INCOMPLETE PRIOR INFORMATION IN A CLASSIFICATION PROCEDURE

by

Ronald H. Randles* and Myles Hollander**

FSU Statistics Report M184
ONR Technical Report No. 45

July 1970
The Florida State University
Department of Statistics
Tallahassee, Florida

* Supported by a National Institute of Health Traineeship (5-T01-GM00913).
Now at the State University of Iowa, Iowa City.

** Supported by the Office of Naval Research under contract (NONR-988(08)NR 042-004).
ABSTRACT

The classification of specimens into one of N populations is considered when the observation on each specimen is a multivariate discrete random variable and when incomplete prior information is available concerning the occurrence of specimens from the various populations. The T-minimax decision criterion is applied and it is shown that an optimal classification procedure is the solution of a finite matrix game. The consistency of the T-minimax procedure is demonstrated when certain population probabilities are estimated by sample proportions. A medical example is considered.
INCOMPLETE PRIOR INFORMATION IN A CLASSIFICATION PROCEDURE
by
Ronald H. Randles and Myles Hollander**

1. STATEMENT OF THE PROBLEM

Important classification problems are found in many disciplines, for example, the diagnosis of diseases in medical science. Consider the classification of specimens (patients) into one of N populations (disease or health state groups) denoted by $\theta_1, \ldots, \theta_N$. Observations on each specimen (symptoms or clinical test results) are assumed to take on only a finite number of values, that is, we observe $X = (X_1, \ldots, X_r)$ where $X_i$ takes on $R_i$ different values. This will occur, for example, when the measurements are qualitative in nature. The vector $X$ can therefore assume $R = \prod_{i=1}^{r} R_i$ values or the observation on any specimen can fall into one of $R$ cells. Given that $X$ is in the $i$-th cell we must then decide into which population the specimen is to be classified.

The Bayes procedure for this classification problem was considered by Cochran and Hopkins [3]. Let $p_{ui}$ denote the $P[\text{specimen from } \theta_u \text{ has its } X \text{ in cell } i]$ and let $C_{uv}$ be the cost of classifying a specimen from $\theta_u$ into $\theta_v$, with $C_{uu} = 0$. If $\pi_u$ is the $P[\text{occurrence of a specimen from } \theta_u]$, then the Bayes rule classifies all specimens whose observations fall in the $i$-th cell into the population $\theta_u$ corresponding to the minimum of

$$t_u = \sum_{v \neq u} \pi_v p_{vi} C_{vu}.$$  

The Bayes rule thus requires the complete specification of $\pi_1, \ldots, \pi_N$. In certain problems, these probabilities will be unknown or

---

*Supported by a National Institute of Health Traineeship (5-T01-GM00913). Now at the State University of Iowa, Iowa City.

**Supported by the Office of Naval Research under contract (N000-988(08) NR 042-004).
only partially known. This absense of information may, for example, result from incomplete follow-up on patients under treatment. Suppose, however, that the incomplete prior information does enable us to specify prior probabilities for groups of the populations of interest. That is, assume we are able to form:

$$\theta_1 = \{\theta_{j_0 + 1}, \ldots, \theta_{j_1}\}, \ldots, \theta_k = \{\theta_{j_{k-1} + 1}, \ldots, \theta_{j_k}\}$$

with $1 \leq k \leq N$ and $J_0 = 0 < J_1 < \ldots < J_k = N$, and that the incomplete prior information consists of the specification of $\pi(t)$, the probability of the occurrence of a specimen from one of the populations in $\theta_t$, for $t = 1, \ldots, k$. Define the collection of prior distributions,

$$T = \{\pi_u, u = 1, \ldots, N | 0 \leq \pi_u \leq 1 \text{ for each } u \text{ and} \}$$

(1)

$$\sum_{t=1}^{J_t} \sum_{u=J_{t-1} + 1}^{J_t} \pi_u = \pi(t) \text{ for } t = 1, \ldots, k.$$  

We would like to use this prior information in determining an optimal classification rule.

Denote a classification procedure by $s = \{s_{ui}\} i = 1, \ldots, R$ and $u = 1, \ldots, N$, where $s_{ui}$ is the conditional probability of classifying a specimen with observation vector in cell $i$ into population $\theta_u$, $u = 1, \ldots, N$. Then $R_u(s) = \sum_{i=1}^{R} \sum_{v=1}^{N} p_{ui} c_{uv} s_{vi}$ is the expected cost of misclassification of specimens from population $\theta_u$. Also, $r(\pi, s) = \sum_{u=1}^{N} \pi_u R_u(s)$ is the expected risk of classification rule $s$ with respect to the prior probability vector $\pi$. 
A general method of utilizing incomplete prior information in a decision rule is given by Blum and Rosenblatt [1]. It is stated here using the notation of the classification problem.

DEFINITION 1. The rule \( \bar{s}^0 \) is said to be a T-minimax decision rule, if for any rule \( \bar{s} \),

\[
\sup_{\pi \in T} r(\pi, \bar{s}^0) < \sup_{\pi \in T} r(\pi, \bar{s}),
\]

where T denotes a class of prior distributions.

Note that the above decision criterion is a generalization which includes both the Bayes and minimax decision rules. If the class T consists of only one prior distribution \( \pi^0 \), a resulting T-minimax rule would be a Bayes rule with respect to \( \pi^0 \). On the other hand, if T consisted of all prior distributions over the populations, a T-minimax decision rule would be a minimax rule. Bunke [2] recognized the potential for applying the T-minimax criterion to classification problems. He did not, however, propose general methods for finding T-minimax procedures and, in fact, in the only example presented, he failed to admit the possibility of a randomized procedure.

A T-minimax classification rule may be found as the optimal strategy in a two person, zero-sum, matrix game. To describe this matrix game, we first specify the role of both players. The strategies (classification rules) for player II are denoted by the vectors, \( \bar{s} \). There are \( N^R \) pure strategies for player II for which \( s_{ui} = 0 \) or 1 for each \( u \) and \( i \). Let
$s^{(m)}$ for $m = 1, \ldots, M$ denote these pure strategies, where $M = N^R$. A randomized strategy for player II will then be denoted by $(\xi_1, \ldots, \xi_M)$, where $\sum_{i=1}^{M} \xi_i = 1$ and $\xi_i$ is the probability that player II uses pure strategy $s^{(i)}$.

Pure strategies for player I are denoted by $Y_j$, $j = 1, \ldots, N$ where strategy $Y_j$ implies that player I selects a specimen from population $\theta_j$. Randomized strategies for player I are denoted by $\pi = (\pi_1, \ldots, \pi_N)$, where $\sum_{j=1}^{N} \pi_j = 1$ and $\pi_j$ is the probability that player I uses pure strategy $Y_j$.

If player I uses pure strategy $Y_j$ and player II uses pure strategy $s^{(i)}$, the payoff from player II to player I is denoted by $w_{ij} = R(\theta_j, s^{(i)})$.

The prior information corresponds to player I using a $\pi$ in $T$ given by (1). Moreover, this fact is known to player II. Hence a $T$-minimax classification rule is an optimal strategy for player II in the restricted matrix game $(T, \Xi, Q)$ where

$$\Xi = \{\xi | 0 \leq \xi_i \leq 1 \text{ for each } i \text{ and } \sum_{i=1}^{M} \xi_i = 1\},$$

(2)

$T$ is given by (1) and $Q(\xi, \pi) = \xi^T W \pi$ when $W$ is the $M \times N$ matrix with elements $w_{ij}$.

2. SOLUTION OF A RESTRICTED MATRIX GAME

The solutions to the restricted matrix game described in section 1 are determined by finding the optimal strategies for a related matrix game. In the related matrix game player II has the same set of strategies,
\[ E, \text{ as player II in the restricted matrix game. A pure strategy for player I}^* \text{ in the related matrix game will be denoted by } (Y_{j_1} \times \ldots \times Y_{j_k}) \text{ where } J_t \text{ is an integer such that } J_{t-1} + 1 \leq J_t \leq J_t \text{ for } t = 1, \ldots, k. \text{ The payoff from II}^* \text{ to I}^* \text{ when I uses a strategy of this form and II}^* \text{ uses a pure strategy } s^{(i)} \text{ is defined to be } \]

\[ K(i;j_1, \ldots, j_k) = \sum_{t=1}^{k} \pi(t) v_{j_t}^{i}. \]

Without loss of generality we assume that \( \pi(t) > 0 \) for \( t = 1, \ldots, k. \) If \( \pi(t) = 0, \) the \( t \)-th group of strategies for player I can be deleted from the problem.

Randomized strategies for player I* in the related matrix game will be denoted by \( \nu = \{v_{j_1}, \ldots, v_{j_k}\} \) where \( v_{j_1}, \ldots, v_{j_k} \) is the probability that player I* plays the pure strategy \( (Y_{j_1} \times \ldots \times Y_{j_k}) \) and \( \nu \) is a \( [\prod_{t=1}^{k} (J_t - J_{t-1})] \times 1 \) column vector. Let

\[ N = \{\nu | 0 \leq v_{j_1}, \ldots, v_{j_k} \leq 1 \text{ for each } (j_1, \ldots, j_k) \text{ and such that } \sum_{(j_1, \ldots, j_k)} v_{j_1}, \ldots, v_{j_k} = 1\}, \]

where the summation \( \sum_{(j_1, \ldots, j_k)} \) is over all \( [\prod_{t=1}^{k} (J_t - J_{t-1})] \) pure strategies for player I*. The payoff function is given by \( Q^* (\xi, \nu) = \xi^T K \nu \text{ where } K \text{ is } k \times \text{finite matrix game by } (N, E, Q^*). \)
Theorem 1 shows that all the solutions of the restricted matrix game 
(T, E, Q) are found by solving the related matrix game (N, E, Q*). Hence 
by Von Neumann's minimax theorem for finite matrix games (See, e.g., 
Karlin [5]), the restricted game has a value and optimal strategies for 
both players. This restricted game and the related matrix game were 
considered by Wesler [6]. He proved that playing the restricted game as 
player II is equivalent to playing the related matrix game as player II*.

THEOREM 1. The strategies *j and *k are optimal for the restricted game 
(T, E, Q) if and only if 

\[ *j = \pi_{j_t} \left( \sum_{j_1, \ldots, j_k} v_{j_1, \ldots, j_k} \right) \]

\[ j_t \text{ fixed} \]

and *k and *k are optimal strategies for the game (N, E, Q*). The games 
(T, E, Q) and (N, E, Q*) have the same value.

Proof: Let *k and *k be optimal strategies for the game (N, E, Q*) with 
value v. Then,

\[ Q^*(k, v) \leq Q^*(k, v) = v \leq Q^*(k, v) \]

for each v ∈ N and k ∈ E. It follows that \[ \sum_{i=1}^{M} \left( \sum_{t=1}^{k} \pi_{i}^{(t)} w_{i, j_t} \right) \xi_{i} \leq v \]

for each \( j_1, \ldots, j_k \) and hence

\[ \sum_{t=1}^{k} \max_{j_t} \left( \sum_{i=1}^{M} \pi_{i}^{(t)} w_{i, j_t} \xi_{i} \right) \leq v, \]
where \( j_t \) is always assumed to be a positive integer such that
\[
J_{t-1} + 1 \leq j_t \leq J_t.
\]
Equation (4) also implies that
\[
\sum_{t=1}^{k} \sum_{i=1}^{M} \pi(t) w_{ij}^{*} v_{j_{1}, \ldots, j_{k}} \geq v
\]
for \( i = 1, \ldots, M \). If we define \( \pi^{*} = \pi(t) \left( \sum_{j_{1}, \ldots, j_{k}} v_{j_{1}, \ldots, j_{k}} \right) \),
\( j_{t} \) fixed
it follows that
\[
\min_{j_{t}} \sum_{i=1}^{k} w_{ij}^{*} \pi_{j_{t}}^{*} \geq v.
\]
(6)

Note that (5) and (6) are really equalities, since
\[
\sum_{t=1}^{k} \max_{j_{t}} \left( \sum_{i=1}^{M} \pi(t) w_{ij}^{*} \xi_{i}^{*} \right) \geq \sum_{t=1}^{k} \sum_{i=1}^{M} \pi(t) w_{ij}^{*} \xi_{i}^{*} \geq \sum_{t=1}^{k} \sum_{i=1}^{M} w_{ij}^{*} \pi_{j_{t}}^{*}.
\]
Hence, we have
\[
\sum_{t=1}^{k} \max_{j_{t}} \left( \sum_{i=1}^{M} \pi(t) w_{ij}^{*} \xi_{i}^{*} \right) = \min_{j_{t}} \sum_{t=1}^{k} \sum_{i=1}^{M} w_{ij}^{*} \pi_{j_{t}}^{*}.
\]
(7)

Let \( \pi \) be an arbitrary member of \( T \), then
\[
v = \sum_{t=1}^{k} \max_{j_{t}} \pi(t) \left( \sum_{i=1}^{M} w_{ij}^{*} \xi_{i}^{*} \right) \geq \xi^{*} W \pi.
\]
Let $\xi$ be an arbitrary member of $\Xi$, then

$$v = \min_{i} \sum_{t=1}^{k} \sum_{j_t = J_{t-1} + 1}^{J_t} w_{ij_t} \pi_j \pi_{ij_t} \leq \xi W_{\pi}^*.$$ 

Therefore $\pi^*$ and $\xi^*$ are optimal strategies for $(T, \Xi, Q)$ and this game has value $v$.

Let $\pi^*$ and $\xi^*$ be optimal strategies for the restricted game $(T, \Xi, Q)$ with value $v$. Then

$$\xi^* W_{\pi^*}^* \geq v \geq \xi^* W_{\pi}^*$$ \hspace{1cm} (8)

for each $\pi \in T$ and $\xi \in \Xi$. Now defining

$$v_{j_1, \ldots, j_k}^* = \frac{\pi_{j_1}^* \cdots \pi_{j_k}^*}{\pi(1) \cdots \pi(k)},$$

it follows that

$$\sum_{(j_1, \ldots, j_k)} v_{j_1, \ldots, j_k}^* = \frac{\pi_{j_t}^*}{\pi(t)}.$$ \hspace{1cm} \text{fixed } j_t

Assume there exists a $(j_1^0, \ldots, j_k^0)$ such that $\sum_{i=1}^{M} \left( \sum_{t=1}^{k} \sum_{i,j_t^0} \pi(t) \right) \xi_i^* > v$.

Then if $\pi^{(0)}$ is the vector having $\pi(t)$ as its $j_t^0$ component $t = 1, \ldots, k$ and 0 as its other components, $\xi^* W_{\pi^{(0)}}^* > v$. This contradicts (8). Thus

$$\sum_{i=1}^{M} \sum_{t=1}^{k} \pi(t) w_{ij_t} \xi_i^* < v \text{ for each } (j_1, \ldots, j_k) \text{ and hence},$$

$$\sum_{i=1}^{M} \sum_{t=1}^{k} \pi(t) w_{ij_t} \xi_i^* < v \text{ for each } (j_1, \ldots, j_k) \text{ and hence},$$
\[ \xi^* K \nu \leq \nu. \]  

(9)

Thus it is seen that we actually solve for a T-minimax classification rule by finding an optimal strategy for the matrix game \((N, \varepsilon, Q^*)\).

3. CONSISTENCY OF CLASSIFICATION RULES

In practice the \(p_{ui}\)'s are not known and must be estimated by the sample proportions. Suppose \(n_u\) specimens are sampled from the \(u\)-th population and of these \(n_{ui}\) have \(X\) in the \(i\)-th cell. Then

\[ n = \sum_{u=1}^{N} n_u \text{ and } \sum_{i=1}^{R} n_{ui} = n_u. \]

Estimate \(p_{ui}\) by \(\hat{p}_{ui} = n_{ui}/n_u\). Let \(n \to \infty\) in such a way that \((n_u/n) \to \lambda_u > 0\) for \(u = 1, \ldots, N\) simultaneously. Then,

\[ P[\lim_{n} \hat{p}_{ui} = p_{ui} \text{ u = 1, \ldots, N, i = 1, \ldots, R}] = 1 \]

and it follows that

\[ P[\lim_{n} \hat{\bar{p}}_{m,j_1,\ldots,j_k} = \bar{p}_{m,j_1,\ldots,j_k} \text{ m = 1, \ldots, N, all (j_1, \ldots, j_k)}] = 1, \]

where

\[ \bar{p}_{m,j_1,\ldots,j_k} = \sum_{i=1}^{R} \sum_{t=1}^{k} \sum_{r=1}^{N} \sum_{r \neq j_t}^{s(m)} \sum_{t=1}^{N} \pi(t) \cdot p_{j_t,i}. \]

For simplicity in notation the "\(\cdot\)" will indicate dependence on the sample proportions \(\hat{p}_{ui}\) based on a sample of size \(n\).
For each \( \xi \) define

\[
M_{j_1, \ldots, j_k} (\xi) = \sum_{m=1}^{N^R} \xi_m P_{m, j_1, \ldots, j_k} \quad \text{and} \quad \rho(\xi)
\]

\[
= \max_{(j_1, \ldots, j_k)} M_{j_1, \ldots, j_k} (\xi). \quad \text{Likewise define} \quad \hat{M}_{j_1, \ldots, j_k} (\xi) \quad \text{and} \quad \hat{\rho}(\xi).
\]

Given the true \( p_{u_i} \)'s, a T-minimax rule is any \( \xi^* \) satisfying

\[
\rho(\xi^*) = \inf_{\xi \in \Xi} \rho(\xi). \quad \text{But the rule which is actually used is a} \quad \hat{\xi} \quad \text{satisfying}
\]

\[
\hat{\rho}(\hat{\xi}) = \inf_{\hat{\xi} \in \hat{\Xi}} \hat{\rho}(\hat{\xi}). \quad \text{The true maximum expected cost of using this rule is}
\]

\[
\rho(\hat{\xi}). \quad \text{One estimate of this is the apparent maximum expected cost which}
\]

\[
\hat{\rho}(\hat{\xi}). \quad \text{The consistency of}
\]

\[
\hat{\rho}(\hat{\xi}) \quad \text{implies that the}
\]

\[
\hat{\rho}(\hat{\xi}) \quad \text{is a}
\]

uniformly consistent estimate of \( \rho(\xi) \), that is, \( \lim_{n} \hat{\rho}(\hat{\xi}) = \rho(\xi) \)

uniformly in \( \xi \) = 1. It then follows that both the true maximum expected cost and the apparent maximum expected cost converge to the T-minimax expected cost, that is,

\[
P[\lim_{n} \hat{\rho}(\hat{\xi}) = \rho(\xi^*)] = 1
\]

and

\[
P[\lim_{n} \hat{\rho}(\hat{\xi}) = \rho(\xi^*)] = 1.
\]

We note that this consistency implies that any accumulation point \( \hat{\xi}^* \) of the sequence \( \{\hat{\xi}\} \) will satisfy \( \rho(\hat{\xi}^*) = \rho(\xi^*) \) and have a true maximum expected cost which equals the T-minimax expected cost. Hence it is
clear that if $\rho(\hat{\xi})$ has a unique minimum in $\xi$, then $P[\lim_{n} \hat{\xi} = \xi^*] = 1$.

If $\rho(\hat{\xi})$ does not have a unique minimum then $\{\hat{\xi}\}$ may not converge. This fact is illustrated by the following example.

Let $k = 1$, $N = J_1 = 2$, and $R = 2$. Take $C_{uv} = 1$ if $u \neq v$ and 0 if $u = v$. If $a_u$ denotes the action of classifying the specimen into $\theta_u$, the strategies and the expected costs are displayed in Table 1.

**TABLE 1**

PROBABILITIES, PURE STRATEGIES, AND EXPECTED COSTS OF A MATRIX GAME

<table>
<thead>
<tr>
<th>Population</th>
<th>Cell</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$p_{11}$</td>
<td>$p_{21}$</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>$p_{12}$</td>
<td>$p_{22}$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Action Taken When Observation is in Cell</th>
<th>Expected Probability of Misclassification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s^{(1)}$</td>
<td>$a_1$</td>
<td>$a_1$</td>
</tr>
<tr>
<td>$s^{(2)}$</td>
<td>$a_2$</td>
<td>$a_2$</td>
</tr>
<tr>
<td>$s^{(3)}$</td>
<td>$a_1$</td>
<td>$a_2$</td>
</tr>
<tr>
<td>$s^{(4)}$</td>
<td>$a_2$</td>
<td>$a_1$</td>
</tr>
</tbody>
</table>
Suppose that $p_{ui} = 1/2$ for $u = 1, 2$ and $i = 1, 2$ and that

$\pi = P[\text{occurrence of a specimen from } \theta_1]$. Then with $\pi^*$ denoting the optimal strategy for nature, the graph of the expected costs of the respective pure strategies is shown in Figure 1.

![Graph of expected costs of pure strategies](image)

**Fig. 1.**--Graph of the expected costs of the pure strategies of Table 1

It is clear that a T-minimax rule is any convex combination of the strategies $\underline{s}(3)$, $\underline{s}(4)$ and $\underline{s}(5)$ where $\underline{s}(5)$ is the strategy which plays $\underline{s}(1)$ with probability $1/2$ and $\underline{s}(2)$ with probability $1/2$.

Suppose that $n^*$ specimens are observed from each population so that $n = 2n^*$. A solution is then found for the estimated game.

**LEMMA 1.** Let $X^n_\sim$ and $Y^n_\sim$ be independent and identically distributed (iid) random variables, each binomial $(n, 1/2)$. If $p_n = \frac{X^n_\sim}{n}$ and $q_n = \frac{Y^n_\sim}{n}$ and $B_n = \{p_n < q_n < 1/2\}$ then $P[\lim_{n \to \infty} B_n] = 1$.

**Proof:** Write $X^n_\sim = X_1 + \ldots + X_n$ and $Y^n_\sim = Y_1 + \ldots + Y_n$ where $X_i$ and $Y_j$ are
iid and each is 1 with probability 1/2 and 0 with probability 1/2. Let 
\[ S_n = 2X_n - n \] and \[ S'_n = 2Y_n - n. \] Then \[ S_n = W_1 + \ldots + W_n \] and \[ S'_n = W'_1 + \ldots + W'_n \] where \( W_1, \ldots, W_n, W'_1, \ldots, W'_n \) are iid, each equal to 1 with probability 1/2 and -1 with probability 1/2. Then \( B_n = \{ S_n < -S'_n < 0 \} \).

Define a two dimensional symmetric random walk \( \omega_n = (S_n, S'_n) \). Let 
\[ \omega_0 = (0, 0) \] and if \( \omega_{n-1} = (s_{n-1}, s'_{n-1}) \), then \( \omega_n \) equals each of 
\[ (s_{n-1}+1, s'_{n-1}+1), (s_{n-1}+1, s'_{n-1}-1), (s_{n-1}-1, s'_{n-1}+1) \] and \( (s_{n-1}-1, s'_{n-1}-1) \) with probability 1/4, independently of \( s_{n-1}, s'_{n-1} \) and \( n \). Let 
\[ A_n = \{ \omega_n = (-2, 0) \}. \] Then \( A_n \subset B_n \). By Polya's Theorem (see, e.g., Feller [4], pg. 360) \( P[\lim \sup A_n] = 1 \), and hence \( P[\lim \sup B_n] = 1 \).

Identify \( p_n \) with \( \hat{p}_{11} \) and \( q_n \) with \( \hat{p}_{22} \). Then Lemma 1 shows that the event \( \hat{p}_{11} < \hat{p}_{22} < 1/2 \) will occur infinitely often with probability one. When this event does occur the graph of the apparent expected cost resembles the graph shown in Figure 2(a) and the T-minimax strategy is of the form \( \hat{\xi}_1^{(1)} = (0, \xi, 0, 1-\xi) \) with \( 0 < \xi < 1 \). Now identify \( p_n \) with \( \hat{p}_{21} \) and \( q_n \) with \( \hat{p}_{12} \). Lemma 1 implies that \( \{ \hat{p}_{21} < \hat{p}_{12} < 1/2 \} \) will occur infinitely often with probability one. When this event does occur the graph of the apparent expected cost is of the form given by the graph in Figure 2(b). The T-minimax strategy is then some \( \hat{\xi}_2^{(2)} = (\xi, 0, 1-\xi, 0) \) with \( 0 < \xi < 1 \). Combining the above results shows that in this example with probability one the T-minimax strategy will be of the form \( \hat{\xi}_1^{(1)} \) infinitely often and of the form \( \hat{\xi}_2^{(2)} \) infinitely often.

Thus \{\( \hat{\xi}_2 \)\} will not converge with probability one.
Fig. 2.—Illustration that the sequence \( \{\hat{x}\} \) need not converge with probability one

4. A MEDICAL EXAMPLE

The following example illustrates the application of Theorem 1 to a classification problem in which incomplete prior information is available. Patients experiencing renal hypertension may undergo surgery for either vascular repair in the area of the kidney or nephrectomy. Following surgery, the patient will be a member of one of three populations. Population \( \theta_1 \) consists of those patients who die within the week following
surgery. Population $\theta_2$ consists of those not in $\theta_1$ who do not show improvement in their condition of renal hypertension within the period of one year following surgery. Population $\theta_3$ will then consist of patients who survive the week following surgery and do show improvement during the year following surgery.

On the basis of the patient's medical history and a few clinical trials we would like to be able to predict his subsequent health state if he does undergo surgery. The variables used to classify the patient are presence of atherosclerosis, duration of hypertension, age at the onset of hypertension, and the serum creatinine level. The range of each of the last three variables was partitioned into 3, 2 and 2 regions respectively, so that the patient's observation vector will fall in one of $2^4$ cells. (Ideally, a finer partition of these variables would be desirable, but this was impractical here due to the relatively small number of patients available.) The partitioning used was $X_1 = 1 (0)$, if atherosclerosis is present (absent); $X_2 = 1, (2), (3)$ if the duration of hypertension is $\leq 40$ months, ($> 40$ and $< 130$), ($\geq 130$); $X_3 = 1 (0)$, if the age at the onset of hypertension is $\geq (<) 40$ years; $X_4 = 1 (0)$ if the serum creatinine level is $\geq (<) 1.5$.

The number of deaths occurring within the week following surgery is available through hospital records. However, the frequency of the occurrence of patients in the remaining two populations is not well documented since many of these patients are lost in follow-up. In the study conducted the frequency of patients from population $\theta_1$ was $1/10$. Thus we specified the incomplete prior information as
\[ T = \{ \pi | \pi_1 = 1/10, \pi_1 \geq 0 \text{ for } i = 2,3 \text{ and } \pi_2 + \pi_3 = 9/10 \}. \]

Data was available on 47 patients experiencing renal hypertension, 5, 18 and 24 of whom were known to have been members of populations \( \theta_1, \theta_2, \text{ and } \theta_3 \) respectively following surgery. This data enabled estimation of the \( p_{ui} \) probabilities. Theorem 1 was then applied assuming \( C_{uv} = 0 \) if \( u = v \) and 1 if \( u \neq v \) so that the expected cost of a classification rule was the probability of misclassification. Player II* has only two pure strategies in the related matrix game. Thus, an optimal classification rule was found by the simple graphical methods used in the example at the end of section 3. (See Karlin [5], page 41.)

A T-minimax procedure (it is not unique) for this problem, calculated from the data available, would classify the patients as follows. On the basis of \( \bar{X} = (X_1, X_2, X_3, X_4) \), classify as \( \theta_1 \) those patients falling into cell \((1,2,1,1)\); classify as \( \theta_2 \) those patients falling into cells \((0,1,0,1), (0,1,1,0), (0,2,1,0), (0,2,1,1), (0,3,0,0), (0,3,1,0), (1,3,0,0), (1,3,0,1), (1,3,1,1)\); classify as \( \theta_3 \) those patients falling into cells \((0,1,0,0), (0,1,1,1), (0,2,0,0), (1,1,1,0), (1,1,1,1), (1,2,0,0), (1,2,0,1), \text{ and } (1,2,1,0).\) The decision criterion does not tell us what to do for empty cells. (The data available fell into 18 of the 24 possible cells with \((0,2,0,1), (0,3,0,1), (0,3,1,1), (1,1,0,0), (1,1,0,1) \text{ and } (1,3,1,0) \) empty.) For any prior distribution over the population with a probability of 1/10 on \( \theta_1 \), and 9/10 spread over the other two populations, this procedure has an (estimated) expected probability of misclassification of .17. Applying the procedure to the 47 patients from which it was calculated shows that only 8 of 47 patients would be misclassified. Of course, this is
a very biased method of estimating the true expected probability of
misclassification. If we use the jackknife method of estimating the
probability of misclassification (for each patient, calculate a T-minimax
rule based on the other 46 patients, then see whether the deleted patient
is correctly classified) we find that 17 of the 47 patients are misclassified.
The difference of 8/47 and 17/47 is accounted for by the fact that many
of the cells contained only one patient. Thus, before one could strongly
recommend using such a rule, much more data is needed. (In particular, it
is impossible to get a good picture of the distribution in population \theta_1
from only 5 patients.) The authors are hopeful that this work will
encourage both the use of T-minimax classification rules as diagnostic
aids, and further development of such procedures.

ACKNOWLEDGMENTS

The authors would like to thank Professor I. R. Savage for his
helpful comments, Professor D. L. Solomon for pointing out reference [2],
and Professor P. E. Leaverton for providing the data used in section 4.
This paper is based on part of a dissertation written by R. H. Randles
under the direction of M. Hollander at Florida State University.
REFERENCES


The classification of specimens into one of \( N \) populations is considered when the observation on each specimen is a multivariate discrete random variable and when incomplete prior information is available concerning the occurrence of specimens from the various populations. The T-minimax decision criterion is applied and it is shown that an optimal classification procedure is the solution of a finite matrix game. The consistency of the T-minimax procedure is demonstrated when certain population probabilities are estimated by sample proportions. A medical example is considered.