Since the distribution of $\Sigma(X_i - \bar{X})^2/\sigma^2$ does not depend on ξ or σ, the probability $P\{\Sigma(X_i - \bar{X})^2 \geqq C|\xi, \sigma\}$ is independent of ξ and increases with σ, so that the conditions of Corollary 5 are satisfied. The test (32), being independent of ξ_1 and σ_1, is UMP for testing $\sigma \leqq \sigma_0$ against $\sigma > \sigma_0$. It is also seen to coincide with the likelihood ratio test (29). On the other hand, the most powerful test (31) for testing $\sigma \geqq \sigma_0$ against $\sigma < \sigma_0$ does depend on the value ξ_1 of ξ under the alternative.

It was tacitly assumed so far that $n > 1$. If $n = 1$, the argument applies without change with respect to H_1, leading to (31) with $n = 1$. However, in the discussion of H_2 the statistic U now drops out, and Y coincides with the single observation X. Using the same λ as before one sees that X has the same distribution under H_λ as under K, and the test ϕ_λ therefore becomes $\phi_\lambda(x) \equiv \alpha$. This satisfies the conditions of Corollary 5 and is therefore the most powerful test for the given problem. It follows that a single observation is of no value for testing the hypothesis H_2 as seems intuitively obvious, but that it could be used to test H_1 if the class of alternatives were sufficiently restricted.

The corresponding derivation for the hypothesis $\xi \leqq \xi_0$ is less straightforward. It turns out* that Student's test given by (30) is most powerful if the level of significance α is $\geqq 1/2$, regardless of the alternative $\xi_1 > \xi_0, \sigma_1$. This test is therefore UMP for $\alpha \geqq 1/2$. On the other hand, when $\alpha < 1/2$ the most powerful test of H rejects when $\Sigma(x_i - a)^2 \leqq b$, where the constants a and b depend on the alternative (ξ_1, σ_1) and on α. Thus for the significance levels that are of interest, a UMP test of H does not exist. No new problem arises for the hypothesis $\xi \geqq \xi_0$ since this reduces to the case just considered through the transformation $Y_i = \xi_0 - (X_i - \xi_0)$.

10. SEQUENTIAL PROBABILITY RATIO TESTS

According to the Neyman-Pearson fundamental lemma, the best procedure for testing the simple hypothesis H that the probability density of X is p_0 against the simple alternative that it is p_1 accepts or rejects H as

$$\frac{p_{1n}}{p_{0n}} = \frac{p_1(x_1) \cdots p_1(x_n)}{p_0(x_1) \cdots p_0(x_n)}$$

is less or greater than a suitable constant C. However, further improvement is possible if the sample size is not fixed in advance but is permitted to depend on the observations. The best procedure, in a certain sense, is then the following *sequential probability ratio test*. Let $A_0 < A_1$ be

* See Lehmann and Stein, "Most powerful tests of composite hypotheses. I. Normal distributions," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 495–516.

J. Rojo (ed.), *Selected Works of E. L. Lehmann*, Selected Works in Probability and Statistics, DOI 10.1007/978-1-4614-1412-4_18, © Springer Science+Business Media, LLC 2012

two given constants and suppose that observation is continued as long as the probability ratio p_{1n}/p_{0n} satisfies the inequality

$$A_0 < \frac{p_{1n}}{p_{0n}} < A_1. \tag{33}$$

The hypothesis H is accepted or rejected at the first violation of (33) as $p_{1n}/p_{0n} \leqq A_0$ or $\geqq A_1$.

The usual measures of the performance of such a procedure are the probabilities, say α_0 and α_1, of rejecting H when $p = p_0$ and of accepting it when $p = p_1$ and the expected number of observations $E_i(N)$ when $p = p_i$ $(i = 0, 1)$.

Theorem 8. *Among all tests (sequential or not) for which*

$$P_0 \text{ (rejecting } H) \leqq \alpha_0, \qquad P_1 \text{ (accepting } H) \leqq \alpha_1$$

and for which $E_0(N)$ and $E_1(N)$ are finite, the sequential probability ratio test with error probabilities α_0 and α_1 minimizes both $E_0(N)$ and $E_1(N)$.

In particular, the sequential probability ratio test therefore requires on the average fewer observations than the fixed sample size test which controls the errors at the same levels. The proof of this result will be deferred to Section 12. In this and the following sections some of the basic properties of sequential probability ratio tests will be sketched.

Because of the difficulty of determining exactly the boundaries A_0 and A_1 for which α_0 and α_1 take on preassigned values, the following inequalities are useful. Let R_n be the part of n-space defined by the inequalities

$$A_0 < \frac{p_{1k}}{p_{0k}} < A_1 \quad \text{for} \quad k = 1, \cdots, n-1 \qquad \text{and} \qquad A_1 \leqq \frac{p_{1n}}{p_{0n}}.$$

This is the set of points $(x_1, \cdots, x_n)$ for which the procedure stops with $N = n$ observations and rejects H. Then

$$\alpha_0 = \sum_{n=1}^{\infty} \int_{R_n} p_{0n} \leqq \frac{1}{A_1} \sum_{n=1}^{\infty} \int_{R_n} p_{1n} = \frac{1-\alpha_1}{A_1}.$$

Similarly, if S_n denotes the part of n-space in which $N = n$ and H is accepted, one has

$$1 - \alpha_0 = \sum_{n=1}^{\infty} \int_{S_n} p_{0n} \geqq \frac{\alpha_1}{A_0}.$$

Here it has been tacitly assumed that

$$\sum_{n=1}^{\infty} P_i\{N = n\} = \sum_{n=1}^{\infty} \int_{R_n \cup S_n} p_{in} = 1 \quad \text{for} \quad i = 0, 1,$$

that is, that the probability is 0 of the procedure continuing indefinitely. For a proof of this fact see Problems 34 and 35. The inequalities

$$A_0 \geqq \frac{\alpha_1}{1-\alpha_0}, \qquad A_1 \leqq \frac{1-\alpha_1}{\alpha_0} \tag{34}$$

suggest the possibility of approximating the boundaries A_0 and A_1 that would yield the desired α_0 and α_1 by

$$A_0' = \frac{\alpha_1}{1-\alpha_0}, \qquad A_1' = \frac{1-\alpha_1}{\alpha_0}.$$

By (34) the error probabilities of the approximate procedure then satisfy

$$\frac{\alpha_1'}{1-\alpha_0'} \leqq A_0' = \frac{\alpha_1}{1-\alpha_0} \quad \text{and} \quad \frac{1-\alpha_1'}{\alpha_0'} \geqq A_1' = \frac{1-\alpha_1}{\alpha_0}$$

and hence

$$\alpha_0' \leqq \frac{\alpha_0}{1-\alpha_1} \quad \text{and} \quad \alpha_1' \leqq \frac{\alpha_1}{1-\alpha_0}.$$

If typically α_0 and α_1 are of the order .01 to .1, the amount by which α_i' can exceed α_i $(i = 1, 0)$ is negligible so that the probabilities of the two kinds of error are very nearly bounded above by the specified α_0 and α_1. This conclusion is strengthened by the fact that $\alpha_0' + \alpha_1' \leqq \alpha_0 + \alpha_1$, as is seen by adding the inequalities $\alpha_1'(1-\alpha_0) \leqq \alpha_1(1-\alpha_0')$ and $\alpha_0'(1-\alpha_1') \leqq \alpha_0(1-\alpha_1')$.

The only serious risk in using the approximate boundaries A_0', A_1' is therefore that α_0' and α_1' are much smaller than required, which would lead to an excessive number of observations. There is some reason to hope that this effect is also moderate. For let

$$z_i = \log\,[p_1(x_i)/p_0(x_i)]. \tag{35}$$

Then (33) becomes

$$\log A_0 < \sum_{i=1}^{n} z_i < \log A_1,$$

and when H is rejected the z's satisfy

$$z_1 + \cdots + z_{n-1} < \log A_1 \leqq z_1 + \cdots + z_n.$$

The approximation consists in replacing $z_1 + \cdots + z_n$ by $\log A_1$. The error will usually be moderate since after $n - 1$ observations Σz_i is still $< A_1$ and the excess has therefore had no possibility to accumulate, but is due to a single observation. An analogous argument applies to the other boundary.

Example 9. Consider a sequence of binomial trials with constant probability p of success, and the problem of testing $p = p_0$ against $p = p_1 (p_0 < p_1)$. Then

$$\frac{p_{1n}}{p_{0n}} = \frac{p_1^{\Sigma x_i}(1 - p_1)^{n - \Sigma x_i}}{p_0^{\Sigma x_i}(1 - p_0)^{n - \Sigma x_i}} = \left(\frac{p_1 q_0}{p_0 q_1}\right)^{\Sigma x_i}\left(\frac{q_1}{q_0}\right)^n.$$

In the case that $\log (p_1 p_0^{-1})/\log (q_0 q_1^{-1})$ is rational, exact formulas have been obtained† for the error probabilities and expected sample size which make it possible to compute the effects involved in the approximation of A_0, A_1 by A_0', A_1'. As an illustration,‡ suppose that $p_0 = .05$, $p_1 = .17$, $\alpha_0 = .05$, $\alpha_1 = .10$. It then turns out that $\alpha_0' = .031$, $\alpha_1' = .099$, and that the expectations of the sample size for the approximate procedure are $E_0'(N) = 31.4$, $E_1'(N) = 30.0$. There is an alternate plan, determined by trial and error, with $\alpha_0^* = .046$, $\alpha_1^* = .097$, $E_0^*(N) = 30.5$, $E_1^*(N) = 26.1$. On the other hand, the fixed sample size procedure with error probabilities .05 and .10 requires 57 observations.

In order to be specific, we assumed in the definition of a sequential probability ratio test that observation continues only as long as the probability ratio is strictly between A_0 and A_1. The discussion applies equally well to the rule of continuing as long as $A_0 < p_{1n}/p_{0n} < A_1$, coming to the indicated conclusion the first time that $p_{1n}/p_{0n} < A_0$ or $> A_1$, and deciding on the boundaries according to any fixed probabilities. The term *sequential probability ratio test* is applied also to this more general procedure. If the probability ratio $p_1(X)/p_0(X)$ has a continuous distribution, all these procedures are equivalent. However, in case of discrete probability ratios the possibility of randomization on the boundary is necessary to achieve preassigned error probabilities. If randomization is permitted also between taking at least one observation or reaching a decision without taking any observations, it can be shown that actually any preassigned error probabilities can be achieved.§

11. POWER AND EXPECTED SAMPLE SIZE OF SEQUENTIAL PROBABILITY RATIO TESTS

The preceding section is somewhat misleading in that it discusses the problem in a setting, that of testing a simple hypothesis against a simple alternative, which is interesting mainly because of its implications for the more realistic situation of a continuous parameter family of distributions.

† Girshick, "Contributions to the theory of sequential analysis, II, III," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 282–298, and Polya, "Exact formulas in the sequential analysis of attributes," *Univ. Calif. Publs. Mathematics*, New Series, Vol. 1 (1948), pp. 229–240.

‡ Taken from Robinson, "A note on exact sequential analysis," *Univ. Calif. Publs. Mathematics*, New Series, Vol. 1 (1948), pp. 241–246.

§ This result is contained in an as yet unpublished paper by Stein, "Existence of sequential probability ratio tests." See also the abstract by Wijsman, "On the existence of Wald's sequential test," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 938–939.

Unfortunately, the property of being uniformly most powerful, which the fixed sample size probability ratio test possesses for families with monotone likelihood ratio (Theorem 2), does not extend to the sequential case. More specifically, consider the sequential probability ratio test for testing $H: \theta_0$ against $K: \theta_1$, and let its power function be $\beta(\theta) = P_\theta$ (rejecting H). Then if θ_2 is some other alternative, the sequential probability ratio test for testing θ_0 against θ_2 with error probabilities α_0 and α_1 does not in general coincide with the original test, which therefore does not minimize $E_{\theta_2}(N)$. It seems in fact likely that from an over-all point of view the sequential probability ratio test is not the best sequential procedure in the continuous parameter case, although it is usually better than the best competitive test with fixed sample size.

When the probability density depends on a real parameter θ and one is testing the hypothesis $\theta \leqq \theta_0$, one is usually not concerned with the power of the test against alternatives θ close to θ_0, but would like to be able to control the probability of detecting alternatives sufficiently far away. The test should therefore satisfy

$$\begin{aligned} \beta(\theta) &\leqq \alpha \quad \text{for} \quad \theta \leqq \theta_0 \\ \beta(\theta) &\geqq \beta \quad \text{for} \quad \theta \geqq \theta_1 \end{aligned} \qquad (\theta_0 < \theta_1), \tag{36}$$

which it will do in particular if

$$\beta(\theta_0) = \alpha, \qquad \beta(\theta_1) = \beta,$$

and if $\beta(\theta)$ is a nondecreasing function of θ. The sequential probability ratio test for testing θ_0 against θ_1 with error probabilities $\alpha_0 = \alpha, \alpha_1 = 1 - \beta$ thus is a solution of the stated problem provided its power function is nondecreasing.

Lemma 4. *Let $X_1, X_2, \cdots$ be independently distributed with probability density $p_\theta(x)$, and suppose that the densities $p_\theta(x)$ have monotone likelihood ratio in $T(x)$. Then any sequential probability ratio test for testing θ_0 against θ_1 ($\theta_0 < \theta_1$) has a nondecreasing power function.*

Proof. Let $Z_i = \log [p_{\theta_1}(X_i)/p_{\theta_0}(X_i)] = h(T_i)$, where h is nondecreasing, and let $\theta < \theta'$. By Lemma 2, the cumulative distribution function $F_\theta(t)$ of T_i satisfies $F_{\theta'}(t) \leqq F_\theta(t)$ for all t, and by Lemma 1 there exists therefore a random variable V_i and functions f and f' such that $f(v) \leqq f'(v)$ for all v and that the distributions of $f(V_i)$ and $f'(V_i)$ are F_θ and $F_{\theta'}$ respectively. The sequential test under consideration has the following graphical representation in the $(n, \sum_{i=1}^n h(t_i))$ plane. Observation is continued as long as the sample points fall inside the band formed by the parallel straight lines

$$\sum_{i=1}^{n} h(t_i) = \log A_j, \qquad j = 0, 1.$$

The hypothesis is rejected if the path formed by the points $(1, h(t_1))$, $(2, h(t_1) + h(t_2)), \cdots, (N, h(t_1) + \cdots + h(t_N))$ leaves the band through the upper boundary. The probability of this event is therefore the probability of rejection, for θ when each T_i is replaced by $f(V_i)$ and for θ' when T_i is replaced by $f'(V_i)$. Since $f(V_i) \leq f'(V_i)$ for all i, the path generated by the $f'(V_i)$ leaves the band through the upper boundary whenever this is true for the path generated by the $f(V_i)$. Hence $\beta(\theta) \leq \beta(\theta')$, as was to be proved.

In the case of monotone likelihood ratios, the sequential probability ratio test with error probabilities $\alpha_0 = \alpha$, $\alpha_1 = 1 - \beta$ therefore satisfies (36). It follows from the optimum property stated in Section 10 that among all tests satisfying (36) the sequential probability ratio test minimizes the expected sample size for $\theta = \theta_0$ and $\theta = \theta_1$. However, one is now concerned with $E_\theta(N)$ for all values of θ. Typically, the function $E_\theta(N)$ has a maximum at a point between θ_0 and θ_1, and decreases as θ moves away from this point in either direction. It frequently turns out that the maximum is $< n_0$, the smallest fixed sample size for which there exists a test satisfying (36). On the other hand, this is not always the case. Thus, in Example 9 for $p_0 = .4$, $p_1 = .6$, $\alpha_0 = \alpha_1 = .005$ for example, the fixed sample size n_0 is 160, and $E_p(N)$, while below this for most values of p, equals 170 for $p = 1/2$. The important problem of determining the test that minimizes $\sup E_\theta(N)$ subject to (36) is still unsolved.

An exact evaluation of the power function $\beta(\theta)$ and the expected sample size $E_\theta(N)$ of a sequential probability ratio test is in general extremely difficult. However, a simple approximation is available provided the equation

$$E_\theta \{[p_{\theta_1}(X)/p_{\theta_0}(X)]^h\} = 1 \tag{37}$$

has a nonzero solution $h = h(\theta)$, as is the case under mild assumptions. (See Problem 38.) Then

$$p_\theta^*(x) = \left[\frac{p_{\theta_1}(x)}{p_{\theta_0}(x)}\right]^h p_\theta(x)$$

is again a probability density. Suppose now that $h > 0$—the other case can be treated similarly—and consider the sequential probability ratio test with boundaries A_0^h, A_1^h for testing p_θ against p_θ^*. With this procedure observation is continued as long as

$$A_0^h < \frac{p_\theta^*(x_1)}{p_\theta(x_1)} \cdots \frac{p_\theta^*(x_n)}{p_\theta(x_n)} < A_1^h .$$

If α_0^* and $1 - \alpha_1^*$ denote the probability of rejection when p_θ and p_θ^* are the true densities, it is seen from (34) that the boundaries are given approximately by

$$A_0^h \sim \frac{\alpha_1^*}{1 - \alpha_0^*}, \qquad A_1^h \sim \frac{1 - \alpha_1^*}{\alpha_0^*} .$$

However, the test under consideration is exactly the same as the sequential probability ratio test with error probabilities $\alpha_0 = \alpha$, $\alpha_1 = 1 - \beta$ for testing θ_0 against θ_1. Hence α_0^* and $\beta(\theta)$, the probability of rejection for the two tests when p_θ is the true density, must be equal. Solving for α_0^* from the above two approximate equations one therefore finds

$$\beta(\theta) \sim \frac{1 - A_0^h}{A_1^h - A_0^h} . \tag{38}$$

An approximation for $E_\theta(N)$ can be based on *Wald's equation*

$$E_\theta(Z_1 + \cdots + Z_N) = E_\theta(N) E_\theta(Z), \tag{39}$$

which is valid whenever the Z's are identically and independently distributed and the procedure is such that the expected sample size $E_\theta(N)$ is finite. For a proof of this equation see Problem 37. If the Z's are defined by (35) and the procedure is a sequential probability ratio test, $Z_1 + \cdots + Z_N$ can be approximated as before by $\log A_1$ and $\log A_0$ when H is rejected and accepted respectively, so that from (39) one obtains

$$E_\theta(N) \sim \frac{\beta(\theta) \log A_1 + [1 - \beta(\theta)] \log A_0}{E_\theta(Z)} \tag{40}$$

provided $E_\theta(Z) \neq 0$.

Example 10. In the binomial problem of Example 9, equation (37) becomes

$$p\left(\frac{p_1}{p_0}\right)^h + q\left(\frac{q_1}{q_0}\right)^h = 1. \tag{41}$$

Since the left-hand side is a convex function of h which is 1 for $h = 0$, it is seen that the equation has a unique nonzero solution except when $p = \log(q_0/q_1)/\log(p_1q_0/p_0q_1)$, in which case the left-hand side has its minimum at $h = 0$. Equations (38) and (41) provide a parametric representation of the approximate power function, which can now be computed by giving different values to h and obtaining the associated values p and β from (38) and (41). (For $h = 0$,

β can be obtained by continuity.) The following is a comparison of the approximate with the exact values of $\beta(p)$ and $E_p(N)$ in the numerical case considered in Example 9, with $p_0 = .05$, $p = .099$, $p_1 = .17$:*

$\beta(p_0)$	$\beta(p)$	$\beta(p_1)$	$E_{p_0}(N)$	$E_p(N)$	$E_{p_1}(N)$	
.05	.44	.90	30	39	25	Approx.
.031	.409	.901	31.4	46.8	30.0	Exact

12. OPTIMUM PROPERTY OF SEQUENTIAL PROBABILITY RATIO TESTS†

The main part of the proof of Theorem 8 is contained in the solution of the following auxiliary problem. For testing the hypothesis H that p_0 is the true probability density of X against the alternative that it is p_1, let the losses resulting from false rejection and acceptance of H be w_0 and w_1, and let the cost of each observation be c. The risk (expected loss plus expected cost) of a sequential procedure is then

$$\alpha_i w_i + cE_i(N)$$

when p_i is the true density, where

$$\alpha_0 = P_0\,(\text{rejecting } H), \qquad \alpha_1 = P_1\,(\text{accepting } H)$$

are the two probabilities of error. If one supposes that the subscript i of the probability density is itself a random variable, which takes on the values 0 and 1 with probability π and $1 - \pi$ respectively, the total average risk of a procedure δ is

$$r(\pi, \delta) = \pi[\alpha_0 w_0 + cE_0(N)] + (1 - \pi)[\alpha_1 w_1 + cE_1(N)]. \tag{42}$$

We shall now determine the Bayes procedure for this problem, that is, the procedure that minimizes (42). Here the interpretation of (42) as a Bayes risk is helpful for an understanding of the proof and gives the auxiliary problem independent interest. However, from the point of view of Theorem 8, the introduction of the w's, c, and π is only a mathematical device, and the problem is simply that of minimizing the formal expression (42).

The Bayes solutions involve two numbers $\pi' \leqq \pi''$ which are uniquely determined by w_0, w_1, and c through equations (44) and (45) below, and which are independent of π. It will be sufficient to restrict attention to the case that $0 < \pi' < \pi'' < 1$ and to a priori probabilities π satisfying $\pi' \leqq \pi \leqq \pi''$.

* Taken from Robinson, *loc. cit.*, where a number of further examples are given.

† This section treats a special topic to which no reference is made in the remainder of the book.

Lemma 5. *Let π', π'' satisfy the equations* (44). *If $0 < \pi' < \pi'' < 1$, then for all $\pi' \leqq \pi \leqq \pi''$ the Bayes risk* (42) *is minimized by any sequential probability ratio test with boundaries*

$$(43) \qquad A_0 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi''}{\pi''}, \qquad A_1 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi'}{\pi'}.$$

Proof. (1) We begin by investigating whether at least one observation should be taken, in which case the resulting risk will be at least c, or whether it is better to come to a decision immediately. Let δ_0 denote the procedure that rejects H without taking any observations, and δ_1 the corresponding procedure that accepts H, so that

$$r(\pi, \delta_0) = \pi w_0 \qquad \text{and} \qquad r(\pi, \delta_1) = (1-\pi)w_1.$$

Let

$$\rho(\pi) = \inf_{\delta \in \mathscr{C}} r(\pi, \delta)$$

where $\mathscr{C}$ is the class of all procedures requiring at least one observation. Then for any $0 < \lambda < 1$ and any π_0, π_1,

$$\rho[\lambda\pi_0 + (1-\lambda)\pi_1] = \inf_{\delta\in\mathscr{C}} [\lambda r(\pi_0, \delta) + (1-\lambda) r(\pi_1, \delta)]$$
$$\geqq \lambda\rho(\pi_0) + (1-\lambda)\rho(\pi_1).$$

Hence ρ is concave, and since it is bounded below by zero it is continuous in the interval (0, 1).* If

$$\rho\left(\frac{w_1}{w_0 + w_1}\right) < \frac{w_0 w_1}{w_0 + w_1},$$

define π' and π'' by

$$(44) \qquad r(\pi', \delta_0) = \rho(\pi') \qquad \text{and} \qquad r(\pi'', \delta_1) = \rho(\pi'').$$

(See Figure 3.) Otherwise let

$$(45) \qquad \pi' = \pi'' = \frac{w_1}{w_0 + w_1}.$$

In the case $0 < \pi' < \pi'' < 1$ with which we are concerned, δ_0 minimizes (42) if and only if $\pi \leqq \pi'$, and δ_1 minimizes (42) if and only if $\pi \geqq \pi''$. This establishes the following uniquely as an optimum first step for $\pi \neq \pi', \pi''$: If $\pi < \pi'$ or $> \pi''$, no observation is taken and H is rejected or accepted respectively; if $\pi' < \pi < \pi''$ the variable X_1 is observed.

(2) The proof is now completed by induction. Suppose that $\pi' < \pi < \pi''$ and that n observations have been taken with outcomes $X_1 = x_1, \cdots, X_n = x_n$, and that one is faced with the alternatives of not taking another

* See, for example, section 3.18 of Hardy, Littlewood, Pólya, *Inequalities*, Cambridge Univ. Press, 1934.

observation and rejecting or accepting H with losses w_0, w_1 for possible wrong decisions, or of going on to observe X_{n+1}. The situation is very similar to the one analyzed in part (1). An unlimited supply of observations $X_{n+1}, X_{n+2}, \cdots$ is available. The fact that one has already incurred the expense of nc units does not affect the problem, since once this loss has been sustained no future action can retrieve it. The procedure is therefore as before: No further observation is taken if the probability of H

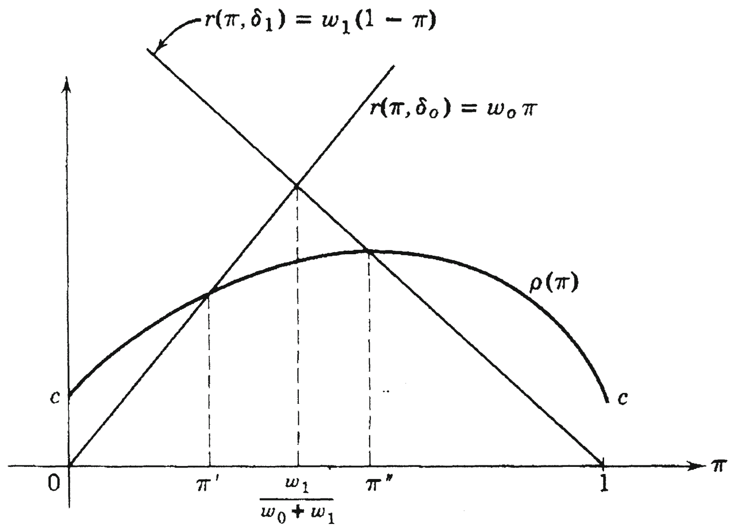

Figure 3.

being true is $< \pi'$ or $> \pi''$, whereas X_{n+1} is observed if this probability is strictly between π' and π''.

One aspect of the situation has changed as a result of observing $x_1, \cdots, x_n$. The probability of H being true is no longer π but has become

$$\pi(x_1, \cdots, x_n) = \frac{\pi p_{0n}}{\pi p_{0n} + (1 - \pi) p_{1n}},$$

the conditional (a posteriori) probability of H given $X_1 = x_1, \cdots, X_n = x_n$. A complete procedure therefore consists in continuing as long as

$$\pi' < \pi(x_1, \cdots, x_n) < \pi''$$

or equivalently as long as

$$A_0 = \frac{\pi}{1 - \pi} \cdot \frac{1 - \pi''}{\pi''} < \frac{p_{1n}}{p_{0n}} < \frac{\pi}{1 - \pi} \cdot \frac{1 - \pi'}{\pi'} = A_1.$$

H is accepted if, at the first violation of these inequalities, p_{1n}/p_{0n} is $< A_0$ and rejected if it is $> A_1$.

(3) In part (1) of this proof the first step of the procedure was uniquely determined as δ_0 for $\pi < \pi'$, as δ_1 for $\pi > \pi''$, and as taking at least one observation when $\pi' < \pi < \pi''$. For $\pi = \pi'$, the procedure δ_0 still minimizes (42) but it is no longer unique, that is, there also exists a procedure $\delta \in \mathscr{C}$ for which $r(\pi', \delta) = \rho(\pi')$. In order to belong to $\mathscr{C}$, such a procedure must require at least one observation. Once X_1 has been observed, it follows from part (2) that the best procedure in $\mathscr{C}$ is obtained by continuing observation as long as $\pi' < \pi(x_1, \cdots x_n) < \pi''$. At the first step it is therefore immaterial whether on the boundary experimentation is continued or the indicated decision is taken. The same is then true at the subsequent steps. This establishes in particular that for $\pi' \leqq \pi \leqq \pi''$ the procedure of taking a first observation and then following the sequential probability ratio test with boundaries (43) is Bayes.

The required connection between the auxiliary problem and the original one is established by the following lemma.

Lemma 6. *Given any $0 < \pi_0' < \pi_0'' < 1$, there exist numbers $0 < w < 1$, $0 < c$ such that the Bayes solution of the auxiliary problem defined by $w_0 = 1 - w$, $w_1 = w$, c, and an a priori probability π satisfying $\pi_0' < \pi < \pi_0''$ is a sequential probability ratio test with boundaries*

$$A_0 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi_0''}{\pi_0''}, \qquad A_1 = \frac{\pi}{1-\pi} \cdot \frac{1-\pi_0'}{\pi_0'} .$$

*Proof.** (1) By Lemma 5, the quantities π' and π'' are functions of w and c, and it is therefore sufficient to find w and c such that $\pi'(w, c) = \pi_0'$, $\pi''(w, c) = \pi_0''$. For fixed w, let $\pi'(c) = \pi'(w, c)$ and $\pi''(c) = \pi''(w, c)$. If c_0 is the smallest value of c such that $\pi'(c_0) = \pi''(c_0)$, then for $0 < c < c_0$ the quantities $\pi'(c)$ and $\pi''(c)$ are determined by the equations

$$(1 - w)\pi' = \rho(\pi', c), \qquad (1 - \pi'')w = \rho(\pi'', c),$$

where $\rho(\pi, c)$ stands for the quantity previously denoted by $\rho(\pi)$. The function $\rho(\pi', c)$ considered as a function of c for fixed π' has the following properties. (i) It is continuous. This follows as before from its being concave. (ii) It is strictly increasing, since for any $\delta \in \mathscr{C}$ the risk $r(\delta, \pi')$ increases strictly with c and since the minimum risk $\rho(\pi', c)$ is taken on by a procedure $\delta \in \mathscr{C}$. (iii) As c tends to zero, so do $\rho(\pi', c)$ and $\rho(\pi'', c)$. This follows from the fact that for n sufficiently large there exists a test of fixed sample size n for which the two error probabilities are arbitrarily small.

* This proof was communicated to me by L. LeCam.

These properties of the function ρ imply that for $0 < c < c_0$ the functions π' and π'' are also continuous, strictly increasing and decreasing respectively, and that $\pi'(c) \to 0$, $\pi''(c) \to 1$ as $c \to 0$. On the other hand, as $c \to c_0$, $\pi''(c) - \pi'(c) \to 0$ so that both quantities tend to the solution $\pi' = \pi'' = w$ of the equation $\pi'(1 - w) = (1 - \pi')w$. It follows from these properties that for fixed w

$$\lambda(c) = \frac{\pi'(c)}{1 - \pi'(c)} \cdot \frac{1 - \pi''(c)}{\pi''(c)}$$

is a continuous, strictly increasing function of c, which increases from 0 to 1 as c varies from 0 to $c_0 = c_0(w)$.

(2) Let

$$\lambda(w, c) = \frac{\pi'(w, c)}{1 - \pi'(w, c)} \cdot \frac{1 - \pi''(w, c)}{\pi''(w, c)}, \qquad \gamma(w, c) = \frac{\pi''(w, c)}{1 - \pi''(w, c)} \cdot$$

Instead of working with the variables π' and π'', it is equivalent and more convenient to work with λ and γ, and to prove the existence of w, c such that

$$\lambda(w, c) = \frac{\pi_0'}{1 - \pi_0'} \cdot \frac{1 - \pi_0''}{\pi_0''} = \lambda_0, \qquad \gamma(w, c) = \frac{\pi_0''}{1 - \pi_0''} = \gamma_0.$$

For any w, there exists by part (1) a unique cost $c = c(w)$ such that $\lambda(w, c) = \lambda_0$. It will be shown below that $\gamma(w) = \gamma[w, c(w)]$ is a 1:1 mapping of the interval $0 < w < 1$ onto $0 < \gamma < \infty$, and hence that there exists a unique value w such that $\gamma(w) = \gamma_0$. This will complete the proof of the lemma.

(3) For the auxiliary problem defined by w, $c = c(w)$, and $\pi = \pi'[w, c(w)]$ there exists by Lemma 5 a Bayes solution δ' which is a sequential probability ratio test with boundaries

$$A_0' = \frac{\pi'[w, c(w)]}{1 - \pi'[w, c(w)]} \cdot \frac{1 - \pi''[w, c(w)]}{\pi''[w, c(w)]} = \lambda[w, c(w)] = \lambda_0, \qquad A_1' = 1.$$

Let δ'' be the corresponding solution of the problem defined by w, $c = c(w)$, and $\pi = \pi''[w, c(w)]$, so that its boundaries are

$$A_0'' = 1, \qquad A_1'' = \frac{\pi''[w, c(w)]}{1 - \pi''[w, c(w)]} \cdot \frac{1 - \pi'[w, c(w)]}{\pi'[w, c(w)]} = \frac{1}{\lambda_0} \cdot$$

Then the error probabilities and the expectations of the sample size α_0', α_1', $E_0'(N)$, $E_1'(N)$ of δ' and α_0'', α_1'', $E_0''(N)$, $E_1''(N)$ of δ'' depend on w and c only through λ_0 and not through γ, so that for fixed λ_0 they are fixed numbers. The Bayes risks for $\pi = \pi'[w, c(w)]$ and $\pi = \pi''[w, c(w)]$ are given by

$$\rho(\pi') = r(\pi', \delta') \qquad \text{and} \qquad \rho(\pi'') = r(\pi'', \delta'')$$

and it follows from (44) that

$$r(\pi', \delta_0) = r(\pi', \delta') \qquad \text{and} \qquad r(\pi'', \delta_1) = r(\pi'', \delta'').$$

These equations can be written more explicitly as

$$\pi'(1 - w) = \pi'[\alpha_0'(1 - w) + cE_0'(N)] + (1 - \pi')[\alpha_1'w + cE_1'(N)]$$

and

$$(1 - \pi'')w = \pi''[\alpha_0''(1 - w) + cE_0''(N)] + (1 - \pi'')[\alpha_1''w + cE_1''(N)].$$

If one substitutes $\lambda_0\gamma$ for $\pi'/(1 - \pi')$ and γ for $\pi''/(1 - \pi'')$ and eliminates c, this reduces to a single equation connecting γ and w:

$$\begin{aligned}\{\lambda_0\gamma(1 - \alpha_0') - w[\lambda_0\gamma(1 - \alpha_0') + \alpha_1']\}\{\gamma E_0''(N) + E_1''(N)\} \\ = \{-\gamma\alpha_0'' + w[(1 - \alpha_1'') + \gamma\alpha_0'']\}\{\lambda_0\gamma E_0'(N) + E_1'(N)\}.\end{aligned}$$

This is linear in w and for any $\gamma > 0$ has a solution $0 < w < 1$. As a function of γ it is quadratic, and the coefficients of the constant and quadratic terms have opposite signs provided $0 < w < 1$. In this case there exists therefore a unique positive solution γ, which establishes the required 1:1 relation between γ and w.

To complete the proof of Theorem 8, consider now any sequential probability ratio test with $A_0 < 1 < A_1$, and any constant $0 < \pi < 1$. Let

$$\pi' = \frac{\pi}{A_1(1 - \pi) + \pi}, \qquad \pi'' = \frac{\pi}{A_0(1 - \pi) + \pi}.$$

These values satisfy (43) and $0 < \pi' < \pi < \pi'' < 1$, and by Lemma 6 there exist therefore constants $0 < w < 1$ and $c > 0$ such that the given test is a Bayes solution for the auxiliary problem with an a priori probability π of p_0 being the true density, with losses $w_0 = 1 - w$ and $w_1 = w$, and cost c. Let the error probabilities and expectations of the sample size be α_0, α_1, $E_0(N)$, $E_1(N)$ for the given test, and consider any competitive procedure δ^*, with error probabilities $\alpha_i^* \leqq \alpha_i$ and expectations of sample size $E_i^*(N) < \infty$ $(i = 0, 1)$. Since the given test minimizes the Bayes risk, it satisfies

$$\begin{aligned}\pi[(1 - w)\alpha_0 + cE_0(N)] + (1 - \pi)[w\alpha_1 + cE_1(N)] \\ \leqq \pi[(1 - w)\alpha_0^* + cE_0^*(N)] + (1 - \pi)[w\alpha_1^* + cE_1^*(N)]\end{aligned}$$

and hence

$$\pi E_0(N) + (1 - \pi)E_1(N) \leqq \pi E_0^*(N) + (1 - \pi)E_1^*(N).$$

The validity of this inequality for all $0 < \pi < 1$ implies

$$E_0(N) \leqq E_0^*(N) \qquad \text{and} \qquad E_1(N) \leqq E_1^*(N),$$

as was to be proved.