. Thus, we are left with checking (3.24) for odd n. To this end, we use Hölder's inequality (2.10) with $X = |\xi|^k$, $Y = |\xi|^{k+1}$ and $p = q = 2$ to obtain

$$\begin{aligned}\nu_{2k+1} = \mathbb{E}|\xi|^k|\xi|^{k+1} \le (\mathbb{E}|\xi|^{2k})^{1/2}(\mathbb{E}|\xi|^{2k+2})^{1/2} &= (m_{2k})^{1/2}(m_{2k+2})^{1/2} \\ &\le (2k)^{2k}(r+\varepsilon)^k(2k+2)^{k+1}(r+\varepsilon)^{k+1}.\end{aligned}$$

Here, the last inequality is implied by (3.23). Using the trivial inequalities

$$2k(2k+2) \le (2k+1)^2 \quad \text{and} \quad 2k+2 \le 2(2k+1), \quad k \in \mathbb{N}$$

we infer (3.24) for $n = 2k+1$. A combination of (3.24) and a crude estimate $n^n/n! \le \sum_{k\ge 0}(n^k/k!) = e^n$ yields

$$\frac{\nu_n}{n!} \le 2e^n(r+\varepsilon)^n \quad \text{for all } n \ge n_0. \tag{3.25}$$

To proceed, let us prove the following inequality

$$\Big|e^{\mathrm{i}x} - \sum_{m=0}^{n-1}\frac{(\mathrm{i}x)^m}{m!}\Big| \le \frac{|x|^n}{n!}, \quad x \in \mathbb{R}, \tag{3.26}$$

where $\mathrm{i} = \sqrt{-1}$ is the imaginary unit.

Integrating by parts we obtain, for $n \in \mathbb{N}_0$,

$$\int_0^x (x-y)^n e^{\mathrm{i}y}\mathrm{d}y = \frac{1}{n+1}\int_0^x e^{\mathrm{i}y}\mathrm{d}(-(x-y)^{n+1}) = \frac{x^{n+1}}{n+1} + \frac{\mathrm{i}}{n+1}\int_0^x (x-y)^{n+1}e^{\mathrm{i}y}\mathrm{d}y.$$

When $n = 0$, this says

$$\int_0^x e^{\mathrm{i}y}\mathrm{d}y = x + \mathrm{i}\int_0^x (x-y)e^{\mathrm{i}y}\mathrm{d}y.$$

The left-hand side is $(e^{\mathrm{i}x}-1)/i$, so rearranging gives

$$e^{\mathrm{i}x} = 1 + \mathrm{i}x + \mathrm{i}^2 \int_0^x (x-y)e^{\mathrm{i}y}\mathrm{d}y.$$

Using the result for $n=1$ now leads to

$$e^{\mathrm{i}x} = 1 + \mathrm{i}x + \frac{\mathrm{i}^2x^2}{2} + \frac{\mathrm{i}^3}{2}\int_0^x (x-y)^2 e^{\mathrm{i}y}\mathrm{d}y$$

and iterating we arrive at

$$e^{\mathrm{i}x} - \sum_{m=0}^{n-1} \frac{\mathrm{i}^m x^m}{m!} = \frac{\mathrm{i}^n}{(n-1)!}\int_0^x (x-y)^{n-1}e^{\mathrm{i}y}\mathrm{d}y.$$

It remains to estimate the term on the right-hand side. Since $|e^{\mathrm{i}y}|=1$ we infer

$$\frac{\mathrm{i}^n}{(n-1)!}\int_0^x (x-y)^{n-1}e^{\mathrm{i}y}\mathrm{d}y \le \frac{1}{(n-1)!}\int_0^x (x-y)^{n-1}\mathrm{d}y = \frac{|x|^n}{n!}.$$

Thus, (3.26) does indeed hold true.

Setting in (3.26) $x = t\xi$ for $t \in \mathbb{R}$ and multiplying it by $e^{\mathrm{i}s\xi}$ for $s \in \mathbb{R}$ we have

$$\left| e^{\mathrm{i}s\xi}\left(e^{\mathrm{i}t\xi} - \sum_{m=0}^{n-1} \frac{(\mathrm{i}t\xi)^m}{m!}\right)\right| \le \frac{|t\xi|^n}{n!}.$$

According to (3.20), for each $m \in \mathbb{N}$,

$$\varphi^{(m)}(s) = \mathrm{i}^m \int_{\mathbb{R}} x^m e^{\mathrm{i}sx}\mathrm{d}\mathbb{P}\{\xi \le x\} = \mathrm{i}^m \mathbb{E}\xi^m e^{\mathrm{i}s\xi}.$$

Taking now expected values and recalling that $|\mathbb{E}\cdot| \le \mathbb{E}|\cdot|$ gives

$$\left|\varphi(s+t) - \varphi(s) - t\varphi'(s) - \ldots - \frac{t^{n-1}}{(n-1)!}\varphi^{(n-1)}(s)\right| \le \frac{|t|^n}{n!}\nu_n$$

In view of (3.25) we conclude that, for any $s \in \mathbb{R}$,

$$\varphi(s+t) = \varphi(s) + \sum_{m\ge 1} \frac{t^m}{m!}\varphi^{(m)}(s), \quad |t| < \frac{1}{er}. \tag{3.27}$$

Assume there is another distribution with the same moments $(m_n)_{n\in\mathbb{N}}$ and let ψ be its characteristic function. Since $\varphi(0) = \psi(0) = 1$, it follows from (3.27) and induction that $\varphi(t) = \psi(t)$ for $|t| < k/er$ for all $k \in \mathbb{N}$, so the two characteristic functions coincide. By the uniqueness theorem for characteristic functions (Theorem 113) the corresponding distributions are the same. □

3.6.3. Some results on symmetric distributions

Definition 139. The distribution of a random variable ξ is called *symmetric* if the distributions of ξ and $-\xi$ are the same.

In terms of distribution functions this reads $F(x) = 1 - F((-x)-)$ for all $x \in \mathbb{R}$ because $\mathbb{P}\{\xi \le x\} = F(x)$ and $\mathbb{P}\{-\xi \le x\} = \mathbb{P}\{\xi \ge -x\} = 1 - F((-x)-)$. If the distribution of ξ is symmetric and has density h, then $h(x) = h(-x)$ for all $x \in \mathbb{R}$.

Here are some examples of symmetric distributions.

Example 140. (a) A standard normal distribution with density h given by

$$h(x) = (2\pi)^{-1/2} e^{-x^2/2}, \quad x \in \mathbb{R}$$

and characteristic function $\varphi(t) = e^{-t^2/2}$ for $t \in \mathbb{R}$.

(b) A uniform distribution on $(-a, a)$ with density h given by

$$h(x) = (2a)^{-1} \mathbb{1}_{(-a,a)}(x), \quad x \in \mathbb{R}$$

and characteristic function $\varphi(t) = \sin(at)/(at)$ for $t \in \mathbb{R}$.

(c) A bilateral exponential distribution with parameter $\beta > 0$ having density h given by

$$h(x) = (\beta/2) e^{-\beta|x|}, \quad x \in \mathbb{R}$$

and characteristic function $\varphi(t) = \beta^2/(t^2 + \beta^2)$ for $t \in \mathbb{R}$.

(d) A two point distribution $\mathbb{P}\{\xi = \pm 1\} = 1/2$ with characteristic function $\varphi(t) = \cos t$ for $t \in \mathbb{R}$.

We see that all the characteristic function above are real and even. Is it coincidence? The answer is NO, as the following result confirms.

Lemma 141. *Let ξ be a random variable with characteristic function φ. The distribution of ξ is symmetric if, and only if, φ is real and even, and given by*

$$\varphi(t) = \int_{\mathbb{R}} \cos(tx) \mathrm{d}\mathbb{P}\{\xi \le x\}, \quad t \in \mathbb{R}.$$

Proof. Assume that the distribution of ξ is symmetric, that is, ξ has the same distribution as $(-\xi)$. Then, according to Lemma 109(d) and (c), we must have $\varphi(t) = \varphi(-t) = \overline{\varphi(t)}$ which shows that φ is a real-valued function. Hence, $\varphi(t) = \mathbb{E}\cos(t\xi)$ for $t \in \mathbb{R}$. It is clear that $\varphi(-t) = \mathbb{E}\cos(-t\xi) = \mathbb{E}\cos(t\xi) = \varphi(t)$, that is, φ is even.

Conversely, assume that $\varphi(t) = \mathbb{E}\cos(t\xi)$ for $t \in \mathbb{R}$. Then $\varphi(t) = \overline{\varphi(t)} = \varphi(-t)$ for $t \in \mathbb{R}$, and we conclude that ξ has the same distribution as $(-\xi)$ by the uniqueness theorem for characteristic functions (Theorem 113). □

In what follows we shall use the following technical result. Its remarkable feature is that while investigating finiteness of moments one can assume that the underlying distribution is symmetric.

Proposition 142. *Let $\gamma > 0$ and ξ_1 and ξ_2 be independent copies of a random variable ξ. Then $\mathbb{E}|\xi|^\gamma < \infty$ if, and only if, $\mathbb{E}|\xi_1 - \xi_2|^\gamma < \infty$.*

For the proof we need a couple of auxiliary statements.

Lemma 143. *Let $g : [0,\infty) \to [0,\infty)$ be a concave function. Then g is subadditive, that is,*

$$g(x+y) \geq g(x) + g(y), \quad x, y \geq 0.$$

Proof. By definition, $g(\alpha x + \beta y) \geq \alpha g(x) + \beta g(y)$ for any $x, y \geq 0$ and any nonnegative α and β such that $\alpha + \beta = 1$. Putting $y = 0$ we obtain $g(\alpha x) \geq \alpha g(x)$. Set now $x = u + v$ and $\alpha = u/(u+v)$ for any $u, v \geq 0$, $u + v > 0$. Then $g(u) \geq u/(u+v)g(u+v)$. Set now $x = u+v$ and $\alpha = v/(u+v)$ for any $u, v \geq 0$, $u+v > 0$. Then Then $g(v) \geq v/(u+v)g(u+v)$. Summing up the last two inequalities yields $g(u+v) \leq g(u) + g(v)$ when $u, v \geq 0$ and $u + v > 0$. The last inequality trivially holds when $u = v = 0$ which completes the proof. □

Lemma 144. *Let $\gamma > 0$. Then, for $x, y \geq 0$*

$$(x+y)^\gamma \leq \max(2^{\gamma-1}, 1)(x^\gamma + y^\gamma).$$

Proof. When $\gamma \in (0, 1]$, the inequality reads

$$(x+y)^\gamma \leq x^\gamma + y^\gamma.$$

This follows from Lemma 143 applied to a concave function $x \mapsto x^\gamma$ on $[0, \infty)$.

Assume now that $\gamma > 1$. Then the function $x \mapsto x^\gamma$ is convex on $[0, \infty)$. In particular, for $x, y \geq 0$,

$$(x/2 + y/2)^\gamma \leq (1/2)x^\gamma + (1/2)y^\gamma.$$

Rearranging this gives

$$(x+y)^\gamma \leq 2^{\gamma-1}(x^\gamma + y^\gamma)$$

as was claimed. □

Lemma 145. *Let $\gamma > 0$ and ξ be a random variable. Then $\mathbb{E}|\xi|^\gamma < \infty$ if, and only if, $\mathbb{E}|\xi - a|^\gamma < \infty$ for any (some) $a \in \mathbb{R}$.*

Proof. Assume that $\mathbb{E}|\xi|^\gamma < \infty$. By the triangle inequality and Lemma 144, for any $a \in \mathbb{R}$,

$$|\xi - a|^\gamma \leq (|\xi| + |a|)^\gamma \leq \max(2^{\gamma-1}, 1)(|\xi|^\gamma + |a|^\gamma).$$

Passing to expectations we obtain

$$\mathbb{E}|\xi - a|^\gamma \leq \max(2^{\gamma-1}, 1)(\mathbb{E}|\xi|^\gamma + |a|^\gamma) < \infty.$$

Assume now that, for some $a \in \mathbb{R}$, $\mathbb{E}|\xi - a|^\gamma < \infty$. Using $|\xi| \leq |\xi - a| + |a|$ and then Lemma 144 we infer $|\xi|^\gamma \leq \max(2^{\gamma-1}, 1)(|\xi - a|^\gamma + |a|^\gamma)$ and thereupon

$$\mathbb{E}|\xi|^\gamma \leq \max(2^{\gamma-1}, 1)(\mathbb{E}|\xi - a|^\gamma + |a|^\gamma).$$

□

Proof of Proposition 142. Assume that $\mathbb{E}|\xi|^\gamma < \infty$. By the triangle inequality and Lemma 144,

$$|\xi_1 - \xi_2|^\gamma \le (|\xi_1| + |\xi_2|)^\gamma \le \max(2^{\gamma-1}, 1)(|\xi_1|^\gamma + |\xi_2|^\gamma).$$

Passing to expectations we obtain

$$\mathbb{E}|\xi_1 - \xi_2|^\gamma \le 2\max(2^{\gamma-1}, 1)\mathbb{E}|\xi|^\gamma < \infty.$$

Assume now that $\mathbb{E}|\xi_1 - \xi_2|^\gamma < \infty$. Let m be any median[1] of the distribution of ξ. Then, for any $x > 0$,

$$\mathbb{P}\{\xi_1 - \xi_2 > x\} \ge \mathbb{P}\{(\xi_1 - m) - (\xi_2 - m) > x, \xi_2 - m \le 0\} \ge \mathbb{P}\{\xi_1 - m > x, \xi_2 - m \le 0\}$$
$$\ge (1/2)\mathbb{P}\{\xi - m > x\}.$$

Arguing similarly we obtain, for $x > 0$,

$$\mathbb{P}\{\xi_1 - \xi_2 < -x\} \ge (1/2)\mathbb{P}\{\xi - m < -x\},$$

whence

$$\mathbb{P}\{|\xi_1 - \xi_2| > x\} \ge (1/2)\mathbb{P}\{|\xi - m| > x\}.$$

Using formula (3.20) we conclude that $\infty > \mathbb{E}|\xi_1 - \xi_2|^\gamma \ge (1/2)\mathbb{E}|\xi - m|^\gamma$. It remains to use Lemma 145 to obtain $\mathbb{E}|\xi|^\gamma < \infty$. □

3.6.4. Moments of arbitrary positive orders Our next task is to investigate the finiteness of n_γ, when $\gamma > 0$ is not necessarily integer.

Theorem 146. *Let ξ be a random variable with characteristic function φ and $(t_n)_{n\ge1}$ a sequence of real numbers satisfying $\lim_{n\to\infty} t_n = 0$. Denote by c' a positive constant whose value may change from line to line.*
(a) If $|t_n|^{-\lambda} \log|\varphi(t_n)| \le -c' < 0$ for some $\lambda < 2$, then $n_\gamma = \mathbb{E}|\xi|^\gamma = \infty$ for all $\gamma > \lambda$.
(b) If $|t_n|^{-\lambda} \log|\varphi(t_n)| \ge -c' > -\infty$ for a sequence (t_n) which additionally satisfies
1) $\sum_{n\ge1} |t_n|^\varepsilon < \infty$ for any $\varepsilon > 0$;
2) $(|t_{n-1}/t_n|)_{n\ge1}$ is a bounded sequence,
then $n_\gamma < \infty$ for all $\gamma \in (0, \lambda)$. In particular, if the function $|t|^{-\lambda}\log|\varphi(t)|$ is bounded at some punctured vicinity of 0, then $n_\gamma < \infty$ for all $\gamma \in (0, \lambda)$.
(c) If $t_n^{-2} \log|\varphi(t_n)| \ge -c' > -\infty$, then $n_2 < \infty$.
(d) If $\lim_{n\to\infty} t_n^{-2} \log|\varphi(t_n)| = 0$, then the distribution of ξ is degenerate.

Proof. (a) By assumption, for some $c > 0$, $|\varphi(t_n)|^2 \le \exp(-c|t_n|^\lambda)$. Let ξ_1 and ξ_2 be independent copies of ξ. Denote by G the distribution function of $\xi_1 - \xi_2$. Then

$$|\varphi(t)|^2 = \mathbb{E}e^{\mathrm{i}t(\xi_1-\xi_2)} = \int_{\mathbb{R}} \cos(tx)\mathrm{d}G(x), \quad t \in \mathbb{R}, \tag{3.28}$$

where the last equality is justified by Lemma 141.

[1] Recall that a real number m is called *median* of the distribution of ξ if $\mathbb{P}\{\xi \le m\} \ge 1/2$ and $\mathbb{P}\{\xi < m\} \le 1/2$.

Let $\lambda < \gamma \le 2$ and assume that $\mathbb{E}|\xi|^\gamma < \infty$. Put $u_n = t_n/2$ for $n \in \mathbb{N}$. According to Proposition 142, $\mathbb{E}|\xi_1 - \xi_2|^\gamma < \infty$ and thereupon

$$\begin{aligned}\infty > 2\int_{\mathbb{R}} |x|^\gamma \mathrm{d}G(x) &= 2\int_{\mathbb{R}} \limsup\nolimits_{n\to\infty} |\sin(u_n x)/u_n|^\gamma \mathrm{d}G(x)\\ &\ge 2\limsup\nolimits_{n\to\infty} \int_{\mathbb{R}} |\sin(u_n x)/u_n|^\gamma \mathrm{d}G(x) \ge \limsup\nolimits_{n\to\infty} |u_n|^{-\gamma} \int_{\mathbb{R}} 2\sin^2(u_n x)\mathrm{d}G(x)\\ &\ge \limsup\nolimits_{n\to\infty} 2^\gamma |t_n|^{-\gamma} \int_{\mathbb{R}} (1-\cos(t_n x))\mathrm{d}G(x) = \limsup\nolimits_{n\to\infty} 2^\gamma |t_n|^{-\gamma}(1-|\varphi(t_n)|^2)\\ &\ge 2^\gamma \lim_{n\to\infty} |t_n|^{-\gamma}(1-\exp(-c|t_n|^\lambda)) = \infty,\end{aligned}$$

a contradiction. Here, the second inequality is a consequence of Fatou's lemma, and the second equality follows from (3.28).
(b) Without loss of generality we can assume that $1 \ge t_{n-1} > t_n > 0$ for integer $n \ge 2$. Then, for some $c_1 > 0$ and sufficiently large n,

$$1 - |\varphi(t_n)|^2 \le 1 - \exp(-c_1 t_n^\lambda) \le c_1 t_n^\lambda.$$

Put again $u_n = t_n/2$ and let $c_2, c_3, \ldots$ denote some constants whose values are of no importance. Then

$$\int_{\mathbb{R}} \sin^2(u_n x)\mathrm{d}G(x) = 2^{-1}\int_{\mathbb{R}} (1-\cos(t_n x))\mathrm{d}G(x) = 2^{-1}(1-|\varphi(t_n)|^2) \le 2^{-1}c_1 t_n^\lambda = c_2 t_n^\lambda.$$

Using the inequality $\sin^2\theta \ge \theta^2 \sin^2 1$ for $0 \le \theta \le 1$ and putting $x_n = 1/u_n$ we obtain

$$\int_{x_{n-1}}^{x_n} x^2 \mathrm{d}G(x) \le c_3 \int_{x_{n-1}}^{x_n} u_n^{-2}\sin^2(u_n x)\mathrm{d}G(x) \le c_4 u_n^{\lambda-2},$$

having utilized the inequality $\sin^2 u_n x \ge u_n^2 x^2 \sin^2 1$ which holds because $x \in [u_{n-1}^{-1}, u_n^{-1}]$, hence $u_n x \in [u_n u_{n-1}^{-1}, 1]$. Further,

$$\int_{x_{n-1}}^{x_n} x^\gamma \mathrm{d}G(x) = \int_{x_{n-1}}^{x_n} (x^\gamma/x^2)x^2 \mathrm{d}G(x) \le (x_n^\gamma/x_{n-1}^2)c_4 u_n^{\lambda-2} = c_4(u_{n-1}/u_n)^2 u_n^{\lambda-\gamma} \le c_5 u_n^{\lambda-\gamma}$$

because $(u_n/u_{n-1})_{n\ge1}$ is a bounded sequence. By assumption, $\sum_{n\ge1}|u_n|^\varepsilon < \infty$ for any $\varepsilon > 0$. Therefore, the last inequality entails $\mathbb{E}|\xi_1-\xi_2|^\gamma = \int_{\mathbb{R}} |x|^\gamma \mathrm{d}G(x) < \infty$ for all $\gamma \in (0,\lambda)$. By Proposition 142, $\mathbb{E}|\xi|^\gamma < \infty$ for all $\gamma \in (0,\lambda)$. In particular, if the function $|t|^{-\lambda}\log|\varphi(t)|$ is bounded in some punctured vicinity of 0 we can take $t_n = \beta^n$ for any $\beta \in (0,1)$.
(c) Assume first that, for some $c > 0$, $|\varphi(t_n)|^2 \ge \exp(-ct_n^2)$. Then, by Fatou's lemma,

$$\begin{aligned}\int_{\mathbb{R}} x^2 \mathrm{d}G(x) &= 2\int_{\mathbb{R}} \liminf\nolimits_{n\to\infty}(1-\cos(t_n x))t_n^{-2}\mathrm{d}G(x)\\ &\le 2\liminf\nolimits_{n\to\infty}\int_{\mathbb{R}}(1-\cos(t_n x))t_n^{-2}\mathrm{d}G(x) \le 2\liminf\nolimits_{n\to\infty}(1-|\varphi(t_n)|^2)t_n^{-2}\\ &\le 2\liminf\nolimits_{n\to\infty}(1-\exp(-ct_n^2)t_n^{-2} = 2c.\end{aligned}$$

This shows that $\mathbb{E}(\xi_1-\xi_2)^2 < \infty$. Hence, $\mathbb{E}\xi^2 < \infty$ by Proposition 142.

(d) In view of $\lim_{n\to\infty} t_n^{-2} \log|\varphi(t_n)| = 0$, we have $t_n^{-2}\log|\varphi(t_n)| \geq -c' > -\infty$ for any $c' > 0$. From the proof of part (c) it follows that $\int_{\mathbb{R}} x^2 \mathrm{d}G(x) \leq c$ for any $c > 0$, whence $\mathbb{E}(\xi_1 - \xi_2)^2 = \int_{\mathbb{R}} x^2 \mathrm{d}G(x) = 0$. Thus, $\mathbb{P}\{\xi_1 - \xi_2 = 0\} = 1$ which implies that $\mathbb{P}\{\xi = a\} = 1$ for some $a \in \mathbb{R}$.

The proof of the theorem is complete. □

Theorem 146 is mainly interesting from theoretical point of view. As far as applications are concerned the following corollary is more useful.

Corollary 147. Let $\lambda \in (0, 2)$ and ξ be a random variable with characteristic function φ.
(a) If the function $|t|^{-\lambda}\log|\varphi(t)|$ is bounded away from 0, then $\mathbb{E}|\xi|^\gamma = \infty$ for all $\gamma > \lambda$.
(b) If the function $|t|^{-\lambda}\log|\varphi(t)|$ is bounded in some punctured neighborhood of 0, then $\mathbb{E}|\xi|^\gamma < \infty$ for all $\gamma \in (0, \lambda)$.

3.7. Characteristic functions and continuity properties

3.7.1. Discrete distributions

Definition 148. Recall from Section 3.1. that the distribution of a random variable ξ is called *discrete* if ξ takes at most countable number of values $\ldots, x_{-1}, x_0, x_1, \ldots$.

The distribution function of a random variable ξ with discrete distribution is a piecewise constant function with jumps of size $\mathbb{P}\{\xi = x_k\}$ at points x_k.

Let us show how to recover the discrete distribution from the corresponding characteristic function.

Theorem 149. *Let ξ be a random variable with characteristic function φ. Then, for $x \in \mathbb{R}$,*

$$\mathbb{P}\{\xi = x\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t.$$

Proof. We first show that

$$\int_{-T}^{T} e^{\mathrm{i}t(y-x)}\mathrm{d}t = \begin{cases} 2T, & \text{if } y = x, \\ \frac{2\sin(T(y-x))}{y-x}, & \text{if } y \neq x. \end{cases} \tag{3.29}$$

Indeed, if $y = x$, then $\int_{-T}^{T} e^{\mathrm{i}t(y-x)}\mathrm{d}t = \int_{-T}^{T}\mathrm{d}t = 2T$, whereas if $y \neq x$, then

$$\int_{-T}^{T} e^{\mathrm{i}t(y-x)}\mathrm{d}t = \int_{-T}^{T}\cos(t(y-x))\mathrm{d}t + \mathrm{i}\int_{-T}^{T}\sin(t(y-x))\mathrm{d}t = 2\int_0^T \cos(t(y-x))\mathrm{d}t$$
$$= \frac{2\sin(T(y-x))}{y-x}.$$

Here, the first equality is a consequence of Euler's formula, and the second follows from the facts that $t \mapsto \cos(t(y-x))$ is even and $t \mapsto \sin(t(y-x))$ is odd.

Further, write

$$\int_{-T}^{T} e^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t = \int_{-T}^{T} e^{-\mathrm{i}tx}\int_{\mathbb{R}} e^{\mathrm{i}ty}\mathrm{d}F(y)\mathrm{d}t = \int_{\mathbb{R}}\mathrm{d}F(y)\int_{-T}^{T} e^{\mathrm{i}t(y-x)}\mathrm{d}t = 2T\int_{\{x\}}\mathrm{d}F(y)$$
$$+\int_{\mathbb{R}\setminus\{x\}}\frac{2\sin(T(y-x))}{y-x}\mathrm{d}F(y) = 2T\mathbb{P}\{\xi = x\} + \int_{\mathbb{R}\setminus\{x\}}\frac{2\sin(T(y-x))}{y-x}\mathrm{d}F(y),$$

where F is the distribution function of ξ. While the second equality follows from

$$\Big|\int_{\mathbb{R}}\mathrm{d}F(y)\int_{-T}^{T} e^{\mathrm{i}t(y-x)}\mathrm{d}t\Big| \le 2T$$

and Fubini's theorem, the third is a consequence of (3.29). Thus, it remains to show that

$$\lim_{T\to\infty}\int_{\mathbb{R}\setminus\{x\}}\frac{\sin(T(y-x))}{T(y-x)}\mathrm{d}F(y) = 0.$$

To see this, first observe that

$$\lim_{T\to\infty}\frac{\sin(T(y-x))}{T(y-x)} = 0.$$

In view of this we have to justify the interchange of the limit and the integral. Since

$$\Big|\frac{\sin(T(y-x))}{T(y-x)}\Big| \le 1$$

and $\int_{\mathbb{R}\setminus\{x\}}\mathrm{d}F(y) = \mathbb{P}\{\xi \neq x\} \le 1$, the claim follows by Lebesgue's dominated convergence theorem. The proof of the theorem is complete. □

Our next task is to discuss the asymptotic behavior at ∞ of characteristic functions of discrete distributions.

Theorem 150. *Let ξ be a random variable with a discrete distribution and characteristic function φ. Then* $\limsup_{|t|\to\infty}|\varphi(t)| = 1$.

To prove this result we need some preparations.

Definition 151. A function $f:\mathbb{R}\to\mathbb{C}$ is called *almost periodic* if given $\varepsilon > 0$ there exists $L(\varepsilon) > 0$ such that each interval of length $L(\varepsilon)$ contains a point τ satisfying $\sup_{z\in\mathbb{R}}|f(z+\tau)-f(z)| \le \varepsilon$. The point τ is called an *ε-translation number* for f.

Every periodic function f with period r is also almost periodic because every interval of length r, $[\ell, \ell+r)$, say, contains a unique point nr for some $n\in\mathbb{Z}$ such that $\sup_{z\in\mathbb{R}}|f(z+nr)-f(z)| = 0$. Thus, the definition above holds with $\tau = nr$.

Lemma 152. *The sum of finite number of almost periodic functions is almost periodic.*

The proof of this lemma which is highly nontrivial can be found on pp. 188-189 in [13].

Proof of Theorem 150. For any fixed $n\in\mathbb{N}$, set

$$f_n(t) = \sum_{j=1}^{n} p_j e^{\mathrm{i}tx_j}, \quad t\in\mathbb{R},$$

where, for $j = 1, 2, \dots, n$, $p_j > 0$ and $x_j \in \mathbb{R}$. We first prove that

$$\limsup_{|t|\to\infty} |f_n(t)| = |f_n(0)| = \sum_{j=1}^{n} p_j. \tag{3.30}$$

In particular, if $\sum_{j=1}^{n} p_j = 1$, then $f_n(t)$ is the characteristic function of a random variable taking finite number of values $x_1, \dots, x_n$.

For each $j = 1, \dots, n$, the function $t \mapsto p_j e^{\mathrm{i}tx_j}$ is periodic with period $2\pi/x_j$, hence also almost periodic. Appealing to Lemma 152 we conclude that f_n is almost periodic, too. Thus, given $\varepsilon > 0$ there exists $L(\varepsilon) > 0$ such that for any $\ell \in \mathbb{R}$ we can find $\tau \in (\ell, \ell + L(\varepsilon))$ satisfying

$$|f_n(0)| - |f_n(\tau)| \le |f_n(0) - f_n(\tau)| \le \sup_{z\in\mathbb{R}} |f_n(z+\tau) - f_n(z)| \le \varepsilon.$$

Thus, for any $\varepsilon > 0$ the inequality $|f_n(t)| \ge |f_n(0)| - \varepsilon$ holds for infinitely many t both positive and negative or equivalently

$$\limsup_{|t|\to\infty} |f_n(t)| \ge |f_n(0)|.$$

Since $|f_n(t)| \le \sum_{j=1}^{n} p_j = |f_n(0)|$ we infer (3.30).

Assume now that ξ takes countably many values x_1, $x_2, \dots$, so that its characteristic function is

$$\varphi(t) = \sum_{j\ge1} p_j e^{\mathrm{i}tx_j}, \quad t \in \mathbb{R},$$

where $p_j := \mathbb{P}\{\xi = x_j\}$ for $j \in \mathbb{N}$ satisfy $\sum_{j\ge1} p_j = 1$. For each $n \in \mathbb{N}$, write

$$|f_n(t)| - |\varphi(t)| \le |\varphi(t) - f_n(t)| = \Big| \sum_{j\ge n+1} p_j e^{\mathrm{i}tx_j} \Big| \le \sum_{j\ge n+1} p_j.$$

This implies that

$$\limsup_{|t|\to\infty} |\varphi(t)| \ge \limsup_{|t|\to\infty} |f_n(t)| - \sum_{j\ge n+1} p_j = \sum_{j=1}^{n} p_j - \sum_{j\ge n+1} p_j$$

having utilized (3.30) for the last equality. The right-hand side converges to 1 as $n \to \infty$. Since the left-hand side does not depend on n we infer

$$\limsup_{|t|\to\infty} |\varphi(t)| \ge 1.$$

Since $|\varphi(t)| \le 1$ we arrive at $\limsup_{|t|\to\infty} |\varphi(t)| = 1$. The proof of Theorem 150 is complete. □

3.7.2. Lattice distributions In this section we discuss an important subclass of discrete distributions.

Definition 153. A discrete distribution is called *lattice* if the values x_k belong to some arithmetic progression $(a + jh)_{j\in\mathbb{Z}}$, where $a \in \mathbb{R}$ and $h > 0$. The number h is called *span* of lattice distribution if the values x_k does not belong to an arithmetic progression $(b+jh_1)_{j\in\mathbb{Z}}$ for any $b\in\mathbb{R}$ and any $h_1>h$. The corresponding lattice distribution is then called h-lattice. A degenerate at $a\in\mathbb{R}$ distribution is assumed $|a|$-lattice. A distribution (not necessarily discrete) is called nonlattice if it is not h-lattice for any $h>0$.

Assume, for instance, that ξ takes positive odd values. These values belong to the arithmetic progression $(a+jh)_{j\in\mathbb{N}_0}$ with $a=0$ and $h=1$. However, $h=1$ is not a span, for these values also belong to the arithmetic progression $(b+jh_1)_{j\in\mathbb{N}_0}$, where $b=1$ and $h_1=2$. Thus, the span is 2.

In Section 3.1. five examples of discrete distributions are given. Among them a Hurwitz zeta distribution is nonlattice, whereas all the other are lattice (degenerate at a distribution is $|a|$-lattice; the distribution of a random variable taking values -1 and 1 is 2-lattice; a Poisson and a binomial distributions are 1-lattice).

Now we characterize characteristic functions of lattice distributions.

Theorem 154. *Let ξ be a random variable with characteristic function φ. The distribution of ξ is lattice if, and only if, there exists $t_0\in\mathbb{R}\backslash\{0\}$ such that $|\varphi(t_0)|=1$.*

Proof. Assume that the distribution of ξ is h-lattice. Thus, for $j\in\mathbb{Z}$, $p_j:=\mathbb{P}\{\xi=a+jh\}$ satisfy $p_j\geq 0$ and $\sum_{j\in\mathbb{Z}}p_j=1$. Then the characteristic function of ξ takes the form

$$\varphi(t)=\sum_{j\in\mathbb{Z}}p_je^{\mathrm{i}t(a+jh)}=e^{\mathrm{i}ta}\sum_{j\in\mathbb{Z}}p_je^{\mathrm{i}jht}.$$

Thus,

$$\begin{aligned}\varphi(2\pi/h)&=\exp(\mathrm{i}a(2\pi/h))\sum_{j\in\mathbb{Z}}p_j\exp(\mathrm{i}2\pi j)=\exp(\mathrm{i}a(2\pi/h))\sum_{j\in\mathbb{Z}}p_j\cos(2\pi j)\\&\quad+\exp(\mathrm{i}a(2\pi/h))\mathrm{i}\sum_{j\in\mathbb{Z}}p_j\sin(2\pi j)=\exp(\mathrm{i}a(2\pi/h)).\end{aligned}$$

We infer $|\varphi(2\pi/h)|=1$, so that we can take $t_0=2\pi/h$.

Assume now that $|\varphi(t_0)|=1$ for some $t_0\in\mathbb{R}\backslash\{0\}$. We have to show that the distribution of ξ is lattice. By assumption,

$$\varphi(t_0)=\int_{\mathbb{R}}e^{\mathrm{i}t_0x}\mathrm{d}\mathbb{P}\{\xi\leq x\}=e^{\mathrm{i}t_0\alpha}$$

for some $\alpha\in\mathbb{R}$. This entails

$$\begin{aligned}0&=\int_{\mathbb{R}}(1-e^{\mathrm{i}t_0(x-\alpha)})\mathrm{d}\mathbb{P}\{\xi\leq x\}=\int_{\mathbb{R}}(1-\cos(t_0(x-\alpha)))\mathrm{d}\mathbb{P}\{\xi\leq x\}\\&\quad-\mathrm{i}\int_{\mathbb{R}}\sin(t_0(x-\alpha))\mathrm{d}\mathbb{P}\{\xi\leq x\}=\int_{\mathbb{R}}(1-\cos(t_0(x-\alpha)))\mathrm{d}\mathbb{P}\{\xi\leq x\},\end{aligned}$$

where the last equality follows from the fact that the left-hand side is real. The integrand $x\mapsto 1-\cos(t_0(x-\alpha))$ is nonnegative and continuous. Therefore, the last integral can only be equal to 0 if the distribution function of ξ increases at points x satisfying $\cos(t_0(x-\alpha))=1$, that is, $x=\alpha+(2\pi/t_0)k$, $k\in\mathbb{Z}$. We have proved that the distribution of ξ is lattice, for ξ only takes values in the set $(\alpha+(2\pi/t_0)k)_{k\in\mathbb{Z}}$. The proof of Theorem 154 is complete. $\square$

Corollary 155. Let ξ be a random variable with characteristic function φ. The distribution of ξ is h-lattice if, and only if, $|\varphi(2\pi/h)| = 1$ and $|\varphi(t)| < 1$ when $0 < |t| < 2\pi/h$.

Proof. From the proof of the direct part of Theorem 154 we know that if the distribution of ξ is h-lattice, then $|\varphi(2\pi/h)| = 1$. Assume that $|\varphi(t_1)| = 1$ for some $t_1 \in (0, 2\pi/h)$. From the proof of the converse part of Theorem 154 it follows that ξ takes values in the set $(\alpha + (2\pi/t_1)k)_{k\in\mathbb{Z}}$ for some $\alpha \in \mathbb{R}$. Then h is not a span, for $h < 2\pi/t_1$, a contradiction. The case $-2\pi/h < t_1 < 0$ can be analyzed similarly. A symmetric argument proves the other implication of the corollary. □

Let φ be the characteristic function of an h-lattice distribution. Let us show that the function $|\varphi|$ is periodic with period $2\pi/h$ (hence, $4\pi/h$, $6\pi/h, \dots$ are also periods). Indeed, for some $a \in \mathbb{R}$ and any $t \in \mathbb{R}$,

$$\begin{aligned}\varphi(t+2\pi/h) &= e^{\mathrm{i}a(t+2\pi/h)}\sum_{j\in\mathbb{Z}} p_j e^{\mathrm{i}(2\pi j+tjh)} = e^{\mathrm{i}a(t+2\pi/h)}\sum_{j\in\mathbb{Z}} p_j(\cos(2\pi j+tjh)+\mathrm{i}\sin(2\pi j+tjh)) \\ &= e^{\mathrm{i}a(t+2\pi/h)}\sum_{j\in\mathbb{Z}} p_j(\cos(tjh)+\mathrm{i}\sin(tjh)) = e^{\mathrm{i}a(t+2\pi/h)}\sum_{j\in\mathbb{Z}} p_j e^{\mathrm{i}tjh},\end{aligned}$$

whence $|\varphi(t+2\pi/h)| = |\varphi(t)|$ for any $t \in \mathbb{R}$ because $|\varphi(t)| = |\sum_{j\in\mathbb{Z}} p_j e^{\mathrm{i}tjh}|$. Looking at the proof of Theorem 154 we also conclude that $2\pi/t_0$ is a period of $|\varphi|$ whenever $|\varphi(t_0)| = 1$.

According to Problem 191 there exists characteristic functions which take the constant value $p \in (0,1)$ at some interval. An interesting question is: does there exist a characteristic function φ taking the value -1 at some interval? To answer this question we formulate an auxiliary result.

Lemma 156. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous periodic function with periods t_1 and t_2 such that the ratio t_1/t_2 is irrational. Then f is a constant function.*

Proof. By Kroneker's theorem (Theorem 438 in Section 23.1 of [5]), the set $A := \{n_1t_1+n_2t_2 : n_1, n_2 \in \mathbb{Z}\}$ is dense in $\mathbb{R}$. Therefore, given $t \in \mathbb{R}$ there exists a sequence $(s_n)_{n\in\mathbb{N}}$ with $s_n \in A$ for each $n \in \mathbb{N}$ such that $\lim_{n\to\infty} s_n = t$. Since f is continuous, it holds that $\lim_{n\to\infty} f(s_n) = f(t)$. Since $s_n \in A$, $f(s_n) = f(0)$, whence $f(t) = f(0)$. Repeating the argument for each $t \in \mathbb{R}$ we infer $f(t) = f(0)$ for all $t \in \mathbb{R}$. □

Getting back to our question assume that a characteristic function φ exists such that $\varphi(t) = -1$ for $t \in [a, b]$ for some $a < b$. Then there exist $t_1, t_2 \in [a, b]$ such that t_1/t_2 is an irrational number and that $|\varphi(t_1)| = |\varphi(t_2)| = 1$. In particular, $|\varphi|$ is a periodic function with periods $2\pi/t_1$ and $2\pi/t_2$ whose ratio is irrational. By the preceding lemma, $|\varphi(t)| = 1$ for all $t \in \mathbb{R}$ because $|\varphi|$ is a continuous function. The latter means that the distribution of ξ is degenerate, that is, $\varphi(t) = e^{\mathrm{i}at}$, $t \in \mathbb{R}$ for some $a \in \mathbb{R}$, a contradiction. We conclude that no characteristic functions can take the value -1 on an interval.

3.7.3. Continuous distributions Let ξ be a random variable with distribution function F, that is, $F(x) = \mathbb{P}\{\xi \le x\}$ for $x \in \mathbb{R}$.

Definition 157. The distribution of a random variable ξ is called *continuous*, if $\mathbb{P}\{\xi = x\} = F(x) - F(x-) = 0$ for all $x \in \mathbb{R}$, that is, the distribution function F is continuous.

The result given next is useful for checking that the given distribution is continuous. Indeed, the latter is the case whenever the limit in (3.31) is equal to 0. The sum $\sum_x \Big(\mathbb{P}\{\xi = x\}\Big)^2$ appearing below contains at most countably many nonzero terms. This follows from the fact that each $x \in \mathbb{R}$ for which $\mathbb{P}\{\xi = x\} > 0$ is a discontinuity point of F the distribution function of ξ. Since F is monotone, the number of its discontinuity points is at most countable.

Proposition 158. *Let ξ be a random variable with characteristic function φ. Then*

$$\sum_x \Big(\mathbb{P}\{\xi = x\}\Big)^2 = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|^2 \mathrm{d}t. \tag{3.31}$$

Proof. Let ξ_1 and ξ_2 be independent random variables with characteristic function φ. According to Corollary 119, the characteristic function of $\xi_1 - \xi_2$ is $t \mapsto |\varphi(t)|^2$. Use now Theorem 149 with $x = 0$ and $\xi_1 - \xi_2$ replacing ξ to obtain

$$\mathbb{P}\{\xi_1 - \xi_2 = 0\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|^2 \mathrm{d}t.$$

It remains to write

$$\begin{aligned}\mathbb{P}\{\xi_1 - \xi_2 = 0\} &= \int_{\mathbb{R}} \mathbb{P}\{\xi_1 - \xi_2 = 0 | \xi_2 = x\} \mathrm{d}\mathbb{P}\{\xi_2 \le x\} = \int_{\mathbb{R}} \mathbb{P}\{\xi_1 = x\} \mathrm{d}\mathbb{P}\{\xi_2 \le x\} \\ &= \sum_x \mathbb{P}\{\xi_1 = x\}\mathbb{P}\{\xi_2 = x\} = \sum_x \Big(\mathbb{P}\{\xi = x\}\Big)^2.\end{aligned}$$

□

Definition 159. The distribution of ξ is called *absolutely continuous* (with respect to the Lebesgue measure), if $\mathbb{P}\{\xi \in C\} = 0$ for each Borel set C of the Lebesgue measure zero. Equivalently, the distribution function F is called *absolutely continuous*, if there exists a nonnegative Lebesgue integrable function h, which is called *density* of the distribution, such that $F(b) - F(a) = \int_a^b h(x)\mathrm{d}x$ for all $a < b$.

Definition 160. The distribution of ξ is called *continuous singular* (with respect to the Lebesgue measure), if it is continuous and $\mathbb{P}\{\xi \in C\} = 1$ for some Borel set C of the Lebesgue measure zero. Equivalently, the distribution function F is called *continuous singular*, if it is continuous and its derivative F' is equal to zero almost everywhere.

Given next is the result which is called *Lebesgue's decomposition theorem* for continuous distributions.

Theorem 161. *Every continuous distribution function F can be uniquely represented as a mixture (convex linear combination) of an absolutely continuous distribution function F_{ac} and a continuous singular distribution function F_{cs}, that is,*

$$F(x) = \alpha F_{ac}(x) + (1-\alpha)F_{cs}(x), \quad x \in \mathbb{R} \tag{3.32}$$

for some $\alpha \in [0,1]$.

Proof. It is known (see Theorems 4 and 5 on pp. 211-212 in [8]) that a nondecreasing function G is almost everywhere differentiable and satisfies

$$\int_a^b G'(x)\mathrm{d}x \leq G(b) - G(a) \quad \text{for all} \quad a < b. \tag{3.33}$$

Thus, the derivative $h(x) := F'(x)$ exists almost everywhere. Put

$$\widehat{F}_{ac}(x) := \int_{-\infty}^x h(y)\mathrm{d}y, \quad x \in \mathbb{R}.$$

The so defined function is nondecreasing (because $h \geq 0$ almost everywhere) and absolutely continuous (by definition). Using (3.33) with $G = F$, $b = x$ and $a = -\infty$ gives $\widehat{F}_{ac}(x) \leq F(x)$ which ensures that $\widehat{F}_{ac}(-\infty) = 0$. However, it may happen that $\widehat{F}_{ac}(+\infty) < 1$. Thus, $\widehat{F}_{ac}$ is a (possibly improper) absolutely continuous distribution function. Set

$$\widehat{F}_{cs}(x) := F(x) - \widehat{F}_{ac}(x), \quad x \in \mathbb{R}.$$

The function $\widehat{F}_{cs}$ is continuous as a difference of two continuous functions. Furthermore, from $\widehat{F}_{cs}(x) \leq F(x)$, $x \in \mathbb{R}$ it follows that $\widehat{F}_{cs}(-\infty) = 0$. Also, $\widehat{F}_{cs}$ is nondecreasing because, for $a < b$,

$$0 \leq F(b) - F(a) - \int_a^b F'(x)\mathrm{d}x = F(b) - \widehat{F}_{ac}(b) - (F(a) - \widehat{F}_{ac}(a)) = \widehat{F}_{cs}(b) - \widehat{F}_{cs}(a)$$

as a consequence of (3.33). Finally, $\widehat{F}'_{cs}(x) = 0$ almost everywhere. It may happen that $\widehat{F}_{cs}(+\infty) < 1$. Thus, $\widehat{F}_{cs}$ is a (possibly improper) continuous singular distribution function.

If $\widehat{F}_{ac}(+\infty) = 1$, then (3.32) holds with $\alpha = 1$ and $F_{ac} := \widehat{F}_{ac}$. If $\widehat{F}_{cs}(+\infty) = 1$, then (3.32) holds with $\alpha = 0$ and $F_{cs} := \widehat{F}_{cs}$. In the remaining cases (3.32) holds with $\alpha = \widehat{F}_{ac}(+\infty)$, $F_{ac}(x) := \widehat{F}_{ac}(x)/\widehat{F}_{ac}(+\infty)$ and $F_{cs}(x) := \widehat{F}_{cs}(x)/\widehat{F}_{cs}(+\infty)$. □

3.7.4. Absolutely continuous distributions In this section we investigate properties of characteristic functions which correspond to absolutely continuous distributions.

Theorem 162 is called the theorem on *absolutely integrable characteristic functions.*

Theorem 162. *Let ξ be a random variable with characteristic function φ. If φ is absolutely integrable on $\mathbb{R}$, that is, $\int_{\mathbb{R}} |\varphi(t)|\mathrm{d}t < \infty$, then the distribution of ξ is absolutely continuous with bounded and continuous density h given by*

$$h(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t, \quad x \in \mathbb{R}. \tag{3.34}$$

Proof. We start by observing that, for any $T > 0$,

$$\int_{-T}^{T} |\varphi(t)|^2 \mathrm{d}t \le \int_{\mathbb{R}} |\varphi(t)|^2 \mathrm{d}t \le \int_{\mathbb{R}} |\varphi(t)| \mathrm{d}t < \infty.$$

Hence,

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |\varphi(t)|^2 \mathrm{d}t = 0,$$

and we conclude with the help of (3.31) that the distribution of ξ is continuous.

Let $x \in \mathbb{R}$ be arbitrary. Using inversion formula (3.9) with $x_1 = x$ and $x_2 = x+\delta$ for $\delta > 0$ and recalling that $\mathbb{P}\{\xi = x\} = \mathbb{P}\{\xi = x + \delta\} = 0$ in view of continuity of the distribution of ξ we obtain

$$\begin{aligned} F(x+\delta) - F(x) &= \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-\mathrm{i}tx}(1 - e^{-\mathrm{i}t\delta})}{\mathrm{i}t} \varphi(t) \mathrm{d}t \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-\mathrm{i}tx}(1 - e^{-\mathrm{i}t\delta})}{\mathrm{i}t} \varphi(t) \mathrm{d}t, \end{aligned} \tag{3.35}$$

where F is the distribution function of ξ. The last equality follows from

$$\Big|\frac{e^{-\mathrm{i}tx}(1 - e^{-\mathrm{i}t\delta})}{\mathrm{i}t} \varphi(t)\Big| \le \frac{2|\sin(t\delta/2)|}{|t|} |\varphi(t)| \le \delta |\varphi(t)| \tag{3.36}$$

and the fact that $|\varphi|$ is Lebesgue integrable on $\mathbb{R}$. The same formula also holds if we take in (3.9) $x_1 = x + \delta$ for $\delta < 0$ and $x_2 = x$. Thus, dividing equality (3.35) by δ and passing to the limit as $\delta \to 0$ we obtain

$$\begin{aligned} \lim_{\delta\to 0} \frac{F(x+\delta) - F(x)}{\delta} = \lim_{\delta\to 0} \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-\mathrm{i}tx}(1 - e^{-\mathrm{i}t\delta})}{\mathrm{i}t\delta} \varphi(t) \mathrm{d}t = \frac{1}{2\pi} \int_{\mathbb{R}} \lim_{\delta\to 0} \frac{e^{-\mathrm{i}tx}(1 - e^{-\mathrm{i}t\delta})}{\mathrm{i}t\delta} \varphi(t) \mathrm{d}t \\ = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\mathrm{i}tx} \varphi(t) \mathrm{d}t. \end{aligned}$$

Thus, for all $x \in \mathbb{R}$, the limit on the left-hand side exists and is equal to F' the derivative of F which is also one version of the density h (recall that densities are determined up to sets of Lebesgue measure 0). The second equality in the last centered formula is justified by Lebesgue's dominated convergence theorem as follows: divide both sides of (3.36) in which $\delta \in \mathbb{R}$ by $|\delta|$ and use the fact that $|\varphi|$ is integrable.

Boundedness of h follows from

$$|h(x)| = \frac{1}{2\pi} \Big| \int_{\mathbb{R}} e^{-\mathrm{i}tx} \varphi(t) \mathrm{d}t \Big| \le \frac{1}{2\pi} \int_{\mathbb{R}} |\varphi(t)| \mathrm{d}t.$$

Thus, it remains to prove continuity of h. To this end, write, for any fixed $x \in \mathbb{R}$, any $\delta \in \mathbb{R}$ and any $A > 0$,

$$|h(x+\delta) - h(x)| = \Big| \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\mathrm{i}tx}(e^{-\mathrm{i}t\delta} - 1) \varphi(t) \mathrm{d}t \Big| \le \frac{1}{\pi} \int_{-A}^{A} |\sin(t\delta/2)| \mathrm{d}t + \frac{1}{\pi} \int_{|t|>A} |\varphi(t)| \mathrm{d}t.$$

Here, we have used $|\varphi(t)| \le 1$ for all $t \in \mathbb{R}$ and $|1 - e^{-\mathrm{i}x}| = 2|\sin(x/2)|$ for $x \in \mathbb{R}$. Therefore,

$$\limsup_{\delta\to 0} |h(x+\delta) - h(x)| \le \frac{1}{\pi} \int_{|t|>A} |\varphi(t)| \mathrm{d}t. \tag{3.37}$$

The limit relation $\lim_{\delta\to 0}\frac{1}{\pi}\int_{-A}^{A}|\sin(t\delta/2)|\mathrm{d}t=0$ holds by Lebesgue's dominated convergence theorem because $\lim_{\delta\to 0}|\sin(t\delta/2)|=0$ and $|\sin(t\delta/2)|\leq 1$ for all $t\in[-A,A]$ and $\delta\in\mathbb{R}$. Since the right-hand side in (3.37) converges to zero as $A\to\infty$ (as a consequence of absolute integrability of φ), and the left-hand side of (3.37) does not depend on A we infer $\lim_{\delta\to 0}|h(x+\delta)-h(x)|=0$, thereby proving the desired continuity. The proof of Theorem 162 is complete. □

Remark 163. The conditions of Theorem 162 are not necessary for absolute continuity. For instance, the uniform distribution on $(-1,1)$ is absolutely continuous with density $x\mapsto 2^{-1}\mathbb{1}_{(-1,1)}(x)$, yet its characteristic function $t\mapsto \sin t/t$ for $t\in\mathbb{R}$ is not absolutely integrable. Remark 184 contains more examples of this flavor.

Theorem 164. *Let ξ be a random variable with characteristic function φ and density h. The function $|\varphi|^2$ is Lebesgue integrable if, and only if, so is the function h^2. Furthermore, the* **Plancherel identity** *holds*

$$\int_{\mathbb{R}}(h(y))^2\mathrm{d}y=\frac{1}{2\pi}\int_{\mathbb{R}}|\varphi(t)|^2\mathrm{d}t. \tag{3.38}$$

For the proof we need an auxiliary result which is of independent interest.

Proposition 165. *Let ξ be a random variable with characteristic function φ satisfying $\varphi(t)\geq 0$ for all $t\in\mathbb{R}$. The function φ is Lebesgue integrable if, and only if, the distribution of ξ is absolutely continuous with bounded density.*

Proof. Assume that φ is Lebesgue integrable. Then, by Theorem 162 the distribution of ξ is absolutely continuous with bounded (and also continuous) density.

Assume now that the distribution of ξ is absolutely continuous with bounded density h, that is, there exists a positive constant M such that $h(t)\leq M$ for all $t\in\mathbb{R}$. We first need a general fact. Let h_1 and h_2 be densities of some absolutely continuous distributions. Set

$$\hat{h}_i(t):=\int_{\mathbb{E}}e^{\mathrm{i}tx}h_i(x)\mathrm{d}x,\quad t\in\mathbb{R}, i=1,2.$$

The following *Parseval identity* holds

$$\int_{\mathbb{R}}\hat{h}_1(t)h_2(t)\mathrm{d}t=\int_{\mathbb{R}}\hat{h}_2(t)h_1(t)\mathrm{d}t.$$

Indeed,

$$\int_{\mathbb{R}}\hat{h}_1(t)h_2(t)\mathrm{d}t=\int_{\mathbb{R}}\int_{\mathbb{R}}e^{\mathrm{i}tx}h_1(x)\mathrm{d}xh_2(t)\mathrm{d}t=\int_{\mathbb{R}}\int_{\mathbb{R}}e^{\mathrm{i}tx}h_2(t)\mathrm{d}th_1(x)\mathrm{d}x=\int_{\mathbb{R}}\hat{h}_2(x)h_1(x)\mathrm{d}x.$$

The second equality is obtained by the interchange of the two integrals. This is allowed by Fubini's theorem, for $|\int_{\mathbb{R}}\hat{h}_1(t)h_2(t)\mathrm{d}t|\leq\int_{\mathbb{R}}h_2(t)\mathrm{d}t=1$.

We are going to use the Parseval identity with $h_1=h$, so that $\hat{h}_1=\varphi$. Fix any $a>0$. In the role of h_2 we take $h^{(a)}$ the density of a random variable η/a, where η has a standard normal distribution. Thus,

$$h^{(a)}(x)=\frac{a}{\sqrt{2\pi}}e^{-a^2x^2/2},\quad x\in\mathbb{R}.$$

Recall from Section 3.1. that the characteristic function of η is $t \mapsto e^{-t^2/2}$ for $t \in \mathbb{R}$. Hence, by Lemma 109 (d),

$$\hat{h}^{(a)}(t) = e^{-t^2/(2a^2)}, \quad t \in \mathbb{R}.$$

Thus, we shall use a specialization of Parseval's identity of the form

$$\frac{a}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(t) e^{-a^2t^2/2} \mathrm{d}t = \int_{\mathbb{R}} e^{-t^2/(2a^2)} h(t) \mathrm{d}t$$

or equivalently

$$\frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) e^{-a^2t^2/2} \mathrm{d}t = \frac{1}{a\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/(2a^2)} h(t) \mathrm{d}t.$$

The last transformation makes sense because $x \mapsto \frac{1}{a\sqrt{2\pi}} e^{-x^2/(2a^2)}$ for $x \in \mathbb{R}$ is the density of the random variable $a\eta$ which particularly implies that $\frac{1}{a\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/(2a^2)} \mathrm{d}t = 1$. This together with the last centered formula ensures that

$$\frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) e^{-a^2t^2/2} \mathrm{d}t \leq \frac{M}{a\sqrt{2\pi}} \int_{\mathbb{R}} e^{-t^2/(2a^2)} \mathrm{d}t = M.$$

Assume now that φ is not Lebesgue integrable. Using Fatou's lemma (this is the only place in the proof where nonnegativity of φ is of principal importance) we then arrive at

$$\liminf_{a \to 0+} \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) e^{-a^2t^2/2} \mathrm{d}t \geq \frac{1}{2\pi} \int_{\mathbb{R}} \varphi(t) \mathrm{d}t = \infty,$$

a contradiction to the preceding centered formula. The proof of Proposition 165 is complete.

□

Now we are ready to prove Theorem 164.

Proof of Theorem 164. Let ξ_1 and ξ_2 be independent copies of ξ. According to Corollary 119, the characteristic function of $\xi_1 - \xi_2$ is $|\varphi|^2$. Also, the distribution of $\xi_1 - \xi_2$ is absolutely continuous with density $\bar{h}$ given by

$$\bar{h}(x) = \int_{\mathbb{R}} h(x+y) h(y) \mathrm{d}y, \quad x \in \mathbb{R}. \tag{3.39}$$

To check this, recall from (3.16) that

$$\mathbb{P}\{\xi_1 - \xi_2 \leq x\} = \int_{\mathbb{R}} \mathbb{P}\{\xi \leq x+y\} \mathrm{d}\mathbb{P}\{\xi \leq y\} = \int_{\mathbb{R}} \mathbb{P}\{\xi \leq x+y\} h(y) \mathrm{d}y.$$

Since the distribution of ξ has density h it holds that $\mathbb{P}\{\xi \leq x\} = \int_{-\infty}^{x} h(z) \mathrm{d}z$ for $x \in \mathbb{R}$. Hence,

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{P}\{\xi \leq x+y\} h(y) \mathrm{d}y &= \int_{\mathbb{R}} \int_{-\infty}^{x+y} h(z) \mathrm{d}z h(y) \mathrm{d}y \\ &= \int_{\mathbb{R}} \int_{-\infty}^{x} h(z+y) \mathrm{d}z h(y) \mathrm{d}y = \int_{-\infty}^{x} \Big(\int_{\mathbb{R}} h(z+y) h(y) \mathrm{d}y \Big) \mathrm{d}z = \int_{-\infty}^{x} \bar{h}(z) \mathrm{d}z \end{aligned}$$

which proves (3.39). The third equality above is justified by Fubini's theorem and the fact that the integrands are nonnegative.

Assume now that $|\varphi|^2$ is Lebesgue integrable. Then, by Theorem 162,

$$\bar{h}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}tx}|\varphi(t)|^2\mathrm{d}t.$$

Setting $x = 0$ we obtain

$$\bar{h}(0) = \frac{1}{2\pi}\int_{\mathbb{R}} |\varphi(t)|^2\mathrm{d}t.$$

On the other hand, an appeal to (3.39) gives

$$\bar{h}(0) = \int_{\mathbb{R}} (h(y))^2\mathrm{d}y,$$

and (3.38) follows.

Assume now that h^2 is Lebesgue integrable. An application of the Cauchy-Schwarz inequality to (3.39) yields

$$\bar{h}(x) \leq \Big(\int_{\mathbb{R}} (h(x+y))^2\mathrm{d}y\Big)^{1/2}\Big(\int_{\mathbb{R}} (h(y))^2\mathrm{d}y\Big)^{1/2} = \int_{\mathbb{R}} (h(y))^2\mathrm{d}y, \quad x \in \mathbb{R}$$

which shows that the density $\bar{h}$ is bounded. Since the corresponding characteristic function $|\varphi|^2$ is real and nonnegative we can use Proposition 165 to conclude that $|\varphi|^2$ is Lebesgue integrable. The proof of Theorem 164 is complete. □

Theorem 166 given next is called *Riemann-Lebesgue theorem*.

Theorem 166. *Let $g : \mathbb{R} \to \mathbb{R}$ be a Lebesgue integrable function. Then*

$$\lim_{|t|\to\infty} \int_{\mathbb{R}} e^{\mathrm{i}tx} g(x)\mathrm{d}x = 0. \tag{3.40}$$

Proof. Assume first that g is a finite linear combination of indicators, that is, for some $n \in \mathbb{N}$, $g(x) = \sum_{j=1}^n c_j \mathbb{1}_{[a_j,b_j]}(x)$ for $x \in \mathbb{R}$, where, for $j = 1,\ldots,n$, c_j and $a_j < b_j$ are real. Then

$$\int_{\mathbb{R}} e^{\mathrm{i}tx} g(x)\mathrm{d}x = \sum_{j=1}^n c_j \int_{a_j}^{b_j} e^{\mathrm{i}tx}\mathrm{d}x = \sum_{j=1}^n c_j \frac{e^{\mathrm{i}tb_j} - e^{\mathrm{i}ta_j}}{\mathrm{i}t} \quad \to \quad 0$$

as $|t| \to \infty$. The theorem on approximation in L_1 states that given $\varepsilon > 0$ and a Lebesgue integrable function g there exists a function g_1 which is a finite linear combination of indicators such that $\int_{\mathbb{R}} |g(x) - g_1(x)|\mathrm{d}x \leq \varepsilon$. Using first the triangle inequality and then the latter inequality we obtain

$$\Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g(x)\mathrm{d}x\Big| \leq \Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g_1(x)\mathrm{d}x\Big| + \Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g(x)\mathrm{d}x - \int_{\mathbb{R}} e^{\mathrm{i}tx} g_1(x)\mathrm{d}x\Big| \leq \Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g_1(x)\mathrm{d}x\Big|$$
$$+ \int_{\mathbb{R}} |g(x) - g_1(x)|\mathrm{d}x \leq \Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g_1(x)\mathrm{d}x\Big| + \varepsilon.$$

According to the first part of the proof,

$$\lim_{|t|\to\infty} \Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g_1(x)\mathrm{d}x\Big| = 0,$$

whence

$$\limsup_{|t|\to\infty} \Big|\int_{\mathbb{R}} e^{\mathrm{i}tx} g(x)\mathrm{d}x\Big| \leq \varepsilon$$

and thereupon (3.40) because $\varepsilon > 0$ was arbitrary. The proof of Theorem 166 is complete. □

Corollary 167. Let ξ be a random variable with characteristic function φ. If the distribution of ξ is absolutely continuous, then

$$\lim_{|t|\to\infty} \varphi(t) = 0.$$

If the distribution of ξ is absolutely continuous with a density whose derivative of the nth order is Lebesgue integrable, then

$$|\varphi(t)| = o(|t|^{-n}), \quad |t| \to \infty.$$

Proof. Assume that the distribution of ξ is absolutely continuous with density h. Then

$$\varphi(t) = \int_{\mathbb{R}} e^{\mathrm{i}tx} h(x)\mathrm{d}x,$$

and the claim follows from Theorem 166 because h is Lebesgue integrable ($\int_{\mathbb{R}} h(x)\mathrm{d}x = 1$).

If h' is Lebesgue integrable, then integrating by parts yields

$$\varphi(t) = \frac{1}{\mathrm{i}t} \int_{\mathbb{R}} e^{\mathrm{i}tx} h'(x)\mathrm{d}x.$$

We infer $|\varphi(t)| = o(|t|^{-1})$ as $|t| \to \infty$ by another appeal to Theorem 166. The proof for derivatives of higher orders is analogous. □

3.7.5. Continuity properties of the distribution of random power series Let $Q_1, Q_2, \ldots$ be independent copies of a positive random variable Q. In this section we are interested in continuity properties of convergent random power series $Z = \sum_{k\geq 0} c^k Q_{k+1}$, where $c \in (0,1)$ is fixed. Using Problem 188 one can check that $\mathbb{P}\{Z < \infty\} = 1$ if, and only if, $\mathbb{E}\log^+ Q < \infty$. However, we will not pursue this here. Obviously, $\mathbb{P}\{Z < \infty\} = 1$ whenever $\mathbb{E}Q < \infty$, for then $\mathbb{E}\sum_{k\geq 0} c^k Q_{k+1} = \mathbb{E}Q/(1-c) < \infty$.

Theorem 168. *Suppose that* $\mathbb{P}\{Z < \infty\} = 1$. *Then the distribution of* Z *is either degenerate, or absolutely continuous, or continuous singular.*

Proof. Observe that

$$Z = Q_1 + cZ_1 \quad \text{with probability one,} \tag{3.41}$$

where the random variable $Z_1 := Q_2 + cQ_3 + \ldots$ has the same distribution as Z and is independent of Q_1.

From the very beginning we note that the distribution of Z is degenerate if, and only if, the distribution of Q is degenerate. Indeed, if $\mathbb{P}\{Q = a\} = 1$ for some $a \geq 0$, it follows that $\mathbb{P}\{Z = a(1-c)^{-1}\} = 1$. For the proof in the converse direction, observe that, under $\mathbb{P}\{Z = b\} = 1$, equality (3.41) reads $b = cb + Q_1$ almost surely, whence $\mathbb{P}\{Q = b(1-c)\} = 1$.

We proceed by showing that if the distribution of Z has atoms, it must be degenerate, that is, concentrated at one point. Denote by $b_1, \ldots b_d$ the atoms of maximal weight ρ. This means that $\mathbb{P}\{Z = b_j\} = \rho$, $j = 1, \ldots, d$ and $\mathbb{P}\{Z = x\} < \rho$ for all $x \neq b_j$, $j = 1, \ldots, d$. Using (3.41) we obtain

$$\rho = \mathbb{P}\{Z = b_j\} = \mathbb{P}\{cZ_1 + Q_1 = b_j\} = \sum_{a \in A} \mathbb{P}\{Z = c^{-1}(b_j - a)\}\mathbb{P}\{Q = a\}, \quad j = 1, \ldots, d, \tag{3.42}$$

where A is the set of atoms of the distribution of Q. Since $\mathbb{P}\{Z = c^{-1}(b_j - a)\} \leq \rho$ and $\sum_{a \in A} \mathbb{P}\{Q = a\} \leq 1$, equality (3.42) can only hold in the case where the distribution of Q is discrete and concentrated at points $a_1, \ldots, a_k$, $k \leq d$ such that $c^{-1}(b_j - a_i) \in \{b_1, \ldots, b_d\}$ for all $j = 1, \ldots, d$ and $i = 1, \ldots, k$. If $k = 1$, then the distribution of Q is degenerate and so is the distribution of Z. Suppose that $k = 2, 3, \ldots$ The values $c^{-1}(b_1 - a_1), \ldots, c^{-1}(b_d - a_1)$ belong to the set $\{b_1, \ldots, b_d\}$ and are distinct. Hence,

$$c^{-1}(b_1 - a_1) + \ldots + c^{-1}(b_d - a_1) = b_1 + \ldots + b_d.$$

Analogously,

$$c^{-1}(b_1 - a_2) + \ldots + c^{-1}(b_d - a_2) = b_1 + \ldots + b_d$$

and thereupon $a_1 = a_2$ which means that the distribution of Q is degenerate.

Assuming now that the distribution of Z is continuous we shall show that it is either absolutely continuous or continuous singular. Denote by $\varphi(t) := \mathbb{E}e^{\mathrm{i}tZ}$, $t \in \mathbb{R}$ the characteristic function of Z. In terms of characteristic functions the Lebesgue decomposition (see Theorem 161) reads

$$\varphi(t) = \alpha_1\varphi_1(t) + \alpha_2\varphi_2(t), \quad t \in \mathbb{R}, \tag{3.43}$$

where $\varphi_1(t)$ is the characteristic function of an absolutely continuous distribution, $\varphi_2(t)$ is the characteristic function of a continuous singular distribution, $\alpha_1, \alpha_2 \geq 0$ and $\alpha_1 + \alpha_2 = 1$.

If $\alpha_1 = 0$, the distribution of Z is continuous singular. Assuming that $\alpha_1 > 0$ we shall prove that $\alpha_1 = 1$ or, equivalently, $\varphi(t) = \varphi_1(t)$. This will mean that the distribution of Z is absolutely continuous. Equality (3.41) is equivalent to

$$\varphi(t) = \varphi(ct)\psi(t), \quad t \in \mathbb{R}, \tag{3.44}$$

where $\psi(t) := \mathbb{E}e^{\mathrm{i}tQ}$, $t \in \mathbb{R}$. Substituting in (3.44) the representation for $\varphi(t)$ given in formula (3.43) we obtain

$$\alpha_1\varphi_1(t) + \alpha_2\varphi_2(t) = \alpha_1\varphi_1(ct)\psi(t) + \alpha_2\varphi_2(ct)\psi(t), \quad t \in \mathbb{R}.$$

Let us show that $\varphi_1(ct)\psi(t)$ is the characteristic function of an absolutely continuous distribution. Let X be a random variable which is independent of Q and has the characteristic function $\varphi_1(t)$ (hence, X has an absolutely continuous distribution). Since $\varphi_1(ct)\psi(t)$ is the characteristic function of the random variable $cX + Q$, we have to check that the distribution of the latter random variable is absolutely continuous. For arbitrary Borel set D of the Lebesgue measure zero the Lebesgue measure of the set $c^{-1}(D - x) := \{c^{-1}(d - x) : d \in D\}$

is equal to zero for any $x \in \mathbb{R}$. This implies $\mathbb{P}\{X \in c^{-1}(D-x)\} = 0$ because the distribution of X is absolutely continuous. Hence, for each such set D,

$$\mathbb{P}\{cX + Q \in D\} = \int_{[0,\infty)} \mathbb{P}\{X \in c^{-1}(D-x)\}\mathrm{d}\mathbb{P}\{Q \le x\} = 0$$

which proves absolute continuity of the distribution of $cX + Q$.

Invoking once again the Lebesgue decomposition (Theorem 161) gives $\varphi_2(ct)\psi(t) = \alpha_3\varphi_3(t) + \alpha_4\varphi_4(t)$, $t \in \mathbb{R}$, where $\varphi_3(t)$ is the characteristic function of an absolutely continuous distribution, $\varphi_4(t)$ is the characteristic function of a continuous singular distribution, and $\alpha_3, \alpha_4 \ge 0$, $\alpha_3 + \alpha_4 = 1$. Since the Lebesgue decomposition of the characteristic function φ is unique, it follows that $\alpha_1\varphi_1(t) = \alpha_1\varphi_1(ct)\psi(t) + \alpha_2\alpha_3\varphi_3(t)$, $t \in \mathbb{R}$. Setting $t = 0$ and recalling that the value of a characteristic function at zero is equal to one, we arrive at $\alpha_2\alpha_3 = 0$. Hence, $\varphi_1(t) = \varphi_1(ct)\psi(t)$, $t \in \mathbb{R}$. Since φ satisfies (3.44), we infer, for $t \ne 0$ and any $n \in \mathbb{N}$,

$$|\varphi(t) - \varphi_1(t)| = |\varphi(ct) - \varphi_1(ct)||\psi(t)| \le \ldots \le |\varphi(c^n t) - \varphi_1(c^n t)|.$$

Letting $n \to \infty$ yields $\varphi(t) = \varphi_1(t)$, $t \ne 0$. From $\varphi(0) = \varphi_1(0) = 1$ we conclude that $\varphi(t) = \varphi_1(t)$ for all $t \in \mathbb{R}$.

So far we have only proved that the distribution of Z cannot be other than degenerate, absolutely continuous or continuous singular. However, we have not shown yet that the distribution could indeed be absolutely continuous or continuous singular. The second option is secured by Example 169 (and many others). The first option follows from the fact that if the distribution of Q is absolutely continuous, so is the distribution of Z. Indeed, using (3.41) we conclude that

$$\mathbb{P}\{Z \in D\} = \mathbb{P}\{cZ_1 + Q_1 \in D\} = \int_{[0,\infty)} \mathbb{P}\{Q \in D - cx\}\mathrm{d}\mathbb{P}\{Z \le x\} = 0$$

for any Borel set D of the Lebesgue measure zero. □

Example 169. Suppose that $c = 1/2$ and that Q has a Poisson distribution. Then the distribution of Z is continuous singular.

Proof. The characteristic function of Q takes the form

$$\psi(t) := \mathbb{E}e^{\mathrm{i}tQ} = \exp(\lambda(e^{\mathrm{i}t} - 1)), \quad t \in \mathbb{R}$$

for some $\lambda > 0$. Since $\mathbb{E}Q < \infty$, it follows that $\mathbb{P}\{Z < \infty\} = 1$. Furthermore,

$$\varphi(t) := \mathbb{E}e^{\mathrm{i}tZ} = \prod_{k\ge 0} \psi(2^{-k}t) = \exp\Big(\lambda \sum_{k\ge 0}(e^{\mathrm{i}2^{-k}t} - 1)\Big), \quad t \in \mathbb{R}.$$

Using Euler's formula $e^{\mathrm{i}x} = \cos x + \mathrm{i}\sin x$ yields

$$|\varphi(t)| = \exp\Big(-\lambda \sum_{k\ge 0}(1 - \cos(2^{-k}t))\Big), \quad t \in \mathbb{R}.$$

Setting $t_n := 2^n\pi$, $n \in \mathbb{N}$ and replacing t with t_n we obtain

$$|\varphi(t_n)| = \exp\Big(-\lambda \sum_{k\geq 0}(1-\cos(2^{n-k}\pi))\Big) = \exp\Big(-\lambda \sum_{j\geq 1}(1-\cos(2^{-j}\pi))\Big) > 0 \quad (3.45)$$

for each $n \in \mathbb{N}$, having utilized the relation $1-\cos x \sim 2^{-1}x^2$, $x \to 0$ for the inequality. If the distribution of Z were absolutely continuous, we would have $\lim_{t\to\infty}|\varphi(t)| = 0$ by the Riemann-Lebesgue theorem (Theorem 166). However, this is the not the case in view of (3.45). Therefore, by Theorem 168, the distribution of Z is continuous singular. □

Remark 170. For absolute continuity of the distribution of Z it is not necessary that the distribution of Q be absolutely continuous. If, for instance, Z has an exponential distribution with parameter 1, that is, $\mathbb{E}e^{-sZ} = (1+s)^{-1}$, $s \geq 0$, then

$$\mathbb{E}e^{-sQ} = \frac{\mathbb{E}e^{-sZ}}{\mathbb{E}e^{scZ}} = \frac{1+cs}{1+s} = c\cdot 1 + (1-c)\frac{1}{1+s}, \quad s \geq 0.$$

Thus, the distribution of Q is a mixture (convex linear combination) of the atom at zero and the standard exponential distribution.

3.8. Criteria for characteristic functions

3.8.1. First orientation The following question arises often: is the given complex-valued function of real argument the characteristic function of a random variable?

Definition 171. A characteristic function φ is called *selfdecomposable* if for any $c \in (0,1)$ there exists a characteristic function φ_c such that $\varphi(t) = \varphi(ct)\varphi_c(t)$ for all $t \in \mathbb{R}$. The corresponding distribution is also called selfdecomposable.

To check selfdecomposability of the characteristic function φ one thus needs to verify that $t \mapsto \varphi(t)/\varphi(ct)$, $t \in \mathbb{R}$ is a characteristic function for each $c \in (0,1)$.

Example 172. Let us show that the characteristic function $\varphi(t) = (1-\mathrm{i}t)^{-1}$, $t \in \mathbb{R}$ of an exponential distribution with parameter 1 is selfdecomposable.

Proof. For $c \in (0,1)$,

$$\varphi_c(t) = \frac{\varphi(t)}{\varphi(ct)} = \frac{1-\mathrm{i}ct}{1-\mathrm{i}t} = c + \frac{1-c}{1-\mathrm{i}t} = c\varphi^{(0)}(t) + (1-c)\varphi(t).$$

This is a convex linear combination of φ and the characteristic function $\varphi^{(0)}$ of the degenerate at 0 distribution. By Lemma 121, φ_c is a characteristic function. □

To prove that the given (absolutely integrable) function ψ is not a characteristic function one may use the following argument. For each $x \in \mathbb{R}$, calculate the integral $p(x) := \int_{\mathbb{R}} e^{-\mathrm{i}tx}\psi(t)\mathrm{d}t$. If $p(x) < 0$ for some $x \in \mathbb{R}$, then ψ is not a characteristic function.

Example 173. Let us check that, for $\alpha \in (0,1/2)$, the function ψ_α defined by $\psi_\alpha(t) = (1-t^2)e^{-\alpha t^2}$ for $t \in \mathbb{R}$ is not a characteristic function.

Proof. Write

$$p_\alpha(x) := \int_{\mathbb{R}} e^{-\mathrm{i}tx}\psi_\alpha(t)\mathrm{d}t = \int_{\mathbb{R}} e^{-\mathrm{i}tx}e^{-\alpha t^2}\mathrm{d}t - \int_{\mathbb{R}} e^{-\mathrm{i}tx}t^2e^{-\alpha t^2}\mathrm{d}t =: I_1(x,\alpha) - I_2(x,\alpha).$$

Recalling that $\int_{\mathbb{R}} e^{-\mathrm{i}tx}e^{-t^2/2}\mathrm{d}t = \sqrt{2\pi}e^{-x^2/2}$ for $x \in \mathbb{R}$ we infer $I_1(x,\alpha) = \sqrt{\pi/\alpha}e^{-x^2/(4\alpha)}$ for $x \in \mathbb{R}$. With this at hand,

$$I_2(x,\alpha) = -\frac{\partial}{\partial\alpha}I_1(x,\alpha) = \sqrt{\frac{\pi}{\alpha}}e^{-x^2/(4\alpha)}\Big(\frac{1}{2\alpha} - \frac{x^2}{4\alpha^2}\Big), \quad x \in \mathbb{R}.$$

Thus,

$$p_\alpha(x) = \sqrt{\frac{\pi}{\alpha}}e^{-x^2/(4\alpha)}\Big(1 - \frac{1}{2\alpha} + \frac{x^2}{4\alpha^2}\Big), \quad x \in \mathbb{R}.$$

If $\alpha \in (0,1/2)$, then $p_\alpha(0) < 0$. Hence, ψ_α is not a characteristic function. □

In some situations the following result may be useful.

Lemma 174. *The only characteristic function φ satisfying $\varphi(t) = 1 + o(t^2)$ as $t \to 0$ is $\varphi(t) \equiv 1$ for all $t \in \mathbb{R}$.*

Proof. Let φ be the characteristic function of a random variable. It admits an expansion

$$\varphi(t) = 1 + t\varphi'(0) + t^2\varphi''(0)/2 + o(t^2), \quad t \to 0$$

with $\varphi'(0) = \varphi''(0) = 0$. By Corollary 133, $\mathbb{E}\xi$ and $\mathbb{E}\xi^2$ are finite. By Corollary 134, $\varphi'(0) = \mathrm{i}\mathbb{E}\xi$ and $\varphi''(0) = -\mathbb{E}\xi^2$, whence $\mathbb{E}\xi = \mathbb{E}\xi^2 = 0$. Thus, $\mathbb{P}\{\xi = 0\} = 1$ or equivalently $\varphi(t) = 1$ for all $t \in \mathbb{R}$. □

Example 175. The function $t \mapsto e^{-t^4}$ is not a characteristic function because $e^{-t^4} = 1 - t^4 + o(t^4) = 1 + o(t^2)$ as $t \to 0$.

We shall now show how Lévy's continuity theorem can be used for checking whether the given function is characteristic.

Example 176. Let ξ be a random variable with a gamma distribution with parameters $\alpha > 0$ and 1, that is, its density h is given by

$$h(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}\mathbb{1}_{(0,\infty)}(x).$$

Then the random variable $\log\xi$ has a selfdecomposable distribution.

Proof. For $p > 0$ we have

$$\mathbb{E}\xi^p = \frac{1}{\Gamma(\alpha)}\int_0^\infty x^{p+\alpha-1}e^{-x}\mathrm{d}x = \frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}.$$

It is known (see, for instance, p. 148 in [16]) that the gamma function Γ is analytic on $\{z \in \mathbb{C} : \operatorname{Re} z > 0\}$. Thus, in the formula above we can replace positive p with any complex z with $\operatorname{Re} z > -\alpha$. Taking this into account

$$\varphi(t) = \mathbb{E}e^{\mathrm{i}t\log\xi} = \mathbb{E}\xi^{\mathrm{i}t} = \frac{\Gamma(\alpha+\mathrm{i}t)}{\Gamma(\alpha)}, \quad t \in \mathbb{R}.$$

To prove that the distribution of $\log \xi$ is selfdecomposable we have to show that, for each $c \in (0,1)$, the function φ_c given by

$$\varphi_c(t) = \frac{\varphi(t)}{\varphi(ct)} = \frac{\Gamma(\alpha + \mathrm{i}t)}{\Gamma(\alpha + \mathrm{i}ct)}, \quad t \in \mathbb{R}$$

is a characteristic function. To this end, we shall use a representation of the gamma function as an infinite product (see, for instance, formula (1) on p. 150 in [16])

$$\frac{1}{\Gamma(u)} = ue^{\gamma u} \prod_{k \geq 1} \Big(1 + \frac{u}{k}\Big) e^{-\frac{u}{k}},$$

where γ is the Euler-Mascheroni constant defined by $\gamma := \lim_{m\to\infty} \Big(\sum_{i=1}^m \frac{1}{i} - \log m\Big)$. This yields

$$\varphi_c(t) = \left(c + \frac{1-c}{1+\frac{\mathrm{i}t}{\alpha}}\right) e^{\mathrm{i}\gamma t(c-1)} \prod_{k\geq 1} \left(c + \frac{1-c}{1+\frac{\mathrm{i}t}{\alpha+k}}\right) e^{\frac{\mathrm{i}t}{k}(1-c)}.$$

For $t \in \mathbb{R}$, set $\phi_0(t) = 1$ and $\phi_1(t) = (1 + \mathrm{i}t/\beta)^{-1}$, where $\beta > 0$. The function ϕ_0 is the characteristic function of a random variable θ with degenerate at 0 distribution, that is, $\mathbb{P}\{\theta = 0\} = 1$. The function ϕ_1 is the characteristic function of $(-\eta)$, where η has an exponential distribution with parameter β, that is, its density is $x \mapsto \beta e^{-\beta x} \mathbb{1}_{(0,\infty)}(x)$. The function

$$\phi_\beta(t) := c + \frac{1-c}{1+\frac{\mathrm{i}t}{\beta}} = c\phi_0(t) + (1-c)\phi_1(t)$$

is characteristic by Lemma 121 as the finite mixture of the characteristic functions ϕ_0 and ϕ_1. Thus, $\varphi_c(t) = \phi_\alpha(t) e^{\mathrm{i}\gamma t(c-1)} \prod_{k\geq 1} \phi_{\alpha+k}(t) e^{\frac{\mathrm{i}t}{k}(1-c)} := \lim_{n\to\infty} \Phi_n(t)$. The function Φ_n given by

$$\Phi_n(t) = \phi_\alpha(t) e^{\mathrm{i}\gamma t(c-1)} \prod_{k=1}^n \phi_{\alpha+k}(t) e^{\frac{\mathrm{i}t}{k}(1-c)}, \quad t \in \mathbb{R}$$

is a characteristic function as the product of finite number of characteristic functions (see Proposition 117). Here, it should be pointed out that $t \mapsto e^{\mathrm{i}\gamma t(c-1)}$ and $t \mapsto e^{\mathrm{i}t(1-c)/k}$ are characteristic functions of random variables taking values $\gamma(c-1)$ and $(1-c)/k$, respectively, with probability 1. Since φ_c is continuous at 0 it must be a characteristic function by Levy's continuity theorem (Theorem 115). □

In the next two sections we shall discuss two more fundamental classical results.

3.8.2. Bochner's theorem

Definition 177. A function $f : \mathbb{R} \to \mathbb{C}$ is called *positive semidefinite* if (a) f is continuous; (b) for any positive integer N, any real $t_1, t_2, \ldots, t_N$ and any complex $c_1, c_2, \ldots, c_N$,

$$S_N := \sum_{j=1}^N \sum_{k=1}^N f(t_j - t_k) c_j \bar{c}_k \geq 0. \tag{3.46}$$

Lemma 178. *Let f be positive semidefinite. Then*
(I) $f(0) \in \mathbb{R}$ and $f(0) \geq 0$;
(II) $f(-t) = \overline{f(t)}$ (this is called Hermitian property);
(III) $|f(t)| \leq f(0)$.

Proof. (I) Put in (3.46) $N = 1$, $t_1 = 0$ and $c_1 = 1$. Then $S_1 = f(0) \geq 0$.
(II) Put in (3.46) $N = 2$, $t_1 = 0$, $t_2 = t$ and fix any complex c_1 and c_2. Then

$$S_2 = f(0)(|c_1|^2 + |c_2|^2) + f(-t)c_1\bar{c}_2 + f(t)\bar{c}_1 c_2 \geq 0. \tag{3.47}$$

From part (I) it follows that the first summand is real. Hence, (3.47) entails that $T := f(-t)c_1\bar{c}_2 + f(t)\bar{c}_1 c_2$ is real. Write $f(-t) = \alpha_1 + \mathrm{i}\beta_1$, $f(t) = \alpha_2 + \mathrm{i}\beta_2$, $c_1\bar{c}_2 = \gamma + \mathrm{i}\delta$ for some real α_i, β_i and any real γ and δ. Since $T = (\alpha_1 + \mathrm{i}\beta_1)(\gamma + \mathrm{i}\delta) + (\alpha_2 + \mathrm{i}\beta_2)(\gamma - \mathrm{i}\delta) \in \mathbb{R}$, we must have $(\beta_1 + \beta_2)\gamma + (\alpha_1 - \alpha_2)\delta = 0$ for any real γ and δ. This is only possible if $\beta_1 + \beta_2 = \alpha_1 - \alpha_2 = 0$. Thus, indeed, $f(-t) = \overline{f(t)}$.
(III) Put in (3.46) $N = 2$, $t_1 = 0$, $t_2 = t$ and $c_1 = f(t)$, $c_2 = -|f(t)|$. Using (3.47) and $f(t)f(-t) = f(t)\overline{f(t)} = |f(t)|^2$ which is a consequence of (II) we obtain

$$S_2 = 2f(0)|f(t)|^2 - 2|f(t)|^3 \geq 0.$$

Thus, if $|f(t)| > 0$, we infer $|f(t)| \leq f(0)$. If $|f(t)| = 0$, then the inequality holds automatically in view of part (I). The proof of the lemma is complete. □

Theorem 179 is called *Bochner's theorem.*

Theorem 179. *A function $f : \mathbb{R} \to \mathbb{C}$ is characteristic function of some random variable if, and only if, f is positive semidefinite and $f(0) = 1$.*

For the proof we need an auxiliary result which is called the *second Helly theorem.*

Theorem 180. *Let $g : \mathbb{R} \to \mathbb{R}$ be a continuous function and $(G_n)_{n\geq 1}$ the sequence of uniformly bounded and nondecreasing functions which weakly converge to a function G on a finite interval $[a, b]$. Then*

$$\lim_{k\to\infty} \int_a^b g(x)\mathrm{d}G_k(x) = \int_a^b g(x)\mathrm{d}G(x).$$

Proof of Theorem 179. Let f be a characteristic function which corresponds to a distribution function F. It is obvious that $f(0) = 1$. By Lemma 109 (e), f is continuous. Further, we write, for any positive integer N, any real $t_1, t_2, \ldots, t_N$ and any complex $c_1, c_2, \ldots, c_N$,

$$S_N = \sum_{j=1}^N \sum_{k=1}^N c_j\bar{c}_k \int_{\mathbb{R}} e^{\mathrm{i}(t_j - t_k)x}\mathrm{d}F(x) = \int_{\mathbb{R}} \left(\sum_{j=1}^N c_j e^{\mathrm{i}t_j x}\right) \left(\overline{\sum_{k=1}^N c_k e^{\mathrm{i}t_k x}}\right) \mathrm{d}F(x)$$

$$= \int_{\mathbb{R}} \left|\sum_{j=1}^N c_j e^{\mathrm{i}t_j x}\right|^2 \mathrm{d}F(x) \geq 0.$$

Thus, f is positive semidefinite.

Assume now that f is a positive semidefinite function and $f(0) = 1$. Let n and N be positive integers and $x \in \mathbb{R}$. Put in (3.46) $t_j = j/n$, $c_j = e^{-\mathrm{i}jx}$ for $j = 0, 1, \dots, N-1$. Then

$$S_N^n(x) = \frac{1}{N}\sum_{j=1}^{N}\sum_{k=1}^{N} f\left(\frac{j-k}{n}\right) e^{-\mathrm{i}(j-k)x} \geq 0, \quad x \in \mathbb{R}.$$

The difference $j-k$ takes value r (r is an integer ranging between $-N+1$ and $N-1$) in $N - |r|$ summands. Collecting these summands together we obtain

$$S_N^n(x) = \sum_{r=-N}^{N}\left(1 - \frac{|r|}{N}\right) f\left(\frac{r}{n}\right) e^{-\mathrm{i}rx} \geq 0, \quad x \in \mathbb{R}. \tag{3.48}$$

Further, for each integer $s \in [-N, N]$,

$$\int_{-\pi}^{\pi} e^{\mathrm{i}sx} S_N^n(x)\mathrm{d}x = \sum_{r=-N}^{N}\left(1 - \frac{|r|}{N}\right) f\left(\frac{r}{n}\right)\int_{-\pi}^{\pi} e^{\mathrm{i}(s-r)x}\mathrm{d}x = 2\pi\left(1 - \frac{|s|}{N}\right) f\left(\frac{s}{n}\right),$$

where we have used (see (3.29))

$$\int_{-\pi}^{\pi} e^{\mathrm{i}(s-r)x}\mathrm{d}x = \begin{cases} 2\pi & \text{if } r = s, \\ 0, & \text{if } r \neq s. \end{cases}$$

Put $F_N^{(n)}(x) = \begin{cases} 0 & \text{if } x \in (-\infty, -\pi], \\ \frac{1}{2\pi}\int_{-\pi}^{x} S_N^n(y)\mathrm{d}y & \text{if } x \in [-\pi, \pi), \\ 1 & \text{if } x \in [\pi, \infty), \end{cases}$ so that $\mathrm{d}F_N^{(n)}(x) = (1/2\pi)S_N^n(x)\mathrm{d}x$, whence

$$\left(1 - \frac{|s|}{N}\right) f\left(\frac{s}{n}\right) = \int_{-\pi}^{\pi} e^{\mathrm{i}sx}\mathrm{d}F_N^{(n)}(x). \tag{3.49}$$

From (3.48) and the definition of $F_N^{(n)}$ it follows that this function is nondecreasing and right-continuous. Using (3.49) and the assumption $f(0) = 1$ gives $\int_{\mathbb{R}} \mathrm{d}F_N^{(n)}(x) = f(0) = 1$. Thus, $F_N^{(n)}$ is a distribution function. With n fixed, consider the sequence of distribution functions $(F_N^{(n)})_{N\geq 1}$. The selection principle tells us that each sequence of distribution functions contains a convergent subsequence. Therefore, there exists a subsequence $(F_{N_k}^{(n)})_{k\geq 1}$ converging weakly as $k \to \infty$ to a nondecreasing function $F^{(n)}$. Furthermore, $F^{(n)}$ is actually a distribution function because, for any positive integer N and any $\varepsilon > 0$, $F_N^{(n)}(-\pi - \varepsilon) = 0$ and $F_N^{(n)}(\pi + \varepsilon) = 1$. Further, (3.49) entails

$$f\left(\frac{s}{n}\right) = \lim_{k\to\infty}\left(1 - \frac{|s|}{N_k}\right) f\left(\frac{s}{n}\right) = \lim_{k\to\infty}\int_{-\pi}^{\pi} e^{\mathrm{i}sx}\mathrm{d}F_{N_k}^{(n)}(x)$$

and thereupon, by Theorem 180,

$$f\left(\frac{s}{n}\right) = \int_{-\pi}^{\pi} e^{\mathrm{i}sx}\mathrm{d}F^{(n)}(x) \tag{3.50}$$

for all integer s.

The distribution function F_n defined by $F_n(x) = F^{(n)}(x/n)$ for $x \in \mathbb{R}$ has the characteristic function φ_n given by

$$\varphi_n(t) = \int_{-\pi n}^{\pi n} e^{\mathrm{i}tx} \mathrm{d}F_n(x) = \int_{-\pi}^{\pi} e^{\mathrm{i}tnx} \mathrm{d}F^{(n)}(x), \quad t \in \mathbb{R}.$$

Using (3.50) we infer

$$\varphi_n\left(\frac{k}{n}\right) = f\left(\frac{k}{n}\right) \tag{3.51}$$

for any integer k. For any fixed real t there exist integer $k = k(t, n)$ satisfying

$$0 \le t - \frac{k}{n} \le \frac{1}{n}. \tag{3.52}$$

Put $\theta = t - k/n$, so that $0 \le \theta \le 1/n$. Further,

$$\begin{aligned} |\varphi_n(t) - \varphi_n(k/n)| &= \left| \int_{-\pi n}^{\pi n} e^{\mathrm{i}xk/n}(e^{\mathrm{i}\theta x} - 1)\mathrm{d}F_n(x) \right| \le \int_{-\pi n}^{\pi n} \left|e^{\mathrm{i}\theta x} - 1\right| \mathrm{d}F_n(x) \\ &= 2\int_{-\pi n}^{\pi n} |\sin(\theta x/2)| \mathrm{d}F_n(x) \le \Big(4\int_{-\pi n}^{\pi n} \sin^2(\theta x/2)\mathrm{d}F_n(x)\Big)^{1/2} \Big(\int_{-\pi n}^{\pi n} \mathrm{d}F_n(x)\Big)^{1/2} \\ &= \Big(2\int_{-\pi n}^{\pi n} (1 - \cos(\theta x)\mathrm{d}F_n(x)\Big)^{1/2} = \Big(2\int_{-\pi}^{\pi} (1 - \cos(n\theta x))\mathrm{d}F^{(n)}(x)\Big)^{1/2}. \end{aligned}$$

Since $n\theta \le 1$ guarantees $1 - \cos(n\theta z) \le 1 - \cos z$ for $|z| \le \pi$ we conclude that

$$\begin{aligned} |\varphi_n(t) - \varphi_n(k/n)| \le \Big(2\int_{-\pi}^{\pi} (1 - \cos x)\mathrm{d}F^{(n)}(x)\Big)^{1/2} &= (2\mathrm{Re}\,(1 - \varphi_n(1/n))^{1/2} \\ &= (2\mathrm{Re}\,(1 - f(1/n)))^{1/2}, \end{aligned}$$

where the last equality follows from (3.51). By assumption, f is continuous and $f(0) = 1$. Therefore, the right-hand side converges to 0 as $n \to \infty$. This proves that

$$\lim_{n\to\infty} |\varphi_n(t) - \varphi_n(k/n)| = 0. \tag{3.53}$$

Also, in view of (3.51) and (3.52),

$$f(t) = \lim_{n\to\infty} f(k/n) = \lim_{n\to\infty} \varphi_n(k/n). \tag{3.54}$$

A combination of (3.53) and (3.54) gives

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} ((\varphi_n(t) - \varphi_n(k/n)) + \varphi_n(k/n)) = f(t),$$

thereby showing that a continuous function f is the limit of the sequence of characteristic functions $(\varphi_n)_{n\ge 1}$. By Lévy's continuity theorem for characteristic functions (Theorem 115 (b)), f is a characteristic function. The proof of the theorem is complete. □

3.8.3. Polya's conditions In this section we provide easy-to-verify conditions due to Polya which are sufficient for the given function to be characteristic. Theorem **??** is called *Polya's theorem*.

Theorem 181. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be an even function which is convex for $t > 0$ and satisfies $\varphi(0) = 1$, and $\lim_{t\to\infty} \varphi(t) = 0$. Then φ is the characteristic function of an absolutely continuous distribution.*

For the proof we need the following *Pringsheim theorem.*

Proposition 182. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function which is nonincreasing on $(0, \infty)$, Lebesgue integrable on $(0, a)$ for any $a > 0$ and satisfies $\lim_{t\to\infty} f(t) = 0$. Then*

$$\frac{1}{2}(f(t+0) + f(t-0)) = \frac{2}{\pi}\int_0^\infty \cos(tu) \int_0^\infty f(y)\cos(uy)\mathrm{d}y\mathrm{d}u, \quad t > 0. \tag{3.55}$$

Proof of Theorem 181. Being convex, the function φ has a nondecreasing right derivative φ' on $(0, \infty)$. Further, the assumption $\lim_{t\to\infty} \varphi(t) = 0$ ensures that $\varphi'(t) \le 0$ for $t > 0$. While the integral $\int_{[-1,1]} e^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t$ trivially converges for all $x \in \mathbb{R}$, the integral $\int_{\mathbb{R}\setminus[-1,1]} e^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t$ converges for all $x \ne 0$ by Dirichlet's test. Put

$$p(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\mathrm{i}tx}\varphi(t)\mathrm{d}t, \quad x \ne 0. \tag{3.56}$$

Since φ is even, we conclude that

$$p(x) = \frac{1}{\pi}\int_0^\infty \cos(tx)\varphi(t)\mathrm{d}t, \quad x \ne 0. \tag{3.57}$$

In Proposition 182, take $f = \varphi$. Then all the assumptions of Proposition 182 are satisfied. The left-hand side of (3.55) is equal to φ, for φ, being convex, is continuous. Using (3.57) we obtain from (3.55) that $\varphi(t) = 2\int_0^\infty \cos(tu)p(u)\mathrm{d}u$ for $t > 0$. Since φ is even the last equality actually holds for $t \ne 0$ or, equivalently,

$$\varphi(t) = \int_{\mathbb{R}} e^{\mathrm{i}tx}p(x)\mathrm{d}x, \quad t \ne 0. \tag{3.58}$$

By assumption, $\varphi(0) = 1$. Hence, using (3.58) gives $\int_{\mathbb{R}} p(x)\mathrm{d}x = 1$. If we can show that p is a nonnegative function, then p is the density of some distribution which completes the proof. Put $\phi(t) = -\varphi'(t)$ and note that ϕ is a nonnegative function which is nondecreasing for $t > 0$. Integrating by parts in (3.57) yields

$$p(x) = \frac{1}{\pi x}\varphi(t)\sin(tx)|_0^\infty + \frac{1}{\pi x}\int_0^\infty \phi(t)\sin(tx)\mathrm{d}t = \frac{1}{\pi x}\int_0^\infty \phi(t)\sin(tx)\mathrm{d}t, \quad x \ne 0.$$

Further,

$$p(x) = \frac{1}{\pi x}\sum_{j\ge 0}\int_{j\pi/x}^{(j+1)\pi/x} \phi(t)\sin(tx)\mathrm{d}x = \frac{1}{\pi x}\int_0^{\pi/x}\sum_{j\ge 0}(-1)^j\phi(t + j\pi/x)\sin(tx)\mathrm{d}t, \quad x \ne 0,$$

where for the last equality we have changed the variable $u = t - j\pi/x$. Let $x > 0$. The series $\sum_{j\ge 0}(-1)^j\phi(t + j\pi/x)$ is alternating, and the absolute value of its general term is nonincreasing. Since the first summand of the series is nonnegative, so are the sum of the series and the last integrand. Hence, $p(x) \ge 0$ for $x > 0$. From (3.57) it follows that p is an even function, whence $p(x) \ge 0$ for $x \ne 0$. Thus, p is a density of some distribution, and φ is the corresponding characteristic function. The proof of the theorem is complete. □

Sometimes characteristic functions having properties as in Theorem 181 are said to belong to *Polya's class.* From the proof of Theorem 181 it is clear that the corresponding densities are given by formulae (3.56) or (3.57). Let us give several examples of characteristic functions from Polya's class.

Example 183.

$$\varphi(t) = e^{-|t|}, \quad t \in \mathbb{R};$$

$$\varphi(t) = \begin{cases} 1 - |t|, & \text{if } t \in [-1/2, 1/2], \\ \frac{1}{4t}, & \text{if } t \in \mathbb{R} \backslash [-1/2, 1/2]; \end{cases} \tag{3.59}$$

$$\varphi(t) = \frac{1}{1 + |t|}, \quad t \in \mathbb{R}; \tag{3.60}$$

$$\varphi(t) = (1 - |t|)\mathbb{1}_{(-1,1)}(t), \quad t \in \mathbb{R}. \tag{3.61}$$

Remark 184. According to Theorem 162, absolute integrability of a characteristic function is sufficient for absolute continuity of the corresponding distribution. The characteristic functions defined in (3.59) and (3.60) demonstrate that this condition is not necessary. A similar example can be found in Remark 163.

Remark 185. Remark 114 provides an example of two characteristic functions which coincide on a finite interval, yet correspond to different distributions. Another example of this kind is given by the characteristic functions defined in (3.59) and (3.61).

Remark 186. Let φ be a characteristic function belonging to Polya's class and having a right derivative which is strictly increasing on the positive halfline. On the graph of φ replace any piece located to the right of 0 with a chord, obtaining in this way the graph of a new function φ_1. For $t < 0$, put $\varphi_1(t) = \varphi_1(-t)$. The function φ_1 is a characteristic function belonging to Polya's class. An interesting feature is that $\varphi_1 = \varphi$ everywhere except an interval symmetric about 0. According to the uniqueness theorem for characteristic functions (Theorem 113, see also Remark 114), the characteristic functions φ_1 and φ correspond to different distributions.

We close this section by an application of theoretical material presented above.

Example 187. For $\alpha \in (0, 2]$, φ_α defined by

$$\varphi_\alpha(t) = e^{-|t|^\alpha}, \quad t \in \mathbb{R}$$

is a characteristic function.

Proof. Case $\alpha \in (0, 1]$. For $t > 0$, we have $\varphi'_\alpha(t) = -\alpha t^{\alpha-1} e^{-t^\alpha}$. This function is increasing, hence φ_α is convex on $(0, \infty)$. The other conditions of Theorem 181 trivially hold, whence φ_α is a characteristic function belonging to Polya's class.
Case $\alpha = 2$. We know that $t \mapsto e^{-t^2/2}$, $t \in \mathbb{R}$ is the characteristic function of ξ a random variable having a standard normal distribution. Then the random variable $\sqrt{2}\xi$ has the characteristic function φ_2.

CASE $\alpha \in (1,2)$. Put

$$p_\alpha(x) := \frac{\alpha}{2|x|^{\alpha+1}} \mathbb{1}_{\mathbb{R}\setminus[-1,1]}(x), \quad x \in \mathbb{R}.$$

It is easily seen that $\int_{\mathbb{R}} p_\alpha(x)\mathrm{d}x = 1$, whence p_α is a density of some distribution. Let ψ_α be the corresponding characteristic function, that is, $\psi_\alpha(t) = \int_{\mathbb{R}} e^{itx} p_\alpha(x)\mathrm{d}x$ for $t \in \mathbb{R}$. We have, for $t \in \mathbb{R}$,

$$1 - \psi_\alpha(t) = \int_{\mathbb{R}} (1 - e^{itx}) p_\alpha(x)\mathrm{d}x = \int_{\mathbb{R}} (1 - \cos(tx)) p_\alpha(x)\mathrm{d}x = \alpha \int_1^\infty (1 - \cos(tx)) x^{-\alpha-1}\mathrm{d}x$$

$$= \alpha|t|^\alpha \int_{|t|}^\infty (1-\cos y) y^{-\alpha-1}\mathrm{d}y = \alpha|t|^\alpha \Big(\int_0^\infty (1-\cos y) y^{-\alpha-1}\mathrm{d}y - \int_0^{|t|} (1-\cos y) y^{-\alpha-1}\mathrm{d}y \Big).$$

Put

$$c_\alpha := \int_0^\infty (1 - \cos y) y^{-\alpha-1}\mathrm{d}y.$$

The function $y \mapsto (1 - \cos y) y^{-\alpha-1}$ is integrable at ∞ because $(1 - \cos y) y^{-\alpha-1} \le y^{-\alpha-1}$ for $y > 0$ and the function on the right-hand side is integrable at ∞. Also, the function is integrable at 0 because $(1 - \cos y) y^{-\alpha-1} \sim 2^{-1} y^{-\alpha+1}$ as $y \to 0+$, and the function on the right-hand side is integrable at 0 (recall that $\alpha \in (1,2)$). Arguing similarly but invoking additionally L'Hospital's rule we infer

$$\int_0^{|t|} (1 - \cos y) y^{-\alpha-1}\mathrm{d}y \sim (2(2-\alpha))^{-1} |t|^{2-\alpha}, \quad t \to 0.$$

Combining pieces together we arrive at

$$\psi_\alpha(t) = 1 - \alpha c_\alpha |t|^\alpha + O(t^2), \quad t \to 0. \tag{3.62}$$

Let $\eta_1, \eta_2, \ldots$ be independent identically distributed random variables with the density p_α. By Proposition 117, the characteristic function of $\eta_1 + \ldots + \eta_n$ is $t \mapsto (\psi_\alpha(t))^n$ for $t \in \mathbb{R}$, and by Lemma 109(d), the characteristic function of $n^{-1/\alpha}(\eta_1 + \ldots + \eta_n)$ is $t \mapsto (\psi_\alpha(n^{-1/\alpha} t))^n$ for $t \in \mathbb{R}$. Using (3.62) we now infer

$$(\psi_\alpha(n^{-1/\alpha} t))^n = (1 - \alpha c_\alpha n^{-1} |t|^\alpha + O(n^{-1/\alpha} t))^n \to e^{-\alpha c_\alpha |t|^\alpha}, \quad n \to \infty.$$

The function $t \mapsto e^{-\alpha c_\alpha |t|^\alpha}$ is continuous on $\mathbb{R}$. Therefore, by Lévy's continuity theorem for characteristic functions (Theorem 115(b)) it is a characteristic function as a continuous limit of the sequence of characteristic functions. Finally, φ_α is a characteristic function because if φ is a characteristic function, so is $t \mapsto \varphi(at)$, $t \in \mathbb{R}$ for any $a \in \mathbb{R}$. □

3.9. Problems

Problem 188. Let ξ be a random variable with characteristic function φ. Prove the following. If, for some $\delta > 0$,

$$\int_0^\delta \frac{|1 - \varphi(t)|}{t} \mathrm{d}t < \infty, \tag{3.63}$$

then $\mathbb{E} \log^+ |\xi| < \infty$, where $\log^+ x = \log x$ if $x \ge 1$ and $= 0$ if $x \in (0,1)$. Conversely, if $\mathbb{E} \log^+ |\xi| < \infty$, then (3.63) holds for any $\delta > 0$.

Problem 189. (a) Prove that

$$h(x) = \frac{1-\cos x}{\pi x^2}, \quad x \in \mathbb{R}$$

is the density of a random variable with the characteristic function

$$\varphi(t) = (1-|t|)\mathbb{1}_{(-1,1)}(t).$$

(b) Let φ_1 be the periodic continuation with period 2 of the function $t \mapsto (1-|t|)$, $t \in [-1,1]$. Expanding φ_1 into Fourier's series show that φ_1 is a characteristic function and find the corresponding distribution.

Problem 190. Prove that

$$h(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad x \in \mathbb{R}$$

is the density of a random variable with the characteristic function

$$\varphi(t) = e^{-|t|}, \quad t \in \mathbb{R}$$

(the distribution of ξ is called *Cauchy distribution*).

Problem 191. Let $p \in (0,1)$. Give an example of characteristic function φ such that $\varphi(t) = p$ for $t \in [a,b]$ for some finite $a < b$.

HINT: Use Lemma 121 with $n = 2$ and appropriate characteristic functions φ_1 and φ_2.

Problem 192. Let $\varphi_1, \varphi_2, \dots$ be characteristic functions of random variables and $\alpha_1, \alpha_2, \dots$ be nonnegative numbers satisfying $\sum_{n\geq 1} \alpha_n = 1$. Use Lemma 121 in combination with Levy's continuity theorem for characteristic functions (Theorem 115) to show that $\sum_{n\geq 1} \alpha_n\varphi_n$ is a characteristic function.

Problem 193. Prove that if the characteristic function of a random variable with a symmetric distribution is convex on the positive halfline, then the first absolute moment of the random variable is infinite.

Problem 194. Prove that

$$\varphi_1(t) = \frac{2}{e^t + e^{-t}} \quad \text{and} \quad \varphi_2(t) = \frac{2\pi t}{e^t - e^{-t}}, \quad t \in \mathbb{R}$$

are selfdecomposable characteristic functions (see Definition 171).

HINT: Use the following representations

$$\cos x = \prod_{n\geq 1}\Big(1 - \frac{4x^2}{\pi^2(2n-1)^2}\Big) \quad \text{and} \quad \frac{\sin x}{x} = \prod_{n\geq 1}\Big(1 - \frac{x^2}{\pi^2 n^2}\Big), \quad x \in \mathbb{C}.$$

Problem 195. Let ξ be a random variable with a *geometrically stable distribution*, that is, its characteristic function φ is given by

$$\varphi(t) = \frac{1}{1 + \lambda|t|^\alpha\omega(t,\alpha,\beta) - \mathrm{i}\mu t}, \quad t \in \mathbb{R},$$

where $\alpha \in (0,2]$, $\beta \in [-1,1]$, $\mu \in \mathbb{R}$, $\lambda \geq 0$ and

$$\omega(t,\alpha,\beta) = \begin{cases} 1 - \mathrm{i}\beta \tan(\pi\alpha/2) \operatorname{sgn} t, & \text{if } \alpha \neq 1, \\ 1 + \mathrm{i}\beta(2/\pi) \log|t| \operatorname{sgn} t, & \text{if } \alpha = 1. \end{cases}$$

(a) Prove that $\mathbb{E}|\xi|^\gamma < \infty$ if, and only if, $\gamma \in (0,\alpha)$.

(b) Prove that φ is a selfdecomposable characteristic function (see Definition 171) if either $\alpha = 2$, or $\alpha \in (0,1) \cup (1,2)$ and $\mu = 0$, or $\alpha = 1$ and $\mu = \beta = 0$.

Problem 196. Let ξ be a random variable with characteristic function φ given by

$$\varphi(t) = \exp\Big(\sum_{n\in\mathbb{Z}} (\cos(b^n t) - 1) k_n\Big), \quad t \in \mathbb{R},$$

where $b \geq 2$ is integer, $k_n \geq 0$ for $n \in \mathbb{Z}$ and $\sup_{n\in\mathbb{Z}} k_n < \infty$. Is the distribution of ξ absolutely continuous?

Problem 197. For each $n \in \mathbb{N}$, let F_n be an absolutely continuous distribution function and $\alpha_n \geq 0$. Assume that $\sum_{n\geq 1} \alpha_n = 1$. Prove that $\sum_{n\geq 1} \alpha_n F_n$ is an absolutely continuous distribution function.

LITERATURE

[1] G. G. Bilodeau. The origin and early development of non-analytic infinitely differentiable functions. Archive for History of Exact Sciences. **27** (1982), 115–135.

[2] R. Durrett, *Probability: theory and examples*, 4th edition. Cambridge University Press, 2010.

[3] W. Feller, *An introduction to probability theory and its applications*, Vol. 2. John Wiley and Sons, 1971.

[4] J. Gil-Pelaez, Note on the inversion theorem. Biometrika. **38** (1951), 481–482.

[5] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 5th edition. Oxford University Press, 1979.

[6] Yu. V. Linnik and I. V. Ostrovskii, *Decomposition of random variables and vectors*. American Mathematical Society, 1977.

[7] E. Lukacs, *Characteristic functions*, 2nd edition. Charles Griffin & Company, 1970.

[8] I. P. Natanson, *Theory of functions of a real variable*, Vol. 1, 5th printing. Frederick Ungar Publishing Co., 1983.

[9] B. Ramachandran, *Advanced theory of characteristic functions*. Statistical Publishing Society, 1967.

[10] B. Ramachandran, Characteristic functions taking constant values on intervals of the real line. Statist. Probab. Letters. **28** (1996), 269–270.

[11] S. I. Resnick, *Adventures in stochastic processes*, 4rd printing. Birkhäuser, 2005.

[12] B. Rosén, On the asymptotic distribution of sums of independent identically distributed random variables. Arkiv för matematik. **4** (1962), 323–332.

[13] K. Sato, *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

[14] R. L. Schilling, R. Song and Z. Vondraček, *Bernstein functions. Theory and applications*, 2nd revised and extended edition. de Gruyter, 2012.

[15] J. Stoyanov, *Counterexamples in probability*, 2nd edition. John Wiley and Sons, 1997.

[16] E. C. Titchmarsh, *The theory of functions*, 2nd edition. Oxford university press, 1939.

[17] N. G. Ushakov, *Selected topics in characteristic functions*. VSP: Utrecht, 1999.

[18] D. V. Widder, *The Laplace transform*. Princeton University Press, 1941.