CHARACTERIZATION OF POPULATIONS BY PROPERTIES OF SUITABLE STATISTICS

EUGENE LUKACS
OFFICE OF NAVAL RESEARCH

1. Introduction

In this paper we consider finite sets of independently and identically distributed random variables $X_1, X_2, \cdots, X_n$. Our aim is to characterize their common distribution function $F(x)$ by properties of the set $X_1, X_2, \cdots, X_n$. The problem can best be formulated by using statistical terminology.

We consider a population and a sample $X_1, X_2, \cdots, X_n$ of n independent observations drawn from this population with population distribution function $F(x)$. As usual, a measurable and single-valued function $S = S(X_1, X_2, \cdots, X_n)$ of the observations is called a statistic. Assumptions concerning the properties of the distributions of certain statistics, based on a sample from the population, will in general impose restrictions on the population distribution function $F(x)$. We are interested in assumptions which determine the population distribution function at least to the extent that it belongs to a certain family of distribution functions. Three different types of assumptions are considered.

In the first part we make assumptions which either give explicitly the distribution of S or which relate it in some specified manner to the population distribution $F(x)$.

In the second we suppose that two suitably chosen statistics $S_1 = S_1(X_1, X_2, \cdots, X_n)$ and $S_2 = S_2(X_1, X_2, \cdots, X_n)$ are given. The assumption that the statistics S_1 and S_2 are independently distributed can be used to characterize various populations. We also consider briefly the characterization of a population by the stochastic independence of more than two statistics. Finally we assume that the conditional expectation of S_1, given S_2, equals the unconditional expectation of S_1 and show that this hypothesis can also be used to characterize populations. This property is weaker than complete independence of the two statistics; its use in investigations of this kind seems to be new.

The third part deals with the characterization of populations by means of the property that two different linear statistics are identically distributed.

We denote the population distribution function by $F(x)$. The characteristic function $f(t)$ of $F(x)$ is given by

$$f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) , \tag{1.1}$$

while

$$\varphi(t) = \ln f(t) \tag{1.2}$$

is called the cumulant generating function (c.g.f.) of $F(x)$. Since every characteristic function $f(t)$ is a continuous function such that $f(0) = 1$, we see that $\varphi(t)$ is certainly defined by (1.2) in an interval of the real axis which contains the origin.