Biometrika (1977), **64**, 2, *pp.* 247–54
Printed in Great Britain

Hypothesis testing when a nuisance parameter is present only under the alternative

By R. B. DAVIES

Applied Mathematics Division, Department of Scientific and Industrial Research, Wellington, New Zealand

SUMMARY

Suppose that the distribution of a random variable representing the outcome of an experiment depends on two parameters ξ and θ and that we wish to test the hypothesis $\xi = 0$ against the alternative $\xi > 0$. If the distribution does not depend on θ when $\xi = 0$, standard asymptotic methods such as likelihood ratio testing or $C(\alpha)$ testing are not directly applicable. However, these methods may, under appropriate conditions, be used to reduce the problem to one involving inference from a Gaussian process. This simplified problem is examined and a test which may be derived as a likelihood ratio test or from the union-intersection principle is introduced. Approximate expressions for the significance level and power are obtained.

Some key words: C-alpha test; Hypothesis testing; Likelihood ratio test; Maximum of Gaussian process; Simple hypothesis; Union-intersection principle.

1. INTRODUCTION

Suppose that the outcome of an experiment is represented by a random variable, X, whose distribution depends on two unknown parameters ξ_0 and θ_0. Suppose further that when $\xi_0 = 0$ this distribution does not depend on θ_0; that is, θ_0 does not enter into the model for X when $\xi_0 = 0$.

Now suppose that we wish to test the hypothesis $\xi_0 = 0$ against $\xi_0 > 0$. If X represents a large sample, one might consider applying one of the standard large sample tests. However, these cannot be applied directly. For example, while the generalized likelihood ratio, λ, could be calculated, the standard chi-squared approximation of $2 \log \lambda$ would not apply even when allowance is made for the one-sided nature of the alternative.

On the other hand, under appropriate conditions, large sample theory can be used to reduce the problem to a tractable one involving inference from a Gaussian process. In this section we see how this reduction may be done. The remainder of the paper is concerned with the resulting simplified situation.

Suppose that X represents a random sample $(x_1, \ldots, x_n)$ of observations, each with density $p(x; \xi_0, \theta_0)$ and that $p(x; 0, \theta_0)$ does not depend on θ_0. Consider the locally optimal test statistic for testing $\xi_0 = 0$ against $\xi_0 > 0$ when θ_0 is known:

$$Z_n(\theta_0) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \left[\frac{\partial}{\partial \xi}\{\log p(x_i, \xi, \theta_0)\}/\gamma(\theta_0)\right]_{\xi=0}, \tag{1·1}$$

where

$$\{\gamma(\theta_0)\}^2 = \text{var}\left\{\left[\frac{\partial}{\partial \xi}\log p(x_i; \xi, \theta_0)\right]_{\xi=0}\right\}$$

calculated under the hypothesis. A test which rejects the hypothesis, $\xi_0 = 0$, for large values of $Z_n(\theta_0)$ is then asymptotically optimal for alternatives with the given value of θ_0 and $\xi_0 = O(n^{-\frac{1}{2}})$. Of course θ_0 is unknown and so $Z_n(\theta_0)$ cannot be evaluated.

The approach adopted in this paper is to base a test on the entire random function, or stochastic process, $Z_n(\theta)$, as a function of θ.

Suppose that under the hypothesis

$$\operatorname{cov}\{Z_n(\theta_1), Z_n(\theta_2)\} = \rho(\theta_1, \theta_2).$$

From the definition of $Z_n(\theta)$ we have $\rho(\theta, \theta) = 1$ for each θ. The following results may be derived from arguments similar to those leading to Theorem 2 of Neyman (1959); see also van Eeden (1963). Let $\mathscr{D}(Z_n; \xi_0, \theta_0)$ denote the distribution of the process $Z_n(.)$ when ξ_0 and θ_0 are the parameter values. Then, under suitable regularity conditions, as $n \to \infty$,

$$\mathscr{D}(Z_n; n^{-\frac{1}{2}}\xi_0, \theta_0) \to \mathscr{D}(Z; \xi_0, \theta_0), \tag{1·2}$$

where $Z(.)$ is a Gaussian random process with expectation

$$E_{\xi_0,\theta_0}\{Z(\theta_1)\} = \xi_0 \gamma(\theta_0) \rho(\theta_1, \theta_0) \tag{1·3}$$

and covariance function

$$\operatorname{cov}\{Z(\theta_1), Z(\theta_2)\} = \rho(\theta_1, \theta_2), \tag{1·4}$$

for all ξ_0 and θ_0. Note particularly the presence of $\rho(.,.)$ in both (1·3) and (1·4).

Similar asymptotic formulae may be obtained if, instead of defining $Z_n(\theta_0)$ as in (1·1), we let

$$Z_n(\theta_0) = n^{\frac{1}{2}} \gamma(\theta_0)\, \hat{\xi}_n(\theta_0) \tag{1·5}$$

or

$$Z_n(\theta_0) = \{2 \log \lambda_n(\theta_0)\}^{\frac{1}{2}} \operatorname{sgn}\{\hat{\xi}_n(\theta_0)\}, \tag{1·6}$$

where $\hat{\xi}_n(\theta_0)$ is the maximum likelihood estimate of ξ_0 when θ_0 is given, $\lambda_n(\theta_0)$ is the generalized likelihood ratio, again with θ_0 given, and where it is assumed that the problem remains defined and regular for negative values of ξ_0. The equivalence of these approaches is essentially proved by Moran (1970). If other nuisance parameters are present and are regular, in particular being estimable when $\xi_0 = 0$, then versions (1·5) and (1·6) of $Z_n(.)$ still satisfy (1·2) provided $\gamma(\theta)$ is suitably redefined. Similarly the natural generalization of (1·1), the $C(\alpha)$ test (Neyman, 1959; Moran, 1970), also provides a version of $Z_n(.)$ satisfying (1·2). In practice, the locally optimal test statistic or the $C(\alpha)$ test statistic are the most straightforward to calculate.

The formulae given here can also be obtained in asymptotic situations other than the usual random sample case, the form of (1·3) and (1·4) arising from results such as that given by Hajék & Šidák (1967, § VI.1.4). In some instances a Gaussian stochastic process $Z(.)$ satisfying (1·3) and (1·4) may arise directly.

In most of what follows it is assumed that the test is to be based on one of these versions of $Z_n(.)$ and that n is so large that deviations from normality or equations (1·3) and (1·4) can be ignored.

We conclude the introduction with several examples.

Example 1. Consider random variables X and Y, independently normally distributed with unit variances and expectations $\xi_0 \cos\theta_0$ and $\xi_0 \sin\theta_0$, and $0 \leqslant \theta_0 \leqslant 2\pi$. This problem arises when two asymptotically normal test statistics are being combined and the alternatives are two-sided. The optimal test when $\theta_0 = \theta$ is given rejects the hypothesis for large values of $Z(\theta) = X\cos\theta + Y\sin\theta$. Formulae (1·3) and (1·4) are satisfied if

$$\lambda(\theta) = 1, \quad \rho(\theta_1, \theta_2) = \cos(\theta_1 - \theta_2).$$

Of course, the usual test would be to reject the hypothesis for large values of $X^2 + Y^2 = \sup\{Z(\theta)\}^2$ and not surprisingly the method in this paper yields the same result.

Example 2. This is as in Example 1, but with $0 \leqslant \theta_0 \leqslant \frac{1}{2}\pi$. This problem arises when two asymptotically normal test statistics are being combined and the alternatives are one-sided. It has been considered extensively by van Zwet & Oosterhoff (1967); see also Oosterhoff (1969).

Example 3. Here X represents a random sample of observations with density

$$p(x; \xi_0, \theta_0) = (1-\xi_0)\, e^{-x} + \xi_0 \theta_0 e^{-\theta_0 x},$$

where $1 < \theta_0 < \infty$. Then $Z_n(\theta)$ defined by (1·1) becomes

$$Z_n(\theta) = [\theta_0 \exp\{(1-\theta_0)\,x\} - 1]\,(2\theta_0 - 1)^{\frac{1}{2}}/(\theta_0 - 1).$$

The covariance $\rho(\theta_1, \theta_2)$ may readily be calculated.

Example 4. Suppose that $X = (X_1, \ldots, X_q)$, where X_k are independent Poisson random variables with expectation given by

$$E_{\xi_0,\theta_0}(X_k) = \sum_{j=1}^{s} \alpha_j e^{-\lambda_j k} + \xi_0 e^{-\theta_0 k}.$$

This is not obviously a large sample problem. The large sample results are obtained if $\alpha_1, \ldots, \alpha_s$ tend to infinity together since, as far as the relevant test statistics are concerned, this is equivalent to the repeated replication of the experiment with fixed $\alpha_1, \ldots, \alpha_s$. The unknown parameters α_j and λ_j $(j = 1, \ldots, s)$ can be fitted by maximum likelihood and the problem is to test for the presence of another decay term. If θ_0 is given, a $C(\alpha)$ test can be found and this provides the statistic $Z(\theta)$. This particular problem arises from an experiment involving a mixture of radioactive elements.

Another example involves the testing of the hypothesis that a point process is Poisson (Davies, 1977, §4).

2. The test

We suppose that the parameter θ_0 is limited to the range $[L, U]$ and consider the Gaussian process $\{Z(\theta): L \leqslant \theta \leqslant U\}$ satisfying (1·3) and (1·4).

We also suppose that $Z(\theta)$ has a continuous second derivative. This ensures that $\rho(\theta_1, \theta_2)$ has a continuous second derivative with respect to either argument and that $\rho_{11}(\theta)$ introduced in the next section is continuous.

The test which we consider in this paper has critical region of the form

$$\{\sup_{L \leqslant \theta \leqslant U} Z(\theta) > c\}. \tag{2·1}$$

This test may be derived from Roy's type I principle (Roy, 1953) or its generalization, the union-intersection principle (Roy, Gnanadesikan & Srivastava, 1971, pp. 36–46). It may also be derived as a likelihood ratio test, for if Z is observed at only a finite number of points: $Z(\theta_0), Z(\theta_1), \ldots, Z(\theta_r)$, then the ratio of the likelihood at (ξ_0, θ_0) to the likelihood when $\xi_0 = 0$ may be readily found to be

$$\exp[\xi_0 \gamma(\theta_0)\, Z(\theta_0) - \tfrac{1}{2}\xi_0^2 \{\gamma(\theta_0)\}^2]. \tag{2·2}$$

This does not depend on $\theta_1, \ldots, \theta_r$ and so gives the likelihood ratio when all of the $Z(\theta)$ is observed. Maximization over $\xi_0 > 0$, $\theta_0 \in [L, U]$ leads to a test of the form (2·1).

The test (2·1) does not take account of the function $\gamma(.)$. One possibility is to assign a value to ξ_0 and base a test on the maximum of (2·2). Another is to consider the locally optimal Bayes test when a prior distribution is assigned to θ_0. This has critical region of the form

$$\int_L^U \gamma(\theta)\, Z(\theta)\, B(d\theta) > c,$$

and may be readily calculated together with its significance level and power but may have rather poor performance; see, for example, the linear test discussed by van Zwet & Oosterhoff for Example 2.

3. Distribution of the test statistic

There seems little chance of being able to calculate

$$\mathrm{pr}_{\xi_0,\theta_0}\{\sup_{L\leqslant\theta\leqslant U} Z(\theta) > c\} \tag{3·1}$$

even when $\xi_0 = 0$. However, a bound may be found by the approach of Cramér & Leadbetter (1967, p. 285), or Leadbetter (1972). For generality we consider

$$\mathrm{pr}_{0,\theta}[\sup_{L\leqslant\theta\leqslant U}\{Z(\theta)-c(\theta)\} > 0], \tag{3·2}$$

where $c(\theta)$ is continuously differentiable. Following Cramér & Leadbetter or Leadbetter's formula (3·3), after making a correction, the expected number of 'upcrossings' of zero by $Z(\theta)-c(\theta)$ under the hypothesis $\xi_0 = 0$ is given by

$$\frac{1}{2\pi}\int_L^U \exp[-\tfrac{1}{2}\{c(\theta)\}^2]\{-\rho_{11}(\theta)\}^{\frac{1}{2}}\psi[c_1(\theta)/\{-\rho_{11}(\theta)\}^{\frac{1}{2}}]\,d\theta, \tag{3·3}$$

where

$$c_1(\theta) = \frac{\partial}{\partial\theta}c(\theta), \quad \rho_{11}(\theta) = \left[\frac{\partial^2}{\partial\phi^2}\rho(\phi,\theta)\right]_{\phi=\theta},$$

$$\psi(x) = \exp(-\tfrac{1}{2}x^2) - x\int_x^\infty \exp(-\tfrac{1}{2}t^2)\,dt. \tag{3·4}$$

Expression (3·3) provides an upper bound on the probability that

$$[\sup_{L\leqslant\theta\leqslant U}\{Z(\theta)-c(\theta)\} > 0 \quad\text{and}\quad Z(L)-c(L) < 0], \tag{3·5}$$

since at least one upcrossing must occur if (3·5) is satisfied. Thus

$$\mathrm{pr}_{0,\theta}[\sup_{L\leqslant\theta\leqslant U}\{Z(\theta)-c(\theta)\} > 0]$$
$$\leqslant \Phi\{-c(L)\} + \frac{1}{2\pi}\int_L^U \exp[-\tfrac{1}{2}\{c(\theta)\}^2]\{-\rho_{11}(\theta)\}^{\frac{1}{2}}\psi\left[\frac{c_1(\theta)}{\{-\rho_{11}(\theta)\}^{\frac{1}{2}}}\right]d\theta, \tag{3·6}$$

where Φ denotes the normal distribution function. In particular

$$\mathrm{pr}_{0,\theta}\{\sup_{L\leqslant\theta\leqslant U} Z(\theta) > c\} \leqslant \Phi(-c) + \frac{1}{2\pi}\exp(-\tfrac{1}{2}c^2)\int_{\theta=L}^U \{-\rho_{11}(\theta)\}^{\frac{1}{2}}\,d\theta. \tag{3·7}$$

This provides the bound on the significance level of the test (2·1).

We would like some indication of the sharpness of the bound (3·7). If we regard the event $\{Z(L) > c\}$ as an upcrossing then the right-hand side of (3·7) gives the expected number of upcrossings of c by $Z(.)$. If c is large and $\rho(\theta_1, \theta_2)$ is near zero for $|\theta_1-\theta_2|$ large then we might expect the number of upcrossings to have an approximately Poisson distribution. Indeed Cramér & Leadbetter (1967, p. 257) and Leadbetter (1972) show that under appropriate conditions in the stationary situation this is the case. Then the actual probability (3·1) can be estimated. However, one will, in general, be concerned with probabilities less than 0·1, in which case (3·7) itself will be very close to the probability estimated using the Poisson distribution.

On the other hand, in Example 1, since $\sup Z(\theta) > c$ will occur if and only if exactly one upcrossing occurs, the term $\Phi(-c)$ in (3·7) can be omitted and then one obtains the exact result $\exp(-\frac{1}{2}c^2)$. In example 2 one also obtains the exact result given by van Zwet & Oosterhoff (1967, formulae (5·8) and (5·10)).

4. Properties under the alternative

4·1. *Sensitivity of the test*

A lower bound on the power of the test can be obtained from

$$\text{pr}_{\xi_0,\theta_0}\{\sup Z(\theta) > c\} \geqslant \text{pr}_{\xi_0,\theta_0}\{Z(\theta_0) > c\} = \Phi\{\xi_0\gamma(\theta_0) - c\}. \tag{4·1}$$

An upper bound on the power can be obtained by substituting $c - \xi_0\gamma(\theta_0)\rho(\theta,\theta_0)$ for $c(\theta)$ in (3·6). The following theorem, worked out jointly with Mr R. Littlejohn of the Mathematics Department, Victoria University of Wellington, shows that the difference between these bounds tends to zero as ξ_0 and c tend to infinity.

THEOREM 4·1. *Suppose that $\rho(\theta_1,\theta_2) < 1$ for each $\theta_1 \neq \theta_2$ in $[L, U]$, that the conditions given in § 2 are satisfied, $\theta_0 \in [L, U]$ and that $\gamma(\theta_0) > 0$. Then, as $\xi_0 \to \infty$,*

$$\text{pr}_{\xi_0,\theta_0}\{\sup Z(\theta) > \xi_0\gamma(\theta_0) - \Delta\} \to \Phi(\Delta). \tag{4·2}$$

Proof. From (4·1)

$$\liminf_{\xi_0\to\infty} \text{pr}_{\xi_0,\theta_0}\{\sup Z(\theta) > \xi_0\gamma(\theta_0) - \Delta\} \geqslant \Phi(\Delta).$$

Also

$$\begin{aligned}\limsup_{\xi_0\to\infty} \text{pr}_{\xi_0,\theta_0}\{\sup Z(\theta) > \xi_0\gamma(\theta_0) - \Delta\} &\leqslant \limsup_{\xi_0\to\infty} \Phi[\Delta - \xi_0\gamma(\theta_0)\{1-\rho(L,\theta_0)\}] \\ &+ \frac{1}{2\pi}\int_L^U \exp[-\tfrac{1}{2}\{\Delta - \xi_0\gamma(\theta_0)\{1-\rho(\theta,\theta_0)\}^2]\{-\rho_{11}(\theta)\}^{\frac{1}{2}}\psi\left[-\xi_0\gamma(\theta_0)\frac{\partial\rho(\theta,\theta_0)/\partial\theta}{\{-\rho_{11}(\theta)\}^{\frac{1}{2}}}\right]d\theta.\end{aligned} \tag{4·3}$$

We suppose $L < \theta_0 \leqslant U$; the case $\theta_0 = L$ follows by an argument similar to the present one. Choose ϵ so that $\theta_0 - \epsilon \geqslant L$ and $\partial\rho(\theta,\theta_0)/\partial\theta$ is monotonically decreasing when $\theta_0 - \epsilon \leqslant \theta \leqslant \theta_0$. Now $\psi(x)$ is bounded when $x > 0$ and $\psi(x) + (2\pi)^{\frac{1}{2}}x$ is bounded when $x < 0$. Then (4·3) is equal to

$$\begin{aligned}\limsup_{\xi_0\to\infty} (2\pi)^{-\frac{1}{2}}\int_{\theta_0-\epsilon}^{\theta_0} \exp(-\tfrac{1}{2}[\Delta - \xi_0\gamma(\theta_0)\{1-\rho(\theta,\theta_0)\}]^2)\,\xi_0\gamma(\theta_0)\frac{\partial}{\partial\theta}\rho(\theta,\theta_0)\,d\theta \\ = \limsup_{\xi_0\to\infty} (2\pi)^{-\frac{1}{2}}\int_0^\kappa \exp\{-\tfrac{1}{2}(\Delta - x)^2\}\,dx = \Phi(\Delta),\end{aligned} \tag{4·4}$$

where the upper limit of integration is $\kappa = \xi_0\gamma(\theta_0)\{1-\rho(\theta_0-\epsilon,\theta_0)\}$. Thus the upper and lower limits are equal and this completes the proof.

The principal difference between (4·3) and (4·4) with fixed ξ_0 arises from the replacement of $\psi(x)$ by 0 if $x > 0$ and $-(2\pi)^{\frac{1}{2}}x$ if $x < 0$ and is approximately equal to $0{\cdot}2\exp(-\frac{1}{2}\Delta^2)/\{\xi_0\gamma(\theta_0)\}$. Thus with $\xi_0\gamma(\theta_0)$ at about 3 we would expect (4·1) to provide a useful measure of the sensitivity of the test.

4·2. *An optimality property*

We have not claimed that the test (2·1) has any optimality properties. Indeed it is clear from van Zwet & Oosterhoff (1967) that it will not, in general, be most stringent, this being the most natural optimality property to aim for. However, we can, following van Zwet & Oosterhoff, prove a rather weak optimality property. For Example 2 they show that as the significance level of the test (2·1) tends to zero its power function approaches that of the optimum test when θ_0 is given. A similar result which may be proved in a similar way holds in general.

THEOREM 4·2. *Suppose that the condition given in* § 2 *are satisfied and* $c(\alpha)$ *is the value of* c *required to give the right-hand side of* (3·7) *a value* α. *Then*

$$\lim_{\alpha\to 0} \sup_{\substack{\xi_0>0 \\ L\leqslant\theta_0\leqslant U}} |\text{pr}_{\xi_0,\theta_0}\{\sup Z(\theta) > c(\alpha)\} - \Phi\{\xi_0\gamma(\theta_0)+\Phi^{-1}(\alpha)\}| = 0. \tag{4·5}$$

4·3. *Confidence region for* θ_0

If the test indicates that $\xi_0 > 0$ we might wish to find a confidence interval for θ_0. If the problem in question is an asymptotic one and the test indicates a highly significant effect, then it is probably best to return to the original problem and use one of the standard asymptotic methods for finding a confidence interval. On the other hand, if the significance is only marginal or the exact likelihood difficult to calculate, one might prefer to find a rough confidence interval based on the asymptotic distribution assumed for $Z(.)$.

Since $Z(\theta_0)$ is sufficient for ξ_0 the dependence on ξ_0 can be removed if one conditions on $Z(\theta_0)$. Thus

$$E_{\xi_0,\theta_0}\{Z(\theta)|Z(\theta_0)=z\} = z\rho(\theta,\theta_0).$$

Suppose $c_\alpha(z,\theta_0)$ is defined by

$$\text{pr}_{\xi_0,\theta_0}\{\sup Z(\theta) > z + c_\alpha(z,\theta_0)|Z(\theta_0)=z\} = \alpha. \tag{4·6}$$

Then a $1-\alpha$ confidence region for θ_0 is given by

$$[\theta_0 : Z(\theta_0) > \sup Z(\theta) - c_\alpha\{Z(\theta_0),\theta_0\}].$$

The left-hand side of (4·6) can be written as

$$\text{pr}_{\xi_0,\theta_0}\left\{\sup\left[\frac{Z(\theta)-z\rho(\theta_1,\theta_0)}{\{1-\rho(\theta,\theta_0)^2\}^{\frac{1}{2}}} - \frac{z+c_\alpha(z,\theta_0)-z\rho(\theta,\theta_0)}{\{1-\rho(\theta,\theta_0)^2\}^{\frac{1}{2}}}\right] > 0|Z(\theta_0)=z\right\}, \tag{4·7}$$

where the first term has been normalized to give it zero conditional expectation and unit conditional variance. Expression (4·7) could then be bounded using a formula of the form (3·3) to find the expected number of upcrossings of zero in $[\theta_0, U]$ plus the expected number of downcrossings in $[L,\theta_0]$. Thus values of $c_\alpha(z,\theta_0)$ could be calculated to give a conservative $1-\alpha$ confidence interval. However, the formula is still too complex for a quick calculation.

Consider (4·7). If z is large, the second term will become large as $\rho(\theta,\theta_0)$ moves away from unity and hence only values of θ near θ_0 need be considered when evaluating the supremum. Over this range the first term can be considered approximately independent of θ, but with a sign change at θ_0. Suppose it is approximately equal to $Z_0 \operatorname{sgn}(\theta-\theta_0)$, where Z_0 is a standard normal random variable. Then (4·7) reduces to

$$\text{pr}\,[Z_0^2 > \{z+c_\alpha(z,\theta_0)\}^2 - z^2].$$

Thus we should choose $\{z+c_\alpha(z,\theta_0)\}^2 - z^2 = \chi^2_{1,\alpha}$ the upper α point of a chi-squared distribution with one degree of freedom. A similar result can be obtained by following an argument similar to that used in § 4·1.

The approximate confidence region for θ_0 then becomes

$$[\theta_0 : Z(\theta_0)^2 > \{\sup_\theta Z(\theta)\}^2 - \chi^2_{1,\alpha}], \tag{4·8}$$

a form which is convenient to calculate and not unexpected if one regards $\frac{1}{2}Z(\theta_0)^2$ as a log likelihood. For Example 1 this region reduces to the Fieller solution and is exact if values of θ which differ by π are identified; see, for example, James, Wilkinson & Venables (1974). However, it is not clear how big $Z(\theta_0)$ must be in general for (4·8) to be useable.

5. Practical matters

In order to carry out the test one will need to find

$$S = \sup_{L \leqslant \theta \leqslant U} Z(\theta), \tag{5·1}$$

$$I = \int_L^U \{-\rho_{11}(\theta)\}^{\frac{1}{2}} d\theta. \tag{5·2}$$

One can then calculate the approximation to the significance probability derived from (3·7):

$$\Phi(-S) + I \exp(-\tfrac{1}{2}S^2)/(2\pi). \tag{5·3}$$

One can also find the value of c in (3·1) to give a significance level of 0·05, say, and hence for each value of θ find the value of ξ_0 required to give a power of, for example, $\frac{1}{2}$. It will not be possible, in general, to evaluate (5·1) and (5·2) analytically. One possible approach is to calculate $Z(\theta), \rho_{11}(\theta)$ for $\theta = (L = \theta_1, \theta_2, \ldots, U = \theta_s)$, where for $n = 1, 2, \ldots$

$$\theta_{n+1} = \min[U, \theta_n + 0{\cdot}1\{\rho_{11}(\theta)\}^{-\frac{1}{2}}]$$

to give an approximate correlation of 0·993 between adjacent values of $Z(\theta_n)$. This was carried out for Example 4 and the maximum of the values of $Z(\theta_n)$ used for S and I evaluated by the trapezoidal rule. Between 10 and 20 values of θ_n were generally required and so the computational requirements were not excessive. In addition, when a significant result was obtained, the value of θ_n corresponding to the maximum provided a suitable starting point for the maximum likelihood estimation of ξ_0 and θ_0.

Finally we note that in most situations $Z(\theta)$ will be only asymptotically normally distributed and as the probabilities derived in this paper are likely to be rather sensitive to deviations from normality $Z(\theta)$ should be derived from especially large samples, particularly if the critical point c is large.

6. Generalizations

It has been assumed that the set of possible values of θ is the closed interval $[L, U]$. The theory can be applied for an open or infinite interval, provided that the integral (5·2) converges. However, when the theory is being applied to an asymptotic problem, one should check that the function $Z(\theta)$ cannot have spurious peaks. This can happen in quite reasonable problems in the same way that a likelihood function can have spurious peaks. In Example 3, for instance, a spurious peak will occur when $\theta = 1/\min(x_i)$ and so the range of values of θ should be bounded above.

If the alternative is two-sided, $\xi > 0$ or $\xi < 0$, the analogue of (2·1) is

$$\{\sup_{L \leqslant \theta \leqslant U} |Z(\theta)| > c\},$$

and the bound on the significance level is twice that given by (3·7). The estimate of the power given in § 4 still holds.

Suppose there are several parameters $\theta_1, \ldots, \theta_r$ present only under the alternative. The analogue $Z(\theta_1, \ldots, \theta_r)$ of $Z(\theta)$ can be defined and a test based on its supremum. A natural generalization of (3·3) is given by Adler & Hasofer (1976) for $r = 2$ or 3 for a stationary

Gaussian process and the result may be readily extended to a nonstationary process. Unfortunately one obtains an estimate of the significance level rather than a bound, although one would expect that it would be adequate for most purposes. For $r > 3$ the computation of the analogue of (5·2) would probably not be feasible.

Suppose that the nuisance parameter, θ_0, can take on only a finite number of values. For example, in Example 4, θ_0 may be known to take on a value from a finite set. Another non-asymptotic example is the change-point problem considered by Page (1957), Hinkley (1970) and others. One observes a sequence of independent normal random variables, X_i $(i = 1, \ldots, n)$, with unit variance, X_i having expectation μ if $i \leqslant \theta$ and $\mu + \xi$ otherwise. With

$$Z(\theta) = (n-\theta)^{-\frac{1}{2}} \sum_{i=\theta+1}^{n} (X_i - \mu)$$

if μ is known, otherwise

$$Z(\theta) = \left\{ \sum_{i=\theta+1}^{n} X_i/(n-\theta) - \sum_{i=1}^{\theta} X_i/\theta \right\} \{\theta(n-\theta)/n\}^{\frac{1}{2}},$$

it is easily checked that (1·3) and (1·4) are satisfied.

In general suppose that θ can take on the values $\{1, \ldots, m\}$ and (1·3) and (1·4) are satisfied. Then an analogue of (3·7) is

$$\mathrm{pr}_{0,\theta}\{\sup Z(\theta) > c\} \leqslant \Phi(-c) + \sum_{i=2}^{m} \mathrm{pr}_{0,\theta}\{Z(i-1) < c \text{ and } Z(i) > c\}. \tag{6·1}$$

The bivariate probabilities in (6·1) can be evaluated from tables or by standard computer subroutines. The results of § 4 hold after the obvious changes have been made.

References

ADLER, R. J. & HASOFER, A. M. (1976). Level crossings for random fields. *Ann. Prob.* **4**. 1–12.

CRAMÉR, H. & LEADBETTER, M. R. (1967). *Stationary and Related Stochastic Processes*. New York: Wiley.

DAVIES, R. B. (1977). Testing the hypothesis that a point process is Poisson. *Adv. Appl. Prob.* **9**. To appear.

HÁJEK, J. & ŠIDÁK, Z. (1967). *Theory of Rank Tests*. New York: Academic Press.

HINKLEY, D. V. (1970). Inference about a change-point in a sequence of random variables. *Biometrika* **57**, 1–17.

JAMES, A. T., WILKINSON, G. N. & VENABLES, W. N. (1974). Interval estimates for a ratio of means. *Sankhyā* A **36**, 177–83.

LEADBETTER, M. R. (1972). Point processes generated by level crossings. In *Stochastic Point Processes*, Ed. P. A. W. Lewis, pp. 436–67. New York: Wiley.

MORAN, P. A. P. (1970). On asymptotically optimal tests of composite hypotheses. *Biometrika* **57**, 47–55.

NEYMAN, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In *Probability and Statistics: The Harald Cramér Volume*, Ed. U. Grenander, pp. 213–34. New York: Wiley.

OOSTERHOFF, J. (1969). *Combination of One-sided Statistical Tests*, Mathematical Centre Tracts, 28. Amsterdam: Mathematisch Centrum.

PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika* **44**, 248–52.

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* **24**, 220–38.

ROY, S. N., GNANADESIKAN, R. & SRIVASTAVA, J. N. (1971). *Analysis and Design of Certain Quantitative Multiresponse Experiments*. Oxford: Pergamon Press.

VAN EEDEN, C. (1963). The relation between Pitman's asymptotic relative efficiency of two tests and the correlation coefficient between their test statistics. *Ann. Math. Statist.* **34**, 1442–51.

VAN ZWET, W. R. & OOSTERHOFF, J. (1967). On the combination of independent test statistics. *Ann. Math. Statist.* **38**, 659–80.

[*Received May* 1976. *Revised December* 1976]