A MODIFIED BAYES STOPPING RULE[1]

By Sigmund J. Amster[2]

University of North Carolina

1. Introduction of the modified Bayes rule. This paper describes a stopping rule for sequential sampling that weighs the cost of additional sampling against the expected gain to be derived from additional observations. The modified Bayes rule (MBR) requires one more observation to be taken as long as the posterior risk is larger than the expected posterior risk for some additional fixed-size sample. In this investigation, risk is defined within the Wald framework of statistical decision theory [6], using losses and costs.

A subjective probability justification for the MBR may be found in interpreting the value of ξ_0, the prior distribution on the parameter θ, at any set of θ values, as representing the original relative conviction that the true value of θ lies in this set. Once the first n observations have been taken, the belief has been changed as reflected in the values of the posterior distribution ξ_n. In either case, the distribution ξ determines *which* risk one would like to minimize. The MBR will accept R_n, the present average risk, (i.e., stop sampling) only if the cost of increasing one's convictions, through knowledge of a sample of any *fixed* size, is more than the expected amount to be gained. If not, one more sample will be taken and the same problem posed with the (hoped-for) better knowledge of the true state of nature.

In the context treated above, the defining property of the Bayes sequential rule (BSR), that of minimizing the *original* average risk, does not seem particularly relevant. However, the method which determines this rule, by comparing R_n with the average risk "resulting from a continuation if at each future stage we did the *best* we could with the resulting observations" ([1], page 243), is really the optimal property. The MBR tries to approximate this by considering only the average risk if any fixed-size future sample were taken.

The calculation of the MBR is feasible whenever the Bayes fixed sample procedure can be explicitly obtained. (Part III of [5] can offer assistance in the evaluation of the integrals involved.) However, except for certain cases, as testing two simple hypotheses, when Wald's sequential probability ratio test (SPRT) is such a rule, or if the BSR is truncated or fixed sample, it is not usually possible to carry out the Bayes procedure. Even in these cases, the determination of the appropriate SPRT is not simple and in the truncated cases the necessary computations are exceedingly tedious.

Received July 12, 1962; revised June 17, 1963.

[1] Submitted in partial fulfillment of the requirements for the Ph.D. degree, University of North Carolina. This research was partially supported by the Office of Naval Research under contract No. Nonr-855(09).

[2] Now at Bell Telephone Laboratories, Whippany, New Jersey.

The MBR calculations are readily adaptable to a high-speed computer since the analysis before and after a sample has been observed differs only in the change from a prior to a posterior distribution of θ. The problem is further simplified if the prior and posterior distributions are of the same functional form, with only a change in parameters, since the same kind of calculation is then required at each stage. A distribution satisfying this latter property has been called a *natural conjugate prior distribution* [5] and a *distribution closed under sampling* [8].

During the investigation of properties of the MBR, the book [5] by Raiffa and Schlaiffer was published. Although the authors specifically refrained from discussing the sequential decision problem, it was found that their *preposterior analysis* is exactly the same as the evaluation of an $R_{n(k)}$ (the expected risk of continuing to the $(n + k)$th observation given the datum at stage n). Previous to this, Wald [6] page 151, briefly discussed the computation of $\min_k R_{n(k)}$ to obtain the optimal second stage sample size. In the discussion of [2], Anscombe briefly mentioned a special case of the MBR. A summary of the current state of sequential research is given in [4].

In Section 3, by defining a sequence of stopping rules having the MBR as a limit, the average risk for the MBR is found to be the limit of a non-increasing sequence whose initial value is that for the fixed-sample size Bayes procedure. Since the average risk of the MBR is, of course, not less than that of the BSR, the two rules coincide if the BSR is actually a fixed-sample size rule. It is shown that the MBR is a SPRT for the problem of testing two simple hypotheses. It is also shown that the actual sample size required by the MBR is never larger than that for the BSR. This is a mathematical translation of the fact that the BSR will not stop unless no possible continuation is expected to improve your lot. At the point of termination, the improvement possible over the MBR is identical to the difference between the average risk of a fully sequential Bayes procedure and that of a fixed-sample size Bayes procedure in which the sample size is zero, the expectations taken with respect to the posterior distribution of θ.

In Section 4, the *asymptotic minimax rule* [7] and MBR are compared. In Section 5, a binomial estimation example is presented in some detail. Although the BSR requires a fairly complicated "working-backward" method, the MBR requires only the solution of several second degree equations.

In Section 6, an example is presented in which the particular SPRT equivalent to the MBR is found for a two simple hypothesis testing problem with proportional cost and simple loss function. In Section 7, several areas are indicated where future research may be profitable.

2. Assumptions and definition of MBR. To simplify the exposition the following assumptions are made:

A1. The experiment consists of observing, possibly sequentially, the random variables $x_1, x_2, \cdots$ (real or vector-valued) which are independent with a common probability density $f(\cdot \mid \theta)$ with respect to a given σ-finite measure μ. [Note: The same notation, $f(\cdot \mid \theta)$, is used to denote the (joint) density of any

number of random variables having the (possibly vector-valued) parameter $\theta \in \Omega$.]

A2. For any $\theta \in \Omega$ and any terminal decision d, the loss $L(\theta; d)$ is bounded from below.

A3. Let $c_n^*(\mathbf{y}_k) \equiv c_{n+k}(\mathbf{x}_{n+k}) - c_n(\mathbf{x}_n)$, where $\mathbf{y}_k \equiv (x_{n+1}, \cdots, x_{n+k})$, and $c_r(\mathbf{x}_r)$ is the cost of observing $\mathbf{x}_r \equiv (x_1, \cdots, x_r)$. It is assumed that for all $\mathbf{x}_n$, $c_n^*(\mathbf{y}_{k+1}) > c_n^*(\mathbf{y}_k)$. Also, for any sequence $x_1, x_2, \cdots$; it is assumed that $\lim_k c_n^*(\mathbf{y}_k) = \infty$.

A4. Any measurability assumptions needed to assure the existence (finitely) of the integrals used for the procedure are made.

A5. The existence of a Bayes terminal decision rule is assumed. A Bayes terminal decision is used exclusively here and is denoted by d.

The identical distribution assumption is made solely to avoid notational complexity. A2 is used primarily to justify the use of the Fubini theorem in proving properties of the MBR. A3 implies that $\inf_k R_{n(k)}$ (see below) is actually attained. A5 is used to avoid the detailed analysis (e.g., [1] page 297) or restrictive assumptions (e.g., [6] page 89) otherwise necessary to verify the existence of a Bayes terminal decision rule in each particular case.

The following notation will be used: for any fixed $\mathbf{x}_n$; $n = 0, 1, \cdots$, and $k = 0, 1, \cdots$, $X^k = X \times X \times \cdots \times X$ (k times), (X the spectrum of a single observation), $\xi_0(\theta)$ = the prior distribution of θ, $d\xi_n(\theta) = f(\mathbf{x}_n \mid \theta)\, d\xi_0(\theta)/f_0(\mathbf{x}_n)$,

$$f_0(\mathbf{x}_n) = \int_\Omega f(\mathbf{x}_n \mid \theta)\, d\xi_0(\theta), dF(\mathbf{y}_k \mid \theta) = \prod_1^k f(y_i \mid \theta)\, d\mu(y_i),$$

$$\bar{c}_{n(k)} = \int_\Omega \int_{X^k} c_n^*(\mathbf{y}_k)\, dF(\mathbf{y}_k \mid \theta)\, d\xi_n(\theta),$$

$$\bar{L}_{n(k)} = \int_\Omega \int_{X^k} L[\theta; d(\mathbf{x}_n, \mathbf{y}_k)]\, dF(\mathbf{y}_k \mid \theta)\, d\xi_n(\theta),$$

and $R_{n(k)} = R_{n(k)}(\mathbf{x}_n) = \bar{L}_{n(k)} + \bar{c}_{n(k)}$. (Note: $R_{n(0)} \equiv R_n$ is the $U_n(x)$ and $R_{0(k)} \equiv R_{(k)}$, $R_0 = R_{(0)}$ is the U_0 in [1].) $R_{n(k)}$ may be interpreted as our present "best guess" of the posterior risk, R_{n+k}, if an additional sample of fixed size k would be observed. $R_{(k)}$ is the average risk of a Bayes fixed-sample size (k) procedure.

The formal definition of the MBR: at the start of sampling (if $n = 0$), or after $\mathbf{x}_n$ is observed (if $n > 0$), (i) if $R_n = \inf_k R_{n(k)}$, stop sampling; (ii) if $R_n \neq \inf_k R_{n(k)}$, observe x_{n+1}. (In either case, A1–A5 implies that the infimum is actually achieved.)

NOTE. An equivalent formulation in terms of sets of distributions may sometimes be more useful (see Section 3). For each n a set of distributions (Ξ_n) is defined such that sampling stops if and only if ξ_n is in Ξ_n.

For the special case of $c_k(\mathbf{x}_k) = kc$, the stopping rule can be conveniently expressed in terms of $\gamma_{n(k)} = (\bar{L}_n - \bar{L}_{n(k)})/k$. That is, stop sampling as soon as $c \geqq \tilde{\gamma}_n \equiv \max_k \gamma_{n(k)}$.

3. General results. The following additional notation will be used in this section: δ_B is the *Bayes sequential rule* (BSR). δ_j is the stopping rule which follows the MBR up to and including a sample of size j; if $R_k \neq \min_t R_{k(t)}$ for $k = 0, 1, \cdots, j$ then a sample of fixed size m is taken. (Where m is the smallest t such that $R_{j(t)} = \min_i R_{j(i)}$.) δ_0 is thus the fixed sample-size Bayes rule and δ_∞ the MBR. $\delta_j(\mathbf{x}_t)$ is the stopping rule of type δ_j started after $\mathbf{x}_t$ is observed (if sampling is continued beyond $\mathbf{x}_{t+j}$, it is with a single fixed-size sample). α_B is the average risk of δ_B. $\alpha_B^{(N)}$ is the average risk of the Bayes rule truncated after N observations. α_j is the average risk of δ_j. $\alpha_j(\mathbf{x}_t)$ is the posterior expected risk of $\delta_j(\mathbf{x}_t)$, using ξ_t. N_B is the sample size using δ_B, a random variable. N_j is the sample size using δ_j, for $j = 0, 1, \cdots, \infty$, a random variable for $j > 0$. E_i is the conditional expectation holding $\mathbf{x}_i$ fixed: specifically, for any $h(\mathbf{x}_{i+1})$,

$$E_i[h(\mathbf{x}_{i+1})] = \int_X h(\mathbf{x}_{i+1}) f_i(x_{i+1})\, d\mu(x_{i+1})$$

where $f_i(x_{i+1}) = \int_\Omega f(x_{i+1} \mid \theta)\, d\xi_i(\theta)$. $A_t = \{\mathbf{x}_t : N_\infty = t\}$ is the stopping region for the MBR.

Several properties follow immediately from the definitions. For any n, δ_{n+1} will differ from δ_n only if the sequence of observations $\mathbf{x}_n$ is such that $R_n > \min_k R_{n(k)}$. A necessary condition for the two rules to differ is therefore that $N_n > n$. Also, as soon as a particular $\mathbf{x}_n$ has been observed, the rule $\delta_k(\mathbf{x}_n)$ will act exactly the same as the rule δ_{n+k}. For example, using the rule δ_n ($n \geqq 1$), the observing of x_1 changes ξ_0 into ξ_1 and hence δ_n into $\delta_{n-1}(x_1)$. $\delta_{n-1}(x_1)$ may be regarded as using a different prior distribution for a new decision problem, where the truly sequential portion has been reduced by one.

The additional properties shown here are based primarily on the recognition that:

$$\alpha_B^{(N)} = \min\{R_0, E_0[\min\{R_1, E_1[\cdots E_{N-2} \min\{R_{N-1}, E_{N-1}R_N\}]\cdots\}]\},$$

$$\alpha_0 = \min\{R_0, E_0R_1, E_0E_1R_2, \cdots\}, \qquad \alpha_0(x_1) = \min\{R_1, E_1R_2, E_1E_2R_3, \cdots\},$$

$$\alpha_k = R_0 \text{ if } \alpha_0 = R_0, \text{ and } \alpha_k = E_0\alpha_{k-1}(x_1) \text{ if } \alpha_0 < R_0, k \geqq 1.$$

Except for the first expression which is proved in [1], these follow almost directly from the definitions (but see the lemmas).

The repetitive nature of the formulas for α_k imply a ready adaptability to programming for high-speed computers. It is noted that the computation of α_k is *not* necessary to perform (but only to evaluate) a δ_k.

LEMMA 1. *For all* $k \geqq 1$, $R_{(k)} = E_0E_1 \cdots E_{k-1}R_k$. Prove by applying E_{k-1}, E_{k-2}, $\cdots$, E_0 successively.

LEMMA 2. *If* $\alpha_0 < R_0$, *then for all* $k \geqq 1$, $\alpha_k = E_0\alpha_{k-1}(x_1)$ *and* $\alpha_1 = E_0 \min[R_1, E_1R_2, E_1E_2R_3, \cdots]$. Prove by considering the effect of a particular x_1 having been observed, and then averaging.

LEMMA 3. *Let* $\xi = \lambda\xi' + (1 - \lambda)\xi''$, $\lambda \in [0, 1]$, *and define* $\alpha_0'(\mathbf{x}_m)$, $\alpha_0''(\mathbf{x}_m)$, $\cdots$, *as the* $\alpha_0, \cdots$, *if* ξ', ξ'' *is used. Then,* $\alpha_0(\mathbf{x}_m) \geqq \lambda\alpha_0'(\mathbf{x}_m) + (1 - \lambda)\alpha_0''(\mathbf{x}_m)$.

PROOF. Define i as the first integer such that $\alpha_0 = R_{(i)} = \min R_{(k)}$. Then $\lambda\alpha_0' + (1-\lambda)\alpha_0'' = \lambda \min R_{(k)}' + (1-\lambda)\min R_{(k)}'' \leqq \lambda R_{(i)}' + (1-\lambda)R_{(i)}'' = R_{(i)} = \alpha_0$.

THEOREM 1. $\alpha_B \leqq \alpha_\infty \leqq \cdots \leqq \alpha_N \leqq \cdots \leqq \alpha_1 \leqq \alpha_0$.

PROOF.

1. $\alpha_1 = R_0$, if $\alpha_0 = R_0$; and $\alpha_1 = E_0[\min\{R_1, E_1R_2, \cdots\}]$ if $\alpha_0 < R_0$. $\alpha_1 \leqq \min[R_{(1)}, R_{(2)}, \cdots]$ if $\alpha_0 < R_0$ (using Lemma 1), and therefore $\alpha_1 \leqq \min R_{(k)} = \alpha_0$.

2. $\alpha_2 \leqq \alpha_1$: If $\alpha_0 = R_0$, then $\alpha_0 = \alpha_1 = \alpha_2$ since no sampling is performed. If $\alpha_0 < R_0$, than $\alpha_1(x_1) \leqq \alpha_0(x_1)$ since (1) was proved for any ξ_0 and is thus true for ξ_1. Taking expectations, $E_0\alpha_1(x_1) \leqq E_0\alpha_0(x_1)$, and therefore $\alpha_2 \leqq \alpha_1$ by Lemma 2.

3. $\alpha_2(x_1) \leqq \alpha_1(x_1)$ since (2) was proved for any ξ_0 and is thus true for ξ_1. $\alpha_3 \leqq \alpha_2$ follows by the method used in the second part of (2). $\alpha_m \leqq \alpha_{m-1}$ follows by repetition, i.e., $\alpha_3(x_1) \leqq \alpha_2(x_1)$, etc.

4. $\alpha_B \leqq \alpha_\infty$ is the fundamental property of δ_B.

THEOREM 2. (see Theorem 9.4.3 of [1]).

If Ω is finite, then $\Xi_n(d')$ is convex, where $\Xi_n(d') = \{\xi_n : \mathbf{x}_n \text{ is in } A_n \text{ and } d = d'\}$ is such that ξ_n is in $\Xi_n(d')$ if and only if the MBR requires sampling to stop at stage n and a particular terminal decision (d') to be made.

PROOF. Let $\Omega = \{\theta_i\}_1^m$, $H(i, d) = L[\theta_i, d(\mathbf{x}_n)] + c_n(\mathbf{x}_n)$ for $i = 1, \cdots, m$, and $\xi_n(i) \equiv \xi_n(\theta_i) = \lambda\xi_n'(\theta_i) + (1-\lambda)\xi_n''(\theta_i)$, for $0 \leqq \lambda \leqq 1$. Fix $\mathbf{x}_n$ such that both ξ_n' and ξ_n'' are in $\Xi_n(d')$. If $n = 0$, interpret $\alpha_0(\mathbf{x}_n) = \alpha_0$, etc. Now $\lambda\sum_1^m \xi_n'(i)H(i, d') + (1-\lambda)\sum_1^m \xi_n''(i)H(i, d')$ equals $\lambda\alpha_0'(\mathbf{x}_n) + (1-\lambda)\alpha_0''(\mathbf{x}_n)$, since both ξ_n' and ξ_n'' are in $\Xi_n(d')$. This is not greater than $\alpha_0(\mathbf{x}_n)$ by Lemma 3; which is less than or equal to R_n, since $\alpha_0(\mathbf{x}_n)$ equals $\min_k R_{n(k)}$. This latter value equals $\min_{d^*} \sum_1^m \xi_n(i)H(i, d^*)$ which is not greater than $\sum_1^m \xi_n(i)H(i, d')$. Therefore, $\alpha_0(\mathbf{x}_n) = R_n = \sum_1^m \xi_n(i)H(i, d')$; i.e., ξ_n is in $\Xi_n(d')$.

COROLLARY 2.1. *If Ω contains only two points, $c_n(\mathbf{x}_n) = nc$ and $L(i, d) = s_i$ for $d \neq i$ and zero otherwise, then there exists an* SPRT *equivalent to the* MBR.

PROOF. It is shown in [1], page 267, that the convexity of $\Xi_n(d')$ is sufficient for this conclusion.

NOTE. It can be shown, following [1] page 259, that if, for any $\mathbf{x}_n$, the marginal cost $c_n^*(\mathbf{y}_k)$ depends only on the values in the vector $\mathbf{y}_k$, the sets $\Xi_n(d')$ do not depend on n. This result does not require the finiteness of Ω. For example, $c_k(\mathbf{x}_k) - \sum_1^k c(x_i)$, where $c(x_i)$ is the cost of the observation x_i, is sufficient.

THEOREM 3. $N_\infty \leqq N_B$.

The proof depends entirely on the fact that $E[\min(\cdot)] \leqq \min E(\cdot)$.

COROLLARY 3.1. *Suppose $c_k(\mathbf{x}_k) = kc$.*

(i) *If $\bar{L}_n \to 0$ uniformly in $\mathbf{x}_n$, then δ_∞ is truncated.*

(ii) *If $\bar{L}_n$ is a function of n only, then δ_∞ is the Bayes fixed-sample size procedure.*

PROOF.

(i) By [1], Theorem 9.3.3, δ_B is truncated. But $N_\infty \leqq N_B$.

(ii) By the same theorem, δ_B is a fixed-sample size procedure. But, using Theorem 1, $\alpha_B \leqq \alpha_\infty \leqq \alpha_0 = \alpha_B$.

4. Asymptotic minimax comparison. Wald [7] introduced the Asymptotic Minimax (AM) sequential estimates for $c_k(\mathbf{x}_k) = kc$ and $L(\theta;\ \hat{\theta}_n) = (\hat{\theta}_n - \theta)^2$. "Asymptotic" refers to the limit as c tends to zero and $\hat{\theta}_n$ is the maximum likelihood estimator of θ, given $\mathbf{x}_n$. He exhibited two AM solutions, T_c^0 and T_c^1, which will use the following notation:

$$\tilde{d}_0 = \inf_\theta \tilde{d}(\theta) = \inf_\theta E\{[\partial \log f(y/\theta)/\partial\theta]^2 \mid \theta\}.$$

T_c^0 : Take N_c observations, $N_c \geqq (c\,\tilde{d}_0)^{-\frac{1}{2}}$, and estimate θ by $\hat{\theta}_{N_c}$.
T_c^1 : Stop sampling as soon as $c \geqq \gamma_n' \equiv [n(n+1)\,\tilde{d}(\hat{\theta}_n)]^{-1}$ and estimate θ by $\hat{\theta}_n$.

Since the AM procedures are among the very few general sequential estimating procedures for which explicit solutions are available, a comparison with the

TABLE 1

$f(x \mid \theta)$	$\xi_0(\theta)$	γ_n'	$\tilde{\gamma}_n$	$\hat{\theta}_n$	d
$N(\theta, 1)$	$N(\theta_0, \tau^{-2})$	$\{n(n+1)\}^{-1}$	$\{(n+\tau^2)(n+\tau^2+1)\}^{-1}$	$\bar{x}_n$	$\dfrac{\{\bar{x}_n + \tau^2\theta_0/n\}}{\{1 + (\tau^2/n)\}}$
$B(\theta)$	$Be(a, b)$	$\dfrac{\bar{x}_n(1-\bar{x}_n)}{n(n+1)}$	$\dfrac{\left(\bar{x}_n + \dfrac{a}{n}\right)\left(1 - \bar{x}_n + \dfrac{b}{n}\right)}{\left(1 + \dfrac{a+b}{n}\right)^2 (n+1+a+b)^2}$	$\bar{x}_n$	$\dfrac{\{\bar{x}_n + (a/n)\}}{\{1 + ((a+b)/n)\}}$
$P(\theta)$	$\Gamma(\theta_0)$	$\dfrac{\bar{x}_n}{n(n+1)}$	$\dfrac{\bar{x}_n + (\theta_0/n)}{(n+1)^2\{1 + (2/n)\}}$	$\bar{x}_n$	$\dfrac{\{\bar{x}_n + (\theta_0/n)\}}{\{1 + (1/n)\}}$

where: $B(\theta)$ is the binomial distribution, $P(\theta)$ is the Poisson, $Be(a, b)$ is the beta, and $\Gamma(\theta_0) \propto e^{-t}t^{\theta_0-1}$.

MBR may be of interest. The similarity of the stopping criteria and estimators, as shown in Table 1, is quite striking in view of the apparent disparity between the two methods. In fact, for large n, the procedures effectively coincide.

5. Binomial estimation computations. To indicate two possible forms that the stopping regions for the MBR can take, and to show the numerical computations required to obtain them, two binomial examples will be considered. The second will be given in considerable detail to include the computation of several α_k's. The form of the BSR for these examples is not known, and hand computation does not seem feasible. It is known only that they are truncated [1], at a value not less than 27 and 12 respectively (Corollary 3.1), and that the stopping boundaries are nowhere within those of the MBR (Theorem 3).

EXAMPLE 1. $L(\theta; d) = (d-\theta)^2$, $c_k(\mathbf{x}_k) = k/4000$ and $g_0(\theta) = \theta(1-\theta)$. Let $s = \sum_1^n x_i$, $r = (n+4)^2(n+5)^2$. From Table 1, $N_\infty = n$ if and only if $1/4000 \geqq (s+2)(n-s+2)/r$, i.e.,

$$s \geqq (\tfrac{1}{2})\{n \pm (n+4)[1 - ((n+5)^2/1000)]^{\frac{1}{2}}\} = (\sigma_n', \sigma_n'').$$

In the (n, s) plane, the stopping region is indicated by the shading in Figure 1.

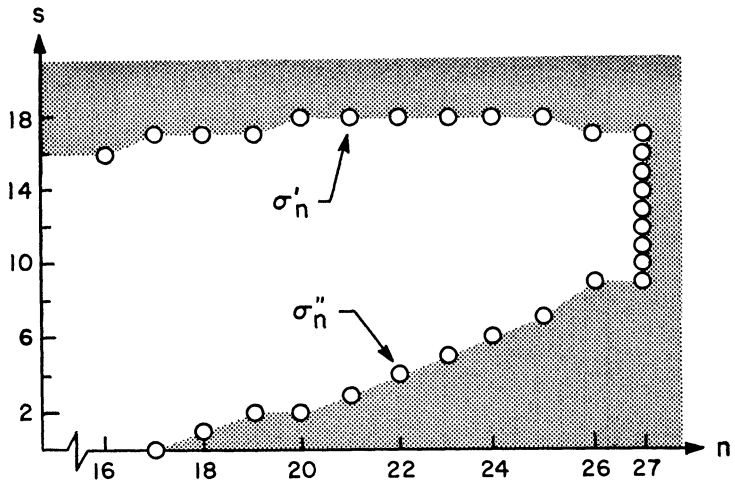

FIG. 1

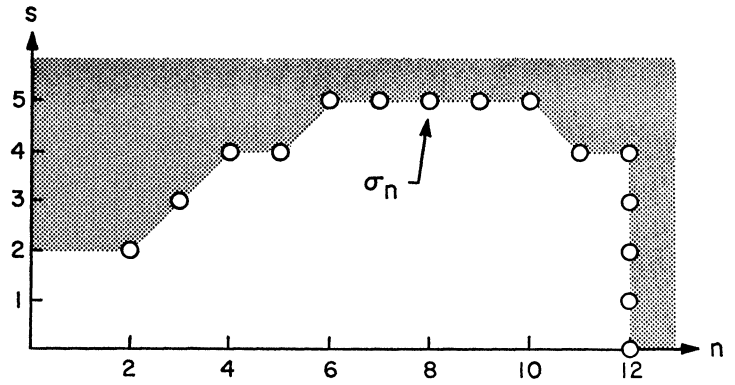

FIG. 2

EXAMPLE 2. $L(\theta; d) = (d - \theta)^2$, $c_k(\mathbf{x}_k) = k/2000$ and $g_0(\theta) = 9\theta^8$. Using Table 1, $N_\infty = n$ if and only if $s \geqq (\frac{1}{2})\{(n-8) + (n+10)[1 - ((n+11)^2/500)]^{\frac{1}{2}}\} = \sigma_n$. For this example, the stopping region is shown in Figure 2. In this example the prior belief that θ is large is sufficiently strong (relative to the cost per observation) that sampling cannot stop for "too few" successes.

Computation of $\alpha_\infty (= \alpha_{12})$. Let (n, λ_k) represent the coordinates in the (n, s) plane of the accessible stopping points where $k(=1, 2, \cdots, 12)$ indexes these points. Let t_k = number of distinct (accessible) paths in the (n, s) plane from $(0, 0)$ to (n, λ_k), (i.e., which do not pass through a (n, σ_n)). Let d_k = the Bayes terminal decision, if sampling stops at the point (n, λ_k). Since this is the mean of the posterior distribution, (see Table 1), $d_k = (9 + \lambda_k)/(n + 10)$.

By definition,

$$\alpha_\infty = 9 \sum_1^{12} \int_0^1 \{(d_k - \theta)^2 + cn(k)\} P_\theta(k) \theta^8 \, d\theta,$$

where $P_\theta(k) = P[\mathbf{x}_n : s = \lambda_k \mid \theta] = t_k \theta^{\lambda_k}(1 - \theta)^{n - \lambda_k}$. After performing the operations indicated, $\alpha_\infty = .00623 + .00138 = .00761$.

Computation of α_j. For any $\mathbf{x}_n$, $g_n(\theta) \propto \theta^{s+8}(1 - \theta)^{n-s}$. Also,

$$R_{n(k)} = \int_0^1 \sum_{i=0}^{k} \left\{ \inf_{d^*} \int_0^1 (d^* - \theta)^2 \, d\xi_{n+k}(\theta) \right\} \binom{k}{i} \theta^i (1 - \theta)^{k-i} \, d\xi_n(\theta) + ck,$$

and therefore,

$$R_{n(k)} = \int_0^1 \sum_{i=0}^{k} \frac{(s + 9 + i)(n - s + 1 + k - i)}{(n + k + 10)^2(n + k + 11)} \binom{k}{i} \theta^i (1 - \theta)^{k-i} \, d\xi_n(\theta) + ck$$

$$= (s + 9)(n - s + 1)/(n + 10)(n + 11)(n + 10 + k) + ck.$$

Assuming that k is a continuous variable, $dR_{n(k)}/dk = 0$ if and only if $k' = [\{2000(s + 9)(n - s + 1)/(n + 10)(n + 11)\}^{\frac{1}{2}} - (n + 10)]'$ and $[y]' \equiv$ integer closest to y.

1°. $\alpha_0 = R_{(3)} = .00629 + .00150 = .00779$.

2°. $\alpha_1 = .00648 + .00125 = .00773$.

Since for $x_1 = 0$, $k' = 6$, and hence $\alpha_0(0) = R_{1(6)}$. Also $\alpha_0(1) = R_{1(1)}$, since $k' = 1$ in this case.

3°. For $\mathbf{x}_2 = (0, 0)$, $k' = 7$; for $\mathbf{x}_2 = (0, 1)$ or $(1, 0)$, $k' = 4$ and for $\mathbf{x}_2 = (1, 1)$, $k' = 0$. Also $f_0(1, 1) = 9/11$, etc. Therefore $\alpha_2 = .00629 + .00138 = .00767$.

4°. For $\mathbf{x}_3 = (1, 0, 1)$ or $(0, 1, 1)$, $k' = 3$, for $\mathbf{x}_3 = (1, 0, 0)$, $(0, 1, 0)$ or $(0, 0, 1)$, $k' = 5$ and $k' = 7$ for $\mathbf{x}_3 = (0, 0, 0)$. Note that $\mathbf{x}_2 = (1, 1) \, \varepsilon \, A_2$. $f_0(1, 0, 1) = 3/44$, $f_0(0, 0, 1) = 3/220$, etc. Therefore,

$$\alpha_3 = \left(\tfrac{9}{11}\right)[\alpha_0(1, 1) + 2/2000] + \sum_{\mathbf{x}_3} [\alpha_0(\mathbf{x}_3) + 3/2000] f_0(\mathbf{x}_3)$$

$$= .00626 + .00141 = .00767.$$

Since the stopping point (2, 2) contributes .00563 to each α_j, the reduction possible as compared with α_0 is rather small in this example. It is noted that although the α_j's are monotonic, the contributions from expected cost and loss are not.

6. Testing hypotheses. The testing of two simple hypotheses is another problem to which the MBR may be applied. By Corollary 2.1, for $c_n(\mathbf{x}_n) = nc$ and $L(i, d) = 0$ if $d = i$ and s_i if $d \neq i$, there exists an equivalent SPRT. From Theorem 3, the particular SPRT must have boundaries at, or within, the SPRT corresponding to the Bayes solution. Since the BSR for this problem is available (by an iterative process [1]), an explicit comparison of the two procedures may be possible. By the use of Wald's well-known formulas [1], approximations of the o.c. function and ASN are obtainable for the MBR in this case.

The actual use of the MBR is not dependent on establishing the equivalent SPRT. If the sample number is expected to be small (for example, if the cost per

observation is high), it may be simpler to determine whether or not sampling should continue directly from the MBR definition.

To illustrate the computations needed to determine the equivalent SPRT, the following numerical example is presented. This same binomial example is found in [1], page 280. The considerable decrease in the computational complexity, when compared with the BSR, is evidence for the remarks made in Section 1.

The problem is to test $H_1 : \theta = \theta_1 = \frac{1}{3}$ against $H_2 : \theta = \theta_2 = \frac{2}{3}$, when $L(\theta_i, d)$ equals zero if $d = \theta_i$, and equals unity if $d \neq \theta_i$; $c_k(\mathbf{x}_k) = k/38.25$, and $\xi_0(\theta_i) = \frac{1}{2}$. Let: a_k = posterior probability that $\theta = \theta_1$, if $\mathbf{x}_k$ is observed, $s = s_k = \sum_1^k x_i$, $a' = 1 - b'$ = probability that $\theta = \theta_1$ and $h = \log(a'/b')$.

Then for any a', $a_k < \frac{1}{2}$ if and only if $s > s^* = h/2 \log 2 + k/2$. Then, if $a' = a_0$, using the notation of Section 2, $\gamma_{(k)} = (b'P_2 - a'P_1)/k$, where

$$P_1 = \sum_{s^*}^{k} \binom{k}{s} \left(\frac{1}{3}\right)^s \left(\frac{2}{3}\right)^{k-s}$$

etc. (it is assumed that $a_0 > \frac{1}{2}$). To establish the critical value of a', c is compared with $\gamma_{(k)}$ for various values of a_0, where $c = 1/38.25 = .0262$.

If $h = 1$, then $a' \doteq .73$, and $\gamma_{(2)} = .020$, $\gamma_{(3)} = .018$ and $\gamma_{(4)} = .021$ indicating that $h = 1$ is too large. If $h = .6$, than $a' \doteq .65$, $\gamma_{(2)} = .042$ showing $h = .6$ to be too small. It is quickly found that $a' = .72$ is approximately the critical value. To find the SPRT boundaries, the notation of [1] page 279 is used and $a_i^* = [\log(.72/.5)(.5/28)/\log(\frac{2}{3})(\frac{1}{3})]' = 2$ i.e., the SPRT has boundaries $(-2, 2)$. Confirming Theorem 3, these boundaries are not outside of the Bayes solutions; in fact, they coincide with the innermost Bayes solution.

For the composite hypothesis testing problem when $c_k(\mathbf{x}_k) = kc$ and $L(\theta, d)$ is a linear function of θ when d is wrong and 0 if d is right, a detailed discussion may be found in Chapter 5 of [5].

7. Discussion. Many questions regarding the MBR remain. In addition to those peculiar to this procedure, there are those applicable to any Bayesian-type rule. The problem remains of establishing a common "numeraire" so that losses and costs can be quantitatively compared. Also, methods are needed for determining the particular functions "appropriate" under any given circumstance. The establishment of "rules" is controversial and requires further investigation.

A brief discussion of several open problems is presented below:

1. *Investigation of* α_m. An indication of the gain achieved by using a sequential procedure would be given by a knowledge of $\alpha_m - \alpha_\infty$. If good bounds were available for $\alpha_\infty - \alpha_B$, a limit to the possible improvement with more sophisticated procedures (see 2.) would be known. At present only lower bounds for α_B [3] can be compared with α_0. In fact, except for the case of testing two simple hypotheses, little is known about the advantage of using a BSR.

2. *Other modifications.* In the same way as the MBR discussed here has been defined in terms of "looking ahead" at fixed samples of any size, a double sequence of such rules can be defined in terms of a truncated Bayes look into the future. That is, let δ_{ij} be such that as long as $N \leqq j$, if

$$R_N > \min \{E_N[\min \{R_{N+1}, E_{N+1}[\cdots R_{N+i}]\} \cdots], R_{N(i+1)}, \cdots\}$$

another sample is taken and otherwise sampling is stopped. Since the right-hand side of the above expression is always not greater than $\min_k R_{N(k)}$, a general theorem of sample size inequality (like Theorem 3) can be shown. Whether the additional computational complexity implied by such rules is worthwhile remains to be investigated.

3. *A single critical* $R_{n(k)}$. A general criterion for determining which $R_{n(k)}$ is the minimum does not seem feasible in the light of the complexity shown in Chapter 5 of [5]. Except for certain members of the exponential class, with their corresponding conjugate prior distributions, the minimizing k seems to be dependent upon $\mathbf{x}_n$. Bounds on R_n or $\tilde{\gamma}_n$ (see Section 2) would provide sufficient conditions for either stopping or continuing.

4. *Properties of the Bayes procedure.* The existence of a general sequential procedure whose average risk is not larger than the fixed sample-size Bayes risk may be useful in evaluating properties of the BSR. For example, since the α_∞ is less than α_0 for the estimation problem found in Section 5, the BSR cannot be a fixed sample-size procedure.

8. Acknowledgments. The author wishes to record his sincere appreciation for the guidance and encouragement given to him by his adviser, Professor W. J. Hall, throughout the course of this investigation. He is also grateful to the referee for his many constructive comments.

REFERENCES

[1] BLACKWELL, D. and GIRSHICK, M. A. (1954). *Theory of Games and Statistical Decisions.* Wiley, New York.

[2] GRUNDY, P. M., HEALY, M. J. and REES, D. H. (1956). Economic choice of the amount of experimentation. *J. Roy. Statist. Soc. Ser. B* **18** 32–55.

[3] HOEFFDING, W. (1957). Lower bound for the expected sample size and the average risk of a sequential procedure. *Ann. Math. Statist.* **28** 57–74.

[4] JOHNSON, N. L. (1961). Sequential analysis: a survey. *J. Roy. Statist. Soc. Ser. A* **124** 372–411.

[5] RAIFFA, H. and SCHLAIFFER, R. (1961). *Applied Statistical Decision Theory.* Graduate School of Business Administration, Harvard Univ.

[6] WALD, A. (1950). *Statistical Decision Functions.* Wiley, New York.

[7] WALD, A. (1951). Asymptotic minimax solutions of sequential point estimation problems. *Proc. Second Berkeley Symp.* 1–11, Univ. of California Press.

8] WETHERILL, G. B. (1961). Bayesian sequential analysis. *Biometrika* **48** 281–292.