APPROXIMATELY OPTIMAL PROPERTIES OF WALD TESTS AND THE PROBLEM OF TESTING STATISTICAL HYPOTHESES

MATHEMATICS

1969

SovietRxiv

Abstract

Full Text

UDC 519.24/27

MATHEMATICS

LI HOANG TU

APPROXIMATELY OPTIMAL PROPERTIES OF WALD TESTS AND THE PROBLEM OF TESTING STATISTICAL HYPOTHESES

(Presented by Academician Yu. V. Linnik on 7 IV 1969)

The content of the note is adjacent to the work of A. Wald ([1]), where a number of approximately optimal properties of tests based on the maximum likelihood estimate were proved; in our work we shall call these Wald tests. The accuracy in Wald' s theorems is estimated; it is shown that it has order $O(N^{-1/2+\varepsilon})$, where N is the sample size and ε is any small positive number.

I. Let X_n $(n = 1, \dots, N)$ be independent identically distributed P-measure random vectors having density $f(X, \theta)$, which satisfies the following conditions:

A. θ belongs to a nondegenerate closed interval Ω of the Euclidean space E_K, and third-order partial derivatives with respect to θ_i $(i = 1, \dots, K)$, $\theta^T = (\theta_1, \dots, \theta_K)$, exist almost everywhere on the space X for all θ. (Here and below the indices i, j, l run through the set of numbers $1, \dots, K$.)

B. $|\partial f(X, \theta)/\partial\theta_i| < F_1(X)$, $|\partial^2 f(X, \theta)/\partial\theta_i\partial\theta_j| < F_2(X)$, where F_1 and F_2 are integrable on $(-\infty, +\infty)$.

C. Almost everywhere on the space X there exist $\partial^2 \ln f(X, \theta)/\partial\theta_i\partial\theta_j$ and $\partial^3 \ln f(X, \theta)/\partial\theta_i\partial\theta_j\partial\theta_l$, and $\| -E_\theta\partial^2 \ln f(X_n, \theta)/\partial\theta_i\partial\theta_j\|$ is a positive definite matrix with absolute values of the determinant (for all $\theta \in \Omega$) not less than K_2 (here and below $K_i, m_i,\ i = 0, 1, 2, \dots,$ denote fixed positive numbers).

D. For fixed $\theta^0 \in \Omega$ and for any $\theta \in \Omega$

$$|\theta - \theta^0| < \frac{\ln^{m_0} N}{\sqrt{N}} \left(|\theta - \theta^0|^2 = \sum_{i=1}^{K} |\theta_i - \theta_i^0|^2 \right) :$$

a) There exists a number a $(0 < a < 1/6)$ such that

$$E_\theta \exp \left| \frac{\partial \ln f(X_n, \theta)}{\partial\theta_i} \right|^{4a/(2a+1)} < K_3.$$

b) There exists a number N_0 such that for $N = N_0$ the random vector

$$\left(\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{\partial \ln f(X_n, \theta)}{\partial \theta_i} \right)$$

has density $p_{0N_0}(y)$ and $\sup_{y \in E_K} |p_{0N_0}(y)| < K_4$.

E. $E_{\theta_1} \left| \dfrac{\partial^3 \ln f(X_n, \theta^2)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right|^2$ and

$$\int_{-\infty}^{+\infty} \cdots \int \left| \frac{\partial^2 \ln f(X, \theta^1)}{\partial \theta_i \partial \theta_j} \frac{\partial f(X, \theta^2)}{\partial \theta_l} \right| dX_1 \cdots dX_p$$

are bounded for $\theta^1, \theta^2 \in \Omega$ and $|\theta^1 - \theta^2| < \ln^{m_2} N / \sqrt{N}$.

II. It is proved that there exists a maximum likelihood estimate $\hat{\theta}^N(X)$ for θ and, as $N \to \infty$,

$$P_\theta\{\sqrt{N}\, |\hat{\theta}^N(x) - \theta| > \ln^4 N\} = O\left(N^{-\sqrt{\ln N}}\right). \tag{1}$$

Consider the system of equations

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{\partial \ln f(X_n, \theta)}{\partial \theta_i} + \sum_{j=1}^{K} \sqrt{N}\, (\dot{\theta}_j^N(X, \theta) - \theta_j)\, E_\theta \frac{\partial^2 \ln f(X_n, \theta)}{\partial \theta_i \partial \theta_j} = 0 \tag{2}$$

$$(i = 1, \dots, K);$$

in C the system has a solution $\dot{\theta}^N(X, \theta)$, which we shall call an auxiliary function for the maximum-likelihood estimate. It can be verified that

$$\sup_{z_0 \in E_K} |p_{\theta N}(z_0) - q_\theta(z_0)| = O\left(N^{-1/2+\varepsilon}\right), \tag{3}$$

where $p_{\theta N}(z_0)$ is the density ($z_0 = \sqrt{N}\, \dot{\theta}^N(X, \theta^0)$) under the true θ; $q_\theta(z_0)$ is the density of the normal law $N(\sqrt{N}\theta,\ \|C_{ij}(\theta^0)\|^{-1})$;

$$|\theta - \theta^0| < \ln^{m_0} N / \sqrt{N}, \qquad C_{ij}(\theta^0) = -E_\theta \partial^2 \ln f(X_n \theta^0) / \partial \theta_i \theta_j.$$

For the subsequent arguments we shall need the equality

$$P_\theta\left\{\left|\frac{1}{N}\sum_{n=1}^{N}\frac{\partial^2\ln f(X_n,\theta)}{\partial\theta_i\partial\theta_j}-E_\theta\frac{\partial^2\ln f(X_n,\theta)}{\partial\theta_i\partial\theta_j}\right|>\frac{\ln^5 N}{\sqrt{N}}\right\}=O\left(N^{-1}\sqrt{\ln N}\right). \tag{4}$$

Assumption G. For a given $\varepsilon > 0$ there exists a number N_ε such that, for $N > N_\varepsilon$, relations (1), (3), (4) hold for all $\theta, \theta^0 \in \Omega$ with

$$|\theta-\theta^0|<\ln^m N/\sqrt{N},$$

where m is a fixed positive number.

III. We first consider a linear hypothesis. Here θ is written in the form

$$\theta=(\theta_1,\dots,\theta_K)^T=({}_1\theta,{}_2\theta)^T,$$

where

$${}_1\theta^T=(\theta_1,\dots,\theta_r),\qquad r<K.$$

As in (1), the problem of testing the linear hypothesis is posed in the form

$$H_1:\ {}_1\theta={}_1\theta^0,$$

where ${}_1\theta^0$ is known ($\omega=\{\theta:\ {}_1\theta={}_1\theta^0\}$), against the alternative

$$K_1:\ {}_1\theta\neq{}_1\theta^0.$$

Let $\underline{\theta}$ be a fixed point belonging to ω; C a constant > 0, and $S_C(\underline{\theta},\theta)$ the surface

$$({}_1\theta-{}_1\theta^0)^T\|\bar{C}_{ij}(\theta)\|({}_1\theta-{}_1\theta^0)=C,\qquad \gamma(\theta)(\theta-\underline{\theta})=0, \tag{5}$$

where

$$\|\bar{C}_{ij}(\theta)\|=\|\sigma_{ij}(\theta)\|^{-1}\quad(i,j=1,\dots,r),$$
$$\|\sigma_{ij}(\theta)\|=\|C_{ij}(\theta)\|^{-1}\quad(i,j=1,\dots,K);$$

$\|C_{ij}(\theta)\|$ is the information matrix defined in C, and $\beta(\theta)$ and $\gamma(\theta)$ are $r\times r$ and $(K-r)\times K$ matrices, respectively, such that

$$\left\|\begin{matrix}\beta(\theta)\ 0\\ \gamma(\theta)\end{matrix}\right\|^T\cdot\|C_{ij}(\theta)\|\cdot\left\|\begin{matrix}\beta(\theta)\ 0\\ \gamma(\theta)\end{matrix}\right\|=E;$$

E is the identity matrix. Transform $S_C(\theta,\underline{\theta})$ into the sphere

$$({}_1\theta' - {}_1\theta^0)^T({}_1\theta' - {}_1\theta^0) = C, \qquad {}_2\theta' = {}_2\theta'.$$

Let

$$\eta(\theta) = \lim_{\rho\to 0} A(\omega'(\theta,\rho))/A(\omega(\theta,\rho)), \tag{6}$$

where

$$\omega(\theta,\rho) = \{\bar\theta : \bar\theta \in S_C(\underline{\theta},\theta),\ |\bar\theta - \theta| < \rho\},$$

ω' is the image of ω;

$$A(\omega) = \int_\omega dA.$$

Theorem 1. The Wald test W'_N

$$N(\hat\theta^N(X) - {}^1\theta^0)^T \left\|\bar C_{ij}(\hat\theta^N(X))\right\| ({}^1\hat\theta^N(X) - {}^1\theta^0) > d_N \tag{7}$$

with level α (among all nonrandomized tests) as $N\to\infty$:

1) has approximately best average power on the family of surfaces $\{S_C(\theta,\underline{\theta})\}$ (5) with weight $\eta(\theta)$ (6), with accuracy up to $O(N^{-1/2+\varepsilon})$,

$$\sup_{S_C(\theta,\underline{\theta}),\,Z_N} \left\{ \int_{S_C(\theta,\underline{\theta})} E_\theta Z_N \frac{\eta(\theta)}{AS_C(\theta,\underline{\theta})}\, dA - \int_{S_C(\theta,\underline{\theta})} E_\theta W'_N \frac{\eta(\theta)}{AS_C(\theta,\underline{\theta})}\, dA \right\} = O(N^{-1/2+\varepsilon}), \tag{8}$$

where Z_N is any test of level α, $AS_C = \int_{S_C} \eta(\theta) dA$.

2) has approximately best constant power on the family $\{S_C(\theta,\underline{\theta})\}$ with accuracy up to $O(N^{-1/2+\varepsilon})$;

3) is an approximately most stringent test with accuracy up to $O(N^{-1/2+\varepsilon})$;

4) is an approximately minimax test with accuracy up to $O(N^{-1/2+\varepsilon})$.

The theorem is proved with the aid of an auxiliary hypothesis on the space $\hat\theta^N(X,\theta^0)$.

If the density of the distribution belongs to the normal law $N(\xi,\Sigma)$, rewrite it in the form $f(X,\theta)$, where θ is a parameter and $\theta^T = (\xi_1,\dots,\xi_p,\sigma_{11},\dots,\sigma_{pp})$; we again obtain the assertion of Theorem 1—among all randomized tests. As a consequence, the approximately optimal properties of the test T^2 hold, and by the method of proof of the theorem formulated above the approximately optimal properties also hold for the test R^2 (under certain restrictions on Ω); these results agree with (5, 7).

IV. We shall consider a general composite hypothesis, where the region ω is given by the equations

$$\xi_1(\theta) = \dots = \xi_r(\theta) = 0, \qquad r < K. \tag{9}$$

Under certain conditions there exist $K-r$ functions $\xi_{r+1}(\theta), \dots, \xi_K(\theta)$, and θ is a function of $(\xi_1, \dots, \xi_K)$.

Let

$$f^*(X,\xi) = f(X,\theta(\xi))$$

satisfy assumptions A, B, C, D, E, F with respect to ξ.

Define

$$C^*_{ij}(\theta) = -E_\xi\, \partial^2 \ln f^*(X_n,\xi)/\partial\xi_i\partial\xi_j. \tag{10}$$

Let $\theta \in \omega$ and let C be a fixed number; construct the surface $S^*_C(\theta, \underline{\theta})$ and define the weight $\eta^*(\theta)$ with the help of $C^*_{ij}(\theta)$ as in (5).

Theorem 2. For testing the hypothesis $H_2 : \theta \in \omega$ against the alternative $K_2 : \theta \in \Omega - \omega$ with fixed level α. Among all nonrandomized tests the Wald test W^2_N

$$N\left({}^1\xi(\hat\theta^N(X))\right)^T \left\|\bar{C}^*_{ij}(\hat\theta^N(X))\right\| \left({}^1\xi(\hat\theta^N(X))\right) > d_N, \tag{11}$$

where $\left({}^1\xi(\theta)\right)^T = (\xi_1(\theta), \dots, \xi_r(\theta))$, d_N is determined by the level α:

1) has approximately best average power on the family of surfaces $\{S^*_C(\theta, \underline{\theta})\}$ with weight $\eta^*(\theta)$, with accuracy up to $O(N^{-1/2+\varepsilon})$;
2) has approximately best constant power on the family $\{S^*_C(\theta, \theta)\}$ with accuracy $O(N^{-1/2+\varepsilon})$;
3) is an approximately most stringent test with accuracy up to $O(N^{-1/2+\varepsilon})$;
4) is an approximately minimax test with accuracy up to $O(N^{-1/2+\varepsilon})$.

Theorem 2 remains valid when the Wald test in (11) is replaced by the likelihood-ratio test under certain restrictions.

In conclusion, I express my deep gratitude to Academician Yu. V. Linnik for posing the problem and for his attention to the work; I also thank N. M. Mitrofanova, A. M. Kagan, and V. V. Petrov for valuable consultations.

Leningrad Branch
of the V. A. Steklov Mathematical Institute
Academy of Sciences of the USSR

Received
7 III 1969

References

1. A. Wald, Trans. Am. Math. Soc., 54 (1943).

2. M. K. Khalikov, Izv. AN UzSSR, No. 2 (1958).

3. A. Bikelis, Lithuanian Mathematical Collection, 8, No. 1 (1968).

4. I. A. Ibragimov, Yu. V. Linnik, *Independent and Stationary Dependent Random Variables*, Nauka, 1965.

5. Yu. V. Linnik, DAN, 169, No. 3, 523 (1966).

6. N. M. Mitrofanova, *Theory of Probability and Its Applications*, 13, issue 3 (1967).

7. Li Hoang Ty, DAN, 180, No. 4, 793 (1968).

Note: Figure translations are in progress. See original paper for figures.