The Annals of Mathematical Statistics
1970, Vol. 41, No. 6, 1845–1852

ESTIMATING THE EMPIRIC DISTRIBUTION FUNCTION OF CERTAIN PARAMETER SEQUENCES[1]

BY RICHARD J. FOX[2]

Michigan State University

1. Introduction and notation. A compound decision problem (see Section 6 of Robbins (1951)) consists of a set of n independent statistical decision problems, each having the basic structure of a so-called component problem. The usual objective is to find procedures, which may use all n observations for each decision, whose average risk across problems converges, as $n \to \infty$, to the component problem Bayes risk versus the empiric distribution of the n underlying parameter values. This latter quantity represents the minimum average risk if one employs a procedure where the same decision rule, which does not depend on the set of observations, is applied independently to the individual problems.

One possible solution technique is to use the set of observations to estimate the empiric distribution of the set of parameter values and then use the Bayes procedure versus this estimate in each individual problem. For a very simple example, the estimators developed in Section 2 and Section 3 of this paper, which deal with two specific uniform distributions, could be used in this way to construct procedures which meet the previously specified objective for the compound test of simple hypotheses problem solved by Hannan and Robbins (1955).

Let $\mathbf{x} = (x_1, x_2, \cdots)$ be a sequence of independent random variables with x_i having distribution function F_{θ_i}, henceforth abbreviated to F_i, $\theta_i \in \Omega$ for $i = 1, 2, \cdots$ and Ω a subset of the real line. Suppose that this family, indexed by the parameter $\theta \in \Omega$, is dominated by Lebesgue measure μ and let f_θ be the density of F_θ with respect to μ. Also abbreviate f_{θ_i} by f_i.

Throughout this paper we will occasionally omit the display of the argument of a function of a real variable. We also adopt the convention that distribution functions are right continuous. Let F be a distribution function; we will also use the letter F to denote the corresponding Lebesgue–Stieltjes measure. If A is an event, $[A]$ will be used to denote the indicator function of A.

Let $\mathbf{F} = \times_{i=1}^{\infty} F_i$, i.e. $\mathbf{F}$ is the product measure on the space of $\mathbf{x}$'s corresponding to a particular $\boldsymbol{\theta}$, $\boldsymbol{\theta}$ denoting a parameter sequence: $(\theta_1, \theta_2, \cdots)$. Let G_n be the empiric distribution function of the first n parameters: $\theta_1, \theta_2, \cdots, \theta_n$. We now define the following functions:

$$\bar{F} = \int F_\theta \, dG_n(\theta) = n^{-1} \sum_{i=1}^{n} F_i; \tag{1.1}$$

$$\bar{f} = \int f_\theta \, dG_n(\theta) = n^{-1} \sum_{i=1}^{n} f_i \tag{1.2}$$

Received March 14, 1969; revised April 29, 1970.

[1] Research supported by National Science Foundation, Grant No. GP-13484.

[2] Now at Procter and Gamble.