MAXIMUM LIKELIHOOD CHARACTERIZATION OF DISTRIBUTIONS

By Henry Teicher

New York University and Purdue University

1. Introduction. It is a commonplace observation that the sample mean and sample variance from a normal population (based on a random sample) are stochastically independent. Considerably less prosaic is the converse proposition, first proved in 1936 by R. C. Geary [2] (under superfluous restrictions), to the effect that the independence of these two statistics entails normality of the underlying population. This, plus the theorem that if two linear combinations (nonzero coefficients) of a pair of independent random variables are themselves independent, the variables are normally distributed, which was proved by Kac in 1939 [4], are harbingers of what are today referred to as characterization theorems. An extensive bibliography of such theorems appears in [5]. Most of these results have the format: if such-and-such statistics are independent (alternatively, if the distribution of such-and-such a statistic is thus-and-so), the underlying population is so-and-so.

The ensuing theorems belong to this genre but adopt a maximum likelihood posture. The first deals with translation (location) parameter and the latter with scale parameter families of distributions.

2. Preliminaries. Since the results expounded here concern maximum likelihood estimators, it would seem appropriate to say a few words concerning these. It is somewhat surprising that major treatises on mathematical statistics and estimation do not define maximum likelihood *estimators* per se but merely a maximum likelihood *estimate*. (Pitman's terminological demarcation between these notions will be made explicit shortly.) The definitions of [8], [9] are closest in spirit to that given here.

In order to pave the way for a discussion of these questions, let $F(x; \theta)$, $-\infty < x < \infty$, $\theta \varepsilon \Omega \subset R^1$ denote a one parameter family of probability distributions on the real line R^1 with spectra S_θ. Define $S = \bigcup_{\theta \varepsilon \Omega} S_\theta$ and $S^n = S \times S \times \cdots \times S$, the n-fold cartesian product of S with itself.[1] If, for each $\theta \varepsilon \Omega$, $F(x; \theta)$ is absolutely continuous, designate its probability density function (p.d.f.) by $f(x; \theta)$; if, for each θ, $F(x; \theta)$ is a step function, the same notation $f(x; \theta)$ will be used to specify the so-called discrete p.d.f., that is, the mass function of the corresponding distribution (positive at the countable set of points constituting S_θ and zero elsewhere).

The customary definition of a maximum likelihood *estimate* of a parameter θ of a population (family of distributions generally restricted to the aforementioned types), based on a (random) sample of n observations $x_1, x_2, \cdots, x_n$, is a value of θ, say $\hat{\theta}_n$, which renders $\prod_{i=1}^n f(x_i; \theta)$ a maximum. A maximum likelihood

Received August 23, 1960; revised January 20, 1961.

[1] When $\bigcup_\theta S_\theta = [a, \infty)$ or $(-\infty, b]$, the points a, b will be deleted in defining S (so as to avoid a special treatment of the origin in Theorem 2).