Z. Wahrscheinlichkeitstheorie verw. Geb. 14, 161–168 (1969)

Asymptotically Minimax Tests of Composite Hypotheses

L. Weiss* and J. Wolfowitz**

Summary. $X_1, \dots, X_n$ are independent, identically distributed random variables with common density function $f(x|\theta_1, \dots, \theta_k, \theta_{k+1})$, assumed to satisfy certain standard regularity conditions. The $k+1$ parameters are unknown, and the problem is to test the hypothesis that $\theta_{k+1}=b$ against the alternative that $\theta_{k+1}=b+c\,n^{-\frac{1}{2}}$. $\theta_1, \dots, \theta_k$ are nuisance parameters. For this problem, the following artificial problem is temporarily substituted. It is known that $|\theta_i - a_i| \leqq n^{-\frac{1}{2}} M(n)$ for $i=1, \dots, k$, where $a_1, \dots, a_k$ are known, and $M(n)$ approaches infinity as n increases but $n^{-\frac{1}{2}} M(n)$ approaches zero as n increases. A Bayes decision rule is constructed for this artificial problem, relative to the a priori distribution which assigns weight A to $\theta_{k+1}=b$, and weight $1-A$ to $\theta_{k+1}=b+c\,n^{-\frac{1}{2}}$, in each case the weight being spread uniformly over the possible values of $\theta_1, \dots, \theta_k$ in the artificial problem. An analysis of the structure of the Bayes rule shows that if estimates of $\theta_1, \dots, \theta_k$ are substituted for $a_1, \dots, a_k$ respectively, the resulting rule is a solution to the original problem, and this rule has the same asymptotic properties as a solution to the artificial problem as the Bayes rule for the artificial problem, no matter what the values $a_1, \dots, a_k$ are.

1. Introduction

In several papers ([1, 2], and the papers cited therein) we have developed a general theory of asymptotically efficient estimators which solves hitherto unsolved problems and includes the maximum likelihood theory as a special case. In the present paper we apply the basic method, which was so successful in the case of estimators, to a study of asymptotic tests of composite hypotheses. Of course, objections made to the theory of testing hypotheses on practical grounds (e.g., [4]) are as valid in the asymptotic case as in the finite sample case. However, the present application can once again serve as an illustration of the power and simplicity of our method. It may also suggest other applications to the reader. This paper is self-contained and no prior familiarity with [1] or [2] is required for its comprehension.

2. A Preliminary Artificial Problem

Let $X_1, \dots, X_n$ be independent, identically distributed chance variables, with the common density function $f(\cdot|\theta^0)$. For simplicity we assume that X_i is one-dimensional, since the extension of our results to the case where X_i is of any finite dimension is immediate. The parameter $\theta^0=(\theta^0_*, \theta^0_{k+1})$, where θ^0_* is the k-vector $(\theta^0_1, \dots, \theta^0_k)$ and $k \geqq 1$. Let $a=(a_1, \dots, a_k)$ be a given k-vector, b a given constant, and c a given positive constant. We shall first consider the following artificial problem: Let H_0 be the (null) hypothesis that $\theta^0_{k+1}=b$, and H_1 the (alternative) hypothesis that $\theta^0_{k+1}=b+c\,n^{-\frac{1}{2}}$. The statistician does not know θ^0_* but does know

* Research supported by NSF Grant GP 7798.

** Research supported by the U.S. Air Force under Grant AF-AFOSR-68-1472.

that

(2.1) $$|\theta_i^0 - a_i| \leqq n^{-\frac{1}{2}} M(n), \qquad i=1, \ldots, k, \text{ and } \theta_{k+1}^0 = b \text{ or } b + c\, n^{-\frac{1}{2}},$$

and wishes to test H_0 against H_1. Here $M(n)$ is a positive function of n such that $M(n) \to \infty$, $n^{-\frac{1}{2}} M(n) \to 0$. Of course, the relation (2.1) implies that θ^0 may depend upon n. We do not exhibit this *now* explicitly in our notation because it would simply complicate things unnecessarily.

What makes this problem more artificial than the usual problem of testing hypotheses is that it is being assumed that the statistician knows a. This assumption will be removed in Section 4.

We make the following "regular" assumptions, the last two of which are themselves the consequences of assumptions standard in the literature for the so-called "regular" case:

(2.2) The second derivatives

$$\frac{\partial^2 \log f(x|\theta)}{\partial \theta_i \, \partial \theta_j} \qquad i, j = 1, \ldots, (k+1)$$

exist and are uniformly continuous in θ for all real x and all θ in some neighborhood of $\bar{\theta} = (a, b)$.

(2.3) The expressions

$$J(i, j | \theta^0) = -E \left\{ \frac{\partial^2 \log f(X_1 | \theta^0)}{\partial \theta_i \, \partial \theta_j} \right\}$$

exist and are continuous in θ^0 in a neighborhood of $\bar{\theta}$, for $i, j = 1, \ldots, (k+1)$. Write $J(i, j) = J(i, j | \bar{\theta})$. The $(k \times k)$ matrix

$$J = \{J(i, j)\} \qquad i, j = 1, \ldots, k$$

and the $((k+1) \times (k+1))$ matrix

$$J^* = \{J(i, j)\} \qquad i, j = 1, \ldots, (k+1)$$

are non-singular.

(2.4) The joint distribution of the chance variables

$$n^{-\frac{1}{2}} \sum_{l=1}^{n} \frac{\partial \log f(X_l | \theta^0)}{\partial \theta_i}, \qquad i = 1, \ldots, (k+1)$$

approaches the normal distribution, with means zero and covariance matrix J^*, uniformly for θ^0 which satisfy (2.1) (i. e., uniformly for any sequence $\theta^0(n)$, $n = 1, 2, \ldots$, such that, for each n, $\theta^0(n)$ satisfies (2.1)).

(2.5) The chance variables

$$-n^{-1} \sum_{l=1}^{n} \frac{\partial^2 \log f(X_l | \theta^0)}{\partial \theta_i \, \partial \theta_j}, \qquad i, j = 1, \ldots, (k+1),$$

converge stochastically to $J(i, j)$, uniformly for θ^0 which satisfy (2.1).

We shall employ the well-known notation o_p, O_p. For example, $o_p(t_i t_j)$ means the following: Let $0<\varepsilon<1$ and $\delta>0$ be arbitrary. Then, for n sufficiently large,

$$P\{|(t_i t_j)^{-1} o_p(t_i t_j)|<\delta\}>1-\varepsilon.$$

The symbols $\bar{o}_p$ and $\bar{O}_p$ will mean that the relationship in question holds *uniformly* for θ^0 which satisfy (2.1). An accent over a vector or a matrix will denote its transpose.

Write $t_i=\sqrt{n}(\theta_i^0-a_i)$, $i=1,\ldots,k$, and let t be the (row) vector $(t_1,\ldots,t_k)$. Let A be the vector whose i^{th} element $(i=1,\ldots,k)$ is the chance variable

$$A(i)=n^{-\frac{1}{2}}\sum_{l=1}^{n}\frac{\partial \log f(X_l\,|\,\bar{\theta})}{\partial\theta_i}. \tag{2.6}$$

For n sufficiently large we have, when (2.1) holds,

$$\begin{aligned}&\sum_1^n \log f(X_i\,|\,\theta_*^0, b+c\,n^{-\frac{1}{2}})\\ &\quad=\sum_1^n \log f(X_i\,|\,\bar{\theta})+t A'-\tfrac{1}{2}t J t'+\sum_{i,j}\bar{o}_p(t_i t_j)\\ &\quad\quad+c A(k+1)-\frac{c^2}{2}J(k+1,k+1)-c\sum_i t_i J(i,k+1).\end{aligned} \tag{2.7}$$

We note that (2.7) also holds if one sets $c=0$ in both members.

Suppose temporarily that the t_i $(i=1,\ldots,k)$ are independent chance variables with the uniform distribution over the interval $[-M(n),M(n)]$, and that $P\{H_0\}=\alpha$, $P\{H_1\}=1-\alpha$. When this is so the solution of this Bayes problem is as follows: Let $S=S_1/S_2$ be the statistic defined by

$$S_1=\int\prod_1^n f(X_i\,|\,y_1,\ldots,y_k,b+c\,n^{-\frac{1}{2}})\,dy_1\ldots dy_k, \tag{2.8}$$

$$S_2=\int\prod_1^n f(X_i\,|\,y_1,\ldots,y_k,b)\,dy_1\ldots dy_k, \tag{2.9}$$

the limits of integration in both cases being from $a_i-n^{-\frac{1}{2}}M(n)$ to $a_i+n^{-\frac{1}{2}}M(n)$ for all variables. When $S>\frac{\alpha}{1-\alpha}$ the statistician accepts H_1, when $S<\frac{\alpha}{1-\alpha}$ he accepts H_0, and when $S=\frac{\alpha}{1-\alpha}$ he takes either action at pleasure.

In (2.8) and (2.9) we perform the transformation $z_i=\sqrt{n}(y_i-a_i)$. Let $z=(z_1,\ldots,z_k)$. Making use of (2.7), we cancel all factors which do not involve the z_i; they will not appear in S because they are the same in S_1 and S_2. Calling the resulting expressions S_1^* and S_2^*, we have that

$$S_2^*=\int\exp\{zA'-\tfrac{1}{2}zJz'+\sum_{i,j}\bar{o}_p(z_i z_j)\}\,dz_1\ldots dz_k, \tag{2.10}$$

$$\begin{aligned}S_1^*=\int\exp\Big\{&zA'-\tfrac{1}{2}zJz'+\sum_{i,j}\bar{o}_p(z_i z_j)\\ &+cA(k+1)-\frac{c^2}{2}J(k+1,k+1)-c\,zJ'(k+1)\Big\}\,dz_1\ldots dz_k,\end{aligned} \tag{2.11}$$

where $J(k+1)$ is the k-vector with i^{th} element $J(i, k+1)$, and all the limits of integration are from $-M(n)$ to $M(n)$. Completing the square we have that

$$S_2^* = \exp\left\{\frac{A J^{-1} A'}{2}\right\} \cdot \int \exp\{-\tfrac{1}{2}(z - A J^{-1}) J (z' - J^{-1} A') + \sum \bar{o}_p(z_i z_j)\} \, dz_1 \ldots dz_k, \tag{2.12}$$

$$\begin{aligned} S_1^* = \exp\Big\{ & c A(k+1) - \frac{c^2}{2} J(k+1, k+1) \\ & + \tfrac{1}{2}[A - c J(k+1)] J^{-1} [A' - c J'(k+1)] \Big\} \\ & \cdot \int \exp\{-\tfrac{1}{2}[z - (A - c J(k+1)) J^{-1}] J [z - (A - c J(k+1)) J^{-1}]' \\ & + \sum \bar{o}_p(z_i z_j)\} \, dz_1 \ldots dz_k \end{aligned} \tag{2.13}$$

with the same limits of integration. Since $M(n) \to \infty$ we would like to conclude that the integrals in (2.12) and (2.13) approach the same limit, $(2\pi)^{k/2} |J|^{-\frac{1}{2}}$, whether H_0 or H_1 holds. Then the limiting distribution of the statistic $\log S$ would be the same as that of

$$\begin{aligned} S^* = & \frac{c^2}{2} [J(k+1) J^{-1} J'(k+1) - J(k+1, k+1)] \\ & + c[A(k+1) - J(k+1) J^{-1} A']. \end{aligned} \tag{2.14}$$

The first of these terms is, incidentally, a constant, say $\dfrac{-c^2 K}{2}$.

3. Asymptotic Distribution of S^*

The integrals in (2.12) and (2.13) are functions of A and hence chance variables. We will prove that, for sufficiently large n, with any probability less than one, each of the integrals is within an arbitrary positive ε of their common limit $(2\pi)^{k/2} |J|^{-\frac{1}{2}}$, provided that, for $1 \leqq i \leqq k$,

$$|\theta_i^0 - a_i| \leqq n^{-\frac{1}{2}} L(n), \tag{3.1}$$

where $0 < L(n) < M(n)$ will be defined suitably below, and will also satisfy

$$\frac{L(n)}{M(n)} \to 1. \tag{3.2}$$

We leave it to the reader to confirm the easily verifiable fact that the $\bar{o}_p(z_i z_j)$ terms will not change the limit, and delete them henceforth. Our result will follow if we prove that, with probability approaching one,

$$M(n) - \max_i |i^{\text{th}} \text{ component of } A J^{-1}| \to \infty \tag{3.3}$$

when (3.1) holds. We have, for $1 \leqq i \leqq k+1$,

$$A(i) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{\partial \log f(X_j | \theta_*^0, b + c\, n^{-\frac{1}{2}})}{\partial \theta_i} - \sum_{l=1}^{k} t_l (1 + \bar{o}_p(1)) J(i, l) + c J(i, k+1), \tag{3.4}$$

when H_1 holds. Call the first term of the right member $C(i)$. By (2.4) the $C(i)$'s are $\bar{O}_p(1)$. The contribution to $A J^{-1}$ of the third term of (3.4) for $1 \leqq i \leqq k$ is $O(1)$, and of the second term is at most

$$\sqrt{n}(1 + \bar{o}_p(1)) \max_l |a_l - \theta_l^0| \leqq (1 + \bar{o}_p(1)) L(n) \tag{3.5}$$

since (3.1) holds. Let $\varepsilon_m \downarrow 0$ as $m \to \infty$ and $\gamma(m, n')$ be an upper bound on the largest of the k quantities $\bar{o}_p(1)$ for sample size $\geqq n'$ and probability $1 - \varepsilon_m$, so that, for fixed m, $\gamma(m, n') \to 0$ as $n' \to \infty$; its existence is the meaning of $\bar{o}_p(1)$. We now define $\varepsilon(n)$ as follows: $\varepsilon(n) = \gamma(1, 1)$ for $n \geqq 1$ until $n = n_1'$, where n_1' is the first integer such that $\gamma(2, n_1') < \frac{1}{2}\gamma(1, 1)$. Thereafter $\varepsilon(n) = \gamma(2, n_1')$ until the first integer n_2' such that $\gamma(3, n_2') < \frac{1}{2}\gamma(2, n_1')$, etc. We now choose $L(n)$ such that (3.2) holds and so that

$$M(n) - (1 + \varepsilon(n)) L(n) \to \infty. \tag{3.6}$$

This is surely possible (eg., $L(n) = (1 + \varepsilon(n))^{-1} [M(n) - \sqrt{M(n)}]$). From (3.6) and (3.5) we obtain (3.3), thus proving the desired result when H_1 holds. The proof when H_0 holds is essentially the same; one sets $c = 0$. The relation (3.2) was not used and will only be used after Theorem 1 is proved.

An examination of the preceding proof shows easily the following fact of importance later: The approach of $\log S$ to S^* is *uniform* in the probability sense for all θ^0 which satisfy (3.1).

We now consider the asymptotic distribution of

$$A(k+1) - J(k+1) J^{-1} A' \tag{3.7}$$

when H_1 holds. Let C be the vector $C(1), \ldots, C(k)$. From (3.4) we obtain that the expression in (3.7) has (uniformly) the same asymptotic distribution as the expression

$$\begin{aligned} & C(k+1) - t J'(k+1) + c J(k+1, k+1) \\ & \quad - J(k+1) J^{-1} (C - tJ + c J(k+1))' \\ & = C(k+1) - t J'(k+1) + c J(k+1, k+1) \\ & \quad - J(k+1) J^{-1} C' + J(k+1) t' - c J(k+1) J^{-1} J'(k+1) \\ & = C(k+1) - C J^{-1} J'(k+1) - c [J(k+1) J^{-1} J'(k+1) - J(k+1, k+1)]. \end{aligned} \tag{3.8}$$

Now the asymptotic distribution of $C(1), \ldots, C(k+1)$ is given in (2.4). Hence we have proved part of the following theorem:

Theorem 1. *The asymptotic distribution of S^* is normal with the same variance $c^2 K$, where*

$$K = J(k+1, k+1) - J(k+1) J^{-1} J'(k+1),$$

when either H_0 or H_1 holds. The mean of the asymptotic distribution is $-(c^2/2)K$ *when H_0 holds and* $(c^2/2)K$ *when H_1 holds. When* (3.1) *also holds the distribution of S^* approaches, uniformly in θ^0, the same limit as that approached, also uniformly in θ^0, by the distribution of the test statistic* $\log S$.

It remains to prove the statements in Theorem 1 which apply when H_0 holds. Then we have, for $1 \leqq i \leqq k+1$,

$$A(i) = n^{-\frac{1}{2}} \sum_{j=1}^{n} \frac{\partial \log f(X_j | \theta_*^0, b)}{\partial \theta_i} - \sum_{l=1}^{k} t_l (1 + \bar{o}_p(1)) J(i, l). \tag{3.9}$$

Write $D(i)$ for the first term of the right member of (3.9) and D for the vector $(D(1), \dots D(k))$. Then the expression (3.7) has the same limit distribution, when H_0 holds, as

$$\begin{aligned} & D(k+1) - t J'(k+1) - J(k+1) J^{-1} (D - tJ)' \\ & \quad = D(k+1) - D J^{-1} J'(k+1). \end{aligned} \tag{3.10}$$

The proof of Theorem 1 is easily completed by using (2.4).

The asymptotic test would break down if $K=0$. This cannot happen since J^* is non-singular.

For the sake of clarity we now write the sequence $\{\theta^0(n), n=1,2,\dots\}$ such that $\theta^0(n)$ satisfies (3.1), and either H_0 or H_1 always holds. Let Z_0 and Z_1 be, respectively, the totalities of these two kinds of sequences. By proper choice of α we can achieve that, for any sequence in Z_0 we have

$$\lim P\{H_0 \text{ is rejected by the } S \text{ test}\} = \beta, \text{ say}, \tag{3.11}$$

uniformly in Z_0, where β is any given number such that $0 < \beta < 1$. Let γ then be the limit, uniformly in Z_1, of

$$P\{H_1 \text{ is rejected by the } S \text{ test}\}. \tag{3.12}$$

Since the test based on S is the particular Bayes test that it is, we have proved

Theorem 2. *Let* $T(1)$, $T(2)$, … *be any sequence of tests for which*

$$\lim_{n \to \infty} P\{H_i \textit{ is rejected by } T(n)\} = \varphi_i(\{\theta^0(n)\}), \tag{3.13}$$

where the limits, which are functions of the sequence $\{\theta^0(n)\}$, are approached uniformly in Z_i, $i=0,1$. Then

$$\max_{Z_0} \varphi_0 \leqq \beta \tag{3.14}$$

implies

$$\max_{Z_1} \varphi_1 \geqq \gamma. \tag{3.15}$$

In a different context, that of estimators, one of us ([3]) has argued that the approach, to its limit, of the distribution of an estimator in its asymptotic application, should be uniform. The same arguments apply here. Hence, in a very reasonable sense, the test based on S^* is asymptotically minimax for all θ^0 which satisfy (3.1).

4. Removal of the Artificial Assumption. Various Remarks

The artificial assumption we made was that the statistician knew a and hence J^*. Suppose that the statistician can estimate J^* and the components of a to within $O_p(n^{-\frac{1}{2}})$, when either H_0 or H_1 holds. This will surely be so in the so-called "regular" case (the one usually treated in the literature), when these quantities can be estimated by maximum likelihood estimators. Instead of the statistic S^* we employ the statistic S^{**}, obtained from S^* by replacing J^* and a by their estimators. Since (3.8) and (3.10) are independent of t, the asymptotic distribution of S^{**} is the same as that of S^*, and is approached uniformly for θ^0 which satisfy (2.1) when either H_0 or H_1 holds.

The statistic S^* was obtained by Neyman [5] by a very ingenious and completely different argument. The latter limits himself to statistics of the special form

$$g(X_1, \ldots, X_n) - VA', \tag{4.1}$$

where A is as above, V is a k-vector and g a function of the arguments exhibited, and g and V are to satisfy certain conditions which are essentially such that, when the unknown parameters are replaced by estimators to within $O_p(n^{-\frac{1}{2}})$, the asymptotic distribution of (4.1) will be unchanged. He then concludes that an optimal choice among g and V which satisfy these conditions is $g = A(k+1)$, $V = J(k+1)J^{-1}$. (We note that $J(k+1)J^{-1}A'$ is the regression of $A(k+1)$ on A.) Thus Neyman was interested in a special class of asymptotically *similar* tests, and Le Cam ([6]) proved that, under certain conditions, the test based on S^{**} is optimal in a larger class of asymptotically similar tests.

Our own approach is completely different. We do not restrict ourselves to a special class of tests, but proceed at once to the Bayes problem. The argument in Section 3 proves that the asymptotic distribution of the test criterion (under (3.1)) does not depend on θ^0_*. The same argument proves that S^{**} has the same distribution as S^*. Since size and power are asymptotically constant the fact that the statistic S^* was obtained as the solution of a Bayesian problem implies that the test is asymptotically minimax.

Bartoo and Puri ([7]) extended Neyman's results to nonidentically distributed X's, and Buhler and Puri [8] to the case where θ_{k+1} is multidimensional. Our method applies to these cases, as well as to some cases where the X's are not independent.

References

1. Weiss, L., Wolfowitz, J.: Maximum probability estimators. Ann. Inst. statist. Math. **19**, 193–206 (1967).
2. — — Maximum probability estimators with a general loss function Proceedings of the International Symposium on Probability and Information Theory, held at McMaster University, Hamilton, Ontario, Canada, April 4 and 5, 1968, 232–256. Lecture Notes in Mathematics, No. 89. Berlin-Heidelberg-New York: Springer 1969.

3. Wolfowitz, J.: Asymptotic efficiency of the maximum likelihood estimator. Teor. Verojatn. Primen. **10**, 267 – 281 (1965).
4. – Remarks on the theory of testing hypotheses. The New York Statistician **18**, No. 7, 1 – 3 (1967).
5. Neyman, J.: Optimal tests of composite statistical hypotheses. The Harald Cramér volume, edited by U. Grenander, p. 213 – 234. Stockholm: Almquist and Wiksell 1959.
6. LeCam, L.: On the asymptotic theory of estimation and testing hypotheses. Proc. Third Berkeley Sympos. math. Statist. Probab. **1**, 129 – 156 (1956).
7. Bartoo, J. B., Puri, P. S.: On optimal asymptotic tests of composite hypotheses. Ann. math. Statistics **38**, 1845 – 1852 (1967).
8. Buhler, W. J., Puri, P. S.: On optimal asymptotic tests of composite hypotheses with several constraints. Z. Wahrscheinlichkeitstheorie verw. Geb. **5**, 71 – 88 (1966).

Professor L. Weiss
Department of Operations Research
Cornell University
Upson Hall
Ithaca, N. Y., USA

Professor J. Wolfowitz
Department of Mathematics
Cornell University
Ithaca, N. Y., USA

(Received November 18, 1968)