Functional Equations and Characterization of Probability Distributions

C. Radhakrishna Rao

1. Introduction. The importance of characterization of probability distributions in problems of statistical inference has been stressed in a recent book by Kagan, Linnik and Rao (Russian ed. 1971, English ed. 1973) which will be referred to as KLR in the rest of the paper. There are different types of characterization problems of which the following two seem to have received considerable attention.

One is to assume a model underlying a stochastic phenomenon and derive the appropriate distribution of an observable random variable. Classical examples are the derivation of the normal distribution from Hagan's hypothesis on errors of measurement and from Maxwell's hypothesis on velocities of molecules in a gas (see Rao [4, pp. 160–161]).

A second type which opened up a rich area of research is what may be described as characterization of probability distributions through properties of sample statistics. More precisely the problem can be stated as follows:

Let \((A, B)\) and \((C, D)\) be two measure spaces and \(T : A \rightarrow C\) be a measurable mapping of \((A, B)\) into \((C, D)\). Further let \(p\) be a probability measure on \((A, B)\) and \(p_T\) the probability measure induced by (statistic) \(T\) on \((C, D)\). Further let \(\pi\) be a specified property of \(p_T\). The problem is to find the class

\[
P = \{p : p_T \text{ has the specified property } \pi\}.
\]

The mathematical problem is interesting when \(\pi\) is a weak property and \(P\) is a small class. A famous example is the Darmois-Skitovic theorem: Let \(X_1, \ldots, X_n\) be independent variables, \(T_1 = a_1X_1 + \cdots + a_nX_n\) and \(T_2 = b_1X_1 + \cdots + b_nX_n\) be linear functions where \(a_i, b_i\) are nonzero, and \(\pi\) be the property that \(T_1\) and \(T_2\) are independently distributed. Then each \(X_i\) is normally distributed.

© 1975, Canadian Mathematical Congress
In most cases the problem of characterization reduces to finding the solution of a functional equation. Indeed, the study of characterization problems has led to several new functional equations, not all of which have been satisfactorily solved. In my article, I shall confine myself to the special area of characterizing the multivariate normal distribution (m.n.d.) on $R^p$, the Euclidean space of $p$ dimensions (with possible generalizations to other spaces), describe the nature of functional equations involved and mention some unsolved problems (listed as Problems 1—7).

2. The Cauchy equation and generalization. Two basic results on functional equations which led to the solution of many characterization problems are given in Lemmas 1 and 2 (see KLR, pp. 29–37, 471–476 for more general results).

**Lemma 1** (KLR, p. 29). Let $f_1, \ldots, f_k$ be continuous complex-valued functions on $R^1$ such that, for given $a$, all different,

$$f_1(t + a_1 u) + \cdots + f_k(t + a_k u) = A(t) + B(u), \forall t, u \in O^1$$

where $O^p$ denotes a region covering the origin in $R^p$. Then $f_i, A, B$ are all polynomials of degree $\leq k$ in $O^1$.

**Lemma 2** (KLR, p. 471). Let $g_i$ be a continuous complex-valued function on $R^p$, and $A_i$ and $B_i$ be matrices of rank $p_i$ and orders $p \times p_i$ and $m \times p_i$ respectively, $i = 1, \ldots, k$, such that $\forall \ t \in O^p$ and $u \in O^m$,

$$g_1(A_1 t + B_1 u) + \cdots + g_k(A_k t + B_k u) = C(t) + D(u).$$

Then $C(t)$ and $D(u)$ are polynomials of degree $\leq k$ in $O^p$ and $O^m$ respectively.

Let us consider a simple case of the equation (2.2) with $k = 1$,

$$g(t + u) = g(t) + g(u) \quad \forall \ t, u \in O^p$$

which is the famous Cauchy equation with a linear function as its solution. Let us restrict the validity of (2.3) to only pairs $t, u \in O^p$ such that the inner product

$$\langle t, u \rangle = 0.$$  

What is the solution of (2.3) with the restriction (2.4)? The answer is given in Lemma 3, which is proved by using Lemma 1.

**Lemma 3.** If $g$ satisfies (2.3) with the restriction (2.4), then $g$ is a polynomial of degree $\leq 2$ in $O^p$.

The solution is no longer linear but is still of the polynomial type.

It may be noted that if the restriction on $(t, u)$ is of the type

$$(t, u) = [(t, t)(u, u)]^{1/2} \cos \alpha,$$

where $\cos \alpha \neq 0$, then again the solution is linear.

It will be of interest to consider other restrictions which may lead to different types of solutions. I mention one such possibility which has applications in characterization problems.
Problem 1. Suppose for any given $t \in \mathcal{O}^p$, there exists $u (\neq 0) \in \mathcal{O}^p$ such that $g(at + bu) = g(at) + g(bu) \forall a, b \in \mathcal{O}^1$. What is the solution for $g$?

Note that no relationship between $t$ and $u$ such as (2.4) is specified. A possible solution for $g$ is

$$(2.5) \quad g(t) = h(t_1, \ldots, t_i) + Q(t_{i+1}, \ldots, t_p)$$

where $t_1, \ldots, t_p$ are components of $t$ in some order, $h$ is an arbitrary function and $Q$ is a quadratic function.

As an application of Lemma 3 we have Theorem 1 characterizing a m.n.d., while a stronger result is true for $p = 2$ as in Theorem 2.

**Theorem 1.** Let $X$ be a $p$-vector r.v. (random variable) such that $a'X$ and $b'X$ are independently distributed for all $a, b \in \mathcal{R}^p$ such that $(a, b) = 0$. Then $X$ has m.n.d.

**Theorem 2.** Let $X$ be a bivariate r.v. (with components $\in \mathcal{R}^1$), $A$ and $B$ be two given nonsingular $2 \times 2$ matrices such that $A^{-1}B$ or $B^{-1}A$ has no zero element. If the components of $BX$ are independent and so are also the components of $AX$, then $X$ has a bivariate n.d.

Theorem 2 shows that to assert bivariate normality of $X$, it is only necessary to find just two pairs of linear functions such that the functions in each pair are independently distributed. For general $p$, Theorem 1 requires independence for a very wide class of pairs of linear functions. We pose the following problems.

Problem 2. Let $X$ be a $p$-vector r.v. Suppose that for any given $a \in \mathcal{R}^p$, there exists $b (\neq 0) \in \mathcal{R}^p$ such that $a'X$ and $b'X$ are independently distributed. Then what can be said about $X$?

If (2.5) is the only solution, then some of the components of $X$ have an arbitrary distribution and the rest have a m.n.d.

Problem 3. What is the smallest class of pairs of vectors $(a, b)$ such that $a'X$ and $b'X$ are independent, which ensures multivariate normality of $X$? (When $p = 2$, just two pairs are sufficient.)

**3. Generalization of Darmois-Skitovic theorem.** Let $X_1, \ldots, X_k$ be $k$ independent $p$-vector variables such that

$$(3.1) \quad A_1X_1 + \cdots + A_kX_k \quad \text{and} \quad B_1X_1 + \cdots + B_kX_k,$$

where $A_i, B_i$ are nonsingular matrices, are independently distributed. Then Ghurye and Olkin [1] showed that each $X_i$ has $p$-variate n.d. This result is obtained by writing down the functional equation satisfied by $g_i$, the log of characteristic function (c.f.) of $X_i$,

$$(3.2) \quad g_1(A_1t + B_1u) + \cdots + g_k(A_kt + B_ku) = C(t) + D(u), \quad \forall t, u \in \mathcal{O}^p$$

where $\mathcal{O}^p$ is a suitable neighbourhood of the origin in which the logs of all functions are well defined, and applying Lemma 2, which shows that $C(t)$ is a polynomial which being a c.f. must be of degree $\leq 2$. Then $\sum A_iX_i$ has m.n.d. and hence $A_iX_i$ and $X_i$ have m.n.d. for each $i$.

The crucial step in the proof is to show that $C(t)$ is a polynomial, which very
much depends on the finite number of terms on the left-hand side of the equation (3.2). It is not clear what happens when the linear forms (3.1) contain an infinite number of terms. We may formulate the problem as follows.

**Problem 4.** Suppose \( \{ X_i \} \) is an infinite sequence of independent \( p \)-vector r.v.'s such that

\[
T_1 = \sum_{i=1}^{\infty} A_i X_i \quad \text{and} \quad T_2 = \sum_{i=1}^{\infty} B_i X_i
\]

are independently distributed. Then what can be said about the distribution of \( X_i \)?

A solution to this problem depends on the nature of the solution for \( C(t) \) in (3.2) when \( k = \infty \). When \( p = 1 \), and the \( A_i, B_i \in \mathbb{R}^1 \) satisfy some conditions it is shown that \( X_i \) are normally distributed (KLR, pp. 34, 94). The proof does not easily generalize to \( p > 1 \).

4. **Characterization through constancy of regression.** Let \( X_1, X_2 \) be independent and identically distributed \( p \)-vector r.v.'s such that the conditional expectation

\[
E(X_1 - AX_2 | X_1 + B'X_2) = 0
\]

for given nonsingular matrices \( A \) and \( B \). What can be said about the distribution of \( X_1 \)? We may suppose that \( X_1 \) has first moment.

A complete solution to the problem is available when \( p = 1 \) (KLR, pp. 158–161). A complete solution when \( p = 2 \) and a partial solution for \( p > 2 \) are given by Khatri and Rao [3]. We shall examine the nature of the functional equation for general \( p \).

Let \( g(t) \) be the log c.f. of \( X_1 \) and define by \( G(t) = \partial g/\partial t \), the vector of partial derivatives of \( g(t) \) with respect to the elements of \( t \). Then it is easy to show that (4.1) implies

\[
G(t) = AG(Bt) \quad \text{or} \quad A^{-1}G(t) = G(Bt).
\]

The problem is to solve (4.2) for \( G \) given \( A \) and \( B \), and eventually to determine \( g \) such that \( G(t) = \partial g/\partial t \).

It is interesting to note that an equation of the type (4.2) occurs in the study of optimization problems and structural stability studied by Andronov and Pontrjagin (see Robbins [5]). In their problem \( A^{-1} (= D \) say) and \( B \) stand for \( C^1 \) diffeomorphisms from a smooth manifold \( M \) onto itself and \( G \) is a homeomorphism such that

\[
D \circ G = G \circ B
\]

in which case \( B \) and \( D \) are said to be topologically conjugate. Theorem 3 considers the special case of (4.2) when \( A = B^{-1} \).

**Theorem 3.** Let

\[
B = \delta_1 Q_1 P_1' + \cdots + \delta_r Q_r P_r'
\]

be the singular value decomposition of \( B \), where \( Q_i \) and \( P_i \) are matrices of order \( p \times m_i \) with orthonormal vectors corresponding to multiplicity \( m_i \) of the root \( \delta_i \). If \( A = B^{-1} \) in (4.2) then \( g(t) \) is of the form

\[
g(Pt) = h_1(t_1) + \cdots + h_r(t_r),
\]

\[
g(Qt) = \delta_1^2 h_1(\delta_1^{-1}t_1) + \cdots + \delta_r^2 h_r(\delta_r^{-1}t_r),
\]
where \( t_i \) is a subvector of \( t \) of order \( m_i \), and \( h_i \) are suitable functions. Then \( h_i \) in (4.4), (4.5) satisfy the equation

\[
\sum h_i(P'_iQ_1t_1 + \cdots + P'_iQ rt_i) = \sum \delta_i^2 h_i(\delta_i^{-1}t_i).
\]

Thus, the solution of the characterization problem (4.1) even in the special case \( A = B^{-1} \) depends on the solution of the functional equation (4.6) which is of the form discussed in KLR (p. 476) but not solved in generality. Solutions have been found for \( p = 1, 2 \) and for general \( p \) when the matrices \( P'_iQ_i \) satisfy the conditions given in KLR (p. 476, Theorem A.5.3), leading to multivariate normality of \( X_t \) in (4.1). A solution to the equation (4.6) in the general case is of interest.

**Problem 5.** Let \( h_i \) be a continuous complex-valued function on \( \mathbb{R}^{m_t} \), \( P_i \) and \( Q_i \) be partitions with \( m_i \) columns of orthogonal matrices \( P \) and \( Q \), and \( \delta_i \) be positive numbers for \( i = 1, \ldots, r \). What are the solutions for \( h_i \) of the functional equation

\[
\sum h_i(P'_iQ_1t_1 + \cdots + P'_iQ rt_i) = \sum \delta_i^2 h_i(\delta_i^{-1}t_i)?
\]

Now we state a general problem:

**Problem 6.** What is the solution for \( G \) or \( g (G = \partial g/\partial t) \) of the equation

\[
A_1G(B_1t) + \cdots + A_kG(B_kt) = 0
\]

where \( A_i \) and \( B_i \) are given matrices and \( k \) may be infinite?

The solution seems to be difficult even for \( k = 2 \). A solution to the general equation (4.8) would enable us to characterize the probability distribution of \( X_t \) by the condition (where \( X_t \) are identically distributed)

\[
E(\sum A_iX_t \mid \sum B_iX_t) = 0.
\]

**5. Generalization to other spaces.** Throughout this paper, the variables like \( t, u \) are considered to belong to \( \mathbb{R}^p \). All the problems could be generalized to variables belonging to other spaces or topological groups. For instance the equation (3.2), in terms of \( f_i \), the c.f. of \( X_t \) is of the form

\[
\prod_{i=1}^{k} f_i(A_i t + B_iu) = h(t) \cdot m(u).
\]

One may generalize the problem as follows:

**Problem 7.** Let \( t, u \in X \), a Hausdorff topological group, \( f_i \) be conditionally positive definite functions and \( A_i, B_i \) be continuous automorphisms of \( X \). Then what are the solutions for \( f_i \) of

\[
\prod_{i=1}^{k} f_i(A_i t B_iu) = h(t) \cdot m(u)?
\]

This problem has been considered by Schmidt [6] and solved when \( k = 2 \). A solution in the general case (including \( k = \infty \)) would be of interest.

Similar generalizations can be made of Lemma 3 where \( t, u \) can belong to a space furnished with an inner product, Problem 1 to a general space, Theorem 2 to a r.v. with components defined on more general spaces instead of \( \mathbb{R}^1 \) and so on.
References


