NOTES

NOTE ON A SEQUENTIAL CLASSIFICATION PROBLEM[1]

By Ester Samuel[2]

Columbia University

1. Summary. This note describes explicitly a minimal complete class of decision rules for the problem of sequentially classifying the individuals of a group which is known to come from one of two completely specified populations.

2. Theorem and discussion. Two populations, π_0 and π_1 are given, and an individual is classified to belong to either π_0 or π_1 on the basis of an observation x on a r.v. X, where the distributions P_θ corresponding to π_θ, for $\theta = 0, 1$ are completely specified. We may without loss of generality assume that P_θ, $\theta = 0, 1$ are specified in terms of their density functions $f(x, \theta)$, $\theta = 0, 1$ with respect to a specified measure space $(\mathfrak{X}, \mathfrak{A}, \mu)$.

Consider the following classification problem: A group of n individuals is known to belong to either π_0 or π_1. The individuals of the group arrive sequentially for inspection and classification, the classification of the ith individual has to be made immediately after he has been inspected, $i = 1, \cdots, n$.

Assume the following loss structure for classifying an individual:

(1)

		decision π_0	decision π_1
true	π_0	0	b
	π_1	a	0

where a and b are two given positive numbers. For any classification procedure, we are interested in the average (or equivalently: total) expected loss due to misclassification.

Let the r.v.'s of the individuals be $X_1, \cdots, X_n$ with observed values $x_1, \cdots, x_n$ and let $\mathbf{x}_i = (x_1, \cdots, x_i)$, $i = 1, \cdots, n$. Then any (randomized) decision rule for the above problem can be written as $T_n = (t_1(\mathbf{x}_1), t_2(\mathbf{x}_2), \cdots, t_n(\mathbf{x}_n))$ with $0 \leqq t_i(\mathbf{x}_i) \leqq 1$ being measurable functions in the ith product space, where $t_i(\mathbf{x}_i)$ and $1 - t_i(\mathbf{x}_i)$ are the probabilities with which one classifies the ith individual to come from π_1 and π_0 respectively, when $\mathbf{X}_i = \mathbf{x}_i$ is observed. Set $f(\mathbf{x}_i, \theta) = \prod_{j=1}^{i} f(x_j, \theta)$, $i = 1, \cdots, n$, $\theta = 0, 1$. Let $R(T_n, \theta)$ denote the risk, i.e. average expected loss, incurred by using T_n when the group actually

Received March 12, 1962; revised March 1, 1963.

[1] This research was sponsored by the Office of Naval Research under Contract Number Nonr-266(33), Project Number NR 042-034. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at The Hebrew University, Jerusalem.

belongs to π_θ, $\theta = 0, 1$. Then

$$
\begin{aligned}
R(T_n, \theta) &= \frac{1}{n}\left[b(1-\theta)\sum_{i=1}^{n}\int t_i(\mathbf{x}_i)f(\mathbf{x}_i, \theta)\,d\mu^i\right.\\
&\qquad\left. + a\theta\sum_{i=1}^{n}\int(1 - t_i(\mathbf{x}_i))f(\mathbf{x}_i, \theta)\,d\mu^i\right] \qquad (2)\\
&= \frac{a\theta}{n} + \frac{b(1-\theta) - a\theta}{n}\sum_{i=1}^{n}\int t_i(\mathbf{x}_i)f(\mathbf{x}_i, \theta)\,d\mu^i
\end{aligned}
$$

where the integral is taken over the ith product space of $\mathfrak{X}$.

We can now state the following

THEOREM. *A complete class of decision rules T_n for the above problem is:* $\{T_n^\eta, 0 \leq \eta \leq 1\}$ *where* $T_n^\eta = (t_1^\eta(\mathbf{x}_1), \cdots, t_n^\eta(\mathbf{x}_n))$ *with*

$$
\begin{aligned}
t_i^\eta(\mathbf{x}_i) &= 1 && \text{if } a\eta f(\mathbf{x}_i, 1) > b(1-\eta)f(\mathbf{x}_i, 0)\\
&= 0 && \text{if } a\eta f(\mathbf{x}_i, 1) < b(1-\eta)f(\mathbf{x}_i, 0) \qquad (3)\\
&= \text{arbitrary in } [0, 1] && \text{if } a\eta f(\mathbf{x}_i, 1) = b(1-\eta)f(\mathbf{x}_i, 0).
\end{aligned}
$$

(*Notice that T_n^η, for fixed η, stands for a whole class of decision rules, and thus the notation is actually incomplete.*)

Moreover, if in (3) *the arbitrary part is taken to be* 1 *for* $\eta = 0$ *and* $f(\mathbf{x}_i, 0) = 0$, *and to be* 0 *for* $\eta = 1$ *and* $f(\mathbf{x}_i, 1) = 0$ *then the class obtained is minimal complete.*

PROOF. Let η be the "a priori probability" that the group comes from π_1, i.e. $P\{\pi_\theta = \pi_1\} = \eta = 1 - P\{\pi_\theta = \pi_0\}$ and let $R(T_n, \eta) = \eta R(T_n, 1) + (1-\eta)R(T_n, 0)$. Then from (1) one has

$$
R(T_n, \eta) = \frac{a\eta}{n} + \frac{1}{n}\sum_{i=1}^{n}\int[b(1-\eta)f(\mathbf{x}_i, 0) - a\eta f(\mathbf{x}_i, 1)]t_i(\mathbf{x}_i)\,d\mu^i \qquad (4)
$$

and for fixed η (4) is minimized by minimizing the integrand in its right hand side, thus, is minimized by any T_n^η defined through (3). The theorem thus follows from the completeness of the class of Bayes rules for our problem, and from the fact that with the modifications mentioned above the rules defined through (3) are admissible. ((3) are the Bayes rules for deciding on θ_i on the basis of $\mathbf{x}_i$, and by chapter 7 of [1] the class of Bayes rules is complete, and with the proper modifications the rules are admissible. Averaging over the components for $i = 1, \cdots, n$ does not change these properties.)

REMARKS.

(a) It should be noticed that the rules T_n^η are "strongly sequential", in the sense that if one lets $T^\eta = (t_1^\eta(\mathbf{x}_1), t_2^\eta(\mathbf{x}_2), \cdots)$ then for every n, $n = 1, 2, \cdots$ T_n^η is the initial n-vector of the sequence T^η. One does therefore *not need advance knowledge of the number n of individuals* belonging to the group, in order to apply T_n^η.

(b) Various interpretations for the rule T_n^η are possible:

I. It is easy to prove that for fixed i the rules $t_i^\eta(\mathbf{x}_i)$, for fixed η, are "Bayes with respect to a priori distribution η", for the following fixed-sample-size i hypothesis testing problem: $X_1, \cdots, X_i$ are independent identically distributed r. v.'s distributed according to P_θ.

$$\begin{cases} H_0 : \theta = 0 \\ H_1 : \theta = 1 \end{cases} \tag{5}$$

with loss structure corresponding to (1) (with π_i there replaced by H_i). Thus T_n^η can be described as a rule which at each stage uses *all previous observations* to conduct a test of (5), and the ith individual is classified to belong to π_θ if and only if H_θ is accepted $\theta = 0, 1$. (The "size" α of the test for the various values of i is however usually not constant in i. It depends on η, b, a, $f(x, 0)$, $f(x, 1)$, and i.)

II. Let

$$\eta_{i-1}(\mathbf{x}_{i-1}) = \eta f(\mathbf{x}_{i-1}, 1)/[\eta f(\mathbf{x}_{i-1}, 1) + (1 - \eta) f(\mathbf{x}_{i-1}, 0)].$$

$\eta_{i-1}(\mathbf{x}_{i-1})$ is the a posteriori probability of $\pi_\theta = \pi_1$ when $\mathbf{X}_{i-1} = \mathbf{x}_{i-1}$ has been observed. Now (3) can also be written as:

$$\begin{aligned} t_i(\mathbf{x}_i) &= 1 && \text{if } a\eta_{i-1}(\mathbf{x}_{i-1})f(x_i, 1) > b[1 - \eta_{i-1}(\mathbf{x}_{i-1})]f(x_i, 0) \\ &= 0 && \text{if } a\eta_{i-1}(\mathbf{x}_{i-1})f(x_i, 1) < b[1 - \eta_{i-1}(\mathbf{x}_{i-1})]f(x_i, 0) \\ &= \begin{matrix}\text{arbitrary} \\ \text{in } [0, 1]\end{matrix} && \text{if } a\eta_{i-1}(\mathbf{x}_{i-1})f(x_i, 1) = b[1 - \eta_{i-1}(\mathbf{x}_{i-1})]f(x_i, 0). \end{aligned} \tag{6}$$

Notice, however, that the Bayes rule, with respect to a priori distribution η, for classifying any single individual, is:

$$\begin{aligned} t(x) &= 1 && \text{if } a\eta f(x, 1) > b(1 - \eta) f(x, 0) \\ &= 0 && \text{if } a\eta f(x, 1) < b(1 - \eta) f(x, 0) \\ &= \text{arbitrary in } [0, 1] && \text{if } a\eta f(x, 1) = b(1 - \eta) f(x, 0). \end{aligned} \tag{7}$$

Comparing (6) and (7) we see that the decision on the ith individual is of structure (7), except that the "a priori distribution η" has in (6) been replaced by the "a posteriori distribution of θ, given $\mathbf{x}_{i-1}$". (Notice that although $\eta_i(\mathbf{x}_i)$ is known at the ith decision, only $\eta_{i-1}(\mathbf{x}_{i-1})$ is used in (6).)

For asymptotically "optimal" solutions of the general sequential 2-way classification problem, where the n individuals need not necessarily come from the same population, see [2].

REFERENCES

[1] Blackwell, David and Girshick, M. A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.

[2] Samuel, Ester (1963). Asymptotic solutions of the sequential compound decision problem. *Ann. Math. Statist.* **34** 1079–1094.