TWO-ACTION COMPOUND DECISION PROBLEMS

by

M. V. Johns, Jr.

TECHNICAL REPORT NO. 87

March 11, 1966

Supported by the Army, Navy, Air Force, and NASA under

Contract Nonr-225(53) (NR-042-002)

with the Office of Naval Research

Gerald J. Lieberman, Project Director

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Two-Action Compound Decision Problems

by

M. V. Johns, Jr.

1. Introduction and Summary:

The compound decision problem considered here consists of a sequence of component problems, in each of which one of two possible actions must be selected. The loss structure is the same for each component decision problem. Each component problem involves independent identically distributed observations whose common distribution function is unknown but belongs to some specified parametric or non-parametric family of distributions (e.g., the family of all Poisson distributions with parameter λ bounded above by some finite number B). This family remains fixed for all component problems. It is assumed that, at the time a decision is made in any particular component problem, the available information includes the data obtained in all previous component decision problems in the sequence.

Compound decision problems of this type arise in situations where routine testing and evaluation programs are in operation. For example, in routine lot by lot acceptance sampling for quality control purposes, each lot of items is sampled, and the lot is either accepted or rejected on the basis of the observations obtained. Another example arises in routine medical diagnosis where a decision between two alternative treatments must be made for each of a continuing sequence of patients on the basis of results obtained from a diagnostic test performed on each patient. In either of these examples records of all past observations could certainly be accumulated.

In the compound decision problem as formulated here, no relationships whatever are assumed to exist among the distributions governing the observations associated with different component decision problems (aside from the requirement that all these distributions are members of a specified general family). A strictly "objective" approach to this situation appears, at first glance, to require that each component problem be treated in isolation with the decision for each problem being based on the observations obtained for that problem alone. It has been known for some time, however, that for certain types of compound decision problems, substantially better performance in terms of average risk incurred for a number of component problems may be obtained by using "compound decision procedures" which make explicit use at each stage of the seemingly irrelevant data from previous component problems. A number of authors have investigated this aspect of compound decision problems, notably Robbins [5], Hannan and Robbins [1], Samuel [8], [9], Hannan and Van Ryzin [2], Van Ryzin [11], and Swain [10]. These references are cited chronologically to indicate stages in the evolution of the subject and are not exhaustive. In the earlier papers [5], [1], and [8] the space of "states of nature," i.e., the family of distribution functions governing the observations, is assumed to be finite, so that these models are not suitable for most applications. In these papers, and in [9] as well, the main results are concerned only with the convergence to zero of the difference between the average risk and a certain "optimal" goal (discussed in detail below) as the number of component problems becomes large. In two of the more recent papers ([2], [11]) the finite state model has been retained but stronger results involving bounds on the deviations of the average risk from the desired goal and rates of convergence to "optimality" are obtained. The papers of Samuel [9] and Swain [10] deal with standard (infinite state) estimation problems with squared error loss, and their results are therefore immediately relevant to applications. In all of these papers except [10] the "optimal" goal asymptotically achieved by the average risk is defined in essentially the same way. For each n, the average risk for the first n component problems is compared to the Bayes optimal risk one could achieve for a single component problem if the parameter of interest had a known <u>a priori</u> distribution equal to the empirical distribution of the parameter values associated with the first n component problems. This criterion does not, however, represent the best that can be achieved by compound decision procedures, and in fact a variety of more stringent criteria may be defined which take into account empirical dependencies of various orders which may occur in the sequence of parameter values. At the suggestion of the present author, these more stringent criteria were considered by Swain in [10] and were shown to be asymptotically achievable for the compound estimation problem. Swain also obtains bounds and rates of convergence for some cases.

The object of the present paper is to find bounds for the deviations of the average risk from various optimal goals for the two-action compound decision problem. Attention is confined to certain classes of loss functions and compound decision procedures, and to the case of discrete-valued observations. Both parametric and non-parametric models are treated and the convergence of the bounds to zero is shown to be ratewise sharp. In order to state these results explicitly the problem must be presented more formally.

The compound decision problem consists of a sequence of component problems where the j^{th} component problem has the following structure:

(a) The distribution governing the observations is denoted by F_j and is a member of a specified family $\mathfrak{F}$ of distribution functions each assigning probability one to a fixed denumerable set of numbers $x_1, x_2, \ldots$.

(b) The statistician obtains k independent observations with common distribution function F_j. The observations are denoted by the vector $X_j = (X_{1j}, X_{2j}, \ldots, X_{kj})$.

(c) For the parametric case the parameter of interest determines F_j completely and is denoted by λ_j. For the non-parametric case, $\lambda_j = Eh(X_{1j})$, where $h(\cdot)$ is a specified function.

(d) On the basis of the observations the statistician selects one of two actions and incurs loss $L_a(\lambda_j)$, $a = 1, 2$, if action a is selected.

A typical compound decision rule for the j^{th} component problem is represented by $\Delta_j(x)$, where $E\Delta_j(x)$ is the probability of taking action one if $X_j = x$. For each value of the vector x, $\Delta_j(x)$ is a random variable depending on the mutually independent random vectors $X_1, X_2, \ldots, X_{j-1}$. The risk for the j^{th} problem is given by

$$r_j = (L_1(\lambda_j) - L_2(\lambda_j))E\,\Delta_j(X_j) + L_2(\lambda_j) .$$

Letting $p_j(x)$ be the probability that $X_j = x$, and

$$(1) \qquad \alpha_j(x) = (L_1(\lambda_j) - L_2(\lambda_j))p_j(x) ,$$

the average risk for the first n component problems is given by

$$(2) \quad \bar{r}_n = \frac{1}{n} \sum_{j=1}^{n} r_j$$

$$= \frac{1}{n} \sum_{j=1}^{n} \sum_{x} (L_1(\lambda_j) - L_2(\lambda_j))E\{\Delta_j(x) | X_j = x\}p_j(x) + \frac{1}{n} \sum_{j=1}^{n} L_2(\lambda_j) ,$$

$$= \frac{1}{n} \sum_{j=1}^{n} \sum_{x} \alpha_j(x)E\,\Delta_j(x) + \frac{1}{n} \sum_{j=1}^{n} L_2(\lambda_j) .$$

The "classical" goal that one attempts to achieve asymptotically, is defined by considering a hypothetical Bayesian version of a typical component problem. Suppose that for such a problem it is known that the sampling distribution F is chosen randomly according to the discrete a priori probability measure on $\mathfrak{F}$ which assigns probability n^{-1} to each element of the set $\{F_1, F_2, \ldots , F_n\}$ of sampling distributions arising in the first n component problems. If one uses the decision rule $\delta(x)$ (based only on the observations obtained for the single component problem under consideration), where $\delta(x)$ is the probability of taking action one when x is observed, the risk incurred is

$$\rho_n = \frac{1}{n} \sum_{j=1}^{n} \sum_{x} \alpha_j(x)\delta(x) + \frac{1}{n} \sum_{j=1}^{n} L_2(\lambda_j) .$$

Letting

$$(3) \qquad m_j(x) = \sum_{i=1}^{j} \alpha_i(x) \ , \ j = 1, 2, \ldots \ ,$$

it is easily seen that the Bayes optimal decision rule is given by

$$\delta^*(x) = \begin{cases} 1, & m_n(x) < 0 \\ 0, & m_n(x) \geq 0 \ , \end{cases}$$

and the optimal Bayes risk is

$$(4) \qquad \rho_n^* = \frac{1}{n} \sum_x m_n(x)^- + \frac{1}{n} \sum_{j=1}^{n} L_2(\lambda_j) \ ,$$

where $m_n(x)^-$ indicates the negative part of $m_n(x)$.

The object is to discover compound decision procedures having the property that the resulting average risks $\bar{r}_n$ satisfy

$$(5) \qquad |\rho_n^* - \bar{r}_n| < b(n) \ , \text{ all } n \ ,$$

where $b(n) \to 0$, as $n \to 0$, and where $b(n)$ is independent of the particular sequence $F_1, F_2, \ldots$, occurring. Theorem 1 of section 2 gives conditions under which a class of compound decision procedures will satisfy (5) with $b(n) = Kn^{-1/2}$, for a certain positive constant K independent of the sequence of F_j's. It is also noted that $n^{-1/2}$ is the best possible rate of convergence for this class of procedures. Typically, of course, neither $\bar{r}_n$ nor ρ_n^* will themselves converge to limits.

In section 3, specific compound decision procedures satisfying the conditions of Theorem 1 are presented for certain parametric cases (Poisson, negative binomial, etc.) involving families of sampling distributions of exponential type. The non-parametric case is also discussed and procedures satisfying Theorem 1 are given. A very simple loss structure is used throughout. In fact it is assumed that

$$L_1(\lambda) - L_2(\lambda) = c(\lambda - b) \ , \tag{6}$$

where b, c are specified constants. It is also assumed that $L_1(\lambda)$ and $L_2(\lambda)$ are bounded on any bounded interval of λ's. The particular loss functions

$$L_1(\lambda) = \begin{cases} 0 \ , & \lambda < b \\ c(\lambda - b) \ , & \lambda \geq b \ , \end{cases}$$

$$L_2(\lambda) = \begin{cases} c(b - \lambda) \ , & \lambda < b \\ 0 \ , & \lambda \geq b \ , \end{cases}$$

where $c > 0$, clearly satisfy (6), and are quite reasonable for many two-action problems of the one-sided hypothesis testing type. The arguments presented extend almost without change to the case where $L_1(\lambda) - L_2(\lambda)$ is any specified polynomial in λ. All of the compound decision procedures considered here are based on the construction of consistent unbiased estimates for each x of the quantities $m_j(x)$, $j = 1, 2, \ldots$, defined by (3). Action one is then chosen in the j^{th} component problem if and only if the estimate of $m_{j-1}(X_j)$ is negative.

The compound decision problem is closely related to the "empirical Bayes" problem where an actual unknown *a priori* distribution is assumed to exist. The empirical Bayes problem corresponding to the compound decision problem considered here is discussed in the non-parametric case by the present author in [3], and in the parametric case by Robbins [6] and Samuel [7]. With the exception of the necessity for a certain amount of auxiliary randomization, the compound decision procedures exhibited in section 3 are essentially the same as those suggested for the corresponding empirical Bayes problems.

The "classical" goal for compound decision problems described above may be generalized to produce a sequence of more stringent goals by extending the definition of the hypothetical Bayes decision problem. Instead of assuming that the present sampling distribution F is selected by a uniform *a priori* measure over $F_1, F_2, \ldots, F_n$, one may assume that the vector $(\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_t)$ of sampling distributions corresponding to the $t - 1$ most recent component problems and the present problem respectively, is a random vector with a discrete *a priori* probability measure on the t-fold product $\mathfrak{F} \times \mathfrak{F} \times \cdots \times \mathfrak{F}$, which assigns probability $(n - t + 1)^{-1}$ to each of the vectors $(F_{j-t+1}, F_{j-t}, \ldots, F_j)$, $j = t, t + 1, \ldots, n$. The optimal Bayes decision rule for such a problem must involve the observations obtained in the $t - 1$ most recent component problems as well as the present one. If the resulting Bayes risk is denoted by $\rho^*_{t,n}$, it is intuitively plausible that this quantity should be decreasing in t since advantage is taken of possible empirical dependencies of higher order as t is increased. Theorem 2 of section 4 shows that for each $t \geq 1$,

$$\rho^*_{t+1,n} < \rho^*_{t,n} + \xi_n ,$$

where $\xi_n = O(n^{-1})$. For "most" sequences of F_j's one would expect $\rho^*_{t+1,n}$ to be significantly smaller than $\rho^*_{t,n}$ when t is small, since "most" sequences will exhibit substantial empirical dependencies of small order. In section 4 certain "t-fold" compound decision procedures are considered and the attainment of the goal $\rho^*_{t,n}$ is discussed. Illustrations of specific t-fold compound decision procedures are given for the problems considered in the "classical" case in section 3.

Some suggestions for further generalizations are given in section 5.

2. General Results:

In this section we assume the existence for each x of an estimator $\hat{\alpha}(x)$, which for any element F of $\mathfrak{F}$ is an unbiased estimator of $\alpha(x) = (L_1(\lambda) - L_2(\lambda))p(x)$, where λ is the parameter value and $p(x)$ the probability mass function associated with F. The estimator $\hat{\alpha}(x)$, which may be randomized, must depend only on observations having F as their common c.d.f., and is assumed to have a finite third absolute moment for each x. For each x, let $\sigma^2(x) = \mathrm{Var}(\hat{\alpha}(x))$ and $\gamma^3(x) = E|\hat{\alpha}(x) - \alpha(x)|^3$.

We now introduce two conditions which impose certain restrictions on $\mathfrak{F}$ and $\hat{\alpha}$.

Condition 1: There exists a finite number B and a function $p_o(x)$ such that (a) $\sum_x p_o(x)^{1/2} < \infty$, and for each element of $\mathfrak{F}$ the corresponding λ and $p(x)$ satisfy (b) $|\lambda| < B$, and (c) $p(x) \leq p_o(x)$ for all x.

Condition 2: There exists a finite number $C > 0$ and a positive function $\epsilon(x) < 1$ such that (a) $\sum_x \epsilon(x) < C$, (b) $\sum p_0(x)\epsilon(x)^{-3} < C$, and for each element of $\mathfrak{F}$ and each x, (c) $\epsilon^2(x) \leq \sigma^2(x) < C(\epsilon^2(x) + p_0(x))$, and (d) $\gamma^3(x) < C$.

For any sequence $F_1, F_2, \ldots,$ of elements of $\mathfrak{F}$ and for each x, let $\hat{\alpha}_j(x)$, $\sigma_j^2(x)$, and $\gamma_j^3(x)$ represent $\hat{\alpha}(x)$, $\sigma^2(x)$, and $\gamma^3(x)$ respectively for the sampling distributions F_j, $j = 1, 2, \ldots$. It is apparent that for fixed x, the sequence $\hat{\alpha}_j(x)$, $j = 1, 2, \ldots,$ is a sequence of independent random variables, provided that any randomization involved is performed independently for each j.

For each x and for $j = 1, 2, \ldots,$ let

$$S_j(x) = \sum_{i=1}^{j} \hat{\alpha}_i(x) \ .$$

We observe that $ES_j(x) = m_j(x)$, and denote the variance of $S_j(x)$ by $s_j^2(x) = \sum_{i=1}^{j} \sigma_i^2(x)$. The compound decision procedure to be evaluated is given for $j > 1$ by

$$\Delta_j(x) = \begin{cases} 1 \ , & S_{j-1}(x) < 0 \\ 0 \ , & S_{j-1}(x) \geq 0 \ . \end{cases} \tag{7}$$

The decision rule $\Delta_1(x)$ for the first component problem may be arbitrary. We now state and prove the following theorem:

Theorem 1. If Conditions 1 and 2 are satisfied then there exists a finite constant K such that the average risk for the compound decision procedure (7) satisfies

$$|\bar{r}_n - \rho_n^*| < Kn^{-1/2} ,$$

for all n, for every sequence of elements of $\mathfrak{J}$.

Proof: Recalling (2), (4), and (7) we have

$$(8) \qquad n|\bar{r}_n - \rho_n^*| < \sum_x \left| \sum_{j=2}^{n} \alpha_j(x) P\{S_{j-1}(x) < 0\} + \xi(x) - m_n(x)^- \right| ,$$

where $\xi(x)$ represents the contribution to the risk due to the arbitrary decision rule $\Delta_1(x)$ used in the first component problem. Since by Condition 1 and (6), $\sum_x |\xi(x)|$ is bounded it will be ignored in the subsequent argument. We now consider an arbitrary fixed value of x and suppress this value whenever it appears as the argument of a previously defined function. Letting $\Phi(\cdot)$ represent the c.d.f. of a standard normal random variable, we know by the Berry-Esseen theorem (see e.g., [4], p. 288) that there exists a constant C_o such that

$$(9) \qquad \left| \sum_{j=2}^{n} \alpha_j P\{S_{j-1} < 0\} - \sum_{j=2}^{n} \alpha_j \Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right) \right|$$

$$\leq \sum_{j=2}^{n} |\alpha_j| \left| P\left\{\frac{S_{j-1} - m_{j-1}}{s_{j-1}} < -\frac{m_{j-1}}{s_{j-1}}\right\} - \Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right) \right|$$

$$\leq C_o \sum_{j=2}^{n} \frac{|\alpha_j|}{s_{j-1}^3} \left| \sum_{i=1}^{j-1} \gamma_i^3 \right| = R_1^{(n)} .$$

We seek a bound on the second sum on the left hand side of (8) under the assumption that $m_n^- = 0$, i.e., $m_n \geq 0$. For any particular sequence $F_1, F_2, \ldots$, such that $m_n \geq 0$, for $y > 1$, let

$$m(y) = m_{j-1} + \alpha_j(y - j + 1) , \quad j - 1 < y \leq j ,$$

$$s(y) = \begin{cases} s_{j-1} , & j - 1 < y \leq j - j^{-2} \\ s_{j-1} + j^2(s_j - s_{j-1})(y - j + j^{-2}) , & j - j^{-2} < y < j , \end{cases}$$

for $j = 2, 3, \ldots$. Thus, we have

$$\int_{j-1}^{j} m'(y)\Phi\left(-\frac{m(y)}{s(y)}\right)dy = \alpha_j \int_{j-1}^{j} \Phi\left(-\frac{m(y)}{s(y)}\right)dy .$$

Also, since $\Phi(\cdot)$ is monotone and bounded by one, and $m(y)/s(y)$ is monotone on the interval $(j - 1, j - j^{-2})$ for each $j > 1$, we have

$$\left|\int_{j-1}^{j} \Phi\left(-\frac{m(y)}{s(y)}\right) dy - \Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right)\right| \leq \left|\Phi\left(-\frac{m_j}{s_{j-1}}\right) - \Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right)\right| + 2j^{-2} .$$

Hence, letting $\varphi(\cdot) = \Phi'(\cdot)$,

$$(10) \qquad \left|\sum_{j=2}^{n} \alpha_j \Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right) - \int_1^n m'(y)\Phi\left(-\frac{m(y)}{s(y)}\right) dy\right|$$

$$= \left|\sum_{j=2}^{n} \alpha_j \left\{\Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right) - \int_{j-1}^{j} \Phi\left(-\frac{m(y)}{s(y)}\right) dy\right\}\right|$$

$$\leq \sum_{j=2}^{n} |\alpha_j| \left|\Phi\left(-\frac{m_j}{s_{j-1}}\right) - \Phi\left(-\frac{m_{j-1}}{s_{j-1}}\right)\right| + 2\sum_{j=2}^{n} |\alpha_j| j^{-2}$$

$$\leq \sum_{j=2}^{n} |\alpha_j| \left|\Phi\left(\frac{|\alpha_j|}{2s_{j-1}}\right) - \Phi\left(-\frac{|\alpha_j|}{2s_{j-1}}\right)\right| + 2\sum_{j=2}^{n} |\alpha_j| j^{-2}$$

$$\leq \varphi(0) \sum_{j=2}^{n} \frac{\alpha_j^2}{s_{j-1}} + 2\sum_{j=2}^{n} |\alpha_j| j^{-2} = R_2^{(n)} .$$

We must now bound the integral appearing on the left hand side of (10) uniformly in all functions $m(y)$ and $s(y)$ corresponding to sequences $F_1, F_2, \ldots,$ such that $m(n) \geq 0$. Let $h(y) = \frac{m(y)}{s(y)}$ so that $m'(y) = s'(y)h(y) + s(y)h'(y)$ (except at the points $y = j - 1$, $j - j^{-2}$, $j = 2, 3, \ldots,$ where $m'(y)$ is not defined). Let

$$(11) \qquad I = \int_1^n m'(y)\Phi\left(-\frac{m(y)}{s(y)}\right) dy$$

$$= \int_1^n s'(y)h(y)\Phi(-h(y))dy + \int_1^n s(y)h'(y)\Phi(-h(y))dy \ .$$

Integrating the first expression by parts and integrating the resulting expression by parts again, we have

$$I = s(n)h(n)\Phi(-h(n)) - s(1)h(1)\Phi(-h(1)) + \int_1^n s(y)h(y)h'(y)\varphi(-h(y))dy$$

$$= s(n)h(n)\Phi(-h(n)) - s(n)\varphi(-h(n)) + \int_1^n s'(y)\varphi(-h(y))dy$$

$$+ s(1)\varphi(-h(1)) - s(1)h(1)\Phi(-h(1)) \ .$$

Now, observing that $\max_{Z>0} Z\Phi(-Z) = C_1\Phi(-C_1)$, where $\Phi(-C_1) = C_1\varphi(C_1)$, we have

$$(12) \qquad |I| \leq (s(n) + s(1))(C_1^2\varphi(C_1) + 2\varphi(0)) = R_3^{(n)} \ .$$

Combining (9), (10), and (12) we see that for any fixed x, the summand on the right hand side of (8) is bounded by $R_1^{(n)} + R_2^{(n)} + R_3^{(n)}$ for any case where $m_n(x) \geq 0$. The same result holds when $m_n(x) < 0$, since then

$$\sum_{j=2}^{n} \alpha_j P\{S_{j-1} < 0\} - m_n^- = \sum_{j=1}^{n} \alpha_j (P\{S_{j-1} < 0\} - 1)$$

$$= - \sum_{j=1}^{n} \alpha_j P\{S_{j-1} \geq 0\} ,$$

and essentially the same argument applies.

We now reintroduce the supressed variable x and undertake to demonstrate that the quantity $\sum_x (R_1^{(n)}(x) + R_2^{(n)}(x) + R_3^{(n)}(x))$ is bounded by $Kn^{1/2}$ where K is independent of the sequence $F_1, F_2, \ldots$. Letting $C_2 = c(B + |b|)$ and recalling (1), we see that by Condition 1 (b) and (c), $|\alpha_j(x)| < C_2 p_j(x) < C_2 p_o(x)$, for each x and j. Thus, referring to (9) we have by Conditions 2 (b), (c), and (d)

$$\sum_x R_1^{(n)}(x) \leq C_0 C_2 C \sum_x \frac{p_o(x)}{\epsilon(x)^3} \sum_{j=2}^{n} (j - 1)^{-1/2}$$

$$\leq 2C_o C_2 C n^{1/2} .$$

Similarly, referring to (10), we have

$$\sum_x R_2^{(n)}(x) \leq \varphi(0) C_2^2 \sum_x \frac{p_o(x)}{\epsilon(x)} \sum_{j=2}^{n} (j - 1)^{-1/2} + 2C_2 \sum_x p_j(x) \sum_{j=2}^{n} j^{-2}$$

$$\leq 2\varphi(0) C_2^2 C n^{1/2} + 2C_2 C_3 ,$$

where $C_3 = \sum_{j=2}^{\infty} j^{-2}$. For $R_3^{(n)}(x)$ given by (12), we note that for each s, $s(n) = s_n(x)$ so that by Condition 2 (c)

$$s_n(x) \leq C^{1/2} n^{1/2} (\epsilon^2(x) + p_o(x))^{1/2} .$$

Hence by Conditions 1 (a) and 2 (a)

$$\sum_x s_n(x) \leq C^{1/2} n^{1/2} \sum_x (\epsilon(x) + p_o(x)^{1/2})$$

$$\leq C^{1/2}(C + B_o) n^{1/2} ,$$

where $B_o = \sum_x p_o(x)^{1/2}$. This completes the proof of the theorem.

Remark 1: The result of Theorem 1 is ratewise sharp since the conditions of the theorem do not, for example, exclude sequences $F_1, F_2, \ldots, F_n$ such that $\lambda_j = b$ (i.e., $\alpha_j(x) = 0$) for $j < n - n^{1/2}$, and $\lambda_j = b_o > b$ (i.e., $\alpha_j(x) = c(b_o - b)p_j(x)$) for $n - n^{1/2} \leq j \leq n$. For such sequences the contribution of the terms $\sum_{j=1}^{n} \alpha_j(x) P\{S_{j-1}(x) < 0\}$ appearing in (8) will typically be of the order of $n^{1/2}$ and positive for each x. Many sequences having this property may be constructed, and such sequences can occur in both the parametric and non-parametric applications discussed in the next section. The constant K appearing in the statement of Theorem 1 is defined implicitly in the proof and the value so determined is not "best" in any sense.

Remark 2: The maximization of the integral I, defined by (11), over the class of all bounded continuous differentiable functions $m(y)$, may be viewed as a classical variational problem whose solution would yield valuable insight concerning "least favorable" sequences $F_1, F_2, \cdots$. Unfortunately, the variational problem is singular and cannot be solved by standard methods.

3. Applications:

A. The parametric case. The parametric families for which estimators $\hat{\alpha}_j(x)$ satisfying the conditions of Theorem 1 can be constructed are essentially those for which the compound estimation problem is tractable (see, e.g., [9]).

The first example, which includes the Poisson and negative binomial families as special cases, is the exponential family with probability mass function

$$p(x) = g(x)\beta(\lambda)\lambda^x , \quad \text{for} \quad x = 0, 1, \ldots ,$$

where $g(x) > 0$ and $g(x)/g(x+1)$ is bounded for all $x \geq 0$. The family $\mathfrak{F}$ consists of all distributions having probability mass functions of this form for a given $g(x)$ with $0 \leq \lambda < B$, where B and $B_1 > B$ are chosen so that $\sum_x g(x)B_1^x < \infty$. For this example, we confine attention to situations where a single observation is obtained for each component problem, i.e., $X_j = X_{1j}$. This observation may be regarded as the value assumed by a sufficient statistic perhaps based on a larger number of observations.

For each x and j let

$$(13) \qquad \hat{\alpha}_j(x) = \begin{cases} \dfrac{cg(x)}{g(x+1)} + Z_j(x) , & \text{for } X_j = x + 1 \\ -cb + Z_j(x) , & \text{for } X_j = x \\ Z_j(x) , & \text{otherwise} , \end{cases}$$

where for each x, $Z_1(x)$, $Z_2(x)$, ... , is a sequence of independent random variables independent of the X_j's, such that $EZ_j(x) = 0$, $EZ_j(x)^2 = \epsilon^2(x)$, and the third absolute moments of the $Z_j(x)$'s are bounded uniformly in x and j. The significance of the $Z_j(x)$'s which represent auxiliary randomization is discussed in Remark 3 below. It is evident that for each x and j $E\hat{\alpha}_j(x) = c(\lambda_j - b)p_j(x) = \alpha_j(x)$, $\epsilon^2(x) < \sigma_j^2(x) < C(\epsilon^2(x) + p_j(x))$, and $\gamma_j^3(x) < C$, for some suitably chosen C. Letting $p_o(x) = g(0)^{-1}g(x)B^x$, and noting that $g(0) \leq \sum_x g(x)\lambda^x = \beta^{-1}(\lambda)$, we see that for each x, $p_o(x) \geq p(x)$ for all elements of $\mathfrak{S}$ and $\sum_x p_o(x)^{1/2} < \infty$ since $\sum_x g(x)B_1^x < \infty$ for $B_1 > B$. Condition 1 and Conditions 2 (c) and (d) are therefore satisfied by estimators of the form (13). To show that Theorem 1 holds for these estimators it remains to exhibit $Z_j(x)$'s satisfying Conditions 2 (a) and (b) with $p_o(x)$ as defined above. For fixed $\delta > 0$, let

$$Z_j(x) = \begin{cases} (x+1)^{-(1+\delta)}, & \text{with probability} = 1/2 \\ \\ -(x+1)^{-(1+\delta)}, & \text{with probability} = 1/2 \,. \end{cases}$$

Then $EZ_j(x)^2 = \epsilon^2(x) = (x+1)^{-2(1+\delta)}$ and Condition 2 (a) is satisfied. Since $\sum_x g(x)B_1^x$ converges for $B_1 > B$, it follows that $\sum_x (x+1)^{3(1+\delta)}p_o(x)$ converges and Condition 2 (b) is satisfied.

Operationally, only one randomization need be performed at each stage since for fixed j, the $Z_j(x)$'s need not be independent for different x's and may be computed on the basis of the outcome of the same randomization experiment.

A second parametric example involves the family of distributions with probability mass functions of the form

$$p(x) = g(x)\beta(\lambda)\left(\frac{\lambda}{a_1 + \lambda}\right)^x \quad , x = 0, 1, \ldots ; \quad \lambda \geq 0 ,$$

where a_1 is a specified positive constant,

$$g(x) = \frac{a_1(a_1 + 1) \ldots (a_1 + (x - 1))}{x!} \quad , \quad x = 1, 2, \ldots ,$$

$g(0) = 1$, and

$$\beta(\lambda) = \left(\frac{a_1}{a_1 + \lambda}\right)^{a_1} .$$

This family possesses the interesting property that $EX = \lambda$ for all λ. These distributions are actually reparameterizations of negative binomial distributions. For each x and j let

$$\hat{\alpha}_j(x) = \begin{cases} \dfrac{ca_1 g(x)}{g(x + t)} + Z_j(x) , & X_j = x + t , t = 1, 2, \ldots \\ -cb + Z_j(x) , & X_j = x \\ Z_j(x) , & \text{otherwise} , \end{cases}$$

where the $Z_j(x)$'s are defined as in the previous example. Again $E\,\hat{\alpha}_j(x) = \alpha_j(x)$, and under the same conditions on the λ's as in the previous example, and with an analogous definition of $p_0(x)$, it is easily verified that Theorem 1 holds for this example also.

The important case of the binomial distribution is treated below as a special case of the non-parametric problem.

B. The non-parametric case: We now consider the situation where the probability mass functions $p(x)$ corresponding to elements of $\mathfrak{S}$ are not assumed to have a known functional form and are not necessarily in a one to one relationship with the values of λ. For this case, $\lambda = Eh(X)$, where $h(\cdot)$ is a specified function and X is a typical observation having probability mass function $p(x)$. Thus, for instance, λ might be EX as in the second parametric example above. Other possibilities are $\lambda = E(X - t)^2$, or $\lambda = P\{X \leq t\}$ for some specified t.

Since so little is assumed about the probability structure of the problem, it is not surprising that the goal which is attainable in this case is slightly less stringent than that achieved in the parametric case. Specifically, if k observations are obtained for each component problem, the procedure discussed below will satisfy the conditions of Theorem 1 with ρ_n^* interpreted as the optimal risk for a hypothetical Bayes problem involving only $k - 1$ observations. Thus, one observation is sacrificed in the interests of generality or as the price of ignorance.

For the case of k observations $(k \geq 2)$, $p(x) = p'(x_1)p'(x_2) \cdots p'(x_k)$, where $x = (x_1, x_2, \cdots, x_k)$ and $p'(\cdot)$ is the probability mass function for a single observation. Letting $x' = (x_1, x_2, \cdots, x_{k-1})$ and recalling (1) and (2) we see that if a compound decision rule $\Delta_j(x')$ based on x' is used, then the expression for $\bar{r}_n$ remains unchanged except that x is replaced by x' throughout. We therefore seek a suitable estimate of $\alpha_j(x')$. Let $y(x') = (y_1, y_2, \cdots, y_{k-1})$, where $y_1 \leq y_2 \leq \cdots \leq y_{k-1}$ are the ordered values of the components of x'. For $t = 1, 2, \cdots, k$, and all

j, let $X_j^{(t)} = (X_{1j}, X_{2j}, \ldots, X_{t-1,j}, X_{t+1,j}, \ldots, X_{jk})$. Finally, for each j and x' let

$$(14) \quad \hat{\alpha}_j(x') = \begin{cases} c(h(X_{tj}) - b)M(x') + Z_j(x'), & y(X_j^{(t)}) = y(x'), \\ & \text{for } t = 1, 2, \ldots, k \\ Z_j(x') & , \text{otherwise} \end{cases}$$

with $M(x') = \dfrac{m_1!m_2! \cdots m_{k-1}!}{k!}$, where m_i is the number of components of x' having the i^{th} smallest distinct value. Even though it is possible for $y(X_j^{(t)})$ to equal $y(x')$ for more than one value of t, $\hat{\alpha}_j(x')$ is still well defined since X_{tj} will have the same value for each such case.

If $EZ_j(x') = 0$, it is evident that $E\,\hat{\alpha}_j(x') = \alpha_j(x')$. If we assume that $E|h(X)|^3 < C < \infty$, for any single observation X with probability mass function corresponding to an element of $\mathfrak{I}$, and if we assume the existence of a function $p_o'(\cdot)$ dominating each $p'(\cdot)$ corresponding to an element of $\mathfrak{I}$ and satisfying $\sum_{x_1} p_o'(x_1)^{1/2} < \infty$, we see that Condition 1 is satisfied with x' replacing x. The choice of the randomizing $Z_j(x')$'s so that Condition 2 is satisfied depends on the particular denumerable set of values which the observations may assume. If this set is the set of integers, then letting

$$Z_j(x') = \begin{cases} -\prod_{i=1}^{k-1} |x_i + 1/2|^{1+\delta}, & \text{with probability} = 1/2 \\ \prod_{i=1}^{k-1} |x_i + 1/2|^{1+\delta}, & \text{with probability} = 1/2, \end{cases}$$

for some $\delta > 0$, we see that Condition 2 is satisfied with x replaced by x' provided $\sum_{x_1} |x_1|^{3(1+\delta)} p'_o(x_1) < \infty$. Under such circumstances, the result of Theorem 1 holds with the interpretation of ρ_n^* given above. It should be noted that the case of the binomial distribution is included in this framework if we allow only the values zero and one for each individual observation, and set $h(x) = x$ so that $\lambda = p'(1) = 1 - p'(0)$. This case is not really "non-parametric" since the value of λ determines the distribution of the observations.

Remark 3: If the λ's are bounded away from zero in the two parametric examples discussed in Part A of this section, it is easily verified that Condition 2, and hence Theorem 1, holds without the introduction of the randomizing $Z_j(x)$'s.

The author knows of no examples within the context of the present paper (parametric or non-parametric) for which randomization can be demonstrated to be necessary for the result of Theorem 1, provided the conditions unrelated to randomization are satisfied. It is conjectured that such randomization is not essential, although because of the form of the Berry-Esseen bound, it is required for the method of proof used here.

4. The t-dependent Case:

In this section criteria based on generalizations of ρ_n^* are introduced.

Consider a hypothetical Bayes decision problem in which one of two actions is chosen on the basis k-dimensional vectors of observations $X_1, X_2, \dots, X_t$, having a random joint probability mass function $\tilde{p}(x_1, x_2, \dots, x_t) = \tilde{p}_1(x_1)\tilde{p}_2(x_2) \cdots \tilde{p}_t(x_t)$, where the $\tilde{p}_i(\cdot)$'s are random functions whose structure is described below. Note that x_i now stands for a k-dimensional vector and not a real component as was the case heretofore.

Now suppose that the vector of random probability mass functions $(\tilde{p}_1(\cdot), \tilde{p}_2(\cdot), \dots, \tilde{p}_t(\cdot))$ corresponds to the random vector of sampling distributions $(\tilde{F}_1, \tilde{F}_2, \dots, \tilde{F}_t)$, chosen according to the discrete _a priori_ probability measure on the t-fold product space $\mathfrak{F} x \mathfrak{F} x \cdots x \mathfrak{F}$ which assigns probability $(n - t + 1)^{-1}$ to each of the vectors $(F_{j-t+1}, F_{j-t+2}, \dots, F_j)$, $j = t, t+1, \dots, n$. Assuming that the losses depend only on the value of the parameter λ_t associated with $\tilde{p}_t(\cdot)$, the risk incurred if the arbitrary decision rule $\delta(x_1, x_2, \dots, x_t)$ is used, is given by

$$\rho_{t,n} = \frac{1}{n - t + 1} \sum_{(x_1,x_2,\dots,x_t)} \delta(x_1, x_2, \dots, x_t) \sum_{j=t}^{n} \alpha_{t,j}(x_1, x_2, \dots, x_t) + \frac{1}{n - t + 1} \sum_{j=t}^{n} L_2(\lambda_j) ,$$

where, for $j \geq t$

$$\alpha_{t,j}(x_1, x_2, \dots, x_t) = (L_1(\lambda_j) - L_2(\lambda_j))p_{j-t+1}(x_1)p_{j-t+2}(x_2) \cdots p_j(x_t).$$

Letting

$$m_{t,j}(x_1, x_2, \dots, x_t) = \sum_{i=t}^{j} \alpha_{t,i}(x_1, x_2, \dots, x_t) ,$$

for $j \geq t$, the optimal Bayes risk is clearly given by

$$\rho^*_{t,n} = \frac{1}{n-t+1} \sum_{(x_1, x_2, \dots, x_t)} m_{t,n}(x_1, x_2, \dots, x_t)^- + \frac{1}{n-t+1} \sum_{j=t}^{n} L_2(\lambda_j),$$

and is achieved by the decision rule

$$\delta^*(x_1, x_2, \dots, x_t) = \begin{cases} 1 , & m_{t,n}(x_1, x_2, \dots, x_t) < 0 \\ 0 , & \text{otherwise} . \end{cases}$$

If the sequence $\tilde{p}_1(\cdot), \tilde{p}_2(\cdot), \dots ,$ were a function valued stochastic process with known probability structure involving dependencies of order $t + 1$, one would expect the Bayes risk based on $t + 1$ vectors of observations to be smaller, in general, than that based on only t vectors of observations. In the present case, the hypothetical _a priori_ probability measure changes as t changes, but an analogous result holds as is shown by the following elementary theorem:

Theorem 2: If $|L_i(\lambda)| < K_o < \infty$, $i = 1, 2$, for all λ's corresponding to elements of $\mathfrak{J}$, then there exists a finite number K_1 such that, for

$$\rho^*_{t+1,n} \leq \rho^*_{t,n} + K_1(n - t)^{-1}$$

for any fixed t and $n > t$, for every sequence of elements of $\mathfrak{J}$.

Proof: The proof is based on the elementary fact that for any $b_1, b_2, \dots, b_n$, $\left(\sum_{j=1}^{n} b_j\right)^- \geq \sum_{j=1}^{n} b_j^-$. Thus

$$\rho^*_{t+1,n} - \rho^*_{t,n}$$

$$\leq \frac{1}{n-t} \sum_{(x_1,\dots,x_{t+1})} \left(\sum_{j=t+1}^{n} \alpha_{t+1,j}(x_1, \dots, x_{t+1})\right)^-$$

$$- \frac{1}{n-t} \sum_{(x_1,\dots,x_t)} \left(\sum_{j=t+1}^{n} \alpha_{t,j}(x_1, \dots, x_t) + \alpha_{t,t}(x_1, \dots, x_t)\right)^-$$

$$+ 2K_0(n - t)^{-1}$$

$$\leq \frac{1}{n-t} \sum_{(x_2,\dots,x_{t+1})} \left(\sum_{j=t+1}^{n} \sum_{x_1} \alpha_{t+1,j}(x_1, \dots, x_{t+1})\right)^-$$

$$- \frac{1}{n-t} \sum_{(x_1,\dots,x_t)} \left\{\left(\sum_{j=t+1}^{n} \alpha_{t,j}(x_1, \dots, x_t)\right)^- - \alpha_{t,t}(x_1, \dots, x_t)^-\right\}$$

$$+ 2K_0(n - t)^{-1} .$$

$$\leq 4K_0(n - t)^{-1} ,$$

since $\sum_{x_1} \alpha_{t+1,j}(x_1, \ldots , x_{t+1}) = \alpha_{t,j}(x_2, \ldots , x_{t+1})$. This establishes the desired result.

As was remarked in Section 1, it is to be expected that many sequences of sampling distributions will exhibit regularities that are equivalent to empirical dependencies. Such sequences will tend to yield values of $\rho^*_{t+1,n}$ substantially smaller than those for $\rho^*_{t,n}$, especially when t is small.

We now consider the use of compound decision rules of the form $\Delta_{t,j}(x_{j-t+1}, \ldots , x_j)$ for the j^{th} component problem for $j \geq t$. It is understood that $\Delta_{t,j}(\cdot)$ may depend on observations obtained for component problems prior to that with index $j-t+1$, and for $i \leq j < t$, $\Delta_{t,j}(\cdot)$ is arbitrary. Letting $x^*_t = (x_1, x_2, \ldots , x_t)$ for notational simplicity, the average risk for the t^{th} to the n^{th} component problems then becomes

$$(15) \qquad \bar{r}_{t,n} = \frac{1}{n-t+1} \sum_{j=t}^{n} \sum_{x^*_t} \alpha_{t,j}(x^*_t) E\, \Delta_{t,j}(x^*_t)$$

$$+ \frac{1}{n-t+1} \sum_{j=t}^{n} L_2(\lambda_j) \ .$$

For $j \geq 2t$ we assume that there exist estimators $\hat{\alpha}_{t,j}(\cdot)$ of $\alpha_{t,j}(\cdot)$ which are unbiased and which depend on the vectors of observations $X_1, X_2, \ldots , X_{j-t}$. Let

$$S_{t,j}(x^*_t) = \sum_{i=2t}^{j} \hat{\alpha}_{t,i}(x^*_t) \ ,$$

for $j \geq 2t$. For $j \geq 2t$ we consider compound decision rules of the form

$$(16) \qquad \Delta_{t,j}(x_t^*) = \begin{cases} 1, & S_{t,j-t}(x_t^*) < 0 \\ 0, & \text{otherwise}. \end{cases}$$

For $t < j < 2t$, $\Delta_{t,j}(\cdot)$ may be arbitrary.

The problem as formulated thus far appears to be essentially the same as that considered in section 2. However, an additional difficulty arises from the fact that, for all cases of interest, the sequence $\hat{\alpha}_{t,2t}(\cdot)\ \hat{\alpha}_{t,2t+1}(\cdot), \ldots$, is a t-dependent sequence of random functions. That is, $\hat{\alpha}_{t,j}(\cdot)$ and $\hat{\alpha}_{t,j'}(\cdot)$ are independent only if $|j - j'| > t$.

The author has been able to show that if compound decision rules of the form (16) are used, then there exist a $\epsilon > 0$ and a finite K such that for all n

$$(17) \qquad |\bar{r}_{t,n} - \rho^*_{t,n}| < Kn^{-\epsilon},$$

for all sequences $F_1, F_2, \ldots$. The conditions for this result to hold are straightforward generalizations to the t-dependent case of Conditions 1 and 2. The proof of (17), which is rather complex, will not be reproduced here since the author is convinced that, in fact, (17) holds with $\epsilon = 1/2$. A "proof" of this conjecture has been produced which requires a suitable version of the Berry-Esseen theorem for t-dependent random variables. Unfortunately, no such theorem seems to be available.

The parametric and non-parametric estimators of the α's given in section 3 are readily adaptable to the t-dependent case. This is illustrated by considering the simplest parametric case, i.e., the case of the geometric distribution. For this case a single observation having probability mass function $p_j(x) = \lambda_j^x(1 - \lambda_j)$, $x = 0, 1, \ldots,$ is obtained for the j^{th} component problem. Thus, recalling that

$x_t^* = (x_1, x_2, \ldots, x_t)$,

$$\alpha_{t,j}(x_t^*) = c(\lambda_j - b)\lambda_{j-t+1}^{x_1}\lambda_{j-t+2}^{x_2}\cdots\lambda_j^{x_t}(1-\lambda_{j-t+1})\cdots(1-\lambda_j) .$$

For $j \geq 2t$ let

$$\hat{\alpha}_{t,j}(x_t^*) = \begin{cases} c + Z_j(x_t^*) \ , & X_j = x_t + 1,\ X_{j-1} = x_{t-1}, \ldots, X_{j-t-1} = x_1 \\ -cb + Z_j(x_t^*), & X_j = x_t \ , \quad X_{j-1} = x_{t-1}, \ldots, X_{j-t+1} = x_1 \\ Z_j(x_t^*) & , \ \text{otherwise} \ , \end{cases}$$

where for some $\delta > 0$,

$$Z_j(x_t^*) = \begin{cases} -\prod_{i=1}^{t}(x_i+1)^{-(1+\delta)} \ , & \text{with probability} = 1/2 \\ \prod_{i=1}^{t}(x_i+1)^{-(1+\delta)} \ , & \text{with probability} = 1/2 \ . \end{cases}$$

If we restrict the possible values of λ to $0 \leq \lambda < B < 1$, then (17) holds for the compound decision rule (16) based on these $\hat{\alpha}_{t,j}$'s . The other parametric and non-parametric cases are disposed of in a similar fashion.

Remark 4: Since the t-dependent case involves the "matching" of t vectors of observations with sequences of t consecutive past observation vectors, it is clear that, if t is much greater than one, the number of component problems must be quite large before good results can be expected. This consideration, together with the fact that the improvement in $\rho^*_{t+1,n}$ compared with $\rho^*_{t,n}$ tends to be greatest when t is small, indicates that in most cases one should use values of t on the order of one, two, or three.

5. Conclusion:

As is customary in papers in this area, we take note of the fact that when the number of component problems is small, the procedures suggested will be relatively ineffective. Thus, as a practical matter, it is necessary to provide some means of orderly transition from "classical" decision procedures to compound decision procedures as the number of component problems increases.

Hopefully, the results of the present paper can be generalized in at least two directions. First, it would be very desirable to find similar results for finite action problems with more than two possible actions. Often such formulations conform more closely to real situations. Furthermore, greater flexibility in the choice of the loss structure can be obtained even under the restriction that the pairwise differences in the loss functions be linear in the parameter of interest.

A second important generalization would be the extension of the present methods to cases involving continuous random variables. Some such results are obtained for both the parametric and non-parametric compound estimation problems in [9] and [10]. It is conjectured that, for sufficiently sophisticated methods, bounds of order arbitrarily close to $n^{-1/2}$ on the difference between the average risk and the appropriate goal can be obtained in the continuous case.

References

[1] Hannan, J. F. and Robbins, H. (1955). Asymptotic solutions of the compound decision problem for two completely specified distributions. Ann. Math. Statist. 26, 37-51.

[2] Hannan, J. F. and Van Ryzin, J. R. (1965). Rate of convergence in the compound decision problem for two completely specified distributions. Ann. Math. Statist. 36, 1743-1752.

[3] Johns, M. V. (1957). Nonparametric empirical Bayes procedures. Ann. Math. Statist. 28, 649-669.

[4] Loève, M., (1960). Probability theory (2nd ed.). Van Nostrand, Princeton.

[5] Robbins, H. (1951). Asymptotic subminimax solution of compound decision problems. Proc. Second Berkeley Symposium of Math. Statistics and Probability. University of California Press, 131-148.

[6] Robbins, H. (1963). The empirical Bayes approach to testing statistical hypotheses. Rev. International Statist. Inst. 31, 195-208.

[7] Samuel, E. (1963). An empirical Bayes approach to the testing of certain parametric hypotheses. Ann. Math. Statist. 34, 1370-1385.

[8] Samuel, E. (1964). Convergence of the losses of certain decision rules for the sequential compound decision problem. Ann. Math. Statist. 35, 1606-1621.

[9] Samuel, E. (1965). Sequential compound estimators. Ann. Math. Statist. 36, 879-889.

[10] Swain, D. D. (1965). Bounds and rates of convergence for the extended compound estimation problem. Statistics Department, Stanford University Technical Report.

[11] Van Ryzin, J. (1965). The sequential compound decision problem with $m \times n$ finite loss matrix. Argonne National Laboratory, Applied Mathematics Division Technical Memorandum No. 54.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)* Stanford University Department of Statistics Stanford, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Two-Action Compound Decision Problems

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(Last name, first name, initial)*

Johns, M. V., Jr.

6. REPORT DATE March 11, 1966	7a. TOTAL NO. OF PAGES 31	7b. NO. OF REFS 11

8a. CONTRACT OR GRANT NO. Contract Nonr-225(53) b. PROJECT NO. NR-042-002 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 87 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Logistics and Mathematical Statistics Branch Office of Naval Research Washington, D. C. 20360

13. ABSTRACT

Asymptotically optimal methods are given for deciding between two alternative actions in situations such as those occurring in routine lot by lot acceptance sampling where essentially the same decision problem arises repeatedly and accumulated past information can be used to advantage. A linear cost structure is used and no assumptions are made concerning the existence of a priori probability distributions of the relevant unknown parameters. Procedures are given which take advantage of empirical dependencies appearing in the sequences of unknown parameter values. Rates of convergence to optimality are obtained.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Compound decisions Acceptance sampling						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
______________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
______________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
______________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

DD FORM 1 JAN 64 1473 (BACK)

UNCLASSIFIED

Security Classification