SEQUENTIAL TESTS FOR COMPOSITE HYPOTHESES

By D. R. COX

Communicated by F. J. ANSCOMBE

Received 14 August 1951

ABSTRACT. A method is given for obtaining sequential tests in the presence of nuisance parameters. It is assumed that a jointly sufficient set of estimators exists for the unknown parameters. A number of special tests are described and their properties discussed.

1. *Introduction.* Wald (9) in his book *Sequential Analysis* gave a comprehensive theory of the likelihood ratio sequential test for deciding between two simple hypotheses. This theory can be used to construct tests for most problems of choosing between a small number of decisions, there being one unknown parameter.

We consider here the construction of sequential tests when there is more than one unknown parameter, e.g. tests for a variance ratio, for a correlation coefficient, for a normal mean (variance unknown), etc. Previous work on this subject includes a general method of constructing such tests—the method of weight functions—due to Wald, leading to a large number of tests in any situation. One of these many tests is usually a natural one to use. Girshick (4) has given an elegant method for problems with two populations each of the form $f(x, \theta)$, it being required to decide which population has the larger θ. A number of special tests have been proposed, and in particular Rushton (8) has given a test for Student's hypothesis based on unpublished theory by C. Stein and G. A. Barnard. This test, which is closely related to a test due to Wald, is obtained by calculating at any stage of the experiment the likelihood ratio of Student's t for the alternative hypothesis against the null hypothesis. Nandi (7) has put forward without proof a similar method for more general situations.

The present paper shows how a method analogous to Rushton's can be used in many problems in which a jointly sufficient set of estimators can be found for the unknown parameters. The tests given are based on a theorem about a jointly sufficient set of estimators (§2). In §3 the principle of the method is explained by a special case, and in §§4, 5 the general method is stated and exemplified. In §§6, 7 properties of the tests are discussed. The statement of the theorem in §2 is rather complicated, and the reader mainly interested in applications should omit this section.

2. *A general theorem.* We first prove a fixed-sample-size theorem on which the method of constructing the sequential tests depends. The theorem asserts the possibility of factorizing a likelihood in certain cases when a jointly sufficient set of estimators exists for the unknown parameters. All applications in the present paper are to random variables with a probability density function, and so the theorem is stated for them only.

The term *functional independence* is used in a rather special sense in the theorem. When it is stated that certain functions of $x_1, \ldots, x_n$ denoted by $t_1, \ldots, t_p, u_1, \ldots, u_m$ are functionally independent, it is meant that there is a transformation from $x_1, \ldots, x_n$ to

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

a set of new variables including $t_1, \ldots, t_p, u_1, \ldots, u_m$, and that the Jacobian of the transformation is different from zero (except possibly for a set of values of total probability zero).

THEOREM 1. *Let* $\{x_1, \ldots, x_n\} = \mathbf{x}$ *be random variables whose probability density function (p.d.f.) depends on unknown parameters* $\theta_1, \ldots, \theta_p$. *The* x_i *may themselves be vectors. Suppose that*

(i) $t_1, \ldots, t_p$ *are a functionally independent jointly sufficient set of estimators for* $\theta_1, \ldots, \theta_p$;

(ii) *the distribution of* t_1 *involves* θ_1 *but not* $\theta_2, \ldots, \theta_p$;

(iii) $u_1, \ldots, u_m$ *are functions of* $\mathbf{x}$ *functionally independent of each other and of* $t_1, \ldots, t_p$;

(iv) *there exists a set* S *of transformations of* $\mathbf{x} = \{x_1, \ldots, x_n\}$ *into* $\mathbf{x}' = \{x_1', \ldots, x_n'\}$ *such that*

(*a*) $t_1, u_1, \ldots, u_m$ *are unchanged by all transformations in* S;

(*b*) *the transformation of* $t_2, \ldots, t_p$ *into* $t_2', \ldots, t_p'$ *defined by each transformation in* S *is* (1, 1);

(*c*) *if* $T_2, \ldots, T_p$ *and* $T_2', \ldots, T_p'$ *are two sets of values of* $t_2, \ldots, t_p$ *each having non-zero probability density under at least one of the distributions of* $\mathbf{x}$, *then there exists a transformation in* S *such that if* $t_2 = T_2, \ldots, t_p = T_p$ *then* $t_2' = T_2', \ldots, t_p' = T_p'$.

Then the joint p.d.f. of $t_1, u_1, \ldots, u_m$ *factorizes into*

$$g(t_1 \mid \theta_1)\, l(u_1, \ldots, u_m, t_1),$$

where g *is the p.d.f. of* t_1 *and* l *does not involve* θ_1.

Proof. The p.d.f. of $\mathbf{x}$ can be written $L(t_1, \ldots, t_p; \theta_1, \ldots, \theta_p)\, M(x_1, \ldots, x_n)$. We can find a transformation of non-vanishing Jacobian J from $\mathbf{x}$ to new variables

$$\{t_1, u_1, \ldots, u_m, t_2, \ldots, t_p, v_1, v_2, \ldots\},$$

where $v_1, v_2, \ldots$ are any suitable functions of $\mathbf{x}$ to complete the transformation. The p.d.f. of the new variables is

$$L(t_1, \ldots, t_p; \theta_1, \ldots, \theta_p)\, M^*(t_1, u_1, \ldots, u_m, t_2, \ldots, t_p, v_1, v_2, \ldots)\, J,$$

where M^* is the function obtained from M by the transformation. An expression of this form holds also if the transformation from $\mathbf{x}$ to new variables is many-one.

By integrating with respect to $v_1, v_2, \ldots$ we get the p.d.f. of the remaining variables in the form

$$L(t_1, \ldots, t_p; \theta_1, \ldots, \theta_p)\, N(t_1, u_1, \ldots, u_m, t_2, \ldots, t_p). \tag{1}$$

This first step of the proof follows Kendall (6).

Now we can obtain the p.d.f. of $t_1, \ldots, t_p$ by integrating out with respect to $u_1, \ldots, u_m$, and we can always arrange that L is exactly this p.d.f. Repeating the argument we can write (1) in the form

$$g(t_1 \mid \theta_1)\, h(t_2, \ldots, t_p \mid \theta_1, \ldots, \theta_p, t)\, l(u_1, \ldots, u_m \mid t_1, \ldots, t_p), \tag{2}$$

where g is the p.d.f. of t_1 and involves only θ_1 by (ii), h is the p.d.f. of $t_2, \ldots, t_p$ given t_1, and l is the p.d.f. of $u_1, \ldots, u_m$ given $t_1, \ldots, t_p$ and does not involve $\theta_1, \ldots, \theta_p$.

Now consider the function l. If we apply a transformation of the set S, $u_1, \ldots, u_m, t_1$ are unaltered and $t_2, \ldots, t_p$ are converted by a (1, 1) transformation into a unique set of values $t_2', \ldots, t_p'$. Therefore

$$l(u_1, \ldots, u_m \mid t_1, t_2, \ldots, t_p) = l(u_1, \ldots, u_m \mid t_1, t_2', \ldots, t_p').$$

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

By condition (iv) (*c*) this holds for all $t_2, \ldots, t_p, t_2', \ldots, t_p'$. Therefore l cannot involve $t_2, \ldots, t_p$ and we may write it $l(u_1, \ldots, u_m, t_1)$. Thus (2) becomes

$$g(t_1 \mid \theta_1)\, h(t_2, \ldots, t_p \mid \theta_1, \ldots, \theta_p, t_1)\, l(u_1, \ldots, u_m, t_1).$$

The p.d.f. of $t_1, u_1, \ldots, u_m$ is obtained by integrating with respect to $t_2, \ldots, t_p$, and the function h, being a probability density, has total integral unity. Therefore the p.d.f. required is

$$g(t_1 \mid \theta_1)\, l(u_1, \ldots, u_m, t_1), \tag{3}$$

and this proves the theorem.

Example. Before discussing the application of the theorem to sequential problems it is best to illustrate its meaning by a simple example. Let $x_1, \ldots, x_n$ be independently and normally distributed with mean $\theta_1\theta_2$ and standard deviation θ_2. Let $\bar{x}, s^2$ denote the sample mean and variance, and let $t_1 = \bar{x}/s, t_2 = s$. Also let $u_1 =$ median/range; in general $u_i =$ (measure of location)/(measure of dispersion). Then the conditions of the theorem are satisfied. The only one that causes any difficulty is condition (iv). To verify this, let S be the set of transformations $\{\mathbf{x}' = a\mathbf{x};\ a > 0\}$. Then (*a*) $t_1, u_1, \ldots, u_m$ are unchanged for all a; (*b*) the transformation $t_2' = at_2$ is (1, 1); (*c*) if T_2, T_2' are any two positive numbers, the transformation with $a = T_2'/T_2$ sends $t_2 = T_2$ into $t_2' = T_2'$. Therefore the theorem holds. Its meaning is that the conditional distribution of $u_1, \ldots$ given t_1 is independent of θ_1; i.e. t_1 has the basic property of a sufficient estimator for a single parameter that, given t_1, the estimators $u_1, \ldots$ give no more information about θ_1. This remark leads to a proof of the result in the Neyman-Pearson theory of testing hypotheses that optimum tests for θ_1 must be based on t_1.

3. *An application to sequential analysis.* Suppose that we want to test a hypothesis about the variance σ^2 of a normal population, the mean μ being unknown. To put the matter at its simplest suppose we have to choose between just two hypotheses about σ^2, $H_0 : \sigma^2 = \sigma_0^2$ and $H_1 : \sigma^2 = \sigma_1^2$ $(\sigma_1^2 > \sigma_0^2)$, the acceptable probabilities α, β of error of the two kinds being given.

Take observations one at a time and let s_n^2 be the usual estimate of variance from the first n observations.† Thus after n steps we have $(n-1)$ estimates of variance $s_2^2, \ldots, s_n^2$. Now we can consider these as 'observations' and apply the likelihood ratio test to them; for as Wald (9) has shown, the likelihood ratio test can be used even when the observations are not independent. Thus after n observations we calculate

$$L_n = \frac{p_n(s_2^2, \ldots, s_n^2 \mid \sigma_1^2)}{p_n(s_2^2, \ldots, s_n^2 \mid \sigma_0^2)}, \tag{4}$$

where $p_n(s_2^2, \ldots, s_n^2 \mid \sigma^2)$ is the joint p.d f. of the estimates of variance in samples of n.

$$\left.\begin{array}{lll} \text{If } \beta/(1-\alpha) < L_n < (1-\beta)/\alpha, & & \text{continue sampling.} \\ \text{If} & L_n \geqslant (1-\beta)/\alpha, & \text{accept } H_1. \\ \text{If } \beta/(1-\alpha) \geqslant L_n, & & \text{accept } H_0. \end{array}\right\} \tag{5}$$

Then the probabilities of error are approximately α and β provided that the probability is one that the test terminates.

† From this point onwards the sample size is denoted by a suffix.

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

Now by Theorem 1

$$p_n(s_2^2, \ldots, s_n^2 \mid \sigma^2) = g_n(s_n^2 \mid \sigma^2)\, l_n(s_2^2, \ldots, s_n^2), \tag{6}$$

where $g_n(s_n^2 \mid \sigma^2)$ is the p.d.f. of s_n^2 in samples of n. To see this, note that: (i) s_n^2 and the sample mean $\bar{x}_n$ are jointly sufficient for μ and σ^2; (ii) s_n^2 has a distribution not involving μ; (iii) $s_2^2, \ldots, s_{n-1}^2$ are functionally independent of s_n^2 and $\bar{x}_n$, and can be taken as the functions $u_1, \ldots, u_m$ in Theorem 1; (iv) the set of transformations $\mathbf{x}' = \mathbf{x} + a$ satisfy conditions (iv) of Theorem 1. Thus the conditions of Theorem 1 are satisfied and (6) follows.

But s_n^2 is distributed as $\sigma^2\chi^2/(n-1)$, where χ^2 has $(n-1)$ degrees of freedom, so that $g_n(s_n^2 \mid \sigma^2)$ is a known function. We find that

$$L_n = \frac{g_n(s_n^2 \mid \sigma_1^2)}{g_n(s_n^2 \mid \sigma_0^2)} = \left(\frac{\sigma_0}{\sigma_1}\right)^{n-1} \exp\left\{-(n-1)\, s_n^2 \left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right\},$$

and the test (5) can be written: continue sampling while

$$\log\frac{\beta}{1-\alpha} < -(n-1)\log\frac{\sigma_1}{\sigma_0} + \frac{(n-1)}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) s_n^2 < \log\frac{1-\beta}{\alpha}, \tag{7}$$

and accept H_0 or H_1 according as the left-hand or the right-hand inequality is the first not satisfied. To complete the justification of the test it remains to show that the probability is one that the test terminates. This can be done very easily by the method explained in §4.

The test (7) is identical with one derived by Stein and Girshick by a different method and is discussed briefly in Wald's book. The essential step in the present derivation is that the apparently complicated likelihood ratio (4) simplifies to an expression depending only on s_n^2.

4. *General statement of method.* To formulate the problem generally, suppose that there are p unknown parameters $\theta_1, \ldots, \theta_p$ and that we want to choose between $H_0 : \theta_1 = \theta_1^0$ and $H_1 : \theta_1 = \theta_1^1$ with assigned probabilities of error α and β, the test being independent of the nuisance parameters $\theta_2, \ldots, \theta_p$. The restriction to the choice between just two hypotheses is not serious because Wald (9) has shown how to apply such tests to the more general problem of choosing between two decisions. Further, Armitage (1) has shown how, by running several such tests simultaneously, it is possible to obtain tests for the choice between more than two decisions.

Suppose that there exists, for all n, a jointly sufficient set of estimators for $\theta_1, \ldots, \theta_p$ after n steps and that one of the set, which we now denote by t_n, has a known distribution $g_n(t_n \mid \theta_1)$ not involving $\theta_2, \ldots, \theta_p$. Suppose also that condition (iv) of Theorem 1 is satisfied with $u_1 = t_1, \ldots, u_{n-1} = t_{n-1}$. Let

$$L_n = g_n(t_n \mid \theta_1^1)/g_n(t_n \mid \theta_1^0). \tag{8}$$

Then the test is defined as follows:

$$\left.\begin{array}{ll} \text{Continue sampling if } \beta/(1-\alpha) < L_n < (1-\beta)/\alpha. & \\ \text{Accept } H_1 \quad \text{if} \quad L_n \geqslant (1-\beta)/\alpha. & \\ \text{Accept } H_0 \quad \text{if } \beta/(1-\alpha) \geqslant L_n. & \end{array}\right\} \tag{9}$$

Provided that the probability is one that the test terminates, the probabilities of error are approximately α and β.

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

The proof is exactly analogous to the proof for the special case given in §3. It depends on factorizing the joint probability of $t_1, \ldots, t_n$ by Theorem 1. It remains to give conditions under which the probability is one that the test (9) terminates. Sufficient conditions are given by the following theorem.

THEOREM 2. *Suppose that*

(i) *the test* (9) *can be written in the form: continue sampling only if* $t_n^- < t_n < t_n^+$, *where* t_n^- *and* t_n^+ *are functions of* $n, \alpha, \beta, \theta_1^0$ *and* θ_1^1;

(ii) t_n *is a function of the sample asymptotically normally distributed with mean* $\bar{t}_n$ *and variance* σ_n^2;

(iii) *either* (a) $(t_n^+ - t_n^-)/\sigma_n \to 0$,

or (b) $(t_n^- - \bar{t}_n)/\sigma_n \to \infty$,

or (c) $(t_n^+ - \bar{t}_n)/\sigma_n \to -\infty$,

as $n \to \infty$.

Then the probability is one that the test (9) *terminates.*

Proof.

$$\text{prob [sample size} > N] \leqslant \text{prob } [t_N^- < t_N < t_N^+] \to G\left[\frac{t_N^+ - \bar{t}_N}{\sigma_N}\right] - G\left[\frac{t_N^- - \bar{t}_N}{\sigma_N}\right], \qquad (10)$$

where $$G(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt.$$

Expression (10) tends to zero under any of the conditions (iii), thus proving the theorem.

For example, the variance test of §3 terminates because it can be written in the form: continue sampling only if $a - b/(n-1) < s_n^{2\prime} < a + c/(n-1)$, where a, b, c are constants, and condition (iii)(a) is applicable.

The condition (ii) that t_n should be asymptotically normally distributed is not necessary but is satisfied in all applications considered in this paper.

5. *Some applications.* We consider the choice between two hypotheses H_0 and H_1. α always denotes the acceptable probability of rejecting H_0 when true, β the acceptable probability of rejecting H_1 when true.

Example 1. *The variance ratio test.* Suppose that we have two normal populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2. Let H_0 be the hypothesis $\sigma_2^2 = \lambda_0 \sigma_1^2$ and H_1 the hypothesis $\sigma_2^2 = \lambda_1 \sigma_1^2$, where λ_0, λ_1 are given constants with, say, $\lambda_1 > \lambda_0$. Take observations in pairs, one from each population. We have four unknown parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2/\sigma_1^2 = \lambda$, say. Now after n pairs have been taken, $\bar{x}_{1(n)}, \bar{x}_{2(n)}, s_{1(n)}^2, s_{2(n)}^2$ form a jointly sufficient set of estimators for the unknown parameters and $F_n = s_{2(n)}^2/s_{1(n)}^2$ is a function of them with a distribution depending only on λ. We can redefine the jointly sufficient set so that F_n is one of them. The other conditions of Theorem 1 hold and therefore the general method applies. We calculate

$$L_n = p_n(F_n \mid \lambda_1)/p_n(F_n \mid \lambda_0),$$

where $p_n(F_n \mid \lambda)$ is the p.d.f. of F_n if the population variance ratio is λ. But F_n/λ has a variance ratio (F) distribution with $(n-1, n-1)$ degrees of freedom. Therefore

$$L_n = \left(\frac{\lambda_0}{\lambda_1}\right)^{\frac{1}{2}(n-1)} \left\{\frac{1 + F_n/\lambda_1}{1 + F_n/\lambda_0}\right\}^{n-1}. \qquad (11)$$

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

Thus the test is defined by (9) with L_n given by (11). This is a test with fixed limits and a complicated criterion. In practice we prefer variable limits which can be tabulated beforehand and a simple criterion. Let F_n^- and F_n^+ be the solutions of

$$\frac{\beta}{1-\alpha}, \quad \frac{1-\beta}{\alpha} = \left(\frac{\lambda_0}{\lambda_1}\right)^{\frac{1}{2}(n-1)} \left\{\frac{1+F_n/\lambda_1}{1+F_n/\lambda_0}\right\}^{n-1},$$

considered as equations for F_n. Then we can write the test: continue sampling while

$$F_n^- < F_n < F_n^+ \tag{12}$$

and accept H_0 or H_1 according as the left-hand or the right-hand inequality is the first not satisfied. Explicit expressions for the limits F_n^- and F_n^+ can easily be obtained and the test can be shown to terminate with probability one for any population variance ratio λ.

This test is of the same form as the test based on the range derived by Wald's method of weight functions (2). Girshick (4) has obtained a different test for comparing two variances. The operating characteristic of his test depends on $\sigma_2^{-2} - \sigma_1^{-2}$, which usually makes it less useful than the present test, whose operating characteristic depends on σ_2^2/σ_1^2.

Example 2. *Sequential analysis of variance.* Suppose we have k normal populations of means $\mu_1, \ldots, \mu_k$ and constant unknown variance σ^2. Let H_0 be the hypothesis: $\mu_1 = \ldots = \mu_k$. Let H_1 be the hypothesis: $\mu_1, \ldots, \mu_k$ are a random sample from a normal superpopulation of variance $\sigma_\mu^2 = \lambda\sigma^2$, where λ is a given constant. At each step in the sequential procedure we take one observation from each population. After the nth step we calculate the variance ratio

$$F_n = \frac{\text{mean square between samples}}{\text{mean square within samples}}$$

with $(k-1)$, $k(n-1)$ degrees of freedom. F_n is a function of a jointly sufficient set of estimators with a distribution depending only on σ_μ^2/σ^2 and condition (iv) of Theorem 1 holds. Therefore F_n can be used to derive a sequential test by calculating

$$L_n = \frac{\text{p.d.f. of } F_n \text{ under } H_1}{\text{p.d.f. of } F_n \text{ under } H_0}.$$

F_n has a distribution of the F form for both H_0 and H_1, and the test reduces to the following.

Let R_n^-, R_n^+ be the solutions of the equations in R_n:

$$\frac{\beta}{1-\alpha}, \quad \frac{1-\beta}{\alpha} = \frac{1}{(1+n\lambda)^{\frac{1}{2}(k-1)}} \left\{\frac{1+R_n}{1+R_n/(1+n\lambda)}\right\}^{\frac{1}{2}(nk-1)}.$$

Calculate at each step $\quad R_n = (k-1)F_n/k(n-1)$

$$= \frac{\text{corrected sum of squares between samples}}{\text{corrected sum of squares within samples}}.$$

Continue sampling while $\quad R_n^- < R_n < R_n^+, \tag{13}$

and accept H_0 or H_1 according as the left-hand or the right-hand inequality is the first not satisfied.

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

The quantities R_n^- and R_n^+ are easily obtained in explicit form and tabulated before the experiment is done. The asymptotic form for large n can also be found and used, with Theorem 2, to show that the probability is one that the test terminates.

An exactly similar method works for more complicated analyses. For example, we might have an experiment in randomized blocks, each step consisting in obtaining the observations from one block. We calculate after each step the appropriate variance ratio, F, and base the test on the likelihood ratio of F.

All this depends on the hypothesis H_1 being expressed in randomized form. The problem is much more difficult if we have to take a non-randomized hypothesis, $H_1: \mu_1, \ldots, \mu_k$ are any constants such that $\Sigma(\mu_i - \bar{\mu})^2/k = \lambda\sigma^2$, where $\bar{\mu} = \Sigma\mu_i/k$ and $\lambda > 0$. For then the variance ratio has a non-central F distribution under H_1, and the likelihood ratio takes a very complicated analytical form. The case most likely to be required is $k = 2$ (comparison of two means), when the problem reduces to the sequential t test considered by Rushton (8).

Example 3. *Sequential t test.* Rushton (8) has given a sequential t test obtained by the likelihood ratio method and an asymptotic expansion for the likelihood ratio for large n. He gives results for the test for a single mean; by a small modification it is possible to obtain sequential t tests for the difference between two means and for the comparison of two treatments in a complex experiment. Rushton's test is considered further in § 7.

Example 4. *Test for correlation coefficient.* Suppose we have samples from a normal bivariate population of correlation coefficient ρ. Let H_0, H_1 be the hypotheses $H_0: \rho = \rho_0$ (usually $\rho_0 = 0$), $H_1: \rho = \rho_1$ ($\rho_1 > \rho_0$, say). Each step in the experiment consists in taking a pair of observations. After n pairs there is a jointly sufficient set of estimators for the unknown parameters which can be chosen so that the sample correlation coefficient, r_n, is one of them. r_n has a distribution depending only on ρ. The condition (iv) of Theorem 1 holds, and therefore the likelihood ratio of r_n can be used to construct a sequential test.

We can either use David's (3) tables of the distribution of the correlation coefficient to find the likelihood ratio, or we can proceed as follows. Let

$$z_n = \tfrac{1}{2}\log\frac{1+r_n}{1-r_n}, \quad \zeta = \tfrac{1}{2}\log\frac{1+\rho}{1-\rho}.$$

By a classical result of Fisher, z_n is nearly normally distributed with variance $1/(n-3)$ and mean ζ. Thus approximately

$$\log L_n = \left(\frac{n-3}{2}\right)(\zeta_1 - \zeta_0)\,(2z_n - \zeta_0 - \zeta_1),$$

and the test becomes: continue sampling while

$$\frac{2}{(\zeta_1-\zeta_0)}\log\frac{\beta}{1-\alpha} < (n-3)\,(2z_n - \zeta_0 - \zeta_1) < \frac{2}{(\zeta_1-\zeta_0)}\log\frac{1-\beta}{\alpha}, \tag{14}$$

and accept H_0 or H_1 according as the left-hand or the right-hand inequality is the first not satisfied.

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

6. *Possible optimum property of the tests.* It is natural to ask whether these tests have an optimum property, i.e. whether out of all possible tests with given control over the probabilities of error, the present tests minimize the mean sample sizes under both H_0 and H_1. The point is not of much practical importance, but is worth discussing because it throws some light on the general likelihood ratio sequential test.

It is almost certain that in general the tests are not optimum. The reason is that the tests are fixed limit likelihood ratio tests (F.L.L.R.) with non-independent† observations. Now if the F.L.L.R. test was always optimum for simple hypotheses the present tests would be optimum. But this is not so. It is possible to find two simple hypotheses such that the F.L.L.R. test is not optimum.

Consider a sequence of random variables $\{x_1, x_2, \ldots\}$, each x_i taking values 0 or 1. Let H_1 be the simple hypothesis

$$\text{prob}\,(x_1 = 0) = 27/59, \quad \text{prob}\,(x_1 = 1) = 32/59, \quad \text{prob}\,(x_1 = x_2 = \ldots) = 1;$$

and H_0 the simple hypothesis

$$\text{prob}\,(x_i = 0) = 1/3, \quad \text{prob}\,(x_i = 1) = 2/3, \quad x_i \text{ independent } (i = 1, 2, \ldots).$$

After n observations the likelihood ratio is

$$L_n = \begin{cases} 0 \text{ if any two observations differ,} \\ {}_0L_n = 27 \cdot 3^n/59 \text{ if all are 0's,} \\ {}_1L_n = 32 \cdot 3^n/59 \cdot 2^n \text{ if all are 1's.} \end{cases}$$

Let T_{ij} be the test: reject H_1 if any two observations differ, reject H_0 if we obtain i 0's or j 1's. Since ${}_0L_2 = {}_1L_5$, T_{25} is a F.L.L.R. test. T_{34} is *not* a F.L.L.R. test, but it is easy to show that we have the following probabilities of error and mean sample sizes.

Test	Probability of rejecting H_0 when true	Probability of rejecting H_1 when true	Mean sample size under H_0	Mean sample size under H_1
T_{25}	59/243	0	714/243	214/59
T_{34}	57/243	0	693/243	209/59

Thus T_{34} gives better control than T_{25} over the probabilities of error, with smaller mean sample sizes.

It is worth trying to explain in non-mathematical terms how this comes about. The two samples (i) 0 0 and (ii) 1 1 1 1 1 have the same likelihood ratio, but the future development in probability of the likelihood ratio is quite different. In case (i) there is a chance of 2/3 that if H_0 is true just one more observation will reveal it. In case (ii) the corresponding chance is only 1/3. Therefore (i) is potentially a 'better' sample for discrimination than (ii). We may expect that if we prolong the test by one observation in case (i) and to compensate reduce the critical sample size to four in case (ii), there will be an improvement in the properties of the tests. This turns out to be so.

† The test for a single variance given in §3 is an exception. As shown by Girshick and Stein this test arises as a test of simple hypotheses about a set of suitably chosen independent variables.

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

When the successive random variables are independent and identically distributed this cannot happen. In fact, suppose S_1 and S_2 are two samples of n_1 and n_2 observations with the same likelihood ratio. Then the distribution, given S_1, of the likelihood ratio in samples of n_1+t is the same under both H_0 and H_1 as the distribution, given S_2, of the likelihood ratio in samples of n_2+t, for all $t = 1, 2, \ldots$. Now it is reasonable, and can be proved formally under certain assumptions, that once sampling has stopped the decision as to which hypothesis to accept should be based on the likelihood ratio in a way independent of sample size. It follows that the reductions in the probabilities of error due to prolonging the experiment by t observations are the same for S_1 as for S_2. Thus if it is best to stop sampling when the sample S_1 is attained it is also best to stop sampling when the sample S_2 is attained, i.e. the critical limits for the likelihood ratio should be independent of sample size. These remarks are an attempt to express in an informal way part of the highly formal work of Wald and Wolfowitz (10).

The example considered above is of course highly artificial, and the actual difference in efficiency between the tests T_{25} and T_{34} is very small. It does, however, show that we may expect some lack of efficiency in the tests given in § 6. Armitage (1) reported some sampling experiments on Wald's sequential t test in which one of the mean sample sizes under the sequential test was slightly *greater* than the corresponding fixed sample size. He suggested that this was because the sequential test was really more powerful than the Wald approximation indicated. Another possibility is that there is an appreciable loss of efficiency due to using fixed limits in the test.

7. *Relation to Wald's method of weight functions.* All the tests given by the method of the present paper can be obtained by Wald's method of weight functions. As an example we discuss the relation between Rushton's sequential t test and the corresponding test obtained by Wald's method ((9), § A 9). Suppose we make observations on independent normally distributed random variables with unknown variance σ^2. Let H_0 be the hypothesis that mean is zero, H_1 the hypothesis that the mean is $\delta\sigma$. Rushton's test (see § 6 above) gives a likelihood ratio

$$L_n = \exp\left[-\frac{\delta^2(n-u_n^2)}{2}\right]\frac{\mathrm{Hh}_{n-1}(-\delta u_n)}{\mathrm{Hh}_{n-1}(0)}, \tag{15}$$

where $u_n = \sum_{i=1}^{n} x_i \Big/ \left\{\sum_{i=1}^{n} x_i^2\right\}^{\frac{1}{2}}$ and $\mathrm{Hh}_{n-1}(-\delta u_n)$ is a standard function, defined for example in Jeffreys and Jeffreys (5). In Wald's method a weight function is introduced for the nuisance parameter σ. Wald takes the weight function to be constant; this leads, for the 'one-sided' test of H_0 against H_1, to the likelihood ratio

$$\exp\left[-\frac{\delta^2(n-u_n^2)}{2}\right]\frac{\mathrm{Hh}_{n-2}(-\delta u_n)}{\mathrm{Hh}_{n-1}(0)}. \tag{16}$$

(This expression can be deduced from formulae given by Armitage (1).)

There is a very close relation between the two tests. Further, if we take for the weight function $1/\sigma$, we get exactly the expression (15). Thus the two methods give identical tests when the weight function is chosen suitably.

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press

I am very grateful to Mr D. V. Lindley for pointing out a serious error in my first statement of Theorem 1, and to him and Mr F. J. Anscombe for helpful comments on the draft of the paper.

Note added in proof. A paper by G. A. Barnard dealing with the above problems is to appear shortly in *Biometrika*.

REFERENCES

(1) Armitage, P. *J.R. statist. Soc.*, Suppl. 9 (1947), 250.
(2) Cox, D. R. *J.R. statist. Soc.* B, 11 (1949), 101.
(3) David, F. N. *Tables of the correlation coefficient* (London, 1938).
(4) Girshick, M. S. *Ann. math. Statist.* 17 (1946), 123.
(5) Jeffreys, H. and Jeffreys, B. S. *Methods of mathematical physics*, 2nd ed. (Cambridge, 1950), §23·081.
(6) Kendall, M. G. *Advanced theory of statistics*, vol. 2 (London, 1946), §17·16.
(7) Nandi, H. K. *Sankhyā*, 8 (1948), 339.
(8) Rushton, S. *Biometrika*, 37 (1950), 326.
(9) Wald, A. *Sequential analysis* (New York, 1947).
(10) Wald, A. and Wolfowitz, J. *Ann. math. Statist.* 19 (1948), 326.

Statistical Laboratory
Cambridge

https://doi.org/10.1017/S030500410002764X Published online by Cambridge University Press