The Annals of Statistics
1977, Vol. 5, No. 6, 1170-1182

GENERALIZED GROUP TESTING PROCEDURES

BY STEVEN F. ARNOLD

Pennsylvania State University

A person wishes to determine which, if any, of $n = \prod_{j=1}^{k} a_j$ i.i.d. random variables, $X(i_1, \cdots, i_k)$, $i_j = 1, \cdots, a_j$, lie in some specified set A. Such observations will be called unsafe. It is assumed that the density of the X's is known and that $Y_j(i_1, \cdots, i_j)$, the sum of all the X's whose first j indices are $i_1, \cdots, i_j$, can be measured as easily as the individual X's. In this paper, search procedures of the following form are studied. The person first measures Y_0, the sum of all the X's. On the basis of Y_0, he decides whether to stop, and classify all the X's as safe, or to continue and measure $Y_1(1), \cdots, Y_1(a_1 - 1)$ (and hence know $Y_1(a_1) = Y_0 - \sum_{i=1}^{a_1-1} Y_1(i)$). If he has decided to continue, he measures $Y_1(j)$. For each of $(Y_0, Y_1(j))$, he must decide whether to stop and classify as safe all X's whose first index is j, or to continue and measure $Y_2(j, 1), \cdots, Y_2(j, a_2 - 1)$ (and hence know $Y_2(j, a_2)$). He continues in this fashion until each X has either been classified safe or has been observed. Unlike most group testing problems, he is not restricted to procedures that will locate all the unsafe observations. Instead there is a loss function $L(x)$ measuring the loss if $X(i_1, \cdots, i_k) = x$ and is not observed. Let V_1 be the expected loss of a procedure (summed over all the X's), and let V_2 be the expected number of measurements. For each $0 \leqq p \leqq 1$, a class of rules $D(p)$ is defined such that if a procedure is in $D(p)$, it minimizes $pV_1 + (1 - p)V_2$, and conversely, if a procedure minimizes $pV_1 + (1 - p)V_2$, then there is a rule in $D(p)$ that leads to the same decisions a.e. The union of the $D(p)$ is shown to be an essentially complete class of rules. A simpler form for the rules in $D(p)$ is derived for the case where the loss function is nondecreasing. More specific calculations are given for the case where the X's are normally distributed, and $L(x)$ is the indicator function for the set $\{x \geqq d\}$.

1. Introduction. In this paper, the following quality control problem is considered. A person wishes to determine which, if any, of $n = \prod_{j=1}^{k} a_j$ independently, identically distributed (i.i.d.) random variables, $X(i_1, \cdots, i_k)$, $i_j = 1, \cdots, a_j$, lie in some specified set A. Such observations will be called unsafe. It is assumed that the density of the X's is known and that $Y_j(i_1, \cdots, i_j)$, the sum of all the X's whose first j indices are $i_1, \cdots, i_j$, can be measured as easily as the individual X's (perhaps by mixing; see the example below). Generalized group testing procedures of the following form are studied. The person first measures Y_0, the sum of all the X's. On the basis of Y_0, he decides whether to stop and classify all the X's as safe or to continue and measure $Y_1(1), \cdots, Y_1(a_1 - 1)$, and hence know $Y_1(a_1) = Y_0 - \sum_{i=1}^{a-1} Y(i)$. He has now completed stage 0. If he has decided to continue, he measures the $Y_1(j)$. For each of $(Y_0, Y_1(j))$, he must decide whether to stop and classify as safe all the X's whose

Received May 1975; revised February 1977.
AMS 1970 *subject classifications*. Primary 62L15; Secondary 62N10, 62C99.
Key words and phrases. Group testing, sequential decision theory, quality control.

first index is j, or to continue and measure $Y_2(j, 1), \cdots, Y_2(j, a_2 - 1)$, and hence know $Y_2(j, a_2)$. He has now completed stage 1. He continues in this fashion until each X has either been classified as safe, or has been observed.

The following example should help to clarify the previous paragraph, as well as motivate the remainder of the paper. Suppose that a state agent has to monitor the algae level in each of twelve lakes. In particular, he would like to know if any of the lakes has level greater than some constant d. He takes a water sample from each lake. The samples are indexed so that $X(i, j, k)$, $i = 1, 2$, $j = 1, 2, 3$, $k = 1, 2$ are the algae concentrations from each water sample. Suppose, in addition, the agent knows from previous experience that the $X(i, j, k)$ are i.i.d. with some known marginal distribution. One possible procedure is to measure each of the 12 water samples separately. The following group testing procedure is suggested as an alternative, which would require fewer measurements. Define

$$Y_0 = \sum_i \sum_j \sum_k X(i, j, k), \qquad Y_1(i) = \sum_j \sum_k X(i, j, k),$$
$$Y_2(i, j) = \sum_k X(i, j, k).$$

If the measurement error is negligible, as will be assumed, he can measure Y_0 as easily as any $X(i, j, k)$ by taking equal quantities from each of the 12 samples, mixing them, and finding the algae content of the mixture. In a similar way, he can determine $Y_1(i)$ and $Y_2(i, j)$ each in one measurement. Therefore, the researcher measures Y_0. If $Y_0 \leqq b_0$ for some constant b_0 he declares that all the water is safe. If not, he measures $Y_1(1)$, and hence knows $Y_1(2) = Y_0 - Y_1(1)$. If $Y_1(1) \leqq b_1$, he declares that the $X(1, j, k)$ are all safe for $j = 1, 2, 3$, $k = 1, 2$. If $Y_1(1) > b_1$, he measures $Y_2(1, 1)$ and $Y_2(1, 2)$, and hence knows $Y_2(1, 3) = Y_1(1) - Y_2(1, 1) - Y_2(1, 2)$. If $Y_2(1, 1) \leqq b_2$ he decides that the $X(1, 1, k)$, $k = 1, 2$ are both safe. If not, he measures $X(1, 1, 1)$, and therefore knows $X(1, 1, 2)$. He follows the same procedure for $Y_2(1, 2)$ and $Y_2(1, 3)$. He then looks at $Y_1(2)$. If $Y_1(2) \leqq b_1$, he declares all the $X(2, j, k)$ to be safe. If not, he measures $Y_2(2, 1)$ and $Y_2(2, 2)$ and proceeds as for $Y_2(1, 1)$ and $Y_2(1, 2)$.

COMMENT 1. If the $X(i, j, k)$ have Bernoulli distributions, $d = 0$, and $b_0 = b_1 = b_2 = 0$, this procedure is a group testing procedure (see Kumar and Sobel (1971)). Therefore these procedures will be called generalized group testing procedures.

COMMENT 2. The generalized group testing procedure takes at most 12 measurements (or n measurements in the general situation), which is the number required to measure all the $X(i, j, k)$ individually.

COMMENT 3. In the traditional group testing procedures, no unsafe X's would go undetected. For many distributions (such as the normal distribution), the only generalized group testing procedure that would guarantee detection of all unsafe observations would be the one that continues at every point (i.e., $b_i = -\infty$). For this procedure there would be no benefit to the group testing.

Because of Comment 3, the restriction that all unsafe observations be detected is replaced with a loss function, $L(x)$, which measures the loss if $X(i_1, \cdots, i_k) = x$ and is undetected. It is assumed that the loss if several unsafe observations go undetected is the sum of the individual losses. For the example given above, a reasonable loss function might be $L(x) = \min(0, x - d)$. Actually for much of this paper, no assumptions are made about $L(x)$, so that there need not be any safe X's, but perhaps some for which the additional information found by measuring them is not worth the cost of measurement.

Let V_1 be the expected loss for a particular procedure, summed over all the X's, and let V_2 be the expected number of measurements for that procedure. In the following sections, sequential decision theory is used to find procedures that make both V_1 and V_2 small. In Section 2 the problem is set up (permitting randomized decisions at each stage). It is shown in Section 3 that the procedures (like those in the example) in which the decisions at the jth stage depend only on Y_j (not on $Y_0, \cdots, Y_{j-1}$) form an essentially complete class. In Section 4, for each $0 \leqq p \leqq 1$, a class of rules $D(p)$ is defined such that:

1. If a procedure is in $D(p)$, it minimizes $pV_1 + (1 - p)V_2$.
2. If a procedure minimizes $pV_1 + (1 - p)V_2$, then there is a rule in $D(p)$ that leads to the same decisions with probability one.
3. $\bigcup_{0 \leqq p \leqq 1} D(p)$ is essentially complete.

COMMENT. In the usual formulation of the sequential decision problems, the goal is to find rules that minimize $V_1 + qV_2$ for some fixed q. Clearly, finding $D(1/(1 + q))$ is equivalent to that problem.

For many problems involving continuous densities, the rules in $D(p)$ for a particular p are equal a.e. In Section 5 it is shown that the rules in $D(p)$ can usually be put into a much simpler form when the loss function is nondecreasing, a natural assumption in problems of water, air or paper purity such as that in the example. In Section 6, the special case in which the X's are normally distributed and $L(x)$ is the indicator function of the set $\{X \geqq d\}$ is considered. Unfortunately, only the case $k = 2$ is considered, because other cases involve the evaluation of trivariate normal integrals. Specific numerical values are given for the case $d = 2$, $a_1 = 2$, $a_2 = 3$ and the case $d = 2$, $a_1 = 3$, and $a_2 = 2$. In Section 7 there is a short discussion about optimal procedures when the i.i.d. assumption is dropped.

In the traditional group testing problem, once the a_j's are selected, the procedure is completely specified, so the goal is to find the optimal choice of the a_j's. However, for each choice of the a_j's, there are infinitely many generalized group testing procedures. In this paper, the optimal rules for a particular sequence of the a_j's are determined. It is assumed that the order of the a_j's is given. For example, if the person has 6 observations, he must decide whether to divide them into 2 sets of 3 observations or into 3 sets of 2. In Section 6 a numerical example is worked 2 ways, once as 2 sets of 3 observations, and once

as 3 sets of 2 observations, and the results are compared. For p equal to each of .95, .99, .995, the Bayes risk of the optimal procedure for 2 sets of 3 observations is about the same as the Bayes risk for 3 sets of 2 observations.

2. Setting up the problem. The specifications for this problem consist of 4 parts.

1. There are $n=\prod_{j=1}^{k} a_j$ i.i.d. random variables $X(i_1, \cdots, i_k)$, $i_j = 1, \cdots, a_j$, having a known distribution. For simplicity, it will be assumed that this distribution is either discrete or continuous with density function $f(x)$. Let

$$(2.1) \qquad \begin{aligned} Y_{k-1}(i_1, \cdots, i_{k-1}) &= \sum_{i_k=1}^{a_k} X(i_1, \cdots, i_k), \\ Y_j(i_1, \cdots, i_j) &= \sum_{i_{j+1}=1}^{a_{j+1}} Y_{j+1}(i_1, \cdots, i_{j+1}), \qquad j = 0, 1, \cdots, k-2. \end{aligned}$$

2. There is a loss function $L(x)$ which represents the loss if $X(i_1, \cdots, i_k)$ is not observed, and is equal to x.

3. Let Ψ be the class of procedures Φ defined in the following manner. Each procedure Φ is a set of k measurable functions,

$$\Phi = (\Phi_0(y_0), \Phi_1(y_0, y_1), \cdots, \Phi_{k-1}(y_0, \cdots, y_{k-1}))$$

where $\Phi_j(y_0, \cdots, y_j)$ represents the probability of continuing at the jth stage (to observe $Y_{j+1}(i_1, \cdots, i_{j+1})$) when $Y_0 = y_0, \cdots, Y_j(i_1, \cdots, i_j) = y_j$.

4. Let

$$V_1(\Phi) = \sum_{i_1=1}^{a_1} \cdots \sum_{i_k=1}^{a_k} E[L(X(i_1, \cdots, i_k))(1 - \prod_{j=0}^{k-1} \Phi_j(Y_0, \cdots, Y_j(i_1, \cdots, i_j)))],$$
$$V_2(\Phi) = 1 + \sum_{m=0}^{k-1} (a_{m+1} - 1) \sum_{i_1=1}^{a_1} \cdots \sum_{i_m=1}^{a_m} E[\prod_{j=0}^{m} \Phi_j(Y_0, \cdots, Y_j(i_1, \cdots i_j,))],$$

$V_1(\Phi)$ represents the expected loss from using the procedure Φ. To see this, note that

$$(1 - \prod_{j=0}^{k-1} \Phi_j(Y_0, \cdots, Y_j(i_1, \cdots, i_j)))$$

is the probability of not observing $X(i_1, \cdots, i_k)$ when the X's are known. Similarly, $V_2(\Phi)$ represents the expected number of measurements needed for procedure Φ, as can be seen by noting that there is one measurement at the first stage, $(a_1 - 1)E\Phi_0(Y_0)$ expected measurements at the second stage (since the last measurement can be determined from Y_0 and the first $a_1 - 1$ measurements). Similarly there are

$$(a_2 - 1) \sum_{i_1=1}^{a_1} E[\Phi_0(Y_0)\Phi_1(Y_0, Y_1(i_1))]$$

expected measurements at the third stage, etc.

The objective of this paper is to find procedures that make V_1 and V_2 small. To simplify the notation, let $Y_j = Y_j(1, \cdots, 1)$, $j = 0, \cdots, k-1$, and let $Y_k = X(1, \cdots, 1)$, and let $\tilde{Y}_j = (Y_0, Y_1, \cdots, Y_j)$. By symmetry, since the $X(i_1, \cdots, i_k)$ are i.i.d.,

$$(2.2) \qquad \begin{aligned} V_1(\Phi) &= c - a_k E[L(Y_k) \prod_{j=0}^{k-1} a_j \Phi_j(\tilde{Y}_j)] \\ V_2(\Phi) &= 1 + \sum_{m=0}^{k-1} (a_{m+1} - 1) E[\prod_{j=0}^{m} a_j \Phi_j(\tilde{Y}_j)] \end{aligned}$$

where $a_0 = 1$, $c = (\prod_{j=1}^{k} a_j) EL(Y_k)$.

2.2. *A preliminary result.* The following lemma and its corollaries are useful in later sections. Let

$$Z_i = Y_i - Y_{i+1}, \qquad i = 0, \cdots, k-1 \qquad\qquad Z_k = Y_k.$$

LEMMA 1. *The Z_i are independent.*

PROOF. By definition, Z_0 is the sum of all the X's whose first index is not 1. Z_1 is the sum of all the X's whose first index is 1 and whose second index is not 1. Z_2 is the sum of all X's whose first 2 indices are 1, and whose third index is not 1, etc. Each Z_i is a sum of different X's. Since the X's are independent the Z's are. □

COROLLARY 1. *The conditional distribution of $\tilde{Y}_{k-2}$ given (Y_{k-1}, Y_k) is the same as the conditional distribution of $(\tilde{Y}_{k-2})$ given Y_{k-1}.*

PROOF. Let Z_i have density $g_i(z_i)$. By Lemma 1, the joint density of the Z_i is $\prod_{i=0}^{k} g_i(z_i)$. The transformation from $(Y_0, \cdots, Y_k)$ to $(Z_0, \cdots, Z_k)$ is an invertible transformation whose Jacobian is 1. The joint density of the Y_i is therefore

$$g_0(y_0 - y_1)g_1(y_1 - y_2) \cdots g_{k-1}(y_{k-1} - y_k)g_k(y_k). \tag{2.3}$$

The corollary follows directly. □

COROLLARY 2. *The conditional distribution of Y_{j+1} given $\tilde{Y}_j$ is the same as the conditional distribution of Y_{j+1} given Y_j.*

PROOF. From (2.3) the density of $\tilde{Y}_{j+1}$ is

$$g_0(y_0 - y_1) \cdots g_j(y_j - y_{j+1})h(y_{j+1})$$

for some function of h. The corollary again follows. □

3. An essentially complete class. In order to find procedures Φ that make both $V_1(\Phi)$ and $V_2(\Phi)$ small, V_1 and V_2 are considered as risks in a 2-parameter decision theory problem. The usual definitions are used. A class D of procedures is *essentially complete* if for any procedure Φ, there exists $\Phi^* \in D$ such that $V_1(\Phi^*) \leqq V_1(\Phi)$ and $V_2(\Phi^*) \leqq V_2(\Phi)$. A procedure Φ is *admissible* if there is no other procedure Φ^* such that $V_1(\Phi^*) \leqq V_1(\Phi)$, $V_2(\Phi^*) \leqq V_2(\Phi)$ and $V_1(\Phi^*) + V_2(\Phi^*) < V_1(\Phi) + V_2(\Phi)$. A procedure Φ is *Bayes with respect to the prior* $(p, 1-p)$ if it minimizes $pV_1(\Phi) + (1-p)V_2(\Phi)$.

COMMENT. This usage of the word Bayes is not standard, since p does not represent a probability. The reason for using the term is the following. It will be shown that the set of all pairs $(V_1(\Phi), V_2(\Phi))$ is a convex set. The complete class theorem can be used to conclude that the Bayes rules form a complete class. This usage is consistent with the usual geometric interpretation for finite θ (see Ferguson (1967), pages 34–43).

DEFINITION. Let Δ be the class of all procedures $\Phi = (\Phi_0, \cdots, \Phi_{k-1})$ such that Φ_j is only a function of Y_j.

THEOREM 1. *Δ is an essentially complete class for Ψ.*

PROOF. Let $\Phi = (\Phi_0, \cdots, \Phi_{k-1})$. Define $\Phi^* = (\Phi_0^*, \cdots, \Phi_{k-1}^*) \in \Delta$ recursively by

$$\Phi_0^*(Y_0) = \Phi_0(Y_0)\,,$$

$$\Phi_j^*(Y_j) = \frac{E(\prod_{i=0}^{j} \Phi_i \mid Y_j)}{E(\prod_{i=0}^{j-1} \Phi_i^* \mid Y_j)} \quad \text{if} \quad E(\prod_{i=0}^{j-1} \Phi_i^* \mid Y_j) \neq 0\,, \tag{3.1}$$

$$= 0 \quad \text{if} \quad E(\prod_{i=0}^{j-1} \Phi_i^* \mid Y_j) = 0\,.$$

Then $E(\prod_{i=0}^{j} \Phi_i \mid Y_j) = E(\prod_{i=0}^{j} \Phi_i^* \mid Y_j), j = 0, \cdots, k-1$. Therefore $V_2(\Phi) = V_2(\Phi^*)$. Similarly, using Corollary 1,

$$\begin{aligned} E(L(Y_k) \textstyle\prod_{j=1}^{k-1} \Phi_j \mid (Y_k, Y_{k-1})) &= L(Y_k) E(\textstyle\prod_{j=1}^{k-1} \Phi_j \mid Y_{k-1}) \\ &= L(Y_k) E(\textstyle\prod_{j=1}^{k-1} \Phi_j^* \mid Y_{k-1}) \\ &= E(L(Y_k) \textstyle\prod_{j=1}^{k-1} \Phi_j^* \mid (Y_k, Y_{k-1}))\,. \end{aligned}$$

Therefore $V_1(\Phi) = V_1(\Phi^*)$. □

4. Bayes procedures. In this section, a class of procedures $D(p)$ is defined for each p, $0 \leqq p \leqq 1$. It is shown that these procedures are all Bayes with respect to the prior $(p, 1-p)$, and conversely, that if any procedure in Δ is Bayes with respect to the prior $(p, 1-p)$, then there is a rule in $D(p)$ that leads to the same results a.e. It is then shown that the union of the $D(p)$ is essentially complete for Ψ. The method used in this paper for finding Bayes rules is very similar to that used for finding Bayes rules for truncated sequential decision problems (see Ferguson (1967), pages 314–318).

Let

$$R(\Phi, p) = pV_1(\Phi) + (1-p)V_2(\Phi) \tag{4.1}$$

be the Bayes risk of the procedure with respect to the prior $(p, 1-p)$. For any procedure $\Phi = (\Phi_0, \cdots, \Phi_{k-1}) \in \Delta$, define U_i recursively by

$$U_{k-1}(p, y) = -(1-p)(a_k - 1) + pa_k E(L(Y_k) \mid Y_{k-1} = y)\,. \tag{4.2}$$

$$\begin{aligned} U_j(\tilde{\Phi}_{j+1}; p, y) = &-(1-p)(a_{j+1} - 1) \\ &+ a_{j+1} E[U_{j+1}(\tilde{\Phi}_{j+1}; p, y_{j+1}) \Phi_{j+1}(Y_{j+1}) \mid Y_j = y] \end{aligned} \tag{4.3}$$

where $\tilde{\Phi}_j = (\Phi_j, \Phi_{j+1}, \cdots, \Phi_{k-1})$.

Although no use is made of this interpretation in the following paragraphs, for motivation it is helpful to think of U_j as representing the expected cost of stopping relative to continuing. The following heuristic argument indicates why this interpretation is correct. If the procedure is continued at stage j, $a_{j+1} - 1$ new measurements must be made, and the cost is therefore $(a_{j+1} - 1)(1-p)$. Now suppose the procedure stops at stage j. It costs nothing if the procedure would have stopped at stage $j+1$. However, the cost is U_{j+1} if it would have continued. Therefore the cost of stopping is $\Phi_{j+1} U_{j+1}$, and the expected cost is $E\Phi_{j+1} U_{j+1} \mid Y_j = y$. The relative cost is the difference between the cost of continuing and the cost of stopping.

LEMMA 2. *Let* $\Phi = (\Phi_0, \cdots, \Phi_{k-1}) \in \Delta$. *For all* $i = 0, 1, \cdots, k-1$,

$$R(\Phi, p) = pc + (1-p) + (1-p)\sum_{m=0}^{i-1}(a_{m+1} - 1)E\prod_{j=0}^{m} a_j\Phi_j - E[U_i(\tilde{\Phi}_{i+1}; p, Y_i)\prod_{j=0}^{i} a_j\Phi_j(Y_j)]\,. \tag{4.4}$$

PROOF. The lemma is proved by induction on $j = k - i$. It is first proved for $j = 1$ ($i = k - 1$). From (2.2)

$$R(\Phi, p) = pc + (1-p) + (1-p)\sum_{m=0}^{k-2}(a_{m+1} - 1)E\prod_{j=0}^{m} a_j\Phi_j + E(E(\prod_{j=0}^{k-1} a_j\Phi_j(\tilde{Y}_j))[(1-p)(a_k - 1) - pa_kL(Y_k)]\,|\,\tilde{Y}_{k-1})\,. \tag{4.5}$$

However,

$$\begin{aligned} &E(\prod_{j=0}^{k-1} a_j\Phi_j(\tilde{Y}_j)[(1-p)(a_k-1) - pa_kL(Y_k)]\,|\,\tilde{Y}_{k-1}) \\ &\quad = \prod_{j=0}^{k-1} a_j\Phi_j(\tilde{Y}_j)[(1-p)(a_k-1) - pa_kEL(Y_k)\,|\,\tilde{Y}_{k-1}] \\ &\quad = -\prod_{j=0}^{k-1} a_j\Phi_j(\tilde{Y}_j)U_{k-1}(p, Y_{k-1}) \end{aligned} \tag{4.6}$$

where the last equality follows from Corollary 2 to Lemma 1. The proof for the case $j = 1$ is completed by substituting the result of (4.6) into (4.5). The induction step follows similarly and is omitted. □

Let a rule $\Phi = (\Phi_0, \cdots, \Phi_{k-1})$ be defined recursively by the following scheme. Let

$$\begin{aligned} \Phi_j(Y_j) &= 1 && \text{if} \quad U_j(\tilde{\Phi}_{j+1}; p, Y_j) > 0 \\ &= b_j(Y_j) && \text{if} \quad U_j(\tilde{\Phi}_{j+1}; p, Y_j) = 0 \\ &= 0 && \text{if} \quad U_j(\tilde{\Phi}_{j+1}; p, Y_j) < 0\,. \end{aligned} \tag{4.7}$$

Let $D(p)$ be the class of all rules satisfying (4.7) for some $b_j(Y_j)$.

COMMENT. For many problems involving continuous distributions, all the rules in $D(p)$ will be equal a.e., since the points of randomization have measure 0 (see Section 6 for an example).

THEOREM 2. *If* $\Phi \in D(p)$, *then* Φ *minimizes* $R(\Phi, p)$. *That is,* Φ *is Bayes with respect to the prior* $(p, 1-p)$.

PROOF. By Lemma 2

$$R(\Phi, p) = pc + (1-p) + (1-p)\sum_{m=0}^{i-1}(a_{m+1} - 1)E\prod_{j=0}^{m} a_j\Phi_j - a_iE[\Phi_i(Y_i)U_i(\tilde{\Phi}_{i+1}; p, Y_i)E(\prod_{j=0}^{i-1} a_j\Phi_j\,|\,Y_i)]\,.$$

Only the last term involves Φ_i, $a_iE(\prod_{j=0}^{i-1} a_j\Phi_j\,|\,Y_i) \geqq 0$, and U_i does not depend on Φ_i. Therefore, by a proof similar to the Neyman–Pearson lemma, Φ_i satisfying (4.7) minimizes $R(\Phi, p)$. □

COROLLARY. *If* $\Phi \in D(p)$, $0 < p < 1$, *then* Φ *is admissible.*

One might expect that any Φ minimizing $R(\Phi, p)$ would have to satisfy (4.7) a.e. Unfortunately, this is not true, because in some situations there are Bayes procedures such that

$$P(\prod_{i=0}^{j}\Phi_i = 0\,|\,Y_{j+1} = y) = 1 \tag{4.8}$$

for some y. If (4.8) is satisfied, it does not matter how $\Phi_{j+1}(y)$ is defined, since $Y_{j+1} = y$ will be observed with probability 0. However, if

$$\prod_{i=0}^{j} \Phi_i = \prod_{i=0}^{j} \Phi_i^* , \qquad \text{for all } j$$

then $\Phi = (\Phi_0, \cdots, \Phi_{k-1})$ and $\Phi^* = (\Phi_0^*, \cdots, \Phi_{k-1}^*)$ really determine the same procedure in that the actions are the same under both procedures.

THEOREM 3. *Suppose* $\Phi = (\Phi_0, \cdots, \Phi_{k-1}) \in \Delta$ *and* Φ *minimizes* $R(\Phi, p)$. *There exists* $\Phi^* = (\Phi_0^*, \cdots, \Phi_{k-1}^*) \in D(p)$ *such that for all* j,

$$\prod_{i=0}^{j} \Phi_i = \prod_{i=0}^{j} \Phi_i^* \quad \text{a.e.} \tag{4.9}$$

$$V_1(\Phi) = V_1(\Phi^*) , \qquad V_2(\Phi) = V_2(\Phi^*) . \tag{4.10}$$

PROOF. By Lemma 2,

$$\begin{aligned} R(\Phi, p) = pc + (1 - p) + (1 - p) \sum_{m=0}^{k-2} (a_{m+1} - 1) E \prod_{j=0}^{m} a_j \Phi_j \\ - a_{k-1} E[\Phi_{k-1}(Y_{k-1}) U_{k-1}(p, Y_{k-1}) (E \prod_{j=0}^{k-2} a_j \Phi_j \mid Y_{k-1})] . \end{aligned} \tag{4.11}$$

Let $A = \{Y_{k-1} : E(\prod_{i=0}^{k-2} \Phi_i \mid Y_{k-1}) \neq 0\}$. Φ minimizes R, only the last term involves Φ_{k-1}, and $E \prod_{i=0}^{k-1} \Phi_i \mid Y_{k-1} > 0$ on A. Therefore, there is a Φ_{k-1}^* of the form of (4.7) such that $\Phi_{k-1}^* = \Phi_{k-1}$ for almost all $Y_{k-1} \in A$, by proof similar to the proof of the converse of the Neyman–Pearson lemma. Therefore

$$|\Phi_{k-1} - \Phi_{k-1}^*| E(\prod_{i=0}^{k-2} \Phi_i \mid Y_{k-1}) = 0 \tag{4.12}$$

for almost all $Y_{k-1} \in A$. But (4.12) holds trivially for $Y_{k-1} \notin A$. Since $\Phi_i \geqq 0$, (4.12) implies that

$$(\prod_{i=0}^{k-2} \Phi_i) \Phi_{k-1} = (\prod_{i=0}^{k-2} \Phi_i) \Phi_{k-1}^* \quad \text{a.e.} \tag{4.13}$$

(4.13) implies that the sequence $(\Phi_0, \cdots, \Phi_{k-1}^*)$ also minimizes R. Therefore, by a similar argument, there is a Φ_{k-2}^* of the form

$$\begin{aligned} \Phi_{k-2}^*(Y_{k-2}) &= 1 && \text{if } U_{k-2}(\Phi_{k-1}^*; Y_{k-2}, p) > 0 \\ &= b_{k-2}(Y_{k-2}) && \text{if } U_{k-2}(\Phi_{k-1}^*; Y_{k-2}, p) = 0 \\ &= 0 && \text{if } U_{k-2}(\Phi_{k-1}^*; Y_{k-2}, p) < 0 \end{aligned}$$

such that

$$(\prod_{j=0}^{k-3} \Phi_i) \Phi_{k-2} = (\prod_{j=0}^{k-3} \Phi_i) \Phi_{k-2}^* \quad \text{a.e.} \tag{4.14}$$

Continuing in this manner, a sequence $\Phi^* = (\Phi_0^*, \cdots, \Phi_{k-1}^*)$ is discovered such that $\Phi^* \in D(p)$ and

$$(\prod_{j=0}^{i-1} \Phi_j) \Phi_i = (\prod_{j=0}^{i-1} \Phi_j) \Phi_i^* \quad \text{a.e.} \tag{4.15}$$

(4.15) implies (4.9) which implies (4.10). □

The last result in this section is to show that

$$\Xi = \bigcup_{0 \leqq p \leqq 1} D(p) \tag{4.16}$$

is essentially complete for Δ and hence for Ψ (see Theorem 1). The only detail remaining to prove that result is the following lemma.

LEMMA 3. *The set of all* $(V_1(\Phi), V_2(\Phi))$ *for* $\Phi \in \Delta$ *is a convex set.*

PROOF. Let $\Phi = (\Phi_0, \cdots, \Phi_{k-1})$ and $\Phi^* = (\Phi_0^*, \cdots, \Phi_{k-1}^*)$ be two rules in Ψ. Define $\Phi^{**} = (\Phi_0^{**}, \cdots, \Phi_{k-1}^{**})$ by

$$\Phi_0^{**} = \alpha\Phi_0 + (1 - \alpha)\Phi_0^* ,$$

$$\Phi_j^{**} = \frac{\alpha \prod_{i=0}^{j} \Phi_i + (1 - \alpha) \prod_{i=0}^{j} \Phi_i^*}{\prod_{i=0}^{j-1} \Phi_i^{**}} \quad \text{if} \quad \prod_{j=0}^{i-1} \Phi_j^{**} \neq 0$$

$$= 0 \quad \text{if} \quad \prod_{j=0}^{i-1} \Phi_j^{**} = 0 .$$

Then $\prod_{i=0}^{j} \Phi_i^{**} = \alpha \prod_{i=0}^{j} \Phi_i + (1 - \alpha) \prod_{k=0}^{j} \Phi_i^*$ for all j, so that

$$V_1(\Phi^{**}) = \alpha V_1(\Phi) + (1 - \alpha)V_1(\Phi^*) ,$$

$$V_2(\Phi^{**}) = \alpha V_2(\Phi) + (1 - \alpha)V_2(\Phi^*) . \qquad \square$$

THEOREM 4. Ξ *is an essentially complete class for* Ψ.

PROOF. By the complete class theorem and Lemma 3, the set of Bayes rules is complete for Δ. By Theorem 3, Ξ is essentially complete for Δ, and by Theorem 1 is also, therefore, essentially complete for Ψ. $\square$

5. Nondecreasing loss function. For many generalized group testing problems, it would be natural to let $L(x)$ be a nondecreasing function (e.g., the examples mentioned in the introduction). Intuitively, it might be expected that $U_j(\Phi; p, y)$ would also be a nondecreasing function of y, and that therefore if $\Phi = (\Phi_0, \cdots, \Phi_{k-1}) \in D(p)$, then

$$\begin{aligned} \Phi_j(Y_j) &= 1 && \text{if} \quad Y_j > c_j \\ &= b_j(Y_j) && \text{if} \quad d_j \leqq Y_j \leqq c_j \\ &= 0 && \text{if} \quad Y_j < d_j , \end{aligned} \tag{5.1}$$

where

$$c_j = \inf (y : U_j(\Phi; p, y) > 0) , \qquad d_j = \sup (y : U_j(\Phi; p, y) < 0) , \tag{5.2}$$

since large values of Y_k would lead to large values of Y_j. In this section, sufficient conditions on the density f of Y_k are given to guarantee that U_j is nondecreasing and hence that if $\Phi \in D(p)$, then Φ satisfies (5.1) and (5.2). This greatly simplifies the calculation of rules in $D(p)$. (See Section 6 for an example.)

The following example shows that U_j need not be nondecreasing for all densities $f(x)$. Let $k = 1$, $a_1 = 2$ and $L(x)$ be the indicator of $\{X \geqq 1\}$. Let $X(1)$ and $X(2)$ be a sample from the density

$$f(0) = .01 , \qquad f(1) = .01 , \qquad f(2) = .02 , \qquad f(3) = .96 .$$

As usual, let $Y_0 = X(1) + X(2)$. By (4.2)

$$U_0(p, y) = -(1 - p) + 2pP(X \geqq 1 \mid Y_0 = y) .$$

Let $g(y) = P(X \geqq 1 \mid Y_0 = y)$. Then $g(y)$ is given by

y	0	1	2	3	4	5	6
$g(y)$	0	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{100}{196}$	1	1	1 .

Therefore U_j would only be a nondecreasing function if $p = 0$. In fact if $\frac{5}{11} \leqq p \leqq \frac{196}{396}$, then Φ_0 would be the indicator of the set $\{2, 4, 5, 6\}$, which does not satisfy (5.1) and (5.2).

We now return to the problem of finding conditions sufficient to guarantee that the Bayes rules satisfy (5.1) and (5.2). Let $W_1, \cdots, W_n$ be a sample from the density function $g(w)$. Let $Z = \sum_{i=1}^{n} w_i$.

DEFINITION. The density $g(w)$ has property $M(n)$ if for all nondecreasing functions $h(w)$, $E(h(W_1) \mid Z = z)$ is a nondecreasing function of z.

THEOREM 5. *Let $L(x)$ be a nondecreasing function of x, and let $\Phi = (\Phi_0, \cdots, \Phi_{k-1}) \in D(p)$. If the density of Y_j has property $M(a_j)$, for $j = 1, \cdots, k$, then $U_j(\Phi; p, y)$ is a nondecreasing function of y (and hence Φ_j satisfies (5.1) and (5.2)).*

PROOF. U_{k-1} is trivially nondecreasing since the density of Y_k has property $M(a_k)$. To see the general case, note that

$$U_j(\tilde{\Phi}_{j+1}; p, y) = a_{j+1} E[\Phi_{j+1}(Y_{j+1}) U_{j+1}(\tilde{\Phi}_{j+2}; p, Y_{j+1}) \mid Y_j = y] - (1 - p)(a_{j+1} - 1),$$

and that

$$\begin{aligned} \Phi_{j+1}(Y_{j+1}) &= 1 && \text{if} \quad U_{j+1}(\tilde{\Phi}; p, Y_{j+1}) > 0 \\ &= b_{j+1}(Y_{j+1}) && \text{if} \quad U_{j+1}(\tilde{\Phi}; p, Y_{j+1}) = 0 \\ &= 0 && \text{if} \quad U_{j+1}(\tilde{\Phi}; p, Y_{j+1}) < 0 . \end{aligned}$$

Therefore, $\Phi_{j+1} U_{j+1}$ is a nondecreasing function of Y_{j+1} if U_{j+1} is. Since Y_{j+1} has property $M(a_{j+1})$, U_j is nondecreasing if U_{j+1} is. □

The following lemmas give sufficient conditions for a density to have property $M(n)$. Let W_1, Z and g be as defined above Theorem 5. Let $W = W_1$.

LEMMA 5. *Let $g(w)$ be a continuous density. If $g(w \mid z_1)/g(w \mid z_2)$ is a nondecreasing function of W for all $z_1 \geqq z_2$, then g has property $M(n)$.*

PROOF. If the conditions of the lemma are satisfied, there is a d such that

$$\begin{aligned} g(w \mid z_1) - g(w \mid z_2) > 0 \quad &\text{implies that} \quad w \geqq d \\ g(w \mid z_1) - g(w \mid z_2) < 0 \quad &\text{implies that} \quad w \leqq d . \end{aligned}$$

Therefore, if $h(w)$ is nondecreasing, then

$$h(w)(g(w \mid z_1) - g(w \mid z_2)) \geqq h(d)(g(w \mid z_1) - g(w \mid z_2))$$

and

$$E(h(w) \mid Z = z_1) - E(h(w) \mid Z = z_2) \geqq h(d) \int (g(w \mid z_1) - g(w \mid z_2))\, dw = 0 .$$

COROLLARY. *The normal and gamma densities have property $M(n)$ for all n.*

LEMMA 5. *Let $g(w)$ be a discrete density with support*

$$\{w : w = a + tb, t = m, m + 1, \cdots, r\} \quad \text{or}$$
$$\{w : w = a + tb, t = m, m + 1, \cdots\} \quad \text{or}$$
$$\{w : w = a + tb, t = 0, \pm 1, \pm 2, \cdots\}$$

for some constants a, b, m and r. If $g(w \mid z + b)/g(w \mid z)$ is a nondecreasing function of w for all z such that z and $z + b$ are in the support of the density of Z, then g has property $M(n)$.

PROOF. The condition of the lemma implies that $g(w \mid z_1)/g(w \mid z_2)$ is a nondecreasing function of w for all $z_1 > z_2$ (and both in support of the density of Z). The remainder of the proof follows that of Lemma 4. □

COROLLARY. *The binomial and Poisson densities have property $M(n)$ for all n.*

COMMENT 1. Note the very strong assumption implicit in these lemmas that the support of g have no holes, i.e., that it be an interval of values.

COMMENT 2. The 4 densities mentioned in the corollaries have the reproductive property (e.g., the sum of i.i.d. normal random variables is a normal random variable). Therefore if the density of $X(i_1, \cdots, i_k)$ is normal, binomial, gamma or Poisson, then Y_j has the property $M(a_j)$, $j = 1, \cdots, k$.

6. An example. Let $k = 2$, $X(i_1, i_2) \sim N(\mu, \sigma^2)$, $i_j = 1, \cdots, a_j$, and let $L(x)$ be the indicator function of the set $\{x \geqq d\}$. The goal of the procedure, therefore, is to find which of the $X(i_1, i_2)$ are greater than d. Without loss of generality, let $\mu = 0$, $\sigma^2 = 1$. By (4.2), (4.3), and (5.1) if $\Phi = (\Phi_0, \Phi_1) \in D(p)$,

$$\begin{aligned} \Phi_i(Y_i) &= 1 \quad \text{if} \quad Y_i > c_i \\ &= b_i \quad \text{if} \quad Y_i = c_i \\ &= 0 \quad \text{if} \quad Y_i < c_i \end{aligned} \tag{6.1}$$

$i = 0, 1$, where c_0 and c_1 satisfy

$$P(Y_2 \geqq d \mid Y_1 = c_1) = \frac{(1 - p)(a_2 - 1)}{pa_2}, \tag{6.2}$$

$$\begin{aligned} &P(Y_2 \geqq d, Y_1 \geqq c_1 \mid Y_0 = c_0) \\ &\quad = \frac{1 - p}{pa_1a_2}(a_1 - 1 + a_1(a_2 - 1)P(Y_1 \geqq c_1 \mid Y_0 = c_0)). \end{aligned} \tag{6.3}$$

In addition, from (2.1),

$$V_1(\Phi) = a_1a_2[P(Y_2 \geqq d) - P(Y_0 \geqq c_0, Y_1 \geqq c_1, Y_2 \geqq d)], \tag{6.4}$$

and

$$V_2(\Phi) = 1 + (a_1 - 1)P(Y_0 \geqq c_0) + a_1(a_2 - 1)P(Y_0 \geqq c_0, Y_1 \geqq c_1). \tag{6.5}$$

It is easily verified that

$$\begin{pmatrix} Y_0 \\ Y_1 \\ Y_2 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a_1a_2 & a_2 & 1 \\ a_2 & a_2 & 1 \\ 1 & 1 & 1 \end{pmatrix}\right), \tag{6.6}$$

$$Y_2 \mid Y_1 \sim N\left(\frac{1}{a_2} Y_1, \left(1 - \frac{1}{a_2}\right)\right), \tag{6.7}$$

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \Big| Y_0 \sim N\left(\begin{pmatrix} \frac{1}{a_1} & Y_0 \\ \frac{1}{a_1 a_2} & Y_0 \end{pmatrix}, \begin{pmatrix} a_2\left(1 - \frac{1}{a_1}\right) & 1 - \frac{1}{a_1} \\ 1 - \frac{1}{a_1} & 1 - \frac{1}{a_1 a_2} \end{pmatrix}\right),$$

$$\begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} a_1 a_2 & a_2 \\ a_2 & a_2 \end{pmatrix}\right), \tag{6.8}$$

$$\begin{pmatrix} Y_0 \\ Y_1 \end{pmatrix} \Big| Y_2 \sim N\left(\begin{pmatrix} Y_2 \\ Y_2 \end{pmatrix}, \begin{pmatrix} a_1 a_2 - 1 & a_2 - 1 \\ a_2 - 1 & a_2 - 1 \end{pmatrix}\right).$$

Therefore, c_1 can be determined from (6.2) and (6.7) using a table of the univariate normal distribution function. Then c_0 can be determined from (6.3) and (6.7) by trial and error, using a table of the bivariate normal distribution function. Once c_0 and c_1 are determined, $V_2(\Phi)$ can be determined from (6.5) and (6.8) using tables of both the univariate and bivariate normal distributions. $V_1(\Phi)$ unfortunately involves a trivariate normal distribution. However it is clear that

$$V_1(\Phi) \leqq (a_1 a_2)(1 - P(Y_0 \geqq c_0, Y_1 \geqq c_1 \mid Y_2 = d))P(Y_2 \geqq d), \tag{6.9}$$

which can be computed using (6.8). In addition $P(Y_0 \geqq c_0, Y_1 \geqq c_1 \mid Y_2 = d)$ is an upper bound for the probability that an unsafe observation will go undetected.

As an example, let $d = 2$, $a_1 = 2$, $a_2 = 3$. The following table was computed using the *Tables of the Bivariate Normal Distribution Function and Related Functions* (1959). In the table, $P_1 = P(Y_0 \geqq c_0, Y_1 \geqq c_1 \mid Y_2 = 2)$, $UB = 6(1 - P_1)(P(Y_2 \geqq 2))$ is the upper bound for $V_1(\Phi)$ given in (6.9), and BR is the upper bound for the Bayes risk, $BR = (1 - p)V_2 + p(UB)$.

p	c_0	c_1	V_2	P_1	UB	BR
.95	1.4	1.57	1.81	.48	.072	.159
.99	−2.4	−.06	3.81	.92	.011	.049
.995	−3.7	−·65	4.49	.97	.004	.026 .

The procedure for $p = .99$, for example, will discover over 92% of the unsafe observations with only 64% as many measurements as the procedure that measures each observation separately.

The following table gives the results of similar calculations for $d = 2$, $a_1 = 3$, $a_2 = 2$.

p	c_0	c_1	V_2	P_1	UB	BR
.95	2.2	1.26	1.63	.41	.081	.159
.99	−1.7	.36	3.63	.91	.012	.048
.995	−3.1	.04	4.23	.97	.004	.025 .

The bivariate normal distribution is only tabulated in intervals of .1 (.05 for

ρ). Therefore, many of the numbers in the above table are approximations. It is interesting that the Bayes risk for the optimal procedure when $a_1 = 2$, $a_2 = 3$ is nearly equal to the Bayes risk when $a_1 = 3$, $a_2 = 2$ for p equals each of .95, .99 and 995. This might indicate that it makes little difference whether the 6 observations are divided into 2 sets of 3 or 3 sets of 2.

7. Generalizations. The assumption that the X's be i.i.d. would seem to be too strong an assumption for most "real-world" problems. However, that assumption was made primarily to keep the notation within reasonable bounds. In this section, we consider generalized group testing problems when the i.i.d. assumption is removed.

There would not seem to be much hope for finding optimal procedures unless the joint distribution of the X's is known. If the joint distribution is known, it should be possible to find Bayes procedures using essentially the same arguments as those in Section 4. Obviously, the U_j would be much more complicated, and the expected loss would depend on $i_1, \cdots, i_j$. If the X's could be assumed independent, then Theorem 1 would still be true and we would only need consider rules in which the action at the jth stage depended only on $Y_j(i_1, \cdots, i_j)$. Otherwise Φ_j would depend on $Y_n(i_1, \cdots, i_n)$ for $n \leqq j$. The notation would obviously become terribly cumbersome. In principle, however, the optimal procedure should be the procedure that continues as long as the expected gain from continuing is greater than the expected loss.

REFERENCES

FERGUSON, THOMAS S. (1967). *Mathematical Statistics, A Decision Theoretic Approach.* Academic Press, New York.

KUMAR, S. and SOBEL, M. (1971). Finding a single defective in binomial group-testing. *J. Amer. Statist. Assoc.* **66** 824–828.

Tables of the Bivariate Normal Distribution Function and Related Functions, (1959). National Bureau of Standards, Applied Mathematics Series **50**, Washington, D.C.

219 POND LABORATORY
PENNSYLVANIA STATE UNIVERSITY
UNIVERSITY PARK, PENNSYLVANIA 16802