The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 286–288

CONSTRUCTION OF SEQUENCES ESTIMATING THE MIXING DISTRIBUTION[1]

BY J. J. DEELY AND R. L. KRUSE

Sandia Laboratory, Albuquerque

In this paper is presented an explicit construction of a sequence of distributions which estimates a mixing distribution or the *a priori* distribution of an empirical Bayes decision problem. In 1964 H. Robbins [3] studied such estimating sequences and proposed the problem of obtaining explicit constructions of the sequences. The sequence constructed in this paper consists of discrete distributions whose weights are determined by the solution of a linear programming problem.

K. Choi [1] has also recently considered this problem. Our approach differs from his in two ways. First, we use the sup norm

$$\|F - G\| = \sup |F(x) - G(x)|$$

as a distance function, whereas Choi uses the Wolfowitz distance

$$W(F, G) = \int |F(x) - G(x)|^2 \, dF(x).$$

Second, our construction can make use of well-known and efficient linear programming algorithms, whereas the computational feasibility of Choi's method has not been clearly established.

We begin with the class $\mathcal{G}$ of distributions $G(\lambda)$ defined on a compact subset Λ of the real line, and a function $F(x \mid \lambda)$, which, for each $\lambda \in \Lambda$, is a distribution on a closed subset $\mathfrak{X}$ of the real line. We define the *mixture* of $F(x \mid \lambda)$ and $G(\lambda)$ by

$$F_G(x) = \int_\Lambda F(x \mid \lambda) \, dG(\lambda).$$

Let $x_1, x_2, \cdots$ be a sequence of independent observations selected according to the mixture $F_G(x)$. We wish to construct a sequence of distributions $\{G_n(\lambda)\}$, where $G_n(\lambda)$ depends only on the observations $x_1, x_2, \cdots, x_n$, which converges weakly to the *a priori* distribution $G(\lambda)$ with probability one. That is,

$$P\{\lim_{n\to\infty} G_n(\lambda) = G(\lambda), \lambda \text{ any continuity point of } G\} = 1.$$

For each natural number n we define $\mathcal{G}_n$ to be the class of discrete distributions on Λ with weights at $\lambda_{1n}, \lambda_{2n}, \cdots, \lambda_{nn}$, where the λ_{in} are chosen so that for any $G \in \mathcal{G}$ there is a sequence $\{G_n\}$ with $G_n \in \mathcal{G}_n$ which converges weakly to G.

Our method is to find $G_n^* \in \mathcal{G}_n$ which minimizes $\|F_H - F_n\|$ for $H \in \mathcal{G}_n$, where $F_n(x)$ is the empirical distribution based upon the first n observations. The following theorem assures the desired convergence.

Received 2 March 1967; revised 11 July 1967.

[1] The authors wish to express their thanks to Professor Herbert Robbins for suggesting this problem. This work was supported by the United States Atomic Energy Commission.

THEOREM. *Suppose*

(i) $F(x \mid \lambda)$ *is continuous on* $\mathfrak{X} \times \Lambda$.
(ii) *If* $F_G = F_H$ *for* $G, H \varepsilon \mathcal{G}$, *then* $G = H$.
(iii) $\{G_n^*(\lambda)\}$ *is a sequence for which* $G_n^*(\lambda) \varepsilon \mathcal{G}_n$ *and*

$$\|F_{G_n^*} - F_n\| = \inf_{H \in \mathcal{G}_n} \|F_H - F_n\|.$$

Then

$$P\{\lim_{n\to\infty} G_n^*(\lambda) = G(\lambda), \lambda \text{ any continuity point of } G\} = 1.$$

To prove this theorem we shall need the following:

LEMMA. *Suppose* $F(x \mid \lambda)$ *is continuous on* $\mathfrak{X} \times \Lambda$ *and* $\{G_n(\lambda)\}$ *is a sequence from* $\mathcal{G}$ *which converges to a given* $G(\lambda) \varepsilon \mathcal{G}$ *at all continuity points of* $G(\lambda)$. *Then* $\lim_{n\to\infty} \|F_{G_n} - F_G\| = 0$.

PROOF. Continuity of $F(x \mid \lambda)$ on $\mathfrak{X}$ and dominated convergence imply that $F_G(x)$ and $F_{G_n}(x)$ are continuous on $\mathfrak{X}$. Furthermore, the Helly-Bray theorem implies the pointwise convergence of $F_{G_n}(x)$ to $F_G(x)$. The result follows, since pointwise convergence of distributions to a continuous distribution implies uniform convergence [2].

PROOF OF THEOREM. By the use of Robbins' Theorem 2 of [3] it suffices to prove that

$$\lim_{n\to\infty} \|F_{G_n^*} - F_G\| = 0$$

with probability 1. Let $\{G_n(\lambda)\}$ be a sequence with $G_n(\lambda) \varepsilon \mathcal{G}_n$ which converges weakly to $G(\lambda)$. By the lemma $\|F_{G_n} - F_G\| \to 0$. Now

$$\begin{aligned} \|F_{G_n^*} - F_G\| &\leqq \|F_{G_n^*} - F_n\| + \|F_n - F_G\| \\ &\leqq \|F_{G_n} - F_n\| + \|F_n - F_G\| \\ &\leqq \|F_{G_n} - F_G\| + 2\|F_n - F_G\|. \end{aligned}$$

Since $\|F_n - F_G\| \to 0$ with probability 1, the theorem follows.

We now outline the construction of the desired distribution $G_n^*(\lambda)$. Let $H(\lambda) \varepsilon \mathcal{G}_n$ have weights, $h_1, \cdots, h_n$ at $\lambda_{1n}, \cdots, \lambda_{nn}$, respectively. Then

$$\|F_H - F_n\| = \max_{1 \leqq i \leqq 2n} |\textstyle\sum_j a_{ij} h_j|,$$

where

$$\begin{aligned} a_{ij} &= F(x_i \mid \lambda_{jn}) - (i - 1)/n, & 1 \leqq i \leqq n, \\ a_{ij} &= F(x_{i-n} \mid \lambda_{jn}) - (i - n)/n, & n + 1 \leqq i \leqq 2n. \end{aligned}$$

Define the $2n \times n$ matrix $A = (a_{ij})$. Finding the desired distribution $G_n^*(\lambda)$ is now equivalent to finding an optimal strategy for the second player in a game with payoff matrix $\binom{A}{-A}$. It is well known that such an optimal strategy always exists and may be found by solving a linear programming problem.

REMARKS. 1. Hypothesis (ii) is needed to use Theorem 2 of [3]. Conditions insuring (ii) have been studied by Teicher [4]. For example, if λ is a scale or translation parameter and the characteristic function of F does not vanish, then (ii) holds.

2. The assumption that Λ is compact is not needed for the results stated here. However, if Λ is not compact, the following additional assumption on $F(x \mid \lambda)$ is required to use Theorem 2 of [3]: $\lim_{\lambda \to \pm\infty, \lambda \varepsilon \Lambda} F(x \mid \lambda)$ exist for each $x \varepsilon \mathfrak{X}$ and are not distributions on $\mathfrak{X}$. That is, $\mathfrak{X}$ has measure less than one relative to these limiting distributions.

3. If Λ is a bounded interval, then the λ_{in} may be equally spaced throughout Λ.

4. Assumption of separate continuity of $F(x \mid \lambda)$ on $\mathfrak{X}$ and on Λ is sufficient, since monotonicity of F on $\mathfrak{X}$ then implies joint continuity.

REFERENCES

[1] CHOI, KEEWHAN (1966). Estimates for the parameters of a finite mixture of distributions. Univ. of Missouri Technical Report No. 18.

[2] PÓLYA, G. (1920). Ueber den zentralen Grenswertsatz der Wahrscheinlichkeitstheorie und das Momentenproblem. *Math. Zeit.* **8** 171.

[3] ROBBINS, HERBERT (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35** 1–20.

[4] TEICHER, HENRY (1963). Identifiability of mixtures. *Ann. Math. Statist.* **34** 1265–1269.