SEQUENTIAL TESTS OF NULL HYPOTHESES

By C. RADHAKRISHNA RAO

Statistical Laboratory, Calcutta

INTRODUCTION

The probability ratio test as developed by Wald (1945) is shown to have the optimum property that it minimises the expected number of observations needed for a decision (Wald & Wolfwitz, 1949). An added attraction of this test is that the limits for the probability ratio can be obtained, at least approximately, without solving any distribution problem. This makes the scope of the test a bit wider because one need not restrict to certain types of distributions which are commonly used in current test procedures.

The probability ratio test seeks to discriminate between two alternative hypotheses H_1 and H_2, sampling being continued till a decision is reached one way or the other. The first modification needed in this problem is that in some cases sampling cannot be continued indefinitely and a decision may have to be made before a specified number of observations are taken. Wald suggested a method of truncating the sequential procedure in such a case. It seems desirable to study this problem in some detail.*

Secondly no satisfactory method is available to test a null hypothesis by sequential sampling. In testing a null hypothesis, it might be noted that nothing is known about the alternative hypothesis and consequently the test procedure should not depend on any alternative hypothesis. A simple solution to this problem is given in this article.

The solution is illustrated with the help of model sampling experiments which supply a proper justification for the suggested test procedures.

1. NULL-HYPOTHESIS

equential tests have been considered to discriminate between two or more alternative hypotheses. This is not strictly applicable in testing a null hypothesis when nothing is known about the alternative hypotheses.

Wald gets over the difficulty by introducing a loss function and determining a suitable range round the null hypothesis. In this case the problem, in a way, reduces to two hypotheses case one comprising the range round the null hypothesis and the other the rest. This procedure has some useful applications but problems arise where the main emphasis is on the null hypothesis and the success of an experiment consists in disproving a null hypothesis. In such problems no single

*A detailed study of this problem will be published in a forthcoming seminar publication edited by H. Wold.

hypothesis can be offered as an alternative nor is it possible to assign any risk function. How can a sequential sampling procedure be adopted with a view to give the observations a fair chance of disproving the null hypothesis?

Obviously large differences from the null hypothesis could be discovered even with a small sample of observations and even small differences from the null hypothesis could be detected with probability approaching unity by choosing an infinitely large number of observations. It is, clearly, not possible to accept the null hypothesis before an indefinitely large sample is observed. If the sequential process is to be terminated only when the null hypothesis is disproved sampling may have to be continued indefinitely specially when the null hypothesis or a closer hypothesis is true. So it is desirable to set an upper limit N to the number of observations and stop sampling if the null hypothesis could be rejected at any earlier stage. It might happen that the null hypothesis stands unrejected even at the Nth stage in which case the situation is same as in the current test procedure when a null hypothesis could not be disproved on the basis of an observed sample. In such a case the null hypothesis may be accepted provisionally or the class of hypotheses from which the observed samples could have reasonably arisen may be determined.

2. A TEST FOR NULL HYPOTHESIS WHEN ALTERNATIVES ARE ONE-SIDED

Consider the problem of testing the hypothesis H_0 that the value of a parameter θ occuring in a probability distribution is θ_0. The current test procedure based on a sample of size N consists in rejecting the hypothesis if

$$P_N(\theta) \geqslant \lambda_N P_N(\theta_0)^*$$

where θ is a possible alternative value and λ_N is determined such that the probability of rejecting H_0 when it is true has an assigned value. When the exact value of the alternative is not known one can construct (Neyman & Pearson, 1936) a test which is best for alternatives near about the null hypothesis and hope that it is fairly good for distant hypotheses also. In particular when the alternatives are known to be one-sided (say $\theta > \theta_0$) the locally most powerful best test is defined by (Rao & Poti, 1946)

$$P'_N(\theta_0) \geqslant \mu P_N(\theta_0)$$

where $P'_N(\theta_0)$ is the first derivative of $P_N(\theta)$ at $\theta = \theta_0$ and μ is suitably determined to keep the first kind of error at an assigned level. This suggests a sequential test of the type

$$P'_n(\theta_0) \geqslant A(N) P_n(\theta_0) \qquad \dots (2.1)$$

where A (N) is a suitably determined constant depending on the level of significance and N is the upper limit to the number of observations, sampling being terminated at the smallest value of n for which the above relationship (2.1) is realised.

*The probability density of the first k observations is represented throughout by $P_k(\theta)$ where θ is any parameter involved in the function.

As observed earlier the sequential process may terminate before N observations are drawn in which case the null-hypothesis is rejected or it may end at the Nth stage without any definite decision being made about the null hypothesis. It is, however, possible to assign the limits between which the true value of θ can be asserted to lie and thus judge by how much a possible alternative hypothesis differs from the true one.

An upper bound to A (N)

Let us define by w_1 a region in the N dimensional space defined by

$$P'_n(\theta_0) \geqslant A(N)\, P_n(\theta_0) \quad \text{for some } n \leqslant N$$

The set of points satisfying this relationship leads to the rejection of the null hypothesis and if the level of significance is α it follows that

$$\int_{w_1} P_N(\theta_0)\, dv = \alpha$$

Consider the set of points C_n inside w_1 which lead to the termination of the sequential process at the n th stage. This consists of an unlimited cylinder in the N dimensional space cutting the n dimensional space $(x_1, \dots, x_n)$ in a region w_{1n} and having its generators parallel to the other axes. Hence

$$\int_{C_n} P'_N(\theta_0)\, dx_1 \dots\dots dx_N = \int_{w_{1n}} P'_n(\theta_0)\, dx_1 \dots\dots dx_n > A(N) \int_{w_{1n}} P_n(\theta_0)\, dx_1 \dots\dots dx_n$$

$$= A(N) \int_{C_n} P_N(\theta_0)\, dx_1 \dots\dots dx_N$$

Summing over all possible n we obtain

$$\int_{w_1} P'_N(\theta_0)\, dv \geqslant A(N) \int_{w_1} P_N(\theta_0)\, dv$$

If the first integral is denoted by r then

$$r \geqslant A(N)\, \alpha$$

which gives an upper limit to A (N) if r is known. The exact determination of r is a difficult problem but its maximum value can be found as follows. Let w be a region such that

$$\int_{w} P_N(\theta_0)\, dv = \alpha \qquad \dots \quad (2.2)$$

and

$$\int_{w} P'_N(\theta_0)\, dv \text{ is a maximum.}$$

The boundry of such a region is given by

$$P'_N(\theta_0) = \lambda\, P_N(\theta_0)$$

where λ is determined from the condition (2.2). Then

$$r_m = max\ r = \int_w P'_N(\theta_0)\, dv$$

Using r_m the upper limit to A(N) is

$$r_m/\alpha$$

Since the relationship

$$P'_n(\theta_0) > \frac{r_m}{\alpha} P_n(\theta_0)$$

implies

$$P'_n(\theta_0) > A(N)\, P_n(\theta_0)$$

it follows that the actual probability of rejection of H_0 when r_m/α is used instead of A(N) is less than α.

If the alternatives are less than θ_0 then the sequential test is defined by

$$P'_n(\theta_0) < B(N)\, P_n(\theta_0)$$

the determination of B(N) being similar to that of A(N).

Test for the mean of a normal distribution when the variance is known

Let μ_0 be the mean of the normal distribution under the null hypothesis. The current test based on N observations is defined by

$$\bar{x} - \mu_0 > \sigma\delta_\alpha/\sqrt{N}$$

where σ is the standard deviation, δ_α is the normal deviate corresponding to upper α percent value.

$$P_N(\mu_0) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N exp\left\{-\frac{\Sigma(x_i-\mu_0)^2}{2\sigma^2}\right\}$$

$$P'_N(\mu_0) = \frac{\Sigma(x_i-\mu_0)}{\sigma^2} P_N(\mu_0)$$

$$r_m = \int P'_N(\mu_0)\, dv = \int_{\bar{x}-\mu_0 > \sigma\delta_\alpha/\sqrt{N}} \frac{N(\bar{x}-\mu_0)}{\sigma^2} \frac{\sqrt{N}}{\sqrt{2\pi}\sigma} e^{-N(\bar{x}-\mu_0)^2/2\sigma^2} d\bar{x}$$

$$= \frac{\sqrt{N}}{\sqrt{2\pi}\sigma} e^{-N(\bar{x}-\mu_0)^2/2\sigma^2}\Big]_{\infty}^{\sigma\delta_\alpha/\sqrt{N}}$$

$$= \frac{\sqrt{N}}{\sqrt{2\pi}\sigma} e^{-\delta_\alpha^2/2}$$

The sequential test for alternative values of $\mu > \mu_0$ is defined by

$$\sum_1^n x_i - n\mu_0 > \frac{\gamma_m}{\alpha}\sigma^2 = \sqrt{N}\,\sigma\Delta_\alpha$$

$$\Delta_\alpha = \frac{1}{\sqrt{2\pi}\,\alpha}\, e^{-\delta_\alpha^2/2}$$

sampling being terminted as soon as the above relationship is realised for some $n < N$. The values of Δ_α for some percentage values of α are given below.

α	Δ_α
1%	2.6655
2.5%	2.3376
5%	2.0626

For example if $N = 100$, $\sigma = 1$ and $\alpha = 5\%$ the sequential test consists in obtaining the successive cumulative sums

$$x_1 - \mu_0,\ \sum_1^2 x_i - 2\mu_0,\ \sum_1^3 x_i - 3\mu_0,\ \ldots ,$$

and terminating the process whenever the sum exceeds $\sqrt{N}\Delta_\alpha = 10 \times 2.0626 = 20.626$. If the alternative values of the mean are less than μ_0 then the sequential test is defined by

$$\sum_1^n (x_i - \mu_0) < -20.626$$

The results of model sampling experiments

To verify the nature of the approximation involved in setting the above limits some model sampling experiments were conducted using the normal samples published by Mahalanobis (1934). Each table (plate) of random samples consists of 10 columns of 50 observations each. One column of 50 values in a plate is combined with the corresponding column of 50 values in the next plate to form a random sample of 100 observations. Starting the first plate 78* series of 100 observations each were obtained. For each series all the 100 cumulative totals are obtained.

In seven of the samples given below the minimum cumulative sum was less than -20.626.

Sample no.	$-\sum_1 x_i$	Value of n
8	21.916	99
12	20.601	100
38	22.959	100
40	27.534	70
53	22.521	60
57	21.570	93
74	24.000	66

* More samples are being constructed to verify the theoretical limits suggested for various problems in this article. These investigations are time consuming and it is hoped that the results will be published as soon as they are ready.

The expected number according to theory is about four and the excess might suggest that samples have a slight negative bias. None of the cumulative sums exceeded 20.626 although in three samples nos. 23, 54 and 63 maximum values were as high as 20.133, 19.186 and 18.985 respectively. The average of maximum values is only 6.4090 whereas the average of minimum values is -9.5128 which again indicates a general negative bias in the samples. For this reason the results obtained for positive and negative deviations from the null hypothesis were combined in table 1. Columns (3) and (5) of this table give the frequency with which various alternative values of μ had been detected in the samples. The values in column (5) exceed the corresponding values in column (3) which may be attributed to a slight negative bias in the published normal tables. The average of columns (3) and (5) is given in column (7). The observed level of significance is 4.49% against the theoretical value of 5%. Column (8) gives the power function values of the best current test based on 100 observations and realised frequencies (col. 7) in the sequential test are more in correspondence with the power function values associated with the best current test based on 81 observations (col. 9). So far as the power is concerned the sequential test based on 100 observations is equivalent to a current test based on 81 observations. But column (6) shows that the average number of observations needed for a sequential test is less than 81 in all the cases considered above. Thus a sequential test is more economical, the actual saving in observations, of course, depends on the nature of the alternative hypotheses, those closer to the null hypothesis requiring more observations.

3. A TEST FOR NULL HYPOTHESIS WHEN ALTERNATIVES ARE TWO SIDED

The locally unbiased most powerful current test based on N observations is defined by

$$P''_N(\theta_0) \geqslant \lambda_1 P'_N(\theta_0) + \lambda_2 P_N(\theta_0)$$

By analogy we set up the sequential test

$$P''_n(\theta_0) \geqslant A_1(N)\, P'_n(\theta_0) + A_2(N)\, P_n(\theta_0)$$

sampling being continued, subject to a maximum of N observations, till the above relationship is realised. The sequential process may terminate without any decision being made about the null hypothesis. If w_1 is the region constituted by points $(x_1, \ldots, x_N)$ leading to a decision then

$$\int_{w_1} P_N(\theta_0)\, dv = \alpha\ , \int_{w_1} P'_N(\theta_0)\, dv = 0$$

and

$$\int_{w_1} P''_N(\theta_0)\, dv \geqslant A_2(N) \int_{w_1} P_N(\theta_0)\, dv = A_2(N)\,\alpha$$

If the first integral is denoted by κ then

$$A_2(N) \geqslant \frac{\kappa}{\alpha}$$

TABLE 1 RESULTS OF 78 SEQUENTIAL SAMPLING TESTS

Value of the alternative hypothesis $\lvert s \rvert$	Positive deviations Test: $P'_s > 20.626P_s$		Negative deviations Test: $P'_s < -20.626P_s$		Average for positive and negative deviations		Expected frequency of detection for the best current test based on	
	Average no. of observations	Frequency of detection	Average no. of observations	Frequency of detection	Average of (2) and (4)	Average of (3) and (5)	100 observations	81 observations
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	21.74	1.0000	21.58	1.0000	21.68	1.0000	.9999+	.9999+
3/4	29.19	1.0000	28.64	1.0000	28.91	1.0000	.9999+	.9999+
1/2	44.96	1.0000	40.23	1.0000	42.59	1.0000	.9996	.9983
1/4	70.11	.6923	64.80	.7564	67.45	.7243	.8037	.7264
1/5	74.85	.5128	66.62	.6026	70.73	.5577	.6376	.5615
1/6	74.87	.3846	68.45	.4743	71.66	.4205	.5080	.4420
1/7	80.09	.2820	72.91	.4359	76.50	.3589	.4141	.3594
1/8	81.39	.2307	72.53	.3590	76.96	.2948	.3465	.3016
1/9	82.63	.2051	73.84	.3205	78.23	.2628	.2966	.2595
1/10	80.33	.1538	74.45	.2820	77.39	.2179	.2595	.2282
0	—	0	79.29	.0897	—	.0449	.0500	.0500

As before the exact value of κ is not known but may be replaced by its maximum value. There seems to be no simple way of determining $A_1(N)$ but if symmetry permits we can equate it to zero. In this case the sequential test is simply

$$P''_n(\theta_0) > A_2(N)\, P_n(\theta_0)$$

The maximum value of κ is given by

$$\int_w P''_N(\theta_0)\, dv$$

where w is the region defined by

$$P''_N(\theta_0) > \lambda P_N(\theta_0)$$

λ being determined from the condition

$$\int_w P_N(\theta_0)\, dv = \alpha$$

it being understood that the region w ensures the condition

$$\int_w P'_N(\theta_0)\, dv = 0$$

4. Test for the mean of a normal population when the variance is known

In this case

$$P'_n(\mu_0) = (\sum_1^n x_i - n\mu_0)\, P_n(\mu_0)/\sigma^2$$

$$P''_n(\mu_0) = \left\{ \frac{(\Sigma x_i - n\mu_0)^2}{\sigma^4} - \frac{n}{\sigma^2} \right\} P_n(\mu_0)$$

The sequential test is defined by

$$(\sum_1^n x_i - n\mu_0)^2 > n\sigma^2 + A_2(N)\, \sigma^4$$

The maximum value of κ is found to be

$$\frac{2N}{\sqrt{2\pi\sigma^2}}\ \delta_\alpha\ e^{-\delta_\alpha^2/2}$$

where δ_α is the α % normal deviate considering positive and negative deviations. The values of $A_2(N)$ for 1% and 5% values of α are given below

α	Value of $A_2(N)$
1%	$7.4497\ N/\sigma^2$
5%	$4.5818\ N/\sigma^2$

The 5% sequential test is

$$(\Sigma x_i - n\mu_0)^2 - n\,\sigma^2 \geqslant 4.5818\, N\sigma^2$$

An alternative test is $|\Sigma x_i - n\,\mu_0| \geqslant A(N)$ where A(N) is suitably determined. No reasonably simple expression for A(N) could be found.

5. Test for a given σ when the mean is known

In this case

$$P_n(\sigma) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\; - \tfrac{1}{2}\left\{\frac{\Sigma(x_i - \mu)^2}{\sigma^2}\right\}$$

$$\frac{P'_n(\sigma)}{P_n(\sigma)} = -\,\frac{n}{\sigma} + \frac{\Sigma(x_i - \mu)^2}{\sigma^3}$$

If the alternatives are $\sigma > \sigma_0$ then the sequential test is defined by

$$\Sigma\,(x_i - \mu)^2 - n\,\sigma_0^2 \geqslant A\,(N)\;\sigma_0^2$$

An upper limit to A (N) is found to be

$$\frac{1}{\alpha}\;\;\frac{\chi_\alpha^N}{\sigma_0}\;\;\frac{2^{(N+2)/2}}{\Gamma\left(\frac{N}{2}\right)}\;e^{-\chi^2_\alpha/2}$$

where χ^2_α is the upper α percent value of χ^2 on N degrees of freedom.

6. Test for σ when the mean is unknown

The previous test can be used in this case with slight modification. We define

$$\chi_2^2 = \tfrac{1}{2}\,(x_2 - x_1)^2$$

$$\chi_3^2 = \chi_2^2 + \tfrac{2}{3}\left(x_3 - \frac{x_1 + x_2}{2}\right)^2$$

$$\cdots\quad\cdots\quad\cdots\quad\cdots$$

$$\chi^2_n = \chi^2_{n-1} + \frac{n}{n+1}\left(x_n - \frac{x_1 + \ldots\ldots + x_{n-1}}{n-1}\right)^2$$

The sequential test is defined by

$$\chi^2_n \geqslant n\sigma_0^2 + A\,(N)\;\sigma_0^2$$

where A (N) is same as that given in the previous section.

It is of some importance to determine the value of N, the upper limit to the number of observat ons in any practical problem. This might be chosen such that the chance of not detecting an alternative differing from the null hypothesis to any specified degree is small say 5% or 1%. The actual solution to this problem is very difficult but some approximate methods are available.

It is suggested that this method of fixing the upper limit and then adopting the sequential sampling procedure may be better than determining an admissible range round the null hypothesis to start with and setting up a sequential procedure without an upper limit to the number of observations to decide whether the true hypothesis is within the admissible range or outside.

References

Mahalanobis, P. C. (1934): Tables of random samples from a normal population. *Sankhyā*, 1, 289.

Neyman, J. & Pearson, E. S. (1936): Contributions to the theory of testing statistical hypotheses. *Stat. Res. Mem.*, 1, 1 & 25.

Rao, C. R. & Poti, S. J. (1946): On locally most powerful tests when alternatives are one-sided. *Sankhyā*, 7, 439.

Wald, A. (1945): Sequential tests of statistical hypotheses. *Ann. Math. Stat.* 16, 1.

Wald, A. & Wolfwitz, J. (1940): Optimum character of the sequential probability ratio test. *Ann. Math. Stat.* 19, 326.