REMARKS ON CHARACTERISTIC FUNCTIONS

G. PÓLYA

STANFORD UNIVERSITY

Introduction

This short paper consists of two parts which have little in common except that in both we discuss characteristic functions of one-dimensional probability distributions. In the first part we consider characteristic functions of a certain special type whose principal merit lies in the fact that it is easily recognizable. In the second part we deal with finite distributions (contained in a certain finite interval) and with finitely different distributions (coinciding outside a certain finite interval).

The notation follows that of Cramér's well-known tract.[1] A distribution function is denoted by a capital letter, as $F(x)$, and the corresponding characteristic function by the corresponding small letter, as $f(t)$. Thus

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\, dF(x)\,. \tag{1}$$

$F(x)$ is real-valued, never decreasing; $F(-\infty) = 0$; $F(\infty) = 1$. Therefore

$$f(0) = 1\,. \tag{2}$$

Moreover $f(t)$ is continuous for all real values of t and has the properties

$$|f(t)| \leqq 1\,, \tag{3}$$

$$f(-t) = \overline{f(t)}\,, \tag{4}$$

that is, $f(-t)$ and $f(t)$ are conjugate complex.

I. A Simple Type of Characteristic Functions

1. A Sufficient Condition for Characteristic Functions

We are given a function, defined for all real values of t; is it the characteristic function of some probability distribution? This question is often important but not often easy to answer. The properties mentioned in the introduction [relations (2), (3), (4), and continuity] constitute simple *necessary* conditions that a given function $f(t)$ should be characteristic. Yet these conditions, taken together, are far from being sufficient. *Necessary and sufficient*

[1] Cramér [3]. Boldface numbers in brackets refer to references at the end of the paper (see p. 123).

conditions have been given,[2] but they are not readily applicable, in which they are similar to any perfectly general necessary and sufficient condition for the convergence of a series. In theorem 1 below we state a simple *sufficient* condition.

THEOREM 1. *A function* f(t), *defined for all real values of the variable* t, *has the following properties:*

$$\text{(1.1)} \qquad f(t) \textit{ is real-valued and continuous,}$$

$$\text{(1.2)} \qquad f(0) = 1\,,$$

$$\text{(1.3)} \qquad \lim_{t\to\infty} f(t) = 0\,,$$

$$\text{(1.4)} \qquad f(-t) = f(t)\,,$$

$$\text{(1.5)} \qquad f(t) \textit{ is convex for } t > 0.$$[3]

Such a function f(t) *is a characteristic function corresponding to a continuous distribution function* F(x) *whose derivative* F′(x), *the probability density, exists, is an even function, and is continuous everywhere except possibly at the point* x = 0.

The essential point of the proof is to show that

$$\text{(1.6)} \qquad \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx} f(t)dt = \frac{1}{\pi}\int_{0}^{\infty} f(t)\cos xt\,dt$$

is *never negative* for $x > 0$. For, under the given conditions (1.3), (1.4), and (1.5), the integral (1.6) visibly converges for any real x different from zero and represents an even function which we may call $F'(x)$. We obtain, then, by Fourier's theorem,

$$\text{(1.7)} \qquad f(t) = \int_{-\infty}^{\infty} e^{itx} F'(x)dx\,,$$

which proves (1). The rest is easy, *if* we know that $F'(x) \geqq 0$.

In order to show the essential point, let us observe that $f'(t)$, defined as the *right-hand* derivative of the convex function $f(t)$, exists and never decreases when $t > 0$. (We could have considered just as well the left-hand derivative.) It follows from (1.3) that $f'(t)$ tends to zero as $t \to \infty$ and $f'(t) \leqq 0$ for $t > 0$. We define

$$\text{(1.8)} \qquad \varphi(t) = -f'(t)\,;$$

[2] The origin of these criteria is Toeplitz's discovery of inequalities, characterizing positive periodical functions, for the Fourier coefficients. By analogy, the positivity of a function defined for all real values of the variables is characterized by inequalities for the Fourier transform; this observation is due to Mathias [8]. For the complete statement of the criteria see Bochner [1], p. 76, and [2], pp. 408–409; also Khintchine [6].

[3] The formal definition of convex functions and the proofs for their intuitive properties used in the sequel are given by Hardy *et al.* [5], pp. 70–72 and pp. 91–96.

$\varphi(t)$ is non-negative and never increasing for $t > 0$, and tends to 0 as $t \to \infty$. Integrating by parts and using (1.3), we obtain that

(1.9)

$$x\int_0^\infty f(t)\cos xt\,dt = -\int_0^\infty f'(t)\sin xt\,dt = \int_0^\infty \varphi(t)\sin xt\,dt$$

$$= \int_0^{\pi/x}\left[\varphi(t) - \varphi\left(t+\frac{\pi}{x}\right) + \varphi\left(t+\frac{2\pi}{x}\right) - \varphi\left(t+\frac{3\pi}{x}\right) + \cdots\right]\sin xt\,dt \geqq 0$$

if $x > 0$. In fact, the series under the last integral sign is alternating, and its terms decrease steadily in absolute value, by the properties of $\varphi(t)$ just mentioned. Thus $F'(x)$, represented by (1.6), is never negative by (1.9), and theorem 1 is proved.[4]

2. Remarks

I. If a function possesses the properties stated in the hypothesis of theorem 1, its graph is symmetrical with respect to the y-axis. The right-hand half of this graph is convex from below, falls steadily, and approaches the x-axis, which is its asymptote. These properties of the curve are immediately visible in many simple cases, as, for example,

(2.1) $$f(t) = e^{-|t|},$$

(2.2) $$f(t) = 1/(1 + |t|),$$

(2.3) $$f(t) = \begin{cases} 1 - |t| & \text{when} \quad |t| \leqq 1, \\ 0 & \text{when} \quad |t| \geqq 1. \end{cases}$$

Therefore these functions are characteristic functions.

II. If two characteristic functions are identical, the corresponding probability distributions are also identical. We may call this well-known proposition the *uniqueness theorem* in the theory of characteristic functions.[5] Could we improve the uniqueness theorem by weakening its hypothesis? Yes, in a trivial way. If two characteristic functions take the same values in a set of points everywhere dense in the infinite interval $t > 0$, the corresponding probability distributions are necessarily equal. In fact, it follows from continuity and (4) that the characteristic functions considered are identical. Yet no further weakening of the condition is possible. If the set of points in which two characteristic functions are supposed to coincide is not everywhere dense in $t > 0$,

[4] Compare the present author's paper [11], p. 378, theorem VII. The present theorem 1 is not formally stated there, or in his later paper [12], p. 104, but has been in fact proved and even applied in establishing an essential point.

[5] See, for example, Cramér [3], p. 27, last lines. The present author is not aware of any explicit statement of the uniqueness theorem or of a completely proved application of it to probability prior to that given in his paper [12], p. 105. This, however, appears to have been forgotten, just as have the remarks in the same paper on pages 106–108.

the functions may be actually different. In fact, take any function of the special type described in theorem 1 whose derivative $f'(x)$ is continuous and strictly increasing for $t > 0$, for instance, the function (2.1) or (2.2) [but not (2.3)]. Replace an arbitrary small arc of the right-hand half of the curve by its chord and change the left-hand half symmetrically. By theorem 1, the graph so obtained, containing two rectilineal stretches, still represents a characteristic function, differing from the former only in two symmetrical, arbitrarily small intervals, and belonging to a different probability distribution.[6]

III. We take for the moment, from now on till the end of the present section, the notation $f(t)$ as defined by (2.3). Let $f_1(t)$ and $f_2(t)$ denote two different characteristic functions which, however, take the same values in the interval $-1 \leqq t \leqq 1$. The existence of such functions has just been shown in the preceding paragraph. Then

$$f_1(t)f(t) = f_2(t)f(t) \tag{2.4}$$

for all real values of t. In going back to the corresponding probability distributions, we obtain two *different* distributions [derived from $f_1(t)$ and $f_2(t)$] which, combined with the same third [derived from $f(t)$], give the same convolution. Or in other terms, knowing the distribution of the sum of two random variables and also the distribution of one of these variables, we may be unable to determine the distribution of the other, for the good reason that it is utterly indeterminate.[7]

II. Characterizing Finite, and Finitely Different, Probability Distributions

3. Finite Distributions

If the distribution function $F(x)$ satisfies the condition

$$F(x)\begin{cases} = 0 & \text{for} \quad x < -h' \\ > 0 & \text{for} \quad x > -h' \\ < 1 & \text{for} \quad x < h \\ = 1 & \text{for} \quad x > h \end{cases} \tag{3.1}$$

we say that the probability distribution is *finite*, contained in the finite interval $-h' \leqq x \leqq h$ but in no smaller interval. The numbers h and h' must satisfy the inequality

$$-h' \leqq h, \tag{3.2}$$

[6] This remark shows that in Cramér [3] theorem 11, p. 29, is erroneous. See also page 121 of the same work, and Lévy [7], p. 49. This error has been recognized before. The first counter-example appears to be due to Gnedenko [4].

[7] Observed by Khintchine [6], in connection with the example given by Gnedenko [4]; see also Lévy [7], p. 190. As mentioned before (footnote 4), the reasoning given in [12], p. 104, can be regarded as another application of theorem 1.

but any one of them can be negative. We wish to call h, for the moment, the *right extremity* of the distribution described by $F(x)$, and denote it briefly thus: rext $[F]$. Similarly $-h'$ is to be denoted by lext $[F]$.

THEOREM 2. *A necessary and sufficient condition that a probability distribution should be finite is that the definition of the characteristic function* f(t) *can be extended to complex values of the variable and this extension shows that* f(t) *is an entire function of exponential type. Moreover, if the distribution function is denoted by* F(x),

$$(3.3)\quad \overline{\lim_{r\to+\infty}}\, r^{-1}\log|f(-ir)| = \text{rext}\,[F], \qquad -\overline{\lim_{r\to+\infty}}\, r^{-1}\log|f(ir)| = \text{lext}\,[F].$$

The proof of this theorem consists of two parts. First we start from $F(x)$ and work toward $f(t)$. Then we start from $f(t)$ and work toward $F(x)$. The first part is easier and will give us an opportunity to recall the definitions of some function-theoretic concepts involved in the statement of theorem 2.

I. We are given a distribution function $F(x)$ satisfying (3.1). Thus

$$(3.4)\qquad h = \text{rext}\,[F], \qquad -h' = \text{lext}\,[F].$$

We define $f(t)$ by (1) which, by virtue of (3.4), reduces to

$$(3.5)\qquad f(t) = \int_{-h'}^{h} e^{itx}\,dF(x).$$

This integral exists and represents a function analytic for all complex values of t, that is, an *entire* function. Let k denote the larger of the numbers $|h|$ and $|h'|$. It follows easily from (3.5) that

$$(3.6)\qquad |f(t)| \leqq Ke^{k|t|}$$

for all complex values of t where K, as k, is a positive constant; an entire function satisfying such an inequality is termed *of exponential type.* (In fact, in our case, K could be chosen as 1.)

Let r denote a positive number. Then, by (3.5),

$$(3.7)\qquad f(-ir) = \int_{-h'}^{h} e^{rx}dF(x) \leqq e^{rh}\int_{-h'}^{h} dF(x) = e^{rh}.$$

Hence, and from (3.4), we obtain that

$$(3.8)\qquad \overline{\lim_{r\to\infty}}\, r^{-1}\log|f(-ir)| \leqq \text{rext}\,[F].$$

II. Now we are given an entire function of exponential type $f(t)$. Ignoring (3.4), we now define h and h' by

$$(3.9)\quad h = \overline{\lim_{r\to+\infty}}\, r^{-1}\log|f(-ir)|, \qquad h' = \overline{\lim_{r\to\infty}}\, r^{-1}\log|f(ir)|.$$

We know that $f(t)$ is a characteristic function, linked to a certain distribution function $F(x)$ by (1). We wish to determine the variation of $F(x)$ in an interval (x_1, x_2) where

$$h < h + \epsilon = x_1 - \epsilon < x_1 < x_2 \, ; \tag{3.10}$$

$2\epsilon = x_1 - h$ is a positive quantity. We assume that x_1 and x_2 are points of continuity for $F(x)$.

Let r denote a positive number. Then, by (1),

$$f(-ir) = \int_{-\infty}^{\infty} e^{rx} dF(x) \geqq \int_{x_1}^{x_2} e^{rx} dF(x) \geqq e^{rx_1}[F(x_2) - F(x_1)] \, . \tag{3.11}$$

It follows from (3.9) and from (3.10) that, for sufficiently large r,

$$|f(-ir)| < e^{(h+\epsilon)r} = e^{(x_1 - \epsilon)r} \, . \tag{3.12}$$

Comparing (3.11) and (3.12), we find that

$$e^{-\epsilon r} \geqq F(x_2) - F(x_1) \geqq 0 \tag{3.13}$$

for a fixed positive ϵ and arbitrarily large r. The left-hand side can be made arbitrarily small by choosing r sufficiently large, and therefore

$$F(x_2) - F(x_1) = 0 \, . \tag{3.14}$$

Thus the variation of $F(x)$ in any interval to the right of h is zero. Recalling the definition of "rext," and that h was defined by (3.9), we can state our result in the form

$$\text{rext}\,[F] \leqq \overline{\lim_{r \to \infty}}\, r^{-1} \log |f(-ir)| \, . \tag{3.15}$$

Comparing (3.8) and (3.15), we obtain the first equation of (3.3). The second can be derived similarly, and so we have proved theorem 2.

4. Finitely Different Distributions

If two distribution functions $F_1(x)$ and $F_2(x)$ are such that their difference vanishes outside a finite interval, we may call the two corresponding probability distributions "finitely different." After the foregoing section, we may suspect that the distributions corresponding to the functions $F_1(x)$ and $F_2(x)$ are finitely different if, and only if, the difference of the respective characteristic functions, $f_1(x) - f_2(x)$, is an entire function of exponential type. This is in fact true, although the proof is much more difficult than that given for theorem 2. Even a little more is true, namely the following:

THEOREM 3. *The function* G(x) *is defined and of bounded variation in the interval* $(-\infty, \infty)$. *The necessary and sufficient condition that* G(x) *should be constant outside a finite interval is that*

$$g(t) = \int_{-\infty}^{\infty} e^{itx} dG(x) \tag{4.1}$$

should be an entire function of exponential type. If (−h′, h) *is the smallest interval outside of which* G(x) *is constant, then*

$$\overline{\lim_{r \to \infty}}\, r^{-1} \log |g(-ir)| = h, \qquad \overline{\lim_{r \to \infty}}\, r^{-1} \log |g(ir)| = h'. \tag{4.2}$$

The phrase "$G(x)$ constant outside $(-h', h)$" means that there are two constants, C and C', such that

$$G(x) = C \quad \text{for} \quad x > h, \qquad G(x) = C' \quad \text{for} \quad x < -h'. \tag{4.3}$$

We see that theorem 2 is contained in theorem 3. This latter is an extension of an important theorem of Paley and Wiener.[8] In fact, the simple proof for the Paley-Wiener theorem given by Plancherel and the present author[9] can be modified so that it yields also theorem 3.

Indeed, the proof of theorem 3 consists of two parts. The first part, which starts from $G(x)$ and works toward $g(t)$, is scarcely different from the corresponding part of the proof as given in section 3, and can be omitted. We concentrate on the more difficult second part that starts from $g(t)$.

We are given an entire function of exponential type, $g(t)$. The numbers h and h' are defined by (4.2). We know that the function $G(x)$ is of bounded variation in $(-\infty, \infty)$. We have to ascertain the behavior of $G(x)$ outside $(-h', h)$.

We know[10] that

$$G(x_2) - G(x_1) = \lim_{r \to \infty} \int_{-r}^{r} \frac{e^{-ix_1 t} - e^{-ix_2 t}}{2\pi i t}\, g(t)\, dt \tag{4.4}$$

if x_1 and x_2 are any two points of continuity of $G(x)$. (In fact, the formula remains valid, and is scarcely harder to prove, even in points of discontinuity, provided that

$$G(x) = \frac{1}{2}\,[G(x-0) + G(x+0)], \tag{4.5}$$

but the weaker fact quoted is enough here.) We wish to evaluate the right-hand side of (4.4), assuming (3.10) and (4.2) and that $g(t)$ is entire and of exponential type.

[8] See [9], pp. 12–13.
[9] See [10], pp. 246–248.
[10] See Cramér [3], p. 28.

We define $f(t)$ by

$$(4.6)\qquad (e^{-ix_1t} - e^{-ix_2t})g(t) = f(t)e^{-i\epsilon t}.$$

By (3.10), this is equivalent to

$$(4.7)\qquad f(t) = g(t)e^{-i(h+\epsilon)t}(1 - e^{-i(x_2-x_1)t}).$$

Let us observe certain properties of $f(t)$.

(i) Since $g(t)$ is an entire function of exponential type, the same is true of $f(t)$, by (4.7).

(ii) By (4.1), $g(t)$ remains bounded for real t. The same is true of $f(t)$, by (4.7).

(iii) By (4.7)

$$f(-ir) = g(-ir)e^{-(h+\epsilon)r}(1 - e^{-(x_2-x_1)r}).$$

The right-hand side remains bounded when the positive variable r tends to infinity, by virtue of the first condition (4.2) and the last inequality (3.10).

(iv) As a well-known function-theoretic argument shows,[11] the results (i), (ii), and (iii) imply that $f(t)$ is bounded in the lower halfplane:

$$(4.8)\qquad |f(t)| \leqq C \quad \text{when} \quad t = re^{-i\varphi}, \qquad 0 \leqq \varphi \leqq \pi.$$

Using (4.6), we write (4.4) in the form

$$(4.9)\qquad 2\pi i[G(x_2) - G(x_1)] = \lim_{r\to\infty}\int_{-r}^{r} f(t)e^{-i\epsilon t}t^{-1}dt.$$

The integrand on the right-hand side is an entire function [$f(0) = 0$ by (4.7)]. We deform the originally straight path of integration into a *half circle*, with center zero and radius r, located in the lower halfplane where (4.8) holds. Therefore

$$(4.10)\qquad \left|\int_{-r}^{r} f(t)e^{-i\epsilon t}t^{-1}dt\right| \leqq C\int_0^{\pi} e^{-\epsilon r\sin\varphi}\,d\varphi.$$

The right-hand side of (4.10) obviously approaches zero when r tends to infinity. Remembering that we have supposed (3.10), we see from (4.9) and (4.10) that

$$(4.11)\qquad G(x_1) = G(x_2) \qquad \text{when} \qquad x_2 > x_1 > h.$$

In other words, $G(x)$ is constant for $x > h$, and we could show by a similar argument that it is also constant for $x < -h'$. Thus we have completed the proof of theorem 3.

[11] See, for example, Pólya and Szegö, [13], vol. 1, p. 149, Nr. 330.

REFERENCES

1. Bochner, S. *Vorlesungen über Fouriersche Integrale.* Leipzig, 1932.

2. ———. "Monotone Funktionen, Stieltjessche Integrale und harmonische Analyse," *Math. Annalen*, vol. 108 (1933), pp. 378–410.

3. Cramér, Harald. *Random Variables and Probability Distributions.* Cambridge, 1937.

4. Gnedenko, B. "Sur les fonctions caractéristiques," *Bull. Math. Univ. Moscou*, Sériè internationale, Sec. A, vol. 1 (1937), pp. 16–17.

5. Hardy, G. H., J. E. Littlewood, and G. Pólya. *Inequalities.* Cambridge, 1934.

6. Khintchine, A. "Zur Kennzeichnung der charakteristischen Funktionen," *Bull. Math. Univ. Moscou*, Sériè internationale, Sec. A, vol. 1 (1937), pp. 1–3.

7. Levy, Paul. *Théorie de l'addition des variables aléatoires*, Paris, 1937.

8. Mathias, M. "Über positive Fourier-Integrale," *Math. Zeitschrift*, vol. 16 (1923), pp. 103–125.

9. Paley, Raymond E. A. C., and Norbert Wiener. *Fourier Transforms in the Complex Domain.* New York, 1934.

10. Plancherel, M., and G. Pólya. "Fonctions entières et intégrales de Fourier multiples," *Commentarii Mathematici Helvetici*, vol. 9, pp. 224–248.

11. Pólya, G. "Über die Nullstellen gewisser ganzer Funktoinen," *Math. Zeitschrift*, vol. 2 (1918), pp. 352–383.

12. ———. "Herleitung des Gauss'schen Fehergesetzes aus einer Funktionalgleichung," *ibid.*, vol. 18 (1923), pp. 96–108.

13. Pólya, G., and G. Szegö. *Aufgaben und Lehrsätze aus der Analysis.* Berlin, 1925.