INTEGRATED RISK OF ASYMPTOTICALLY BAYES SEQUENTIAL TESTS[1]

By Gary Lorden

Cornell University and Northwestern University

1. Introduction. For general multiple-decision testing problems, and even two-decision problems involving more than two states of nature, how to construct sequential procedures which are optimal (e.g. minimax, Bayes, or even admissible) is an open question. In the absence of optimality results, many procedures have been proposed for problems in this category. Among these are the procedures studied in Wald and Sobel (1949), Donnelly (1957), Anderson (1960), and Schwarz (1962), all of which are discussed in the introduction of the paper by Kiefer and Sacks (1963) along with investigations in sequential design of experiments (notably those of Chernoff (1959) and Albert (1961)) which can be regarded as considering, inter alia, the (non-design) sequential testing problem.

The present investigation concerns certain procedures which are asymptotically Bayes as the cost per observation, c, approaches zero and are definable by a simple rule: continue sampling until the *a posteriori* risk of stopping is less than Qc (where Q is a fixed positive number), and choose a terminal decision having minimum *a posteriori* risk. This rule, with $Q = 1$, was first considered by Schwarz and was shown to be asymptotically Bayes, under mild assumptions, by Kiefer and Sacks (whose results easily extend to the case of arbitrary $Q > 0$). Given an *a priori* distribution, F, and cost per observation, c, we shall use $\delta_F(Qc)$ to denote the procedure defined by this rule and $\delta_F^*(c)$ to denote a Bayes solution with respect to F and c. The result of Kiefer and Sacks, for $Q = 1$, states that $r_c(F, \delta_F(c)) \sim r_c(F, \delta_F^*(c))$ as $c \to 0$, where $r_c(F, \delta)$ is the integrated risk of δ when F is the *a priori* distribution and c is the cost per observation. The principal aim of the present work is to construct upper bounds (valid for all $c > 0$) on the difference $r_c(F, \delta_F(Qc)) - r_c(F, \delta_F^*(c))$, so that one can determine values of c (or the probabilities of error) small enough to insure that simple asymptotically optimum procedures are reasonably efficient.

The main result is Theorem 2.1, which states that, when Assumptions I–V are satisfied, there exists a bound on this difference which is of the form Mc, where M depends upon Q. (Assumptions I–V are identical with the assumptions of Kiefer and Sacks except for a greater limitation of the structure of indifference and semi-indifference regions, as indicated in the remark following Assumption II.) Since $r_c(F, \delta_F^*(c)) = O(c|\log c|)$ as $c \to 0$, as shown in (*K-S*), it follows from the main result that the "efficiency" of $\delta_F(Qc)$, $r_c(F, \delta_F^*(c))/r_c(F, \delta_F(Qc))$, is

Received 18 May 1966; revised 14 March 1967.

[1] The present paper is part of the author's Ph.D. thesis at Cornell University. The research was supported in part by the Office of Naval Research contract number Nonr-266(04) (NR407-005) and by the National Science Foundation grant NSF GP-4972.

$1 - O(|\log c|^{-1})$ as $c \to 0$, for every $Q > 0$. Also, the proof of Theorem 2.1 shows how to construct M from certain basic estimates associated with suitable partitions of the parameter space (Lemma 2.1). Bounds on $r_c(F, \delta_F(Qc))$ can also be constructed, as indicated in the remark following the proof of Theorem 2.1.

In Section 3, construction of M is studied in detail for general multiple-decision problems in which the set of possible states of nature is finite. The numerical examples given in Section 4 include the following problem, which Kiefer and Sacks discuss in their introduction, comparing asymptotically the procedures of Donnelly, Anderson, and Schwarz. Given independent normally distributed random variables with unknown mean θ and variance one, test the hypothesis $\theta = 1$ against the alternative $\theta = -1$, with indifference point $\theta = 0$ (the loss structure being "0 -1"). It is shown in the proof of Theorem 2.1 that (in general) $\delta_F(c)$ has smaller integrated risk due to error than $\delta_F^*(c)$, and a calculation in Section 4 shows that, for arbitrary F, the *a priori* expected sample size using $\delta_F(c)$ is at most 7.7 larger than that of $\delta_F^*(c)$ in this example; hence, $M = 7.7$ suffices for the conclusion of Theorem 2.1. (For many common problems with infinite but compact parameter spaces, M can be constructed in a manner indicated below the proof of Theorem 3.3.)

The numerical bounds on $r_c(F, \delta_F(Qc)) - r_c(F, \delta_F^*(c))$ lead to lower bounds on the "efficiency" of $\delta_F(Qc)$, $r_c(F, \delta_F^*(c))/r_c(F, \delta_F(Qc))$, upon computation of suitable lower bounds on $r_c(F, \delta_F^*(c))$, or on the expected sample sizes of procedures attaining prescribed probabilities of error. For example, consider the problem just mentioned, testing the mean, θ, of a normal distribution, and suppose the *a priori* distribution F assigns equal probabilities to $\theta = 1, 0, -1$. Inequality (1.4) of Hoeffding (1960) gives a lower bound on the expected sample size $E_\theta N$ at $\theta = 0$ for any procedure with probabilities of error under $\theta = 1, -1$ at most α, β. For $\theta = 1, -1$ lower bounds on $E_\theta N$ can be given in terms of α, β by using Wald's formulas for the expected sample sizes of a sequential probability ratio test and the optimality property of SPRT's. To simplify and improve the results, we consider the analogous testing problem for a Wiener process, in which case Wald's formulas are exact. From the lower bounds mentioned one obtains a lower bound, $n(\nu)$, on the *a priori* expected sample size of a procedure with integrated probability of error less than ν. One can derive a lower bound on $r_c(F, \delta_F^*(c))$ by using $n(\nu)$ or, more simply, use $n(\nu)$ in connection with the following formulation.

Let $N(\delta)$ denote the *a priori* expected sample size of δ, and without reference to the cost c define the efficiency of $\delta_F(\nu)$ to be $\inf N(\delta)/N(\delta_F(\nu))$, taking the infimum over all δ's whose integrated probability of error is no greater than that of $\delta_F(\nu)$. As shown in the proof of Lemma 2.2, $\delta_F(\nu)$ has integrated risk of error less than ν, the difference between the two coming from "excess over the boundaries." For 0-1 loss the integrated risk of error is simply the integrated probability of error; hence, the "competitors" of $\delta_F(\nu)$ are subject to the lower bound $n(\nu)$, which is therefore a lower bound on the numerator in the above definition of efficiency. In the Wiener process problem, the calculation in Example 2 of Sec-

tion 4 shows that $r_c(F, \delta_F(2c)) - r_c(F, \delta_F^*(c)) \leqq (6.1)c$. By setting $c = \nu/2$, we obtain $N(\delta_F(\nu)) - \inf N(\delta) \leqq 6.1$, since the infimum is taken over the class of procedures having smaller integrated risk of error than $\delta_F(\nu)$. For $\nu = 10^{-4}$, 10^{-8}, 10^{-16}, we obtain $n(\nu) = 6.1, 14.6, 31.3$, respectively, and thus have lower bounds 50%, 70%, 84% on the efficiency of $\delta_F(\nu)$. If $\theta = 1, 0, -1$ is replaced by $\theta = \theta_0, 0, -\theta_0$, both $n(\nu)$ and the bound 6.1 are multiplied by θ_0^2, so that the estimates of efficiency are the same for every ν.

These lower bounds are far from sharp, as is clear from the kinds of estimates used in proving Lemma 2.2 and Theorem 2.1. Improvements can be made in the estimates of $r_c(F, \delta_F(Qc)) - r_c(F, \delta_F^*(c))$ by taking into account the specific distribution F (which is not done in Theorem 2.1) and modifying the proof of Theorem 2.1 to take advantage of this. For simple special problems, like the Wiener process two-decision problem discussed above, one can make a much more direct investigation of the error probabilities and expected sample sizes of $\delta_F(Qc)$'s, using, for example, the extension of Anderson's method described in his paper (1960). Anderson makes an exact computation of the operating characteristics of similar but slightly simpler tests, obtaining results very close to Hoeffding's lower bound.

It is easy to extend our main results, Theorem 2.1 and Corollary 2.2, to a wider class of procedures. Suppose a family of tests $\{\delta(c)\}$ satisfies these requirements: for some $Q > 0$, $\delta(c)$ stops no later than $\delta_F(Qc)$ and chooses a terminal decision whose *a posteriori* risk is at most Kc. Then the integrated risk (under F) of $\delta(c)$ exceeds that of $\delta_F(Qc)$ by at most Kc and, hence, is at most $(K + M)c$ larger than the Bayes risk. In case the number of possible states of nature is finite, it is straightforward to show that for F having full support the above conditions are equivalent to the following: for some $B > 0$, $\delta(c)$ stops if for some decision there is a state of nature θ under which the decision is correct such that the likelihood ratio of θ to θ' exceeds B/c for all θ' where the decision is incorrect; also, $\delta(c)$ chooses a terminal decision satisfying the same requirement with B/c replaced by A/c, $B \geqq A > 0$. Since the conditions just stated do not depend on F, a family $\{\delta(c)\}$ satisfying them has $O(c)$ excess integrated risk for every *a priori* distribution F with full support. Clearly one can meet these requirements using several SPRT's simultaneously on the same observations. Kiefer and Sacks give a discussion of the use of such simultaneous tests.

The key to the present results is Lemma 2.2, which establishes that a Bayes procedure continues sampling whenever the *a posteriori* risk of stopping exceeds M^*c. The argument used to prove this lemma is the heart of the present paper; the proof of Theorem 2.1 is based upon it, and so is the construction of M satisfying Theorem 2.1. This argument proceeds straightforwardly from Lemma 2.1, which establishes the existence of certain finite partitions of the parameter space, permitting it to be treated as if it were a finite point set. For similar reasons, Kiefer and Sacks required a slightly weaker result than Lemma 2.1 (Lemma 3 of their paper); we indicate amendments to their method of proof which yield Lemma 2.1.

Lemma 2.2 is a strengthening of a central result of Schwarz, whose work concerns the problem of testing sequentially between two composite hypotheses $\theta \leqq \theta_1$ and $\theta \geqq \theta_2$ concerning the real parameter θ of a distribution of exponential (Koopman-Darmois) type, with indifference region (θ_1, θ_2). Schwarz investigated the Bayes continuation region, $B(c)$, for a fixed *a priori* distribution and cost per observation c, in the space (n, S_n), where S_n is the usual sufficient statistic after n observations. In order to find the "asymptotic shape" of $B(c)$, Schwarz proved that $C(c) \supset B(c) \supset C(\text{const } c|\log c|)$ for sufficiently small c, where $C(c)$ is the set on which the *a posteriori* risk of stopping exceeds c. (Kiefer and Sacks generalized this result in Lemma 4 of their paper.) Lemma 2.1 yields the improvement $B(c) \supset C(M^*c)$ whenever Assumptions I–V are satisfied; Assumption IV requires, however, that the parameter space be compact (i.e. that the *a priori* distribution in Schwarz's setting have compact support). Unfortunately, this compactness assumption does not seem to be removable by the kind of device used in (K-S) to extend their results to problems of Schwarz's type. (Remark 5 of their paper is insufficient because the compact subset Ω_0 would, in our context, have to vary with the *a posteriori* distribution G in the proof of Lemma 2.2.)

Recently, M. Fushimi (1965) has extended part of Schwarz's results on asymptotic shapes, in the binomial and normal cases, for certain loss functions and *a priori* distributions. Schwarz showed that $C(c) = B_0 \log c^{-1} + o(\log c^{-1})$ (where B_0 is a computable region), whereas Fushimi obtains $C(c) = B_0[\log c^{-1} - (\frac{3}{2}) \log \log c^{-1}]$ to within $o(\log \log c^{-1})$. However, Schwarz's main result, $B(c) = B_0 \log c^{-1} + o(\log c^{-1})$, is not thereby improved to within $o(\log \log c^{-1})$ because $C(\text{const } c \log c^{-1}) = B_0[\log c^{-1} - (\frac{5}{2}) \log \log c^{-1}]$ plus $o(\log \log c^{-1})$. For compactly supported *a priori* distributions, Lemma 2.2 implies $B(c) \supset C(M^*c)$, which leads to $B(c) = B_0[\log c^{-1} - (\frac{3}{2}) \log \log c^{-1}]$ to within $o(\log \log c^{-1})$. Fushimi proposes the use of the stopping region $B_0[\log c^{-1} - (\frac{5}{2}) \log \log c^{-1}]$, on somewhat vague grounds. For reasons discussed in (K-S), "asymptotic shapes" (even when specified up to $o(\log \log c^{-1})$) seem to yield insufficient information about procedures to obtain useful estimates of error probabilities and sample sizes.

2. Risk of asymptotically Bayes procedures. The sequential decision problem is formulated as in the paper of Kiefer and Sacks (1963), with a few changes in notation. Independent and identically distributed random variables $X_1, X_2, \cdots$ are observed sequentially, taking values in a space $(\mathfrak{X}, \mathfrak{A})$ on which a σ-finite measure μ is defined. The parameter space, Ω, is a compact subset of Euclidean space (in the usual topology) and $\{f_\omega ; \omega \in \Omega\}$ is a set of probability densities for X_1, with respect to μ. There are d terminal decisions ($d \geqq 2$) and corresponding non-negative loss functions, $L_i(\omega)$ ($i = 1, \cdots, d$), representing the loss incurred in making the ith decision when f_ω is the true density. The cost per observation is c ($c > 0$). It is assumed throughout that $\min_i L_i(\omega)$ is identically zero; that is, for each ω there is at least one "correct" decision.

Assumptions I, II, III, and V below are taken from Assumptions 1, 2, 3, and

5, respectively, of (K-S). (Assumption IV appears as a hypothesis in each of their theorems.) Let $\Omega_i = \{\omega \in \Omega: L_i(\omega) = 0\}$, $i = 1, \cdots, d$. (Ω_i is the subset of Ω where the ith decision is correct.)

ASSUMPTION I. Put $L_i = \sup_{\omega \in \Omega} L_i(\omega)$ and $b_i = \inf_{\omega \notin \Omega_i} L_i(\omega)$. Assume that each L_i is finite and each b_i is positive, and let $L = \max_i L_i$ and $b = \min_i b_i$.

ASSUMPTION II. Let $\omega, \theta \in \Omega$, with $\omega \in \Omega_i$ and $\theta \in \Omega - \Omega_i$.

(a) $E_\omega[\log f_\omega(X) - \log f_\theta(X)]$ exists and is continuous in $\omega \in \Omega_i$ for every $\theta \in \Omega - \Omega_i$.

(b) $\lambda_i(\omega) = \inf_{\theta \in \Omega - \Omega_i} E_\omega[\log f_\omega(X) - \log f_\theta(X)]$ is continuous in $\omega \in \Omega_i$ and is bounded below by $\lambda_i > 0$.

NOTE. Assumption II insures that Ω_i and its complement in Ω are "separated." For their results, Kiefer and Sacks require only that the complement of Ω_i is separated from a suitable subset of Ω_i (which may be a proper subset if Ω_i contains an indifference region or semi-indifference region).

ASSUMPTION III. For every $\omega \in \Omega_i$ and $\theta \in \Omega - \Omega_i$,

(a) $E_\omega[\log f_\omega(X) - \log f_\theta(X)]^2 < \infty$.

(b) $\lim_{\rho \downarrow 0} E_\omega[\log \sup_{|\theta' - \theta| \leq \rho} f_{\theta'}(X) - \log f_\theta(X)]^2 = 0$.

(c) $\lim_{\omega' \to \omega} E_\omega[\log f_{\omega'}(X) - \log f_\omega(X)]^2 = 0$.

ASSUMPTION IV. For $i = 1, \cdots, d$, Ω_i and $\Omega - \Omega_i$ are compact.

ASSUMPTION V.

(a) $E_\omega[\log f_\omega(X) - \log f_\theta(X)]^2$ and $E_\omega[\log \sup_{|\theta' - \theta| \leq \rho} f_{\theta'}(X) - \log f_\theta(X)]^2$ are continuous in $\omega \in \Omega_i$ for every $\theta \in \Omega - \Omega_i$ and $\rho > 0$.

(b) $E_\omega[\log f_{\omega'}(X) - \log f_\omega(X)]^2$ is jointly continuous in ω and ω' for $\omega, \omega' \in \Omega_i$, $i = 1, \cdots, d$.

LEMMA 2.1. *Under Assumptions* I–V, *there is a finite covering*, $\{V_j, j = 1, \cdots, k\}$, *of* Ω_1, *with corresponding numbers* A_j, *such that the expectation, under every* $\omega \in V_j$, *of the first* n *for which* $X_1, \cdots, X_n$ *satisfy*

$$\int_{V_j} \prod_{m=1}^n f_\omega(X_m) P_1(d\omega) > \zeta \int_{\Omega - \Omega_1} \prod_{m=1}^n f_\theta(X_m) P_2(d\theta) \tag{2.1}$$

is bounded above, for $\zeta \geq 2$, *by* $A_j \log \zeta$, provided $P_1(V_j) \geq P_2(\Omega - \Omega_1) > 0$.

REMARK 1. The stopping variable defined by (2.1) is clearly increasing in ζ, so that, for $\zeta \in [1, 2)$, $A_j \log 2$ bounds from above its expectation under every $\omega \in V_j$. Thus, for $\zeta \geq 1$, we have an upper bound of the form $A_j \log \zeta + B_j$.

In (K-S), the proof of the statement containing (2.31) proves a weaker version of Lemma 2.1 in which the V_j's and the A_j's are not required to be independent of P_1 and P_2. (In that paper, the measures corresponding to our P_1 and P_2, ξ and η, are determined by the *a priori* distribution and are therefore fixed throughout.) However, in the proof of (2.31) (which relies upon the major part of the proofs of Lemmas 2 and 3) the covering $\{V(\omega_1), \cdots, V(\omega_k)\}$ (corresponding to our $\{V_1, \cdots, V_k\}$) clearly is chosen independently of ξ and η, and the loss functions, $L_1(\theta)$ and $L_2(\omega)$, of that paper. Let the loss functions be constant, in θ and ω, respectively, with the two constant values chosen for each $V(\omega_j)$ so that A_c' (2.14) of K-S) is not greater than $|\log c|$ (corresponding to our requirement $P_1(V_j) \geq P_2(\Omega - \Omega_1)$). Then the M_j's (corresponding to our A_j's) depend only

upon estimates of the first two moments of random variables (namely, the S_n's and B_n^i's) whose distributions are independent of η, while the estimates themselves are independent of ξ, and these observations suffice for the present Lemma 2.1.

LEMMA 2.2. *If Assumptions* I–V *are satisfied, there exists an* M^* *such that, for every* a priori *distribution and cost per observation,* c, *a Bayes procedure (with probability one) continues sampling whenever the* a posteriori *risk of stopping is at least* M^*c.

PROOF. For cost per observation c, let $r_c(G, \delta)$ denote the integral with respect to G of the risk function of δ. Also, for $i = 1, \cdots, d$ let δ_i denote the procedure which chooses the ith decision without sampling, let $r(G, \delta_i)$ denote its integrated risk (which is independent of c), and set $r(G) = \min r(G, \delta_i)$ $(i = 1, \cdots, d)$. It is sufficient to prove, for some M^* and Q, that

$$r_c(G, \delta_G(Qc)) < r(G) \quad \text{if} \quad r(G) \geqq M^*c. \tag{2.2}$$

We fix $Q > 0$ in the remainder of the proof and find an M^* satisfying (2.2).

The part of $r_c(G, \delta_G(Qc))$ that comes from wrong decisions can be estimated as follows. It is well-known that the integrated risk of error is equal to the *a priori* expectation of the stopping risk upon stopping, which is less than Qc for $\delta_G(Qc)$.

Therefore, (2.2) follows from

$$\int_\Omega E_\omega N(\delta_G(Qc))G(d\omega) \leqq c^{-1}r(G) - Q \quad \text{if} \quad r(G) \geqq M^*c, \tag{2.3}$$

where $N(\delta_G(Qc))$ is the first n for which $X_1, \cdots, X_n$ satisfy

$$\int_\Omega \prod_{m=1}^n f_\omega(X_m)G(d\omega) > (Qc)^{-1} \min_{v=1,\cdots,d} \int_{\Omega-\Omega_v} L_v(\theta) \prod_{m=1}^n f_\theta(X_m)G(d\theta).$$

Using Lemma 2.1 we will show that, for $i = 1, \cdots, d$, there is a finite covering $\{V_j ; j \in Z(i)\}$ of Ω_i (the $Z(i)$'s are disjoint sets of integers) with corresponding non-negative functions, g_j, concave and strictly increasing on $[0, \infty)$, such that

$$\sup_{\omega \in V_j} E_\omega N(\delta_G(Qc)) \leqq g_j(\log (r(G, \delta_i)/QcG(V_j))), \tag{2.4}$$

$$\text{if } r(G, \delta_i) \geqq Qc \quad \text{and} \quad G(V_j) > 0.$$

Taking $i = 1$ (for convenience) and $\{V_j ; j = 1, \cdots, k\}$ and corresponding A_j's satisfying Lemma 2.1, note first that $N(\delta_G(Qc))$ is less than or equal to the first n for which

$$\int_{V_j} \prod_{m=1}^n f_\omega(X_m)G(d\omega) > (Qc)^{-1} \int_{\Omega-\Omega_1} L_1(\theta) \prod_{m=1}^n f_\theta(X_m)G(d\theta)$$

or, equivalently,

$$\int_{V_j} \left(\prod_{m=1}^n f_\omega(X_m)\right)G(d\omega)/G(V_j) \tag{2.5}$$

$$> [r(G, \delta_1)/QcG(V_j)] \int_{\Omega-\Omega_1} \left(\prod_{m=1}^n f_\theta(X_m)\right)L_1(\theta)G(d\theta)/r(G, \delta_1).$$

Now (2.5) is just (2.1) with $r(G, \delta_1)/QcG(V_j)$ in place of ζ, $P_1 = [G(V_j)]^{-1}G$ and $dP_2(\theta) = [r(G, \delta_1)]^{-1}L_1(\theta)\,dG(\theta)$ (so that $P_1(V_j) = P_2(\Omega - \Omega_1) = 1$); therefore Lemma 2.1 and the remark following it imply that $\{V_j ; j = 1, \cdots, k\}$

with $g_j(x) = A_j x + A_j \log 2$ satisfies (2.4). (We could restrict ourselves to linear g_j's in (2.4); however, other concave g_j's are used in some of the computational results, and the present development has been written so as to apply there as well as to prove Lemma 2.1.)

We now use the estimate in (2.4) to obtain an upper bound, (2.12) below, on the left hand side of (2.3), assuming $r(G) \geqq Qc$. For each $i = 1, \cdots d$, choose a subset, $J(i)$, of $\cup Z(k)$ $(k = 1, \cdots, d)$ containing $Z(i)$ such that $\{V_j\,; j \varepsilon J(i)\}$ covers Ω; of course, $J(i) = \cup Z(k)$ suffices, for all i; but selecting each $J(i)$ as a proper subset of $\cup Z(k)$ (if possible) yields a smaller bound in (2.11) below. Assume, for convenience, that $r(G) = r(G, \delta_1)$, and observe that (2.4) implies that

$$\begin{aligned} &\sup_{\omega \varepsilon V_j} E_\omega N(\delta_G(Qc)) \\ &\leqq g_j(\log(r(G)/QcG(V_j))) \quad \text{if} \quad j \varepsilon Z(1) \quad \text{and} \quad G(V_j) > 0, \\ &\leqq g_j(\log(L/QcG(V_j))) \quad \text{if} \quad j \varepsilon J(1) - Z(1) \quad \text{and} \quad G(V_j) > 0, \end{aligned} \tag{2.6}$$

since $\max_i r(G, \delta_i) \leqq L$, while the g_j's are increasing on $[0, \infty)$. Evidently, each of the functions $g_j^*(x) = g_j(x) - g_j(0), j \varepsilon \cup Z(i)$, is non-negative, concave, and strictly increasing, and therefore satisfies $g_j^*(x + y) \leqq g_j^*(x) + g_j^*(y)$ for all $x, y \geqq 0$, or, equivalently, $g_j(x + y) \leqq g_j(x) + g_j^*(y)$. Letting $g^i = \max_{j \varepsilon J(i)} g_j^*$, we obtain

$$g_j(\log(r(G)/QcG(V_j))) \leqq g_j(|\log G(V_j)|) + g^1(\log(r(G)/Qc)) \tag{2.7}$$

for $j \varepsilon Z(1)$ and $G(V_j) > 0$; while, for $j \varepsilon J(1) - Z(1)$ and $G(V_j) > 0$,

$$\begin{aligned} &g_j(\log(L/QcG(V_j))) \\ &\leqq g_j(|\log G(V_j)|) + g_j^*(\log(L/Qc)) \\ &\leqq g_j(|\log G(V_j)|) + g^1(\log(r(G)/Qc)) + g^1(\log(L/r(G))). \end{aligned} \tag{2.8}$$

Since $\{V_j\,; j \varepsilon J(1)\}$ covers Ω and $\{V_j\,; j \varepsilon Z(1)\}$ covers Ω_1, from (2.6)–(2.8) we obtain the following.

If $r(G) = r(G, \delta_1) \geqq Qc$, then

$$\begin{aligned} \int_\Omega E_\omega N(\delta_G(Qc)) G(d\omega) &\leqq \sum_{j \varepsilon \bar{J}(1)} G(V_j) g_j(|\log G(V_j)|) \\ &\quad + g^1(\log(r(G)/Qc)) + G(\Omega - \Omega_1) g^1(\log(L/r(G))), \end{aligned} \tag{2.9}$$

where $\bar{J}(1) = \{j \varepsilon J(1) : G(V_j) > 0\}$.

Since $r(G) = r(G, \delta_1) \geqq bG(\Omega - \Omega_1)$,

$$\begin{aligned} G(\Omega - \Omega_1) g^1(\log(L/r(G))) &\leqq (r(G)/b) g^1(\log(L/r(G))) \\ &= (L/b)[(r(G)/L) g^1(\log(L/r(G)))]. \\ &\leqq (L/b) \sup_{0 < x \leqq 1} x g^1(|\log x|). \end{aligned} \tag{2.10}$$

The supremum is finite because it is equal to the maximum (over $j \varepsilon J(i)$) of the supremum (over $x \varepsilon (0, 1]$) of the strictly concave function $x g_j^*(|\log x|)$.

(The strict concavity follows from the fact that g_j^* is concave and strictly increasing.)

To obtain an upper bound on the first term on the right hand side of (2.9), we first show that there is no loss of generality in assuming that the V_j's (for $j \in J(1)$) are disjoint. If they are not, it suffices to replace each V_j by its subset $T_j = V_j - \cup\{V_k : k < j;\ k \in J(1)\}$ (assuming the integers in $Z(1)$ are smaller than those in the other $Z(i)$'s). Since $G(T_j) \leqq G(V_j)$ for all j, the T_j's clearly satisfy (2.6); also, the T_j's are disjoint and cover Ω, while $\{T_j\,;\ j \in Z(1)\}$ covers Ω_1.

Assuming, then, that the V_j's, for $j \in J(1)$, are disjoint, their measures, $G(V_j)$, add up to one. Therefore,

$$(2.11) \qquad \sum_{j\in J(1)} G(V_j) g_j(|\log\ G(V_j)|) \leqq \sup \sum_{j\in J(1)} \alpha_j g_j(|\log\ \alpha_j|),$$

where the supremum is taken over all sets of α_j's $\geqq 0$ summing to one, and $\alpha_j g_j(|\log \alpha_j|)$ is defined to be zero for $\alpha_j = 0$.

The supremum in (2.11) is bounded above by the sum over $j \in J(1)$ of $\sup_{0<x\leq 1} x g_j(|\log x|)$, which is finite by the same argument used in connection with (2.10). For $i = 1, \cdots, d$, let

$$D_i = (L/b) \sup_{0<x\leq 1} x g^i(|\log x|) + \sup \sum_{j\in J(i)} \alpha_j g_j(|\log \alpha_j|),$$

where the supremum in the last term is taken as in (2.11). This supremum can be calculated by applying Lemma 2.3. We conclude from (2.9)–(2.11) and similar results for the cases $r(G) = r(G, \delta_i)(i = 2, \cdots, d)$ that

$$(2.12) \qquad \int_\Omega E_\omega N(\delta_G(Qc)) G(d\omega) \leqq \max_{i=1,\cdots,d} [D_i + g^i(\log(r(G)/Qc))], \quad \text{if } r(G) \geqq Qc.$$

We show first that, for all $j \in \cup Z(i)(i = 1, \cdots, d)$, the function $x - g_j^*(\log(x/Q))$ is convex on $[Q, \infty)$ and approaches $+\infty$ as x approaches $+\infty$. Since both g_j^* and $\log(x/Q)$ are concave and strictly increasing, so is $g_j^*(\log(x/Q))$, and therefore $x - g_j^*(\log(x/Q))$ is convex. Since g_j^* is concave and non-negative on $[0, \infty)$, $g_j^*(x)/x$ is monotone decreasing and non-negative, and, hence, as x approaches $+\infty$, $g_j^*(\log(x/Q))$ is $O(\log x)$ and $x - g_j^*(\log(x/Q))$ approaches $+\infty$. It follows that (for $i = 1, \cdots, d$ and $j \in \cup Z(i)$) the equation

$$(2.13) \qquad x - g_j^*(\log(x/Q)) = D_i + Q$$

has a unique solution in $[Q, \infty)$, since the left hand side equals Q at $x = Q$ while the right hand side is larger than Q.

Let M^* be the largest of the solutions of (2.13) for all i and j such that $j \in J(i)$. For $x \geqq M^*$, each of the convex functions $x - g_j^*(\log(x/Q)), j \in \cup Z(i)$ $(i = 1, \cdots, d)$, is increasing; therefore, if $c^{-1}r(G) \geqq M^*$, $c^{-1}r(G) - g_j^*(\log(r(G)/Qc)) \geqq D_i + Q$, for all i, j with $j \in J(i)$, and, hence,

$$(2.14) \qquad \text{if } r(G) \geqq M^*c, \quad c^{-1}r(G) - Q \geqq D_i + g^i(\log(r(G)/Qc)),$$

for $i = 1, \cdots, d$.

Since $M^* > Q$, (2.12) and (2.14) imply that (2.3) holds for any $Q > 0$ and M^* (depending on Q) as defined above, and the proof of the lemma is complete.

THEOREM 2.1. *If Assumptions* I–V *are satisfied, then for any* $Q > 0$ *there is an* M *such that*

$$r_c(F, \delta_F(Qc)) - r_c(F, \delta_F{}^*(c)) \leqq Mc, \tag{2.15}$$

for every a priori *distribution,* F, *cost per observation,* $c > 0$, *and Bayes procedure,* $\delta_F{}^*(c)$.

PROOF. For any *a priori* distribution, F, and procedure, δ, let $e(F, \delta)$ be the integrated risk due to error. We shall first prove

$$\begin{aligned} e(F, \delta_F(Qc)) - e(F, \delta_F{}^*(c)) &\leqq Qc, \quad \text{if} \quad Q > 1, \\ &\leqq 0 \quad \text{if} \quad Q \leqq 1. \end{aligned} \tag{2.16}$$

For $Q > 1$, (2.16) follows from the estimate $e(F, \delta_F(Qc)) < Qc$, which was used in proving Lemma 2.2. For $Q \leqq 1$, we apply the well-known fact that (with probability 1) $\delta_F{}^*(c)$ does not continue when the stopping risk is less than c. It follows that the stopping risk when $\delta_F(c)$ stops (whose *a priori* expectation is $e(F, \delta_F(c))$) is (with probability 1) not larger than the stopping risk when $\delta_F{}^*(c)$ stops (whose *a priori* expectation is $e(F, \delta_F{}^*(c))$). Therefore, $e(F, \delta_F(c)) \leqq e(F, \delta_F{}^*(Qc))$, and the same argument can be applied to $\delta_F(Qc)$ for $Q < 1$.

Let $M^* > 1$ be any number satisfying Lemma 2.2 (for instance, the largest of the solutions of (2.13) obtained for $Q = 1$). Since $\delta_F{}^*(c)$ does not stop until the stopping risk is less than M^*c, it takes at least as many observations as $\delta_F(M^*c)$ (with probability 1). Letting $n_c(F, \delta) = r_c(F, \delta) - e(F, \delta)$, we have, therefore,

$$n_c(F, \delta_F{}^*(c)) \geqq n_c(F, \delta_F(M^*c)) \geqq n_c(F, \delta_F(Qc)), \quad \text{if} \quad Q \geqq M^*. \tag{2.17}$$

If $Q < M^*$, define $\tilde{F}$ as the (chance) *a posteriori* distribution when $\delta_F(M^*c)$ stops. The identity

$$n_c(F, \delta_F(Qc)) - n_c(F, \delta_F(M^*c)) = \int_\Omega E_\omega n_c(\tilde{F}, \delta_{\tilde{F}}(Qc))F(d\omega) \tag{2.18}$$

is proved by the standard type of argument used in connection with stopping variables.

By the definition of $\tilde{F}$, $r(\tilde{F}) < M^*c$, and therefore

$$n_c(\tilde{F}, \delta_{\tilde{F}}(Qc)) \leqq c \max_{i=1,\cdots,d} [D_i + g^i(\log [M^*/Q])] \tag{2.19}$$

follows from (2.12) in case $r(\tilde{F}) \geqq Qc$ and holds trivially in case $r(\tilde{F}) < Qc$ since the left hand member is zero.

Combining (2.17)–(2.19), we have, for $Q < M^*$,

$$n_c(F, \delta_F(Qc)) - n_c(F, \delta_F{}^*(c)) \leqq c \max_{i=1,\cdots,d} [D_i + g^i(\log [M^*/Q])]. \tag{2.20}$$

Putting together (2.16), (2.17), and (2.20), we obtain

$$
\begin{aligned}
& r_c(F, \delta_F(Qc)) - r_c(F, \delta_F{}^*(c)) \\
(2.21) \qquad & \leqq c \cdot Q && \text{if} \quad Q \geqq M^* \\
& \leqq c \cdot Q + c \cdot \max_{i=1,\cdots,d} [D_i + g^i(\log [M^*/Q])] && \text{if} \quad 1 < Q < M^* \\
& \leqq c \cdot c \cdot \max_{i=1,\cdots,d} [D_i + g^i(\log [M^*/Q])] && \text{if} \quad Q \leqq 1,
\end{aligned}
$$

which proves the theorem.

REMARK. Assuming $r(F) \geqq Qc$, $r_c(F, \delta_F(Qc))$ itself is bounded above by $Qc + c \max_{i=1,\cdots,d} [D_i + g^i(\log (r(F)/Qc))]$.

COROLLARY 2.2. *Under Assumptions* I–V, *for every* $Q > 0$ *and* a priori *distribution*, F, *with full support*

$$r_c(F, \delta_F{}^*(c))/r_c(F, \delta_F(Qc)) = 1 - O(|\log c|^{-1}), \quad \text{as} \quad c \to 0.$$

The corollary follows immediately from Theorem 2.1 and the corollary to Theorem 1 in (K-S), which implies that $r_c(F, \delta_F(Qc)) \geqq (\text{constant}) \cdot c\,|\log c|$ for small c.

REMARK 2. Possible choices of Q are $\lim_{x\to\infty} g_j{}^*(x)/x$ (which equals A_j if $g_j(x) = A_j x + B_j$), for $j \in \cup Z(i)$. The limits exist since each $g_j{}^*(x)/x$ is monotone decreasing and non-negative. (In fact, the limits can be shown to be positive.) If all the g_j's are of the form $A_j x + B_j$ it is easy to show that one of these choices of Q yields the smallest M^*, as defined in the proof of Lemma 2.2. In general, there exists a choice of Q for which the M^* defined just below (2.13) is smallest, and it is one of those values of Q which minimize the solution of (2.13) for some i and j such that $j \in J(i)$.

LEMMA 2.3. *Given* $H_j(\cdot), j = 1, \cdots, v (v \geqq 2)$, *strictly concave and continuous on* $[0, 1]$, *let* H_j' *be the left-hand derivative of* H_j, *which exists on* $(0, 1]$ *and is strictly decreasing.*

For $j = 1, \cdots, v$, *there is a function* $\chi_j(\cdot)$ *on* $(-\infty, \infty)$ *with the following properties:*

(1) $H_j'(\chi_j(y)) \geqq y \geqq H_j'(\chi_j(y)+)$ *if* $\chi_j(y) > 0$.

(2) χ_j *is non-increasing.*

(3) $\lim_{y\to\infty} \chi_j(y) = 0$ *and* $\lim_{y\to-\infty} \chi_j(y) = 1$.

There exists a number y^* *such that* $\sum_{j=1}^v \chi_j(y^*) = 1$ *and*

$$
(2.22) \quad \sum_{j=1}^v H_j(\alpha_j) \leqq \sum_{j=1}^v H_j(\chi_j(y^*)) \quad \text{if} \quad \sum_{j=1}^v \alpha_j = 1 \quad \text{and} \quad \alpha_j \geqq 0, \quad j = 1, \cdots, v.
$$

REMARK 3. For application to (2.11) and to evaluate the D_i's, we use $H_j(x) = x g_j(|\log x|)$ for $x \in (0, 1]$ with $H_j(0) = 0$. The required continuity at zero follows from the fact that $g_j(|\log x|) = O(|\log x|)$ as x approaches zero, which is proved by the same argument used just above (2.13).

It should be pointed out that Lemma 2.3 follows at once from an argument using Lagrange multipliers in case H_j' is continuous. The additional generality of Lemma 2.3 is sometimes needed when one obtains a g_j by taking the minimum of two other concave functions (see the discussion following Theorem 3.1).

PROOF OF LEMMA 2.3. Since H_j' is strictly decreasing on $(0, 1]$, its inverse is strictly decreasing on the set $H_j'((0, 1])$ where it is defined, and has range $(0, 1]$; there is an obvious extension of this inverse function to $(-\infty, \infty)$ which satisfies properties (2) and (3). This extension, χ_j, obviously satisfies property (1) in case $\chi_j(y)$ is a point of continuity of H_j'; in case H_j' has a jump at $\chi_j(y)$, it is easy to verify that property (1) holds since H_j' is left-continuous.

Each χ_j is continuous, since it is monotone and has range $(0, 1]$ or $[0, 1]$. Thus, $\sum_{j=1}^{v} \chi_j(y)$ is continuous, has limit zero as y approaches $+\infty$, and limit v ($v \geqq 2$) as y approaches $-\infty$. There is a number y^*, therefore, such that $\sum_{j=1}^{v} \chi_j(y^*) = 1$.

To prove (2.22) we first show that, for each j, $H_j(x) - y^*x$ (which is continuous on $[0, 1]$) attains its maximum when $x = \chi_j(y^*)$. Clearly, the maximum is attained at zero if $H_j'(x) - y^*$ is negative throughout $(0, 1]$; in this case, $\chi_j(y^*)$ is zero also, or else property (1) would be violated. If $H_j'(x) - y^*$ is non-negative at some point in $(0, 1]$, then the maximum of $H_j(x) - y^*x$ is attained at a point x^* in $(0, 1]$, and $H_j'(x) - y^*$ obviously attains its smallest non-negative value at x^* (uniquely); in this case, $\chi_j(y^*) = x^*$, by property (1) and the monotonicity of H_j'.

Now, if $\sum_{j=1}^{v} \alpha_j = 1$ and $\alpha_j \geqq 0, j = 1, \cdots, v$, we have

$$\begin{aligned}
\sum_{j=1}^{v} H_j(\alpha_j) &= y^* + \sum_{j=1}^{v} (H_j(\alpha_j) - y^*\alpha_j) \\
&\leqq y^* + \sum_{j=1}^{v} \max_{0 \leqq x \leqq 1} (H_j(x) - y^*x) \\
&= y^* + \sum_{j=1}^{v} [H_j(\chi_j(y^*)) - y^*\chi_j(y^*)] \\
&= \sum_{j=1}^{v} H_j(\chi_j(y^*)),
\end{aligned}$$

and the proof is complete.

In applying Lemma 2.3 to compute the D_i's, the fact that $\sum_{j=1}^{v} \chi_j(y)$ is non-increasing helps considerably in finding y^*.

REMARK 4. Because they always exist under our assumptions, bounds $g_j(x)$ of the form $A_jx + B_j$ are of particular interest. We have $g_j^*(x) = A_jx$, so that $g^i(x) = A^ix$, where $A^i = \max_{j \in J(i)} A_j$, for $i = 1, \cdots, d$. A simple upper bound on D_i can be obtained as follows: Let $B^i = \max_{j \in J(i)} B_j$ and let $s(i)$ be the number of elements in $J(i)$, for $i = 1, \cdots, d$. The first term in the expression for D_i (just above (2.12)) is L/b times A^i/e, since $\sup_{x \in (0,1]} x\,|\log x|$ equals $1/e$. The second term can be bounded above by replacing each g_j by g^i; the supremum then becomes $A^i \log s(i) + B^i$ by an obvious concavity argument (or by using Lemma 2.3). Thus

$$D_i \leqq (L/b) \cdot (A^i/e) + A^i \log s(i) + B^i.$$

It is easy to show that $D_i = O(\log s(i))$ as $s(i) \to \infty$, under appropriate interpretation and conditions (say, if $g_1, \cdots, g_{s(i)}$ is an initial segment of an infinite sequence $g_1, g_2, \cdots$ satisfying $g_1 \leqq g_k \leqq g_2$ for $k = 1, 2, \cdots$).

3. Bounds in the finite parameter space case. If Ω is a finite set and if, for $i = 1, \cdots, d$, we choose the covering $\{V_j ; j \in Z(i)\}$ of Ω_i to consist of the single-

point subsets of Ω_i, then the equivalent of (2.1), for arbitrary i, is

$$(\prod_{m=1}^{n} f_\omega(X_m))P_1(\omega) > \zeta \sum_{\theta \varepsilon \Omega - \Omega_i} (\prod_{m=1}^{n} f_\theta(X_m))P_2(\theta),$$

where $\{\omega\} = V_j$ (and $j \varepsilon Z(i)$).

Now, if $P_1(\omega) \geqq P_2(\Omega - \Omega_i) > 0$, the first n for which this relation is satisfied is never larger than the first n for which

$$\prod_{m=1}^{n} f_\omega(X_m) > \zeta \max_{\theta \varepsilon \Omega - \Omega_i} \prod_{m=1}^{n} f_\theta(X_m),$$

or, equivalently,

$$\min_{\theta \varepsilon \Omega - \Omega_i} \sum_{m=1}^{n} (\log f_\omega(X_m) - \log f_\theta(X_m)) > \log \zeta. \tag{3.1}$$

Therefore, if the expectation under ω ($\{\omega\} = V_j$) of the first n satisfying (3.1) is bounded above, for $\zeta \geqq 1$, by $g_j(\log \zeta)$ (where g_j is a concave function on $[0, \infty)$), then g_j satisfies (2.4). If, for every $j \varepsilon \cup Z(i) (i = 1, \cdots, d)$, such a g_j can be constructed, then an M (depending on Q) satisfying Theorem 2.1 can be computed by using (2.21). Of course, for $i = 1, \cdots, d$, it suffices to choose $J(i)$ containing $Z(i)$ and such that $\{V_j ; j \varepsilon J(i)\}$ is a covering of Ω by its single-point subsets. (In other words, for each $i = 1, \cdots, d$, one need only assign each point in $\Omega - \Omega_i$ to one of the Ω_k's containing it; of course, these assignments can be made so as to minimize the value of M obtained from (2.21).) Suppose $\Omega - \Omega_i = \{\theta_1, \cdots, \theta_p\}$; then we can write (3.1) in the equivalent form

$$\min_{r=1,\cdots,p} S^r(n) > \gamma, \tag{3.2}$$

where $\gamma = \log \zeta \geqq 0$, $S^r(0) = 0$, and

$$S^r(n) = \sum_{m=1}^{n} (\log f_\omega(X_m) - \log f_{\theta_r}(X_m)) \qquad \text{for} \quad n = 1, 2, \cdots.$$

The problem at hand is to construct a concave function, g, on $[0, \infty)$ satisfying

$$EN_p(\gamma) \leqq g(\gamma), \tag{3.3}$$

where

$$N_p(\gamma) = \inf \{n \mid \min_{r=1,\cdots,p} S^r(n) > \gamma\} \quad \text{for} \quad \gamma \geqq 0.$$

(It follows from (3.3) trivially, using Wald's equation that $g(\gamma) \geqq \gamma(\min_r ES^r(1))^{-1}$ for all γ, so that g is necessarily non-negative, unbounded, and, hence, strictly increasing on $[0, \infty)$.) For $r = 1, \cdots, p$, $S^r(n)$ is the nth partial sum of independent, identically distributed random variables. Letting S^r stand for $S^r(1)$ and re-indexing $S^1(n), \cdots, S^p(n)$ if necessary, we assume in the sequel that

$$0 < ES^1 \leqq ES^2 \leqq \cdots \leqq ES^p,$$

and

$$\operatorname{Var} S^r < \infty \quad \text{for} \quad r = 1, \cdots, p,$$

by virtue of Assumptions I–V.

The special cases $p = 1$ and $p = 2$ are of particular importance in connection with so-called Koopman-Darmois families, which we define as follows: suppose that for ω in an interval containing Ω (on the real line) we have (μ)-densities of the form

$$f_\omega(x) = \Gamma(\omega)h(x)e^{\psi(\omega)x},$$

where $\psi(\omega)$ is strictly monotone. Then (3.1) is equivalent to

$$(\psi(\omega) - \psi(\theta)) \sum_{m=1}^{n} X_m + n(\log \Gamma(\omega) - \log \Gamma(\theta)) > \log \zeta, \tag{3.4}$$

for $\theta = \theta_1, \cdots, \theta_p$.

We fix ω, ζ, and $X_1, \cdots, X_n$, and consider the left hand member of (3.4) as a function of θ (over the interval where ψ and Γ are defined). In case $\psi(\theta) = \theta$, $\log \Gamma(\theta)$ is concave (as is well-known) and, hence, the left-hand member of (3.4) is convex in θ. Since $\theta = \omega$ clearly fails to satisfy the inequality in (3.4), it follows that if $\theta = \theta^*$ satisfies it and $\theta^* > \omega$, then every $\theta \geqq \theta^*$ satisfies it by virtue of the convexity. (The same conclusion follows in the case of arbitrary strictly monotone $\psi(\theta)$ by an obvious argument, using the re-parameterization $\phi = \psi(\theta)$.) A similar conclusion applies for $\theta^{**} < \omega$, and thus if $\theta_1 < \cdots < \theta_q < \omega < \theta_{q+1} < \cdots < \theta_p$, then (3.4) is equivalent to

$$(\psi(\omega) - \psi(\theta)) \sum_{m=1}^{n} X_m + n(\log \Gamma(\omega) - \log \Gamma(\theta)) > \log \zeta, \tag{3.5}$$

for $\theta = \theta_q, \theta_{q+1}$, with the provision that if $\omega < \theta_1$, $\theta_q = \theta_{q+1} = \theta_1$, and if $\theta_p > \omega$, then $\theta_q = \theta_{q+1} = \theta_p$.

We have shown that, for Koopman-Darmois families, the problem to construct g satisfying (3.3) arises only for $p = 1$ or $p = 2$.

We now state the principal theorem we shall use to solve the problem of (3.3).

THEOREM 3.1. *Assume* $S^1, S^2, \cdots, S^p (p \geqq 2)$ *have finite variances and that* $0 < ES^1 \leqq ES^2 \leqq \cdots \leqq ES^p$. *Let* $\tau(\cdot)$ *be a mapping of* $\{2, \cdots, p\}$ *into* $\{1, \cdots, p-1\}$ *satisfying* $\tau(r) < r$ *for* $r = 2, \cdots, p$. *Let* R *and* $\bar{R}$ *be complementary subsets of* $\{2, \cdots, p\}$ *such that* $ES^{\tau(r)} < ES^r$ *for all* $r \in R$ *and* $ES^{\tau(r)} \leqq \nu_r ES^r (\nu_r \leqq 1)$ *for* $r \in \bar{R}$. *Then, for all* $\gamma \geqq 0$,

$$\begin{aligned} ES^1 \cdot EN_p(\gamma) \leqq{}& E \min_{r=1,\cdots,p} S^r(N_p(\gamma)) \\ &+ \sum_{r \in R} E \sup_{n \geqq 0} (S^{\tau(r)}(n) - S^r(n)) \\ &+ \tfrac{1}{2}[EN_p(\gamma)]^{\frac{1}{2}} \sum_{r \in \bar{R}} [\operatorname{Var}(S^{\tau(r)} - \nu_r S^r)]^{\frac{1}{2}}. \end{aligned} \tag{3.6}$$

PROOF. The expected value of $N_p(\gamma)$ is finite by Theorem 3 of Farrell (1964) and, applying Wald's equation, we obtain

$$\begin{aligned} ES^1 \cdot EN_p(\gamma) = ES^1(N_p(\gamma)) = {}& E \min_{r=1,\cdots,p} S^r(N_p(\gamma)) \\ &+ E[S^1(N_p(\gamma)) - \min_{r=1,\cdots,p} S^r(N_p(\gamma))]. \end{aligned} \tag{3.7}$$

Starting with the fact that $\tau(2)$ is necessarily 1, it is straightforward to show that, for all n,

$$S^1(n) - \min_{r=1,\cdots,p} S^r(n) \leqq \sum_{r=2}^{p} [S^{\tau(r)}(n) - S^r(n)]^+. \tag{3.8}$$

By (3.7) and (3.8), clearly

$$(3.9)\quad ES^1 \cdot EN_p(\gamma) \leqq E \min_{r=1,\cdots,p} S^r(N_p(\gamma)) + \sum_{r=2}^{p} E[S^{\tau(r)}(N_p(\gamma)) - S^r(N_p(\gamma))]^+.$$

For $r \in R$, $ES^{\tau(r)} < ES^r$ and, hence,

$$(3.10)\quad E[S^{\tau(r)}(N_p(\gamma)) - S^r(N_p(\gamma))]^+ \leqq E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)] < \infty.$$

For $r = 2, \cdots, p$, $r \notin R$, we let $\Delta = \nu_r ES^r - ES^{\tau(r)} (\Delta \geqq 0)$ and write N as a shorthand for $N_p(\gamma)$. By Wald's equation,

$$E[S^{\tau(r)}(N) - \nu_r S^r(N) + \Delta N]^+ - E[S^{\tau(r)}(N) - \nu_r S^r(N) + \Delta N]^- = 0,$$

and, hence, we have (since $S^r(N) \geqq \gamma \geqq 0$)

$$\begin{aligned} E[S^{\tau(r)}(N) - S^r(N)]^+ &\leqq E[S^{\tau(r)}(N) - \nu_r S^r(N) + \Delta N]^+ \\ &= \tfrac{1}{2} E|S^{\tau(r)}(N) - \nu_r S^r(N) + \Delta N| \\ &\leqq \tfrac{1}{2}[\mathrm{Var}\,(S^{\tau(r)}(N) - \nu_r S^r(N) + \Delta N)]^{\frac{1}{2}}. \end{aligned} \tag{3.11}$$

By Theorem 2 of Chow, Robbins, and Teicher (1965),

$$\begin{aligned} (3.12)\quad \mathrm{Var}\,(S^{\tau(r)}(N) - \nu_r S^r(N) + \Delta N) &= EN \cdot \mathrm{Var}\,(S^{\tau(r)} - \nu_r S^r + \Delta) \\ &= EN \cdot \mathrm{Var}\,(S^{\tau(r)} - \nu_r S^r). \end{aligned}$$

Theorem 3.1 follows from (3.9) by applying (3.10) for $r \in R$ and applying (3.11) and (3.12) for $r \notin R$.

REMARK 5. Useful upper bounds on $E \sup_{n \geqq 0} (S^{\tau(r)}(n) - S^r(n))$ are given in the discussion following Theorem 3.3. These bounds hold under more restrictive assumptions than those of Theorem 3.1. When these assumptions are not satisfied, $\bar{R}$ can be chosen to be $\{2, \cdots, p\}$, so that R is empty.

Clearly, one should choose $\tau(r)$ and ν_r for $r \in \bar{R}$ to minimize $\mathrm{Var}\,(S^{\tau(r)} - \nu_r S^r)$ and, for $r \in R$, should choose $\tau(r)$ so that the best available estimate for $E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)]$ is smaller than that for any other choice. Choosing R advantageously is slightly more complicated. As is illustrated in the examples considered below, the most advantageous choice of R may well depend on γ. (It is easy to verify that the best R does not decrease as γ increases.) In this case, if one uses Theorem 3.1 (with different choices of R) to obtain concave bounds satisfying (3.3), none of them is smallest for all γ, and it pays to use their minimum, which is also concave and, clearly, also solves the problem in (3.3).

The next theorem provides a means of constructing estimates of the term $E \min_r S^r(N_p(\gamma))$ which appears in the right hand side of (3.6). It is a generalization of Theorem 2 of Farrell (1964), which deals with the case $p = 1$. In addition, it contains slight improvements of Farrell's result that require more than the trivial modification of his argument necessary for a straightforward generalization. The nature of the improvements can be seen as follows: for $p = 1$, if Δ is defined to be $EW_0(|S^1|)$ (instead of the smaller value it is given in Theorem 3.2), then (3.11) is equivalent to Farrell's result upon setting $w = 0$ in (3.11).

THEOREM 3.2. *Assume* $ES^r > 0$ *and* $E|S^r| < \infty$ *for* $r = 1, \cdots, p$ $(p \geqq 1)$. *Let* W_0 *be a convex function strictly increasing on* $[0, \infty)$ *and satisfying* $W_0(0) = 0$ *and* $\lim_{x\to\infty} W_0(x)/x = \infty$. *Assume that* $EW_0((S^r)^+) < \infty$, $r = 1, \cdots, p$.

Let w *be the right-hand derivative of* W_0 *at zero and define*

$$W(x) = W_0(x) \quad \text{if} \quad x \geqq 0$$
$$= wx \quad \text{if} \quad x < 0.$$

Set $\Delta = EW(\max_r S^r)$. *Then* $\Delta < \infty$ *and*

$$E \min_{r=1,\cdots,p} S^r(N_p(\gamma)) \leqq \gamma + W_0^{-1}(\Delta EN_p(\gamma) - w\gamma). \tag{3.11}$$

PROOF. Since $w \geqq 0$, $W(x) \leqq W_0(x^+)$ for all x and, hence,

$$\Delta = EW(\max_r S^r) \leqq EW_0((\max_r S^r)^+)$$
$$= EW_0(\max_r (S^r)^+) \leqq E \sum_{r=1}^p W_0((S^r)^+) < \infty.$$

Using N as a shorthand notation for $N_p(\gamma)$, observe that

$$\max_r [S^r(N) - S^r(N-1)] \geqq [\gamma - \max_r S^r(N-1)]^+ + [\min_r S^r(N) - \gamma]. \tag{3.12}$$

Since $W(0) = 0$ and W is convex and non-decreasing, $W(z) \geqq W(x) + W(y)$ provided $z \geqq x + y$ and $x, y \geqq 0$. Thus, noting that the terms in (3.12) are non-negative, we conclude that

$$W(\max_r [S^r(N) - S^r(N-1)]) \geqq W([\gamma - \max_r S^r(N-1)]^+) + W(\min_r S^r(N) - \gamma). \tag{3.13}$$

Now, $W(x) \geqq wx$ for all x. Using this fact and (3.13), we have

$$\begin{aligned} &\sum_{k=1}^N W(\max_r [S^r(k) - S^r(k-1)]) \\ &\quad \geqq W(\min_r S^r(N) - \gamma) + w \sum_{k=1}^{N-1} \max_r [S^r(k) - S^r(k-1)] \\ &\qquad + w[\gamma - \max_r S^r(N-1)]^+ \\ &\quad \geqq W(\min_r S^r(N) - \gamma) + w\gamma. \end{aligned} \tag{3.14}$$

Theorem 3 of Farrell (1964) has already been cited to show that $EN_p(\gamma)$ is finite. Hence, Wald's equation can be applied to the sum on the left hand side of (3.14), and we obtain (using Jensen's inequality)

$$\begin{aligned} \Delta EN = EW(\max_r S^r) \cdot EN &= E \sum_{k=1}^N W(\max_r [S^r(k) - S^r(k-1)]) \\ &\geqq EW(\min_r S^r(N) - \gamma) + w\gamma \\ &= EW_0(\min_r S^r(N) - \gamma) + w\gamma \\ &\geqq W_0(E \min_r S^r(N) - \gamma) + w\gamma. \end{aligned} \tag{3.15}$$

The assumptions on W_0 imply that it is continuous and has range $[0, \infty)$. Since W_0 is strictly increasing, W_0^{-1} is defined and increasing on $[0, \infty)$, and (3.11) follows at once from (3.15).

REMARK 6. Under our assumption that $S^1, \cdots, S^r$ have finite variances, $W_0(x) = x^2$ satisfies the hypotheses of the theorem. If $S^1, \cdots, S^r$ have moment generating functions in a neighborhood of zero, then $W_0(x) = e^{wx} - 1$ (for sufficiently small $w > 0$) may be used.

To show how Theorem 3.1 and Theorem 3.2 can be used to solve the problem in (3.3), we put (3.6) and (3.11) together to obtain

$$\begin{aligned} ES^1 \cdot EN_p(\gamma) &\leqq \gamma \\ &+ W^*(\Delta EN_p(\gamma) - w\gamma) + \textstyle\sum_{r \in R} E \sup_{n \geqq 0} (S^{\tau(r)}(n) - S^r(n)) \\ &+ \tfrac{1}{2}[EN_p(\gamma)]^{\frac{1}{2}} \textstyle\sum_{r \in \bar{R}} [\mathrm{Var}\,(S^{\tau(r)} - \nu_r S^r)]^{\frac{1}{2}}, \end{aligned} \tag{3.16}$$

where W^* is W^{-1} (which is a concave strictly increasing extension of W_0^{-1} to $(-\infty, \infty)$) unless $w = 0$, in which case W^* is W_0^{-1}. By this device, $W^*(\Delta EN_p(\gamma) - w\gamma)$ is well-defined for every ordered pair of non-negative real numbers, $(\gamma, EN_p(\gamma))$, and it is easy to verify that the set of such pairs satisfying (3.16) is a convex subset of the first quadrant of (Euclidean) 2-space and contains $\{(\gamma, 0) | \gamma \geqq 0\}$. Thus, if this convex set is bounded above, it has an upper boundary which can be represented by a concave function $g(\cdot)$ on $[0, \infty)$, and this $g(\cdot)$ satisfies (3.3). (The boundedness follows easily from the assumption in Theorem 3.2 that $\lim_{x \to \infty} W_0(x)/x = \infty$.) Of course, the same observations are valid if in (3.16) Δ, $E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)]$, and $\mathrm{Var}\,(S^{\tau(r)} - \nu_r S^r)$ are replaced by upper bounds on these quantities.

Thus, concave bounds, g, satisfying (3.3) can be determined implicitly by applying Theorems 3.1 and 3.2 to obtain an inequality of the form (3.16). When it is feasible to solve such a relation for $g(\gamma)$ only "one value of γ at a time," a variety of more or less obvious methods can be used to approximate the D_i's and (where necessary) the g's themselves in order to obtain M^* (and bounds on the integrated risk).

We now give an explicit solution, $g(\cdot)$, of the problem in (3.3) which depends only upon two parameters (besides p itself): a (positive) lower bound on the first moments of $S^1, \cdots, S^p$, and an upper bound on the second moments of $S^1, \cdots, S^p$.

THEOREM 3.3. *If $ES^r \geqq \mu > 0$ and $E(S^r)^2 \leqq \eta < \infty$ for $r = 1, \cdots, p$ $(p \geqq 1)$, then*

$$EN_p(\gamma) \leqq \gamma/\mu + \tau(\gamma/\mu + \tau^2/4)^{\frac{1}{2}} + \tau^2/2,$$

where $\tau = (p + p^{\frac{1}{2}} - 1)\eta^{\frac{1}{2}}/\mu$.

PROOF. Applying Theorem 3.1 with $\tau(r) = 1$ for all r and R empty, we obtain

$$\mu EN_p(\gamma) \leqq E \min_r S^r(N_p(\gamma)) + [EN_p(\gamma)]^{\frac{1}{2}}(p - 1)\eta^{\frac{1}{2}}, \tag{3.17}$$

using the obvious estimate $\mathrm{Var}\,(S^1 - S^r) \leqq 4\eta$.

Applying Theorem 3.2 with $W_0(x) = x^2$, we obtain $\Delta = EW(\max_r S^r) \leqq E \sum_{r=1}^{p} (S^r)^2 \leqq p\eta$ and, hence,

$$E \min_r S^r(N_p(\gamma)) \leqq \gamma + (p\eta)^{\frac{1}{2}}[EN_p(\gamma)]^{\frac{1}{2}}. \tag{3.18}$$

By (3.17) and (3.18), $\mu EN_p(\gamma) \leqq \gamma + (p + p^{\frac{1}{2}} - 1)\eta^{\frac{1}{2}}[EN_p(\gamma)]^{\frac{1}{2}}$, or, equivalently, $EN_p(\gamma) \leqq \gamma/\mu + \tau[EN_p(\gamma)]^{\frac{1}{2}}$. The proof is completed by the obvious manipulation of this last relation.

Of course, the crude bounds on Δ and Var $(S^1 - S^r)$ used in obtaining (3.17) can be improved if more information on $S^1, \cdots, S^p$ is available. (The choice of $\tau(r) \equiv 1$ and $\nu_r \equiv 1$ might be improved upon, also.) The chief interest of Theorem 3.3 stems from its usefulness in the general setting of Assumptions I–V. Proving the existence of an upper bound on $EN_p(\gamma)$ which depends only on μ and η (for fixed p, γ) is an important step in the proof of (2.31) in (K-S) and, hence, in the proof of Lemma 2.1. Using Theorem 3.3 with the Kiefer-Sacks argument, it is sometimes possible to carry out the proof of Lemma 2.1 constructively; that is, to obtain the A_j's and V_j's of Lemma 2.1 explicitly, and thus to construct an M^* satisfying Lemma 2.2 and bounds satisfying Theorem 2.1. This possibility hinges on whether one has precise enough information about the functions (like $\lambda_i(\omega)$) which appear in Assumptions I–V so that the Kiefer-Sacks compactness arguments can be carried out constructively. (For common Koopman-Darmois families, such as normal and exponential distributions, it is easy to verify that this can be done.)

We have so far discussed the application of Theorem 3.1 only with R empty. To use non-empty R's, we need to estimate $E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)]$. Theorem 4.1 in Spitzer, (1956) states that

$$(3.19) \quad E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)] = \sum_{k=1}^{\infty} (1/k) E[S^{\tau(r)}(k) - S^r(k)]^+.$$

When good estimates of $E[S^{\tau(r)}(k) - S^r(k)]^+$ can be constructed and the resulting infinite series can be summed, (3.19) yields an estimate, as desired (for instance, when $S^{\tau(r)}$ and S^r are exponentially distributed). If $S^{\tau(r)} - S^r$ has a moment generating function, then one can use the estimate

$$(3.20) \quad E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)] \leqq 1/\sup \{t \mid E \exp [t(S^{\tau(r)} - S^r)] \leqq 1\};$$

(3.20) follows immediately from the well-known relation

$$P[\sup_{n \geqq 0} (S^{\tau(r)}(n) - S^r(n)) \geqq u > 0] \leqq \exp [(-u) \sup \{t \mid E \exp [t(S^{\tau(r)} - S^r)] \leqq 1\}],$$

which is derived from the fundamental identity of Wald for sequential analysis. Note that if the moment generating function that appears in the right hand member of (3.20) exists for some positive t, then the supremum indicated is positive, since $E \exp [t(S^{\tau(r)} - S^r)] = 1 + tE(S^{\tau(r)} - S^r) + o(t)$ as $t \to 0+$ while $E(S^{\tau(r)} - S^r) < 0$ by assumption. We point out that if the moment generating function exists for $t = t_0 > 0$, it exists for all positive $t \leqq t_0$; also, the value of the moment generating function for t equal to this supremum is not necessarily one (see Bahadur and Rao (1960)), although it is indeed one if the function is continuous. When necessary, a sufficiently close upper bound on the moment generating function can be used to find a positive value of t, say t_0, for which

$E \exp [t_0(S^{\tau(r)} - S^r)] \leqq 1$, so that $1/t_0$ is an upper bound on the left hand member of (3.20).

For problems involving normal distributions (with known variance) a very good estimate of $E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)]$ is available because $S^{\tau(r)}(n) - S^r(n)$ in the nth partial sum of normal random variables. For a Weiner process, $X(t)$, with negative mean drift per unit time, μ, and variance per unit time, σ^2,

$$E \sup_{t \geqq 0} X(t) = \sigma^2/2(-\mu),$$

by a result of Doob (1949). It follows easily that

$$(3.21) \quad E \sup_{n \geqq 0} [S^{\tau(r)}(n) - S^r(n)] \leqq \operatorname{Var} (S^{\tau(r)} - S^r)/2E(S^r - S^{\tau(r)}),$$

when $S^{\tau(r)} - S^r$ is normally distributed.

An upper bound due to Wald (1947) can often be used, as an alternative to Theorem 3.2, to estimate $E \min_r S^r(N_p(\gamma))$. We shall next see that, when the right-hand side of (3.22) is finite, it is smaller for large γ than any estimate obtained by Theorem 3.2 for $p = 1$; and for all γ it provides the best estimate known to the author for use in the examples given below.

Wald's bound for the case $p = 1$ is given by

$$(3.22) \qquad E[S^1(N_1(\gamma)) - \gamma] \leqq \sup_{k \geqq 0} E[S^1 - k \mid S^1 > k].$$

Note that the right hand side is independent of γ. For $p = 1$, Theorem 3.2 yields an estimate of the form

$$(3.23) \qquad E[S^1(N_1(\gamma)) - \gamma] \leqq W_0^{-1}(\Delta E N_1(\gamma) - w\gamma).$$

Using the hypotheses of Theorem 3.2 and applying Jensen's inequality, it is easy to show that the right-hand side of (3.23) approaches ∞ as γ becomes large. Hence, (3.22) gives a better estimate of $ES^1(N_1(\gamma))$ than (3.23) for sufficiently large γ, provided the supremum in (3.22) is finite. A sufficient condition for this supremum to be finite is given below. However, even when it is finite, Wald's bound is often not computable and is sometimes a very poor bound; for instance, in the case of distributions of S^1 having large "gaps."

In this connection, a very interesting result is Theorem 9 of Karlin's paper on the renewal equation (1955), which serves to evaluate the limit as $\gamma \to \infty$ of the quantity on the left-hand side of (3.22) in terms of the first two moments of the non-negative "ladder variable" $S^1(N_1(0))$ (both of which are finite if the third moment of S^1 is finite). It follows easily from the existence of the limit as $\gamma \to \infty$ that $E[S^1(N_1(\gamma)) - \gamma]$ is bounded above uniformly in γ. Unfortunately, an explicit bound does not seem to be obtainable from Karlin's proof of Theorem 9.

A sufficient condition for Wald's bound to be finite and a means of approximating it when direct computation is impossible can be given in terms of the so-called "hazard rate," $h(x)(1 - \Phi(x))^{-1}$, where Φ is the distribution function of S^1 and h is its (Lebesgue) density function. By a straightforward computation,

$$(3.24) \qquad \text{if } T \leqq h(x)(1 - \Phi(x))^{-1} \leqq T^* > 0 \quad \text{for} \quad x \geqq k \geqq 0,$$
$$\text{then } 1/T \leqq E[S^1 - k \mid S^1 > k] \leqq 1/T^*.$$

From (3.24), it is easy to show that Wald's bound is finite if $\liminf h(x)(1 - \Phi(x))^{-1}$ (as $x \to \infty$) is positive.

The above discussion of Wald's bound applies to the case $p = 1$ only. For Koopman-Darmois families, only the cases $p = 1$ and $p = 2$ need be considered, as already shown; and we now show how the problem of estimating $E \min_r S^r(N_2(\gamma))$ $(r = 1, 2)$ for such families can be approached by using Wald's bound for the case $p = 1$.

By an argument just like that required for (3.22),

$$(3.25)\quad E[\min_{r=1,2} S^r(N_2(\gamma)) - \gamma] \leqq \operatorname{ess\,sup}_{\{(K,R)\mid \max(K,R)\geqq 0\}} E[\min (S^1 - K, S^2 - R) \mid S^1 \geqq K, S^2 \geqq R]$$

where the essential sup is a restriction to pairs (K, R) for which

$$P\{S^1 \geqq K, S^2 \geqq R\} > 0.$$

Now, for $K \geqq 0$, we have, for some $\tilde{R}$ depending on R,

$$E[\min (S^1 - K, S^2 - R) \mid S^1 \geqq K, S^2 \geqq R] \leqq E[S^1 - K \mid S^1 \geqq K, S^2 \geqq R] = E[S^1 - K \mid S^1 \geqq K, S^1 \leqq \tilde{R}] \leqq E[S^1 - K \mid S^1 \geqq K],$$

where the equality holds because $S^2 \geqq R$ and $S^1 \leqq \tilde{R}$ are identical events, for some $\tilde{R}$ depending on R (in the special case $p = 2$ for K-D families, it is clear from (3.2) and (3.5) that S^1 and S^2 are monotone in the observed variable X, one increasing and the other decreasing). The restriction to pairs for which $P\{S^1 \geqq K, S^2 \geqq R\} > 0$ insures that $\tilde{R} < K$.

Using the above and a similar estimate for $R \geqq 0$, we obtain

$$\operatorname{ess\,sup}_{\{(K,R)\mid \max(K,R)\geqq 0\}} E[\min (S^1 - K, S^2 - R) \mid S^1 \geqq K, S^2 \geqq R] \leqq \max_{r=1,2} \sup_{K\geqq 0} E[S^r - K \mid S^r \geqq K],$$

and, by (3.25),

$$(3.26)\quad E \min_{r=1,2} S^r(N_2(\gamma)) \leqq \gamma + \max_{r=1,2} \sup_{K\geqq 0} E[S^r - K \mid S^r \geqq K].$$

(3.26) is used in the examples involving normal distributions discussed below.

4. Examples and further results. We now apply Theorem 3.1 to obtain an M^* satisfying Lemma 2.2 in several problems involving normal distributions. In a testing problem concerning the mean of a normal distribution with known variance, the densities (with respect to Lebesgue measure) form a Koopman-Darmois family. Therefore, the argument leading to (3.5) shows that only the cases $p = 1$ and $p = 2$ of the problem stated in (3.3) need be considered, provided the parameter space, Ω, is finite (no matter how many decisions, d, are allowed).

Suppose f_1 and f_2 are normal densities with means μ_1 and μ_2, respectively, and common variance σ^2. If f_1 is the (true) density of the (observed) random variable X, then $\log f_1(X) - \log f_2(X)$ is normally distributed with mean α and variance 2α, where $\alpha = (\frac{1}{2})(\mu_1 - \mu_2)^2/\sigma^2$. Thus, we are led to consider the following problems.

PROBLEM I. Find $g(\gamma)$ satisfying (3.3) when $p = 1$ and S^1 is normal with mean α and variance 2α ($\alpha > 0$).

PROBLEM II. Find $g(\gamma)$ satisfying (3.3) when $p = 2$, S^1 is normal with mean α and variance 2α ($\alpha > 0$), and $S^2 = \beta + (\beta/\alpha)^{\frac{1}{2}}(\alpha - S^1)$, where $\beta \geqq \alpha$.

The relationship between S^1 and S^2 stated in Problem II is a direct consequence of the definition of $S^r(n)$ in (3.2).

Problem I can be solved by using Wald's bound (relation (3.22)), which gives (Wald (1947), p. 180),

$$\alpha g(\gamma) = \gamma + \xi(\alpha), \quad \text{where} \quad \xi(\alpha) = \alpha + (2\alpha)^{\frac{1}{2}}\phi((\alpha/2)^{\frac{1}{2}})/\Phi((\alpha/2)^{\frac{1}{2}}), \tag{4.1}$$

and ϕ and Φ are the standard normal density function and distribution function, respectively. (No choice of $W_0(\cdot)$ in Theorem 3.2 known to the author yields a solution to Problem I which is smaller than that given in (4.1) for any $\gamma \geqq 0$.)

From Theorem 3.1 and (3.26), we obtain the following two solutions of Problem II:

$$\alpha g(\gamma) = \gamma + \xi(\alpha) + (1 + \nu^{\frac{1}{2}})/(1 - \nu^{\frac{1}{2}}), \quad \text{if} \quad \nu = \alpha/\beta < 1, \tag{4.2}$$

and

$$\alpha g(\gamma) = \gamma + \xi(\alpha) + \tfrac{1}{4}(1 + \nu^{\frac{1}{2}})^2 + [(1 + \nu^{\frac{1}{2}})/2^{\frac{1}{2}}](\gamma + \xi(\alpha) + \tfrac{1}{8}(1 + \nu^{\frac{1}{2}})^2)^{\frac{1}{2}} \tag{4.3}$$

where $\nu = \alpha/\beta$ and $\xi(\alpha)$ is as defined in (4.1). Relation (3.21) is used to obtain the final term in (4.2), while (4.3) comes from the obvious manipulation of (3.6) where $\bar{R} = \{2\}$ and R is empty, $\tau(2) = 1$, and $\nu_2 = \nu = \alpha/\beta$. In applying (3.26) to obtain the $\xi(\alpha)$ term in (4.2) and (4.3), it is necessary to observe that $\xi(\alpha) \leqq \xi(\beta)$; that is, $\xi(\alpha)$ is increasing in α (in fact $(2\alpha)^{-\frac{1}{2}}\xi(\alpha)$ is increasing in α, by the calculation on pp. 168–169 of Wald (1947)).

The solution given in (4.2) applies to Problem II only when $\alpha < \beta$. When it applies, (4.2) is clearly smaller for large γ than (4.3). (In fact, if $\alpha \geqq \frac{1}{2}$, then for sufficiently small ν, depending on α, (4.2) is smaller than (4.3) for all γ). However, for a wide range of values of α and ν, it turns out that no disadvantage is incurred (i.e. the computed M^* is no larger) if one uses the bound in (4.3) for all γ. An explanation of this phenomenon is given in the discussion following Example 4.

To keep the number of numerical examples small, we will consider only problems where Ω contains three points. (The numerical results do not change dramatically as the number of points in Ω and the number of decisions, d, is increased.) Specifically, we consider the following two problems.

Two-decision Problem. Given three normal distributions with unit variance and means μ, 0, $-\mu$ ($\mu > 0$), decision 1 is correct when μ or 0 is the true mean and decision 2 is correct when $-\mu$ or 0 is the true mean. The loss is zero for correct decisions and one for incorrect decisions.

Three-decision Problem. Given three normal distributions with unit variance

and means $\mu_1, 0, \mu_2$ ($\mu_2 < 0 < \mu_1 \leqq -\mu_2$), decisions 1, 2, and 3 are correct for the true values $\mu_1, 0, \mu_2$, respectively. The loss is "zero-one," as in the two-decision problem.

EXAMPLE 1. (Two-decision Problem; $\mu = 1$). It is easy to verify that for both decisions (that is, $i = 1, 2$) it is sufficient that $\{g_j \mid j \varepsilon J(i)\}$ be taken as $\{g_1, g_2, g_2\}$ where g_1 is given by (4.1) for $\alpha = \frac{1}{2}$ and g_2 is given by (4.1) for $\alpha = 2$. Hence,

$$g_1(\gamma) = 2\gamma + 2.019 \quad \text{and} \quad g_2(\gamma) = \tfrac{1}{2}\gamma + 1.285$$

suffice (the decimals having been rounded off upwards). We obtain $D_1 = D_2 \leqq 3.435$. The choice $Q = 2$ can easily be shown to yield the smallest M^* satisfying Lemma 2.2. For this choice of Q, it turns out to be sufficient that M^* satisfy

$$M^* = 2 \log (M^*/2) + 5.435.$$

The solution of this equation is slightly smaller than 8.3 so that Lemma 2.2 is satisfied for $M^* = 8.3$.

It is evident upon examination of (2.21) and (2.13) that if M^* is obtained from (2.13) for $Q = Q^*$ and then (2.21) is used to obtain M (depending on Q) satisfying Theorem 2.1, then the question as to which Q gives the smallest M in Theorem 2.1 can be answered as follows. The choice $Q = M^*$ gives $M = M^*$ (and so does the choice $Q = Q^*$); the choice $Q = 1$ gives

$$M = \max_{i=1,\cdots,d} [D_i + g^i(\log M^*)]. \tag{4.4}$$

The smallest M is the smaller of M^* and the M given in (4.4).

For the problem in Example 1, the choice $Q = 1$ gives the smallest M; for this Q, $M = 7.7$ satisfies Theorem 2.1. (For the same problem with arbitrary $\mu \geqq 2/e$, $Q = 1$ gives the smallest M; for $\mu < 2/e$, $Q = 2/\mu^2$ gives the smallest M and $Q = 1$ does not.)

EXAMPLE 2. (Wiener process, two decisions). Suppose that in the two-decision problem the three normal distributions are replaced by the corresponding (in the obvious sense) Weiner processes and the experimenter observes $X(t)$ continuously from $t = 0$, stopping whenever he chooses. Then the entire decision problem and all the results above can be formulated with only trivial changes, and all the same results hold, with one exception. The bounds given in (2.16) and (2.21) for $Q \leqq 1$ need no longer apply and we must instead use the bounds given for $1 < Q < M^*$. This change is necessary because the argument used to prove (2.16) for $Q \leqq 1$ relies on the trivial observation that no Bayes procedure continues sampling when the *a posteriori* risk is smaller than c; but this breaks down for the Wiener process problem.

The Wiener process version of the problem treated in Example 1 is of interest because one can use in place of (3.27) the exact relation $\alpha E N_1(\gamma) = \gamma$, so that

$$g_1(\gamma) = 2\gamma \quad \text{and} \quad g_2(\gamma) = \tfrac{1}{2}\gamma$$

suffice for the case $\mu = 1$ considered in Example 1. Carrying out the same kind of computations as were used above, we find $D_1 = D_2 \leqq 1.838$, and for the (still

preferable) choice $Q = 2$, M^* is the solution of

$$M^* = 2 \log (M^*/2) + 3.838.$$

The solution of this equation is slightly smaller than 6.1, which is not a very great reduction of the value $M^* = 8.3$ obtained in Example 1.

EXAMPLE 3. (Three-decision Problem; $\mu_1 = 1$, $\mu_2 = -\frac{3}{2}$). For all three decisions (that is, $i = 1, 2, 3$) it is sufficient that the set $\{g_j \mid j \varepsilon J(i)\}$ be taken as $\{g_1, g_2, g_3\}$ where g_1 and g_2 are given by (4.1) for $\alpha = \frac{1}{2}$ and $\alpha = \frac{2}{3}$, respectively, and g_3 is given by (4.3) for $\alpha = \frac{1}{2}$ and $\nu = \frac{4}{9}$. We use

$$g_1(\gamma) = 2\gamma + 2.019, \qquad g_2(\gamma) = \tfrac{8}{9}\gamma + 1.519,$$

$$g_3(\gamma) = 2\gamma + 3.408 + (\tfrac{5}{9}\gamma + 7.536)^{\frac{1}{2}}.$$

Calculation yields $D_3 \leqq 7.118$. (D_1, D_2 need not be calculated, since they are smaller than D_3 and clearly do not play any genuine role in determining M^*.) The choice $Q = 2.7$ is close to the value of Q for which (2.13) yields the smallest M^*. For $Q = 2.7$, it is easily determined that M^* need only satisfy

$$M^* = 2 \log (M^*/2.7) + (\tfrac{5}{9} \log (M^*/2.7) + 7.536)^{\frac{1}{2}} + 7.072.$$

The solution of this equation is slightly smaller than 14.6.

EXAMPLE 4. (Three-decision Problem; $\mu_1 = 1$, $\mu_2 = -1$). As above, proper selection of the sets of indices required in the proof of Lemma 2.1 results in $\{g_j \mid j \varepsilon J(i)\}$ equal to $\{g_1, g_2, g_3\}$ (for $i = 1, 2, 3$), where for this example g_1 and g_2 are both given by (4.1) for $\alpha = \frac{1}{2}$, and g_3 is given by (4.3) for $\alpha = \frac{1}{2}$ and $\nu = 1$. We have

$$g_1(\gamma) = g_2(\gamma) = 2\gamma + 2.019,$$

$$g_3(\gamma) = 2\gamma + 4.019 + (8\gamma + 12.076)^{\frac{1}{2}}.$$

Calculation yields $D_3 \leqq 9.027$, and for $Q = 2.8$ we obtain

$$M^* = 2 \log (M^*/2.8) + (8 \log (M^*/2.8) + 12.076)^{\frac{1}{2}} + 8.352.$$

The solution of this equation is slightly smaller than 17.2.

It is interesting to note that in Example 3 the bound $g_3(\gamma)$ obtained from (4.3) yields the same smallest M^* (over all possible choices of Q) as the bound obtained by taking the minimum of the right-hand members of (4.2) and (4.3). This phenomenon can be made clearer as follows.

For the problem in Example 3, an easy computation shows that the smallest value, γ_0, for which the bound in (4.2) is not larger than the one in (4.3) is at least ten. On the other hand, one can easily show in this problem that only values of γ smaller than four "count" in comparing $g_3(\gamma)$ with an alternative bound, in the following precise sense. Any alternative concave bound which is larger than $g_3(\gamma)$ for all $\gamma \leqq 4$ is larger for all γ than the minimum of $g_3(\gamma)$ and $(\gamma - 4)/\alpha + g_3(4)$, by an argument using the concavity of the g's and the lower bound on $EN_2(\gamma)$ from Wald's equation, γ/α. Now, with the other bounds g_1 and g_2

used in the example, it is easy to show that the M^* (for the best choice of Q) obtained using $g_3(\gamma)$ is as small as the M^* obtained (for any Q) by using the minimum of $g_3(\gamma)$ and $(\gamma - 4)/\alpha + g_3(4)$. Hence, only a concave bound which is smaller than $g_3(\gamma)$ for some γ in the interval [0, 4] (where 4 is a convenient choice, not the smallest possible) can "be better than $g_3(\gamma)$," as far as determining a smaller M^* is concerned.

More generally, for "competing" concave bounds satisfying (3.3), it is the size of the bounds for small γ that determines their value in yielding small M in Theorem 2.1. Indeed, the bounds obtainable from Theorem 3.1 are often poor for large γ, and there is a quite different approach to the construction of $g(\gamma)$'s which yields bounds that are substantially better for large γ, but not as valuable for Theorem 2.1. This approach is based upon the inequality

$$EN_p(\gamma) \leqq EN_1(\gamma) + EN_1(0) \sum_{n=1}^{\infty} P[S^1(n) > \gamma \geqq \min_{r>1} S^r(n)],$$

which can be proved by an argument using so-called "ladder variables." The terms in the infinite series can be estimated by exponential bounds, assuming S^1 has a moment generating function in a neighborhood of zero. When $ES^1 < ES^2$, this approach can be used to show that $EN_p(\gamma) - EN_1(\gamma)$ approaches zero exponentially as γ becomes large.

The numerical results we have given imply in a simple way numerical results for a very wide class of problems involving three normal distributions with equal variances. It is clear from (2.13) and the argument following it that if (2.13) is used (for some Q) to construct an M^* satisfying Lemma 2.2, then that M^* is determined by the concave g_j's $(j \in \cup Z(i))$, assuming L and b are fixed. Furthermore, multiplying all the g_j's by the same constant multiplies the smallest M^* (over all possible Q) by that constant. (The value of Q yielding the smallest M^* gets multiplied by the same constant.)

Now, we can apply this observation by noting that the bounds in (4.1) and (4.3) have the property that $\alpha g(\gamma)$ is monotone increasing in α for fixed γ and ν. (It was shown above that $\xi(\alpha)$ is monotone increasing.) Considering the problem of Example 1 for $\mu = \mu_0 < 1$, we reason as follows. For $\mu = \mu_0$, we have $\alpha = \alpha_0 = \frac{1}{2}\mu_0^2 < \frac{1}{2}$ (which is the value of α for $\mu = 1$) and, hence, by the monotonicity just observed, the bound in (4.1) for $\alpha = \alpha_0$ is smaller than $1/2\alpha_0$ times the bound for $\alpha = \frac{1}{2}$; therefore, $1/2\alpha_0$ times the bounds used for $\mu = 1$ can be used in (2.13) to determine M^*. Doing this, one obtains M^* (the smallest value, for all Q) equal to $(2\alpha_0)^{-1}$ times the (smallest) M^* for the problem with $\mu = 1$; that is, $M^* = 8.3/\mu_0^2$ suffices.

Example 1 can be extended still further, to the case of means $\mu_1 < \mu_2 < \mu_3$ of normal distributions with common variance, σ^2, satisfying $\mu_3 - \mu_2 \leqq \mu_2 - \mu_1$ and $\mu_3 - \mu_2 \leqq \sigma$ (or, similarly, $\mu_2 - \mu_1 \leqq \mu_3 - \mu_2$ and $\mu_2 - \mu_1 \leqq \sigma$). For this problem, we have $\alpha = (\mu_3 - \mu_2)^2/2\sigma^2$ and $\beta = (\mu_2 - \mu_1)^2/2\sigma^2$, and, hence, $\alpha \leqq \frac{1}{2}$. Besides the monotonicity of $\alpha g(\gamma)$ in (4.1) and (4.3) discussed in the last paragraph, we need to observe that $g(\gamma)$ of (4.1) is decreasing in α and that $\alpha g(\gamma)$ of (4.3) is increasing in ν for fixed α and γ, and is thus, decreasing in β. Using

these observations together with the remarks about (2.13) given above, it is straightforward to derive the bound $M^* = 8.3/2\alpha = 8.3\sigma^2/(\mu_3 - \mu_2)^2$ from the fact that $M^* = 8.3$ suffices when $\alpha = \beta = \frac{1}{2}$.

For this same problem (with $\mu_3 - \mu_2 \leqq \mu_2 - \mu_1$ and $\mu_3 - \mu_2 \leqq \sigma$) one can, of course, use (4.1) and (4.3) directly to compute a sufficient M^*, as was done in Example 1. It is evident that the M^* so obtained will be smaller than the value $8.3\,\sigma^2/(\mu_3 - \mu_2)^2$, derived by the "short-cut" method. In case $\mu_3 - \mu_2 = \mu_2 - \mu_1$, it is easy to see, however, that one cannot obtain in this manner an M^* smaller than $6.1\,\sigma^2/(\mu_3 - \mu_2)^2$; the reason for this is that the bounds in (4.1) and (4.3) are larger than the corresponding bounds for the Wiener process problem involving the same μ_1, μ_2, μ_3, and σ, and the latter yield $M^* = 6.1\,\sigma^2/(\mu_3 - \mu_2)^2$ by a straightforward extension of Example 2.

A simple short-cut derivation of a sufficient M^* for Lemma 2.2 can be made, by means of similar arguments, to extend Examples 3 and 4. Extending Example 3 to the Three-decision Problem with normal means $\mu_1 < \mu_2 < \mu_3$ and common variance, σ^2, we find that $M^* = (17.2)\sigma^2/(\mu_3 - \mu_2)^2$ is sufficient for Lemma 2.2 provided $\mu_3 - \mu_2 \leqq \mu_2 - \mu_1$ and $\mu_3 - \mu_2 \leqq \sigma$. Similarly, for Example 4, $M^* = (14.6)\,\sigma^2/(\mu_3 - \mu_2)^2$ suffices whenever $\mu_3 - \mu_2 \leqq (\frac{2}{3})(\mu_2 - \mu_1)$ and $\mu_3 - \mu_2 \leqq \sigma$.

REFERENCES

ALBERT, A. E. (1961). The sequential design of experiments for infinitely many states of nature. *Ann. Math. Statist.* **32** 774–799.

ANDERSON, T. W. (1960). A modification of the sequential probability ratio test to reduce the sample size. *Ann. Math. Statist.* **31** 165–197.

BAHADUR, R. R. and RAO, R. R. (1960). On deviations of the sample mean. *Ann. Math. Statist.* **31** 1015–1027.

CHERNOFF, H. (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–770.

CHOW, Y. S., ROBBINS, H., and TEICHER, H. (1965). Moments of randomly stopped sums. *Ann. Math. Statist.* **36** 789–799.

DONNELLY, T. G. (1957). A family of truncated sequential tests. Doctoral dissertation, Univ. of North Carolina.

DOOB, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **20** 393–403.

FARRELL, R. H. (1964). Limit theorems for stopped random walks. *Ann. Math. Statist.* **35** 1332–1343.

FUSHIMI, M. (1965). Properties of asymptotically optimal Bayes sequential tests. *Raag Res. Notes* **3** No. 90, 25 pp.

HOEFFDING, W. (1960). Lower bounds for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **31** 352–368.

KARLIN, S. (1955). On the renewal equation. *Pacific J. Math.* **5** 229–257.

KIEFER, J. and SACKS, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34** 705–750.

SCHWARZ, G. (1962). Asymptotic shapes of Bayes sequential testing regions. *Ann. Math. Statist.* **33** 224–236.

SOBEL, M. and WALD, A. (1949). A sequential decision procedure for choosing one of three hypotheses concerning the unknown mean of a normal distribution. *Ann. Math. Statist.* **20** 502–522.

SPITZER, F. (1956). A combinatorial lemma and its application to probability theory. *Trans. Amer. Math. Soc.* **82** 332–339.

WALD, A. (1947). *Sequential Analysis*. Wiley, New York.