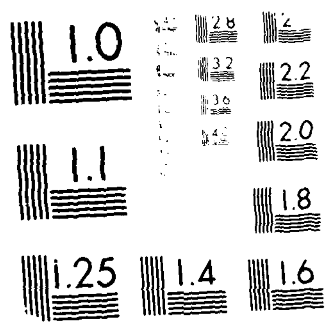
AD-A190 320 STOPPING RULES AND OBSERVED SIGNIFICANCE LEVELS(U) NORTH CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHASTIC PROCESSES J BATHER SEP 87 TR-209 1/1

UNCLASSIFIED AFOSR-TR-87-1013 F49620-85-C-0144 F/G 12/3 NL

END
DATE
FILMED
4 88


1.0
2.8
3.2
2.2
3.6
2.0
1.1
1.8
1.25
1.4
1.6

DTIC FILE COPY

(2)

AD-A190 320

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 209	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 87-1913
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AFOSR/NM
6c. ADDRESS (City, State, and ZIP Code) Statistics Dept. 321-A Phillips Hall 039-A Chapel Hill, NC 27514	7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER ~~AFOSR~~ No. F49620 85C 0144.
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO: 61102F PROJECT NO: 2304 TASK NO: A5 WORK UNIT ACCESSION NO:

11. TITLE (Include Security Classification)
Stopping rules and observed significance levels

12. PERSONAL AUTHOR(S)
Rather, J.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Preprint	FROM 9/87 TO 8/88	September 1987	11

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Key Words & Phrases: hypothesis testing; likelihood ratios; Neyman-Pearson theory; sequential decisions.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

It is well known how to combine the significance levels observed in a number of independent experiments. When this number is a random variable determined by a stopping rule, the observed significance level can still be calculated if there is an acceptable ordering of the points in the extended sample space. But what can be said if the stopping time is ill-defined? This paper obtains explicit lower bounds on the level of significance by considering orderings based on a family of alternative hypotheses. These bounds give some measure of the effect of failing to specify the stopping rule in advance.

DTIC ELECTE JAN 1 1 1988 E

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified/unlimited
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff	22b. TELEPHONE (Include Area Code) (202)767-5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
Unclassified/Unlimited

CENTER FOR STOCHASTIC PROCESSES

AFOSR-TR- 87-1813

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

STOPPING RULES AND OBSERVED SIGNIFICANCE LEVELS

by

John Bather

Technical Report No. 209

September 1987

STOPPING RULES AND OBSERVED SIGNIFICANCE LEVELS

John Bather*

Mathematics Division, University of Sussex
and
Center for Stochastic Processes, University of North Carolina

Key words and Phrases: hypothesis testing; likelihood ratios; Neyman-Pearson theory; sequential decisions.

ABSTRACT

It is well known how to combine the significance levels observed in a number of independent experiments. When this number is a random variable determined by a stopping rule, the observed significance level can still be calculated if there is an acceptable ordering of the points in the extended sample space. But what can be said if the stopping time is ill-defined? This paper obtains explicit lower bounds on the level of significance by considering orderings based on a family of alternative hypotheses. These bounds give some measure of the effect of failing to specify the stopping rule in advance.

*This research was supported by the Air Force Office of Scientific Research Grant No. F49620 85C 0144.

Accession For
NTIS GRA&I
DTIC TAB
Unannounced
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

1. INTRODUCTION

The aim of this paper is to consider what advice a statistician should give in the following situation.

Example. In order to detect a possible difference in performance between two types of subject, a controlled comparison is carried out four times and the significance levels observed in independent, one-sided tests are .04, .15, .01, .05. The investigator consults a statistician about the proper combination of these results and is recommended to use Fisher's method, based on the product of the observed levels. He does this correctly and obtains a nominal significance level α_o=.00131. However, the investigator then reveals that he is uneasy about the method because the number of comparisons was not fixed in advance. He simply continued the sequence of experiments until he thought there was enough evidence to reject the null hypothesis and establish a real difference between the two types of subject.

In his first reaction, the statistician might well regret hearing the last remark since it weakens, if not destroys the foundations of his previous advice. On the other hand, perhaps the situation illustrated by this example is common enough to deserve a more serious response. Can we properly say anything about the combined significance of a sequence of experiments when the stopping rule is ill-defined?

Before we attempt to answer this question, let us consider a more general setting. Suppose that a sequence of m independent experiments is carried out and that the results are used separately to test a certain null hypothesis H_0, the same test statistic t being used in each case. For convenience, let us assume that t is a continuous variable with a distribution function Φ_o, under H_0, such that the observed level of significance $u = 1 - \Phi_o(t)$ is uniformly distributed on the interval [0,1]. Thus, we are assuming that the detailed results of the original experiments have been reduced to a sequence of observed significance levels $u_1, u_2, \ldots, u_m$ and that, under the null hypothesis, this is equivalent to a random sample from the uniform distribution on [0,1].

For a single experiment, rejecting H_0 when t is large is equivalent to using $u=1-\Phi_o(t)$ as a test statistic and there is an implicit ordering of the underlying sample space in which points are treated as "extreme" if the corresponding value of u is small. A justification for this might be an application of the Neyman-Pearson Lemma, for some alternative hypothesis. For example, suppose such a hypothesis restricts the distribution of u to a family of Lehmann alternatives with probability densities of the form :

$$h(u) = \theta u^{\theta-1}, \quad 0 < u < 1, \tag{1}$$

with parameter θ, $0 < \theta < 1$. Then, rejection of H_0 when the observed level u is small is justified by Neyman-Pearson theory. Of course, the same is true for any alternative density h, provided that h(u) is decreasing in u over the unit interval. However, it turns out that the family of Lehmann alternatives is particularly appropriate for dealing with combinations of several observed levels.

Suppose m is fixed, $m \geq 2$, leaving aside the question of stopping rules for the moment. We are tacitly assuming that the experiments are all similar and have similar sample sizes, so it is natural to give equal weights to the components $u_1, u_2, \ldots, u_m$ in assessing their combined significance. The usual method of combining m independent tests is based on the product $v_m = u_1 u_2 \ldots u_m$ and it relies on the fact that $-2\log v_m$ has a χ^2-distribution with 2m degrees of freedom. The method is due to Fisher (1950) and it is also described in standard texts such as Cox and Hinkley (1974), see page 80. For our purpose, it will be more convenient to write $S_n = \sum_1^n w_i$, where $w_i = -\log u_i$. Thus, $\{S_n, n\geq 1\}$ is a random walk and, under the null hypothesis, the independent steps w_i have a common probability density e^{-w}, $w>0$. In particular, it is easily verified that

$$\alpha_o = P_o(S_m \geq s) = \{1+s+\ldots+\frac{s^{m-1}}{(m-1)!}\}\, e^{-s}. \tag{2}$$

In the above example, $m=4$ and $s = 12.72$, so the nominal significance level is $\alpha_o = .00131$. It will be argued later that, in the absence of a stopping rule, a lower bound on reasonable significance levels is $\alpha^* = .00458$: see equation (16). More precisely, $\alpha^* = \min\{\alpha(\theta), 0<\theta<1\}$, where θ represents a simple alternative hypothesis and $\alpha(\theta)$ is determined by stopping the process $\{S_n\}$ as soon as the likelihood ratio exceeds a certain critical value.

The main argument of this paper relies on several principles. The first is to use Fisher's method of combining significance levels when the number is .xed in advance. Other methods have been suggested in the literature, for example, several are mentioned in a discussion of asymptotic optimality by Berk and Cohen (1979). However, they do not seem appropriate here. The second principle is that the ordering of points $(u_1, u_2, \ldots, u_m)$ in $\mathbb{R}^m$ determined by the product v_m can be formally justified by an application of the Neyman-Pearson Lemma, for some alternative hypothesis. We postulate an alternative H_1, which may be composite and assume that the observed levels $u_1, u_2, \ldots, u_m$ are independent and identically distributed not only under the null hypothesis H_0, but also under H_1. Then, as we shall see in Section 2, the only alternative distributions of u with the required properties are those given by (1). The final principle is one of conservatism in the evaluation of $\alpha(\theta)$. When the number of experiments is not prescribed and a range of values is possible, the simple alternatives represented by θ, $0 < \theta < 1$, lead to different orderings of the corresponding sample space. If there is no clearly defined stopping rule, then it seems reasonable to consider one that maximizes the error probability $\alpha(\theta)$ and this corresponds to using a test of power one.

The next section gives a version of the Neyman-Pearson Lemma for independent sampling up to a well-defined stopping time. This is followed by a discussion of alternative hypotheses for our particular problem. The explicit calculation of $\alpha(\theta)$ is described in Section 3 and it is a simple matter to find the minimum

$$\alpha^* = \min\{\alpha(\theta),\ 0 < \theta < 1\}. \tag{3}$$

In general, the results $u_1, u_2, \ldots, u_m$ of m tests determine a final value $S_m = s$ and the corresponding levels α_o and α^* depend only on m and s, with $0 < \alpha_o < \alpha^* < 1$. The interpretation of α_o as an observed significance level is familiar and, provided that m is fixed, it can be justified in terms of Neyman-Pearson theory by appealing to any one of the alternative hypotheses represented by θ, $0 < \theta < 1$. In the absence of any stopping rule, α^* is a useful measure, but its interpretation is more complex, since each value of θ leads to a different ordering of the set of possible stopped sequences. Unfortunately, the choice of θ is arbitrary and all we can say is that α^* is the smallest significance level that can be justified by applying the Neyman-Pearson Lemma, for some alternative hypothesis. Roughly speaking, the ratio α^*/α_o is a measure of the effect of failing to specify the stopping rule in advance. The final section gives some numerical values of the ratio, corresponding to nominal significance levels $\alpha_o = .01$ and $.001$.

2. PRELIMINARIES

Let $y = (y_1, y_2, \ldots)$ be a sequence of i.i.d. random variables and suppose there are two simple hypotheses, H_0 and H_1, which specify their common probability density as f or g, respectively. Suppose further that the stopping rule is given and, for convenience, let this be determined by continuation sets $C_1, C_2, \ldots$, where each C_n is a measurable subset of $\mathbb{R}^n$. For any sequence y, let $y^n = (y_1, y_2, \ldots, y_n)$. It is assumed that, if $y^n \in C_n$, then $y^m \in C_m$ for $m = 1, 2, \ldots, n-1$. The stopping time is

$$N(y) = \min\{n \geq 1:\ y^n \notin C_n\}, \tag{4}$$

so that $N(y) = \infty$ if $y^n \in C_n$ for every $n \geq 1$. Consider the stopping set at time n :

$$D_n = \{y^n \in \mathbb{R}^n : y^m \in C_m,\ m<n,\ y^n \notin C_n\}.$$

For $y^n \in D_n$, the observed sequence that arises from stopping will be written

$(n, y_1, y_2, \ldots, y_n)$. A terminal decision rule is defined by any measurable partitions of the stopping sets : $D_n = A_n \cup B_n$, $n=1,2,\ldots$. Thus, H_0 is accepted if $N(y) = n$ and $y^n \in A_n$, for some $n \geq 1$, and rejected if $N(y) = n$ and $y^n \in B_n$. We are concerned here with comparing terminal decisions for a given stopping rule and finite stopping times, so the case $N(y) = \infty$ can be treated arbitrarily. For simplicity, suppose that H_0 is always accepted if $N(y) = \infty$. Then the error probabilities are

$$\alpha = \sum_{n=1}^{\infty} \int_{B_n} dP_f, \quad \beta = \sum_{n=1}^{\infty} \int_{A_n} dP_g + P_g\,(N = \infty), \tag{5}$$

where P_f and P_g refer to the distributions of y under H_0 and H_1. Now consider the stopped likelihood ratio when $N = n$. For an observed sequence $(n, y_1, y_2, \ldots, y_n)$, the likelihood ratio is

$$\left.\frac{dP_g}{dP_f}\right|_n = \lambda_n(y^n) = \prod_{i=1}^{n} \frac{g(y_i)}{f(y_i)}. \tag{6}$$

<u>Definition</u>. We say that the partitions $\{A_n, B_n,\ n \geq 1\}$ define a likelihood ratio test if there is a positive constant λ such that $\lambda_n(y^n) \leq \lambda$ if $y^n \in A_n$ and $\lambda_n(y^n) \geq \lambda$ if $y^n \in B_n$, $n = 1,2,\ldots$.

Note that the standard proof of the Neyman-Pearson Lemma applies to two measures on any σ-field and, in particular, to a stopped σ-field. This leads to the result that, in general, likelihood ratio tests have the following optimality property.

<u>Lemma</u>. Let $\{A_n, B_n\}$ define a likelihood ratio test with critical value $\lambda > 0$ and error probabilities α, β given by (5). Consider any other terminal decision rule defined by partitions
$D_n = A'_n \cup B'_n$, $n=1,2,\ldots$ and let α', β' be the corresponding error probabilities. Then,

$$\alpha' \leq \alpha \Rightarrow \beta' \geq \beta \quad \text{and} \quad \alpha' < \alpha \Rightarrow \beta' > \beta.$$

The lemma shows that, for any stopping rule, the ordering of observed sequences should be based on the likelihood ratio and this holds whether we are comparing data vectors of the same or different dimensions.

Before we can apply this result, we need to discuss alternative hypotheses for the problem of combining independent significance tests. In the first place, let us return to the case when the number of tests is fixed. There is no given alternative, but it is difficult to justify any method of combination and the implicit ordering of vectors of significance levels without introducing one. In fact, there are many alternatives consistent with the usual method of combination. For a fixed number of tests with observed levels $u_1, u_2, \ldots, u_m$, $m \geq 2$, the ordering based on the product $v_m = u_1 u_2 \ldots u_m$ is equivalent to using the statistic $\sum_i^m w_i$ on the set of points $\{(w_1, w_2, \ldots, w_m) : w_i > 0,\ 1 \leq i \leq m\}$. In other words, $(w_1', w_2', \ldots, w_m')$ is more extreme than $(w_1, w_2, \ldots, w_m)$ if $\sum_1^m w_i' \geq \sum_1^m w_j$. Under the null hypothesis, when the u_i are uniformly distributed on [0,1], we have the probability density $f(w) = e^{-w}$, $w > 0$, for the $w_i = -\log u_i$. Now consider an alternative hypothesis under which their common probability density is $g(w)$ and assume that $g(w) > 0$ and its derivative $g'(w)$ is continuous for $w > 0$. The likelihood ratio is given by

$$\log \lambda_m = \sum_1^m \log g(w_i) + \sum_1^m w_i .$$

It can only produce the same ordering as $\sum_1^m w_i$ if it is a function of this sum. Then, differentiation shows that $\sum_1^m d(\log g(w_i)) = 0$ whenever $\sum_1^m dw_i = 0$. Since $m \geq 2$, it follows that $g'(w_i)/g(w_i)$ is constant for all $w_i > 0$ and, hence, the alternative hypothesis must be represented by an exponential distribution :

$$g(w) = \theta e^{-\theta w},\ w > 0. \tag{7}$$

The parameter θ is positive and, since λ_m is required to be increasing in $\sum_1^m w_i$, we demand that $0 < \theta < 1$. Finally, it is easily verified that (7) is equivalent to the family of Lehmann alternatives (1), for the original significance levels u_i.

3. SIGNIFICANCE LEVELS

We now turn to the question of assessing the significance attained by a series of tests when there is no clearly defined stopping rule. The null hypothesis H_0 is given by (7), with $\theta = 1$, and we first consider a simple alternative corresponding to a fixed value of the parameter θ, $0 < \theta < 1$. Bearing in mind that the investigator might have been seeking to reject the null hypothesis as soon as he felt convinced that this was proper, we ought to use a conservative evaluation of the "observed" significance level $\alpha(\theta)$. In general, the likelihood ratio at any stage is

$$\lambda_n(\theta) = \theta^n \exp\{(1-\theta)\sum_1^n w_i\}. \tag{8}$$

Suppose we are given the results of m tests and let $\lambda_{obs} = \lambda_m(\theta)$ be the final value of the likelihood ratio. The stopping rule underlying these results is unknown and, perhaps, there were intermediate values of the likelihood ratio exceeding λ_{obs}. However, in view of the lemma in Section 2, any other series of tests must be treated as stronger evidence against H_0 if and only if it produces a final value $\lambda_n(\theta) \geq \lambda_{obs}$. The most conservative evaluation of $\alpha(\theta)$ corresponds to rejecting H_0 for any sequence $\{\lambda_n(\theta)\}$ such that $\lambda_n(\theta) \geq \lambda_{obs}$ for some $n \geq 1$. This suggests a stopping rule: any sequence with $\sup_{n\geq 1} \lambda_n(\theta) \geq \lambda_{obs}$ is stopped when it first exceeds the critical level. The terminal decision is to reject H_0 whenever the stopping time is finite. Note that the procedure depends on θ. It leads to a significance level $\alpha(\theta)$ which is a maximum amongst likelihood ratio tests with the same critical level.

In order to evaluate $\alpha(\theta)$, consider the stochastic process $\{S_n\}$, where $S_0 = 0$ and $S_n = \sum_1^n w_i$, $n \geq 1$. We note that the condition $\lambda_n(\theta) \geq \lambda_{obs}$ is equivalent to

$$S_n \geq c + nk, \tag{9}$$

$$c = (1-\theta)^{-1}\log \lambda_{obs}, \quad k = -(1-\theta)^{-1}\log \theta. \tag{10}$$

The stopping time is determined by a linear boundary :

$$N = \min\{n \geq 1 : S_n \geq c + nk\}. \tag{11}$$

The likelihood ratio test with critical value $\lambda = \lambda_{obs}$ always leads to rejection of H_0 if N is finite, so the corresponding error probability is

$$\alpha(\theta) = P_0(N<\infty) = P_0(S_n \geq c + nk \text{ for some } n \geq 1). \tag{12}$$

Any other stopping rule, followed by a likelihood ratio test with the same critical value λ_{obs}, must produce a smaller error probability under H_0. This is a consequence of the definition of N and the terminal decision rule. It is a straightforward matter to evaluate $\alpha(\theta)$.

<u>Proposition.</u> Let N be the stopping time defined by (10) and (11), for a fixed value of the parameter θ, $0 < \theta < 1$, and suppose that $\lambda_{obs} \geq \theta$. Then the likelihood ratio test with this critical value has error probabilities

$$\alpha(\theta) = \theta\lambda_{obs}^{-1}, \quad \beta(\theta) = 0.$$

<u>Proof.</u> These results are consequences of well-known properties of the random walk $\{S_n\}$: see, for example, problem 2.1 in Siegmund's monograph (1985). We first show that $\beta(\theta) = 0$. Under the alternative hypothesis, $\{S_n\}$ has independent steps with mean θ^{-1} and it is easily verified from (10) that the slope $k=k(\theta)$ has $1 < k(\theta) < \theta^{-1}$. It follows from the strong law of large numbers that N is finite with probability 1 and, according to (5), $\beta(\theta) = 0$.

The formula for $\alpha(\theta)$ can now be established by using equations (5) and (6). We note that P_f and P_g are both derived from exponential distributions and that

$$D_n = B_n = \{(w_1, w_2, \ldots, w_n): S_j<c+jk,\ 1\leq j<n,\ S_n\geq c+nk\},$$

$$\alpha(\theta) = \sum_{n=1}^{\infty} \int_{B_n} dP_f = \sum_{n=1}^{\infty} \int_{B_n} \{\lambda_n(\theta)\}^{-1} dP_g.$$

By using (8) and (10), we find that

$$\{\lambda_n(\theta)\}^{-1} = \lambda_{obs}^{-1} \exp\{-(1-\theta)(S_n-c-nk)\}.$$

On the set B_n, $S_n - c - nk = \zeta_n$ is the overshoot beyond the linear boundary when the random walk stops. Thus,

$$\alpha(\theta) = \lambda_{obs}^{-1} \sum_{n=1}^{\infty} \int_{B_n} \exp\{-(1-\theta)\zeta_n\} dP_g.$$

Since $\beta(\theta) = 0$, we know that $P_g(N < \infty) = 1$ and the above series reduces to the expectation of $\exp\{-(1-\theta)\zeta_n\}$ under the alternative hypothesis. This is easy to evaluate for a random walk with exponentially distributed steps, since the overshoot ζ_N must also have the probability density (7). The only exception is when $\lambda_{obs} = \theta$ and, hence, $c + k < 0$. In this case, the event $S_1 > c + k$ always occurs and $P_f(N=1) = P_g(N=1) = 1$, so $\alpha(\theta) = 1$. Otherwise, $\lambda_{obs} \geq \theta$ and the required expectation is

$$\int_0^{\infty} \exp\{-(1-\theta)\zeta\} g(\zeta) d\zeta = \theta.$$

It follows that $\alpha(\theta) = \theta\lambda_{obs}^{-1}$, as required.

The critical level of the likelihood ratio was defined as the final value observed in a particular series of m tests. Let this correspond to stopping the random walk when $S_m = s$, so that $\lambda_{obs} = \theta^m \exp\{(1-\theta)s\}$ by equation (8). The main result of the proposition can be expressed in a more convenient form :

$$\alpha(\theta) = \theta^{1-m} \exp\{-(1-\theta)s\}. \tag{13}$$

The condition that $\lambda_{obs} \geq \theta$ is satisfied provided that

$$s \geq (m-1)k, \quad k = -(1-\theta)^{-1} \log \theta. \tag{14}$$

It can be shown that, as θ increases from 0 to 1, $k=k(\theta)$ decreases from ∞ to 1. Hence, the formula (13) is valid for sufficiently large values of θ, provided that $s > m-1$. It holds whenever it produces a significance level $\alpha(\theta) \leq 1$ and, otherwise, it should be replaced by $\alpha(\theta) = 1$.

Our results, so far, depend on choosing a single value of θ and we must now consider the general hypothesis H_1 represented by (7) for $0 < \theta < 1$. It is easily verified that the inequality (9), which characterises stopping points where the likelihood ratio exceeds λ_{obs}, is equivalent to

$$S_n - s \geq (n-m)k(\theta). \tag{15}$$

This shows how our ordering of the set of possible results depends on θ. There is no single ordering which adequately represents the composite alternative H_1. However, it will be useful to determine the range of values achieved by $\alpha(\theta)$

for any given m and s.

By differentiating (13), we find that $\alpha'(\theta)/\alpha(\theta) = s-(m-1)\theta^{-1}$, so there is a unique minimum of $\alpha(\theta)$ at $\theta=\theta^{*}=(m-1)s^{-1}$ and $\theta^{*}<1$ if $s > m-1$. The case $0 \leq s \leq m-1$ is trivial, since it implies that $\alpha(\theta) = 1$, $0 < \theta < 1$. Let us assume, from now on, that $s > m-1$. Then the minimum level $\alpha^{*} = \alpha(\theta^{*}) < 1$ and

$$\alpha^{*} = \{\frac{s}{m-1}\}^{m-1}\exp(m-1-s), \quad s > m-1. \tag{16}$$

It is also clear that any level in the interval $(\alpha^{*},1)$ can be attained by choosing an appropriate $\theta \in (\theta^{*},1)$.

4. NUMERICAL ILLUSTRATION

Finally, we can compare the formulae (2) and (16). As before, $s = -\sum_{1}^{m} \log u_i$, where $u_1, u_2, \ldots u_m$ are the observed significance levels in m independent tests of the same hypothesis. The nominal combined level of significance α_0 represents these results, assuming that m was fixed in advance. On the other hand, α^{*} is a measure of the evidence against the null hypothesis when no stopping rule was prescribed. The previous arguments can be summarised by saying that α^{*} is a lower bound in the following sense : it is the smallest significance level that can be justified by a sequential likelihood ratio test, for a suitable alternative hypothesis.

It is clear from (2) and (16) that $\alpha_0 = \alpha^{*}$ if m=1. For each $m \geq 2$ and any $s > m-1$, we have $0 < \alpha_0 < \alpha^{*} < 1$. This is implicit in our construction of α^{*}, but it can also be verified directly from the formulae. For convenience, we shall consider the ratio α^{*}/α_0. It is a straightforward exercise to establish the following properties by using (2) and (16): both α_0 and α^{*} are decreasing in s and α^{*}/α_0 is increasing in s, with

$$\lim_{s\to\infty} \frac{\alpha^{*}}{\alpha_0} = \{\frac{e}{m-1}\}^{m-1}(m-1)! \; . \tag{17}$$

When $s = m-1$, $\alpha^{*}=1$ and $\alpha_0 < 1$, which confirms that $\alpha^{*}/\alpha_0 > 1$ for all $s > m-1$. The limit given by (17) is an upper bound. The table below shows values of the ratio α^*/α_0 for m = 2,3,...,10,20, with s chosen so that α_0 = .01, .001. The final row gives the limits as $s \to \infty$ and $\alpha_0 \to 0$.

TABLE I

Ratios of significance levels : α^*/α

m	2	3	4	5	6	7	8	9	10	20
α_0=.01	2.36	2.92	3.26	3.52	3.72	3.89	4.02	4.12	4.22	4.81
α_0=.001	2.45	3.09	3.53	3.82	4.08	4.30	4.46	4.61	4.75	5.56
α_0= 0	2.72	3.70	4.46	5.12	5.70	6.23	6.71	7.16	7.59	10.97

BIBLIOGRAPHY

Berk, R.H. and Cohen, A. (1979). Asymptotically optimal methods of combining tests, Journal of the American Statistical Association, 74, S12-S14.

Cox, D.R. and Hinkley, D.V. (1974). Theoretical statistics, London: Chapman and Hall.

Fisher, R.A. (1950). Statistical methods for research workers (11th ed.), London: Oliver and Boyd.

Siegmund, D. (1985). Sequential analysis, New York: Springer-Verlag.

DATE

FILMED

4 8