TESTS OF STATISTICAL HYPOTHESES CONCERNING SEVERAL PARAMETERS WHEN THE NUMBER OF OBSERVATIONS IS LARGE(1)

BY

ABRAHAM WALD

TABLE OF CONTENTS

1. **Introduction.** In this paper we shall deal with the following general problem: Let $f(x^1, x^2, \cdots, x^r, \theta^1, \cdots, \theta^k)$ be the joint probability density function of the variates (chance variables) $x^1, \cdots, x^r$ involving k unknown parameters $\theta^1, \cdots, \theta^k$. Any set of k values $\theta^1, \cdots, \theta^k$ can be represented by a point θ in the k-dimensional Cartesian space with the coordinates $\theta^1, \cdots, \theta^k$. We shall denote the set of all possible parameter points by Ω. The set Ω is called parameter space. The parameter space Ω may be the whole k-dimensional Cartesian space, or a subset of it. For any subset ω of Ω, we shall denote by H_ω the hypothesis that the parameter point lies in ω. If ω consists of a single point, H_ω is called a simple hypothesis, otherwise H_ω is called a composite hypothesis. In this paper we shall discuss the question of an appropriate test of the hypothesis H_ω based on a large number of independent observations on $x^1, \cdots, x^r$.

For simplicity we shall introduce the following notations: The letter θ or θ_i for any subscript i will denote a point in the parameter space Ω. The letter x

Some of the results contained in this paper were presented to the Society, February 22, 1941 and September 2, 1941; received by the editors March 31, 1943.

(1) Research under a grant-in-aid from the Carnegie Corporation of New York.

will denote the random vector with the components $x^1, \cdots, x^r$, and x_α will denote the random vector with the components $x_\alpha^1, \cdots, x_\alpha^r$ where x_α^i $(i=1, 2, \cdots, r)$ denotes the αth observation on x^i. In general, the components of a vector v in the s-dimensional space will be denoted by $v^1, \cdots, v^s$, that is the components of a vector will always be indicated by superscripts. Throughout this paper all vectors have their initial points at the origin. We denote by E_n a sample point in the rn-dimensional sample space of n independent observations on the random vector x. For any relation R we denote by $P(R|\theta)$ the probability that R holds under the assumption that θ is the true parameter point. A region (subset) of the rn-dimensional sample space will always be denoted by a capital letter with the subscript n. For any region W_n the symbol $P(W_n|\theta)$ will denote the probability that E_n falls within W_n under the assumption that θ is the true parameter point. Throughout this paper the word "region" will be used synonymously with "subset," since in the theory of testing statistical hypotheses it is customary to call the subsets which are used as criterions of rejection, critical regions.

By maximum likelihood estimates $\theta_n^1, \cdots, \theta_n^k$ of $\theta^1, \cdots, \theta^k$ we mean values of $\theta^1, \cdots, \theta^k$ for which $\prod_{\alpha=1}^n f(x_\alpha^1, \cdots, x_\alpha^r, \theta^1, \cdots, \theta^k)$ becomes a maximum. The subscript n in the symbol θ_n^i will indicate that the maximum likelihood estimate is based on n independent observations on $x^1, \cdots, x^r$. A region W_n in the rn-dimensional sample space is called a critical region for testing the hypothesis H_ω if we decide to reject H_ω when and only when the observed sample point falls within W_n. For any θ not in ω the value of $P(W_n|\theta)$ is called the power of the critical region W_n with respect to the alternative hypothesis θ. The least upper bound of $P(W_n|\theta)$ with respect to θ, restricting θ to ω, will be called the size of the critical region W_n. A critical region is considered the better, the smaller its size and the greater its power.

In several previous publications(2) the author has considered the case of a single unknown parameter θ and the problem of testing a simple hypothesis $\theta=\theta_0$. It was shown, among other things, that under certain conditions the critical region given by the inequality $|n^{1/2}(\hat{\theta}_n-\theta_0)| \geqq A_n$ has certain optimum properties. Here the symbol A_n denotes some properly chosen constant. In this paper the general case of several unknown parameters is treated and simple as well as composite hypotheses are considered.

By an equality or inequality among vectors we shall mean that the equality or inequality holds for all components. For example, if $\hat{\theta}_n$ denotes the vector with the components $\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k$, where $\hat{\theta}_n^i$ is the maximum likelihood estimate of θ^i, and if A is a real number, then the inequality

$$n^{1/2}(\hat{\theta}_n - \theta) < A$$

(2) *Asymptotically most powerful tests of statistical hypotheses*, Ann. Math. Statist. vol. 12 (1941). *Some examples of asymptotically most powerful tests*, Ann. Math. Statist. ibid.

denotes the set of inequalities

$$n^{1/2}(\hat{\theta}_n^i - \theta^i) < A \qquad (i = 1, \cdots, k).$$

Or, if t is a vector with the components $t^1, \cdots, t^k$, then

$$n^{1/2}(\hat{\theta}_n - \theta) < t$$

denotes the set of inequalities

$$n^{1/2}(\hat{\theta}_n^i - \theta^i) < t^i \qquad (i = 1, \cdots, k).$$

2. **Assumptions on the density function $f(x,\theta)$.** For any function $\psi(x)$ we shall use the symbol $\int_{-\infty}^{+\infty}\psi(x)dx$ as an abbreviation for $\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\psi(x)\,dx^1\cdots dx^r$. Denote by $E_\theta\psi(x)$ the expected value of $\psi(x)$ under the assumption that θ is the true parameter point, that is

$$E_\theta\psi(x) = \int_{-\infty}^{+\infty}\psi(x)f(x,\theta)dx.$$

For any x, any positive value δ, and any θ_1 denote by $\psi_{ij}(x, \theta_1, \delta)$ the greatest lower bound, and by $\phi_{ij}(x, \theta_1, \delta)$ the least upper bound of $\partial^2 \log f(x, \theta)/\partial\theta^i\partial\theta^j$ with respect to θ in the θ-interval $|\theta-\theta_1| \leqq \delta$.

Throughout this paper the following assumptions on $f(x, \theta)$ are made:

ASSUMPTION I. *Denote by D_n the set of all sample points E_n for which the maximum likelihood estimate $\hat{\theta}_n=(\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k)$ exists and the second order partial derivatives $\partial^2 f(x_\alpha, \theta)/\partial\theta^i\partial\theta^j$ $(\alpha=1, \cdots, n; i, j=1, \cdots, k)$ are continuous functions of θ. It is assumed that*

$$\lim_{n=\infty} P(D_n \mid \theta) = 1 \qquad \textit{uniformly in } \theta.$$

If for a sample point E_n there exist several maximum likelihood estimates, we can select one of them by some given rule. Hence we shall consider $\hat{\theta}_n$ as a single-valued function of E_n defined for all points of D_n.

ASSUMPTION II. *For any positive ϵ*

$$\lim_{n=\infty} P[|\hat{\theta}_n - \theta| < \epsilon \mid \theta] = 1$$

uniformly in θ, where $\hat{\theta}_n$ denotes the vector with the components $\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k$ and $\hat{\theta}_n^i$ is the maximum likelihood estimate of θ^i.

Assumption II is somewhat more than consistency of the maximum likelihood estimate $\hat{\theta}_n$. In fact consistency means only that for any positive ϵ

$$\lim_{n=\infty} P[|\hat{\theta}_n - \theta| < \epsilon \mid \theta] = 1,$$

without requiring that the convergence be uniform in θ. If $\hat{\theta}_n$ satisfies Assump-

tion II we shall say that $\hat{\theta}_n$ is a uniformly consistent estimate of θ. A rigorous proof of the consistency of $\hat{\theta}_n$ (under certain restrictions on $f(x, \theta)$) was given by J. L. Doob(3). The uniform consistency of $\hat{\theta}_n$ together with the uniform consistency of the likelihood ratio test will be proved on the basis of some weak assumptions on $f(x, \theta)$ in a forthcoming paper.

ASSUMPTION III. *The following three conditions are fulfilled*:

(a) *For any sequences* $\{\theta_{1n}\}$, $\{\theta_{2n}\}$, *and* $\{\delta_n\}$ *for which* $\lim_{n=\infty} \theta_{1n} = \lim_{n=\infty} \theta_{2n} = \theta$ *and* $\lim \delta_n = 0$ *we have*

$$\lim_{n=\infty} E_{\theta_{1n}}\psi_{ij}(x, \theta_{2n}, \delta_n) = \lim_{n=\infty} E_{\theta_{1n}}\phi_{ij}(x, \theta_{2n}, \delta_n) = E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j}$$

uniformly in θ.

(b) *There exists a positive* ϵ *such that the expectations* $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ *and* $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ *are bounded functions of* θ_1, θ_2 *and* δ *in the domain* D_ϵ *defined by the inequalities* $|\theta_1 - \theta_2| \leqq \epsilon$ *and* $|\delta| \leqq \epsilon$.

(c) *The greatest lower bound with respect to* θ *of the absolute value of the determinant of the matrix* $\| -E_\theta(\partial^2 \log f(x, \theta))/\partial\theta^i\partial\theta^j \|$ *is positive.*

ASSUMPTION IV. $\int_{-\infty}^{+\infty} \partial f(x, \theta)/\partial\theta^i \, dx = \int_{-\infty}^{+\infty} \partial^2 f(x, \theta)/\partial\theta^i\partial\theta^j \, dx = 0$.

Assumption IV simply means that we may differentiate with respect to θ under the integral sign. In fact

$$\int_{-\infty}^{+\infty} f(x, \theta)dx = 1$$

identically in θ. Hence

$$\frac{\partial}{\partial\theta^i}\int_{-\infty}^{+\infty} f(x, \theta)dx = \frac{\partial^2}{\partial\theta^i\partial\theta^j}\int_{-\infty}^{+\infty} f(x, \theta)dx = 0.$$

Differentiating under the integral sign, we obtain the relations in Assumption IV.

ASSUMPTION V. *There exists a positive* η *such that*

$$E_\theta\left|\frac{\partial}{\partial\theta^i}\log f(x, \theta)\right|^{2+\eta} \qquad (i = 1, \cdots, k)$$

are bounded functions of θ.

3. **The joint limit distribution of** $\hat{\theta}_n$. Denote $n^{1/2}(\hat{\theta}_n^i - \theta^i)$ by $z_n^i(\theta^i)$ and let $z_n(\theta)$ be the vector with the components $z_n^1(\theta^1), \cdots, z_n^k(\theta^k)$. For any constant vector t denote the probability $P[z_n(\theta) < t \,|\, \theta]$ by $\Phi_n(t \,|\, \theta)$. We shall prove the following proposition.

(3) J. L. Doob, *Probability and statistics*, Trans. Amer. Math. Soc. vol. 36 (1934) pp. 759–775.

PROPOSITION I. *The distribution function* $\Phi_n(t|\theta)$ *converges with* $n \to \infty$, *uniformly in* t *and* θ, *towards the cumulative multivariate normal distribution with zero means and covariance matrix*

$$\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$$

where

$$c_{ij}(\theta) = -E_\theta \partial^2 \log f(x, \theta)/\partial\theta^i \partial\theta^j.$$

Proof. Because of Assumption IV, we have

$$E_\theta \frac{\partial \log f(x, \theta)}{\partial \theta^i} = \int_{-\infty}^{+\infty} \frac{\partial f(x, \theta)}{\partial \theta^i} dx = 0. \tag{1}$$

From

$$\frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} = \frac{1}{f(x, \theta)} \frac{\partial^2 f(x, \theta)}{\partial\theta^i \partial\theta^j} - \frac{1}{f^2}\left[\frac{\partial f}{\partial\theta^i} \frac{\partial f}{\partial\theta^j}\right]$$

we obtain because of Assumption IV

$$E_\theta\left\{\left[\frac{\partial \log f}{\partial\theta^i} \frac{\partial \log f}{\partial\theta^j}\right]\right\} = -E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} = c_{ij}(\theta). \tag{2}$$

From (2) it follows that the matrix $\|c_{ij}(\theta)\|$ is positive definite or semidefinite. Because of condition (c) of Assumption III the matrix $\|c_{ij}(\theta)\|$ must be positive definite, For any point E_n of the set D_n defined in Assumption I consider the Taylor expansion

$$\sum_\alpha \frac{\partial \log f(x_\alpha, \theta_1)}{\partial\theta^i} = \sum_\alpha \frac{\partial \log f(x_\alpha, \theta)}{\partial\theta^i} + \sum_j (\theta_1^j - \theta^j)\left[\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j}\right], \tag{3}$$

where $\bar{\theta}$ lies on the segment connecting the points θ and θ_1. Denote $n^{-1/2}\sum_\alpha \partial \log f(x_\alpha, \theta)/\partial\theta^i$ by $y_n^i(\theta)$ and let $y_n(\theta)$ be the vector with the components $y_n^1(\theta), \cdots, y_n^k(\theta)$. Substituting $\hat{\theta}_n$ for θ_1 in (3), the left-hand side of (3) becomes equal to zero and we obtain

$$y_n^i(\theta) + \sum_j \{[n^{1/2}(\hat{\theta}_n^j - \theta^j)]\frac{1}{n}\left[\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j}\right]\} = 0 \tag{4}$$

or

$$y_n^i(\theta) + \sum_j z_n^j(\theta) \frac{1}{n}\left[\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j}\right] = 0. \tag{5}$$

Let ν be an arbitrary positive number and let $Q_n(\theta)$ be the subset of D_n for which the inequality

$$\left| \frac{1}{n} \sum_{\alpha} \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} + c_{ij}(\theta) \right| < \nu \tag{6}$$

holds. We will prove that

$$\lim_{n=\infty} P[Q_n(\theta) \mid \theta] = 1 \tag{7}$$

uniformly in θ.

Let τ_0 be a positive number such that

$$\begin{aligned} &\left| E_\theta \phi_{ij}(x, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \frac{\nu}{2}, \\ &\left| E_\theta \psi_{ij}(x, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \frac{\nu}{2} \end{aligned} \tag{8}$$

for all values of θ. Because of condition (a) of Assumption III, such a τ_0 certainly exists. Denote by $R_n(\theta)$ the subset of D_n consisting of all points E_n for which the inequality

$$|\dot{\theta}_n - \theta| \leqq \tau_0 \tag{9}$$

holds. Because of Assumption II

$$\lim_{n=\infty} P[R_n(\theta) \mid \theta] = 1 \tag{10}$$

uniformly in θ. Since $\bar{\theta}$ lies in the interval $[\dot{\theta}_n, \theta]$ we have for all points of $R_n(\theta)$

$$|\bar{\theta}^i - \theta^i| \leqq \tau_0 \qquad (i = 1, \cdots, k). \tag{11}$$

Hence at any point of $R_n(\theta)$ the inequality

$$\sum_{\alpha} \psi_{ij}(x_\alpha, \theta, \tau_0) \leqq \sum_{\alpha} \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} \leqq \sum_{\alpha} \phi_{ij}(x_\alpha, \theta, \tau_0) \tag{12}$$

holds. Let the region $S_n(\theta)$ be defined by the inequality

$$\left| \frac{1}{n} \sum_{\alpha} \phi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \phi_{ij}(x, \theta, \tau_0) \right| < \frac{\nu}{2} \tag{13}$$

and $T_n(\theta)$ be defined by the inequality

$$\left| \frac{1}{n} \sum_{\alpha} \psi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \psi_{ij}(x, \theta, \tau_0) \right| < \frac{\nu}{2}. \tag{14}$$

It follows from (b) of Assumption III and Tshebysheff's inequality that

$$\lim_{n=\infty} P[S_n(\theta) \mid \theta] = \lim_{n=\infty} P[T_n(\theta) \mid \theta] = 1 \tag{15}$$

uniformly in θ. Denote by $U_n(\theta)$ the common part of the regions $R_n(\theta)$, $S_n(\theta)$ and $T_n(\theta)$. In $U_n(\theta)$ we have because of (8), (13) and (14)

$$\begin{aligned} &\left| \frac{1}{n} \sum_{\alpha} \phi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \nu, \\ &\left| \frac{1}{n} \sum_{\alpha} \psi_{ij}(x_\alpha, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i \partial\theta^j} \right| < \nu. \end{aligned} \tag{16}$$

From this we obtain (6) because of (12). That is to say, the inequality (6) is valid everywhere in $U_n(\theta)$. Since $\lim_{n=\infty} P[U_n(\theta) \mid \theta] = 1$ uniformly in θ and $U_n(\theta)$ is a subset of D_n, our statement about $Q_n(\theta)$ is proved.

Since the determinant $|c_{ij}(\theta)|$ has a positive lower bound, we obtain easily from (5) and (6)

$$z_n^i(\theta) = \sum_j y_n^j(\theta)[\sigma_{ij}(\theta) + \nu\epsilon_{ijn}(E_n, \theta, \nu)], \tag{17}$$

where $\epsilon_{ijn}(E_n, \theta, \nu)$ is a bounded function of E_n, θ, and ν, provided that, for each θ, E_n is restricted to points of $Q_n(\theta)$ and $|\nu|$ is less than a certain positive number ν_0.

Let $\bar{z}_n(\theta)$ be defined as follows: $\bar{z}_n(\theta) = z_n(\theta)$ at any point of $Q_n(\theta)$, and

$$\bar{z}_n^i(\theta) = \sum_j y_n^j(\theta)\sigma_{ij}(\theta)$$

at any point outside $Q_n(\theta)$. It follows from (7) that

$$\lim_{n=\infty} \{P[\bar{z}_n(\theta) < t \mid \theta] - P[z_n(\theta) < t \mid \theta]\} = 0 \tag{18}$$

uniformly in t and θ.

Denote $\sum_j y_n^j(\theta)\sigma_{ij}(\theta)$ by $\dot{z}_n^i(\theta)$ and let $\dot{z}_n(\theta)$ be the vector with the components $\dot{z}_n^1(\theta), \cdots, \dot{z}_n^k(\theta)$. From (1), (2), Assumption V and the general limit theorems it follows that $P[y_n(\theta) < t \mid \theta]$ converges with $n \to \infty$, uniformly in t and θ, towards the k-variate cumulative normal distribution with zero means and covariance matrix $\|c_{ij}(\theta)\|$. From this we easily obtain that $P[\dot{z}_n(\theta) < t \mid \theta]$ converges with $n \to \infty$, uniformly in t and θ, towards the cumulative joint normal distribution with zero means and covariance matrix $\|\sigma_{ij}(\theta)\|$. Since ν can be chosen arbitrarily small, we obviously have because of (17)

$$\lim_{n=\infty} \{P[\dot{z}_n(\theta) < t \mid \theta] - P[\bar{z}_n(\theta) < t \mid \theta]\} = 0 \tag{19}$$

uniformly in t and θ. Proposition I follows from (18) and (19).

4. **Reduction of the general problem to the case of a multivariate normal distribution.** In this section we shall prove two lemmas which will enable us to reduce the general problem of large sample inference to the case where the variates under consideration have a joint normal distribution.

LEMMA 1. *For each positive integer n there exists a set-function $W_n^*(W_n)$ defined over all Borel measurable subsets W_n of the rn-dimensional sample space such that the following two conditions are fulfilled:*

(a) *For each W_n, $W_n^*(W_n)$ is a Borel measurable subset of the rn-dimensional sample space with the following property: For each point of $W_n^*(W_n)$ the maximum likelihood estimate exists and if a sample point E_n lies in $W_n^*(W_n)$ then also all those points E_n' lie in $W_n^*(W_n)$ for which $\hat{\theta}_n(E_n') = \hat{\theta}_n(E_n)$.*

(b) $\operatorname{Lim}_{n=\infty} \{P[W_n^*(W_n) \mid \theta] - P[W_n \mid \theta]\} = 0$ *uniformly in θ and W_n.*

Proof. Let λ be a real variable which takes only non-negative values and consider the region $W_n(\theta, \lambda)$ defined as the common part of the region W_n and the region $|n^{1/2}(\hat{\theta}_n - \theta)| \leqq \lambda$. Similarly let $W_n^*(W_n, \theta, \lambda)$ be the intersection of $W_n^*(W_n)$ and the region $|n^{1/2}(\hat{\theta}_n - \theta)| \leqq \lambda$.

For any function $\Phi(v)$ we will denote by g.l.b.$_v$ $\Phi(v)$ and l.u.b.$_v$ $\Phi(v)$ the greatest lower bound and the least upper bound of $\Phi(v)$ respectively.

Since, on account of Proposition I, for any sequence $\{\lambda_n\}$ for which $\lim_{n=\infty} \lambda_n = \infty$

$$\lim_{n=\infty} \{\text{g.l.b.}_{\theta} [P[|n^{1/2}(\hat{\theta}_n - \theta)| \leqq \lambda_n \mid \theta]]\} = 1,$$

Lemma I is proved if we show that there exists a sequence $\{\lambda_n\}$ not depending on θ and W_n such that $\lim_{n=\infty} \lambda_n = \infty$ and

$$\lim_{n=\infty} \{P[W_n(\theta, \lambda_n) \mid \theta] - P[W_n^*(W_n, \theta, \lambda_n) \mid \theta]\} = 0 \tag{20}$$

uniformly in W_n and θ.

Let q be a real variable restricted to values greater than 1. For any set of k integers $(r_1, \cdots, r_k)$ and for any value q denote by $I_n(r_1, \cdots, r_k, q)$ the region defined by the inequalities:

$$\frac{r_1 - 1/2}{qn^{1/2}} < \hat{\theta}_n^1 < \frac{r_1 + 1/2}{qn^{1/2}}, \cdots, \frac{r_k - 1/2}{qn^{1/2}} < \hat{\theta}_n^k < \frac{r_k + 1/2}{qn^{1/2}}. \tag{21}$$

Furthermore denote by $\theta_n(r_1, \cdots, r_k, q)$ the parameter point with the coordinates $r_1/qn^{1/2}, \cdots, r_k/qn^{1/2}$. We order the system of all sets of k integers $(r_1, \cdots, r_k)$ in a sequence and we shall denote by $I_{ns}(q)$ the interval $I_n(r_1, \cdots, r_k, q)$ where $(r_1, \cdots, r_k)$ is the sth element in the ordered sequence ($s = 1, 2, \cdots$, ad inf.). Similarly $\theta_{ns}(q)$ denotes the parameter point $\theta_n(r_1, \cdots, r_k, q)$ where $(r_1, \cdots, r_k)$ is the sth element in the ordered sequence.

Let $J_{ns}(W_n, q)$ be the common part of the three regions $I_{ns}(q)$, W_n and $Q_n[\theta_{ns}(q)]$ where for any θ, $Q_n(\theta)$ is defined by the inequalities

$$(22)\qquad \begin{aligned} &\left| \frac{1}{n} \sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta, \frac{1}{n^{1/3}}\right) + c_{ij}(\theta) \right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n), \\ &\left| \frac{1}{n} \sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta, \frac{1}{n^{1/3}}\right) + c_{ij}(\theta) \right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n). \end{aligned}$$

The expression $\nu(n)$ is equal to l.u.b.$_\theta$ $[E_\theta \phi_{ij}(x, \theta, 2/n^{1/3}) - E_\theta \psi_{ij}(x, \theta, 2/n^{1/3})]$ and the expression $\bar{\nu}(n)$ is defined as follows: Denote by $c_{ij}(\theta, n)$ the least upper bound of $|c_{ij}(\theta) - c_{ij}(\bar{\theta})|$ with respect to $\bar{\theta}$ where $\bar{\theta}$ can take only values in the interval $[\theta - 1/n^{1/3}, \theta + 1/n^{1/3}]$. Then $\bar{\nu}(n)$ is defined as the least upper bound of $c_{ij}(\theta, n)$ with respect to θ, i and j. Because of condition (a) of Assumption III, we obtain

$$(23)\qquad \lim_{n=\infty} \nu(n) = \lim_{n=\infty} \bar{\nu}(n) = 0.$$

Let $J^*_{ns}(W_n, q)$ be a subset of $I_{ns}(q)$ such that the following two conditions are fulfilled:

(24a) If E_n is an element of $J^*_{ns}(W_n, q)$ then also all those points E_n' for which $\theta_n(E_n') = \theta_n(E_n)$ are elements of $J^*_{ns}(W_n, q)$, that is $J^*_{ns}(W_n, q)$ can be represented as a subset of the space of the maximum likelihood estimates. Furthermore $J^*_{ns}(W_n, q)$ is an interval in this space.

(24b) $\operatorname{Lim}_{n=\infty} \{\text{l.u.b.}_{s, W_n} | P[J_{ns}(W_n, q) | \theta_{ns}(q)] - \mathfrak{P}[J^*_{ns}(W_n, q) | \theta_{ns}(q)]| \} = 0$ where $\mathfrak{P}(V_n | \theta)$ denotes the probability of V_n calculated on the basis that the joint distribution of $n^{1/2}(\theta_n^1 - \theta^1), \cdots, n^{1/2}(\theta_n^k - \theta^k)$ is normal with zero means and covariance matrix $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$.

The existence of such a set $J^*_{ns}(W_n, q)$ can be proved as follows: Obviously there exists a subset $J^*_{ns}(W_n, q)$ of $I_{ns}(q)$ such that (24a) is fulfilled and

$$\mathfrak{P}[J^*_{ns}(W_n, q) | \theta_{ns}(q)] = \min \{P[J_{ns}(W_n, q) | \theta_{ns}(q)], \mathfrak{P}[I_{ns}(q) | \theta_{ns}(q)]\}.$$

Since $J_{ns}(W_n, q)$ is a subset of $I_{ns}(q)$ and since $\lim_{n=\infty} \{P[I_{ns}(q) | \theta_{ns}(q)] - \mathfrak{P}[I_{ns}(q) | \theta_{ns}(q)]\} = 0$ uniformly in s, the above defined subset $J^*_{ns}(W_n, q)$ satisfies also the condition (24b). We define

$$(25)\qquad W^*_n(W_n, q) = \sum_{s=1}^{\infty} J^*_{ns}(W_n, q).$$

Furthermore we define the regions $J_n(W_n, \theta, \lambda, q)$ and $J^*_n(W_n, \theta, \lambda, q)$ as follows:

$$(26)\qquad J_n(W_n, \theta, \lambda, q) = \sum_s J_{ns}(W_n, q),$$

$$(27)\qquad J^*_n(W_n, \theta, \lambda, q) = \sum_s J^*_{ns}(W_n, q).$$

where the summation is to be taken over all values of s for which $|\theta-\theta_{ns}(q)| \leqq \lambda/n^{1/2}$.

Let $\bar{J}_{ns}^*(W_n, q)$ be the intersection of $J_{ns}^*(W_n, q)$ and $Q_n[\theta_{ns}(q)]$, and let $\bar{J}_n^*(W_n, \theta, \lambda, q) = \sum_s \bar{J}_{ns}^*(W_n, q)$ where the summation is to be taken over all values of s for which $|\theta-\theta_{ns}(q)| \leqq \lambda/n^{1/2}$. Furthermore let $W_n^*(W_n, \theta, \lambda, q)$ be the intersection of $W_n^*(W_n, q)$ with the region $|n^{1/2}(\theta_n-\theta)| \leqq \lambda$.

If for a value of s we have $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ then for all points of $I_{ns}(q)$ we have $n^{1/2}|\theta-\theta_n| \leqq \lambda+1/q$. If there exists a point in $I_{ns}(q)$ for which $n^{1/2}|\theta-\theta_n| \leqq \lambda-1/q$ then $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$. Hence $W_n^*(W_n, \theta, \lambda-1/q, q)$ is a subset of $J_n^*(W_n, \theta, \lambda, q)$, and the latter is a subset of $W_n^*(W_n, \theta, \lambda+1/q, q)$. Thus

$$
\begin{aligned}
&|P[W_n^*(W_n, \theta, \lambda, q) \mid \theta] - P[J_n^*(W_n, \theta, \lambda, q) \mid \theta]| \\
&\leqq P[W_n^*(W_n, \theta, \lambda + 1/q, q) \mid \theta] - P[W_n^*(W_n, \theta, \lambda - 1/q, q) \mid \theta] \\
&\leqq P[\lambda - 1/q \leqq n^{1/2}|\theta_n - \theta| \leqq \lambda + 1/q \mid \theta]. \qquad (28)
\end{aligned}
$$

According to condition (b) of Assumption III $E_\theta[\phi_{ij}(x, \theta, \delta)]^2$ and $E_\theta[\psi_{ij}(x, \theta, \delta)]^2$ are bounded functions of θ. Hence also $E_\theta\phi_{ij}(x, \theta, \delta)$ and $E_\theta\psi_{ij}(x, \theta, \delta)$ are bounded functions of θ. Substituting $\delta=0$ we find that $c_{ij}(\theta)$ is a bounded function of θ. From the boundedness of $c_{ij}(\theta)$ and from the fact that the determinant $|c_{ij}(\theta)|$ has a positive lower bound (condition (c) of Assumption III) it follows because of Proposition I that

$$
\lim_{q=\infty} \{\text{l.u.b.}_{\theta} P[\lambda - 1/q \leqq n^{1/2}|\theta_n - \theta| \leqq \lambda + 1/q \mid \theta]\} = 0 \qquad (29)
$$

uniformly in λ. From (28) and (29) we obtain

$$
\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n^*(W_n, \theta, \lambda, q) \mid \theta] - P[J_n^*(W_n, \theta, \lambda, q) \mid \theta]|\} = \epsilon_1(\lambda, q) \qquad (30)
$$

where $\lim_{q=\infty} \epsilon_1(\lambda, q) = 0$ uniformly in λ.

Denote by $R_n(\theta, \lambda, q)$ the common part of the regions $Q_n[\theta_{ns}(q)]$ formed for all values s for which $|\theta-\theta_{ns}(q)| \leqq \lambda/n^{1/2}$. Then for almost all n the region $R_n(\theta, \lambda, q)$ contains the region $T_n(\theta)$ as a subset, where $T_n(\theta)$ is defined by the inequalities

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) + c_{ij}(\theta)\right| < \frac{1}{n^{1/3}} + \nu(n), \\
&\left|\frac{1}{n}\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) + c_{ij}(\theta)\right| < \frac{1}{n^{1/3}} + \nu(n). \qquad (31)
\end{aligned}
$$

In fact, from (31) it follows that

$$
\begin{aligned}
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n), \\
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n).
\end{aligned} \tag{32}
$$

Since

$$
\begin{aligned}
\frac{1}{n}\sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) &\leqq \frac{1}{n}\sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq \frac{1}{n}\sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{n}\sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) &\leqq \frac{1}{n}\sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq \frac{1}{n}\sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right)
\end{aligned}
$$

for almost all n and for all s for which $n^{1/2}\left|\theta - \theta_{ns}(q)\right| \leqq \lambda$, we obtain from (32)

$$
\begin{aligned}
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n), \\
c_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leqq \frac{1}{n}\sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \leqq c_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n),
\end{aligned} \tag{33}
$$

for almost all n and for all s for which $n^{1/2}\left|\theta - \theta_{ns}(q)\right| \leqq \lambda$. Since $\left|c_{ij}[\theta_{ns}(q)] - c_{ij}(\theta)\right| \leqq \bar{\nu}(n)$ for almost all n and for all s for which $n^{1/2}\left|\theta - \theta_{ns}(q)\right| \leqq \lambda$, we obtain from (33)

$$
\begin{aligned}
c_{ij}[\theta_{ns}(q)] - \frac{1}{n^{1/3}} - \nu(n) - \bar{\nu}(n) &\leqq \frac{1}{n}\sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq c_{ij}[\theta_{ns}(q)] + \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n), \\
c_{ij}[\theta_{ns}(q)] - \frac{1}{n^{1/3}} - \nu(n) - \bar{\nu}(n) &\leqq \frac{1}{n}\sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) \\
&\leqq c_{ij}[\theta_{ns}(q)] + \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n)
\end{aligned} \tag{34}
$$

for almost all n and for all s for which $n^{1/2}\left|\theta - \theta_{ns}(q)\right| \leqq \lambda$. The inequalities (34) are equivalent to the inequalities (22) if in (22) $\theta_{ns}(q)$ is substituted for θ. Hence our statement about the region $R_n(\theta, \lambda, q)$ is proved.

Consider the region $U_n(\theta)$ defined by the inequalities

$$(35)\qquad \begin{aligned} &\left| \frac{1}{n} \sum_{\alpha} \phi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) - E_\theta \phi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right) \right| < \frac{1}{n^{1/3}}, \\ &\left| \frac{1}{n} \sum_{\alpha} \psi_{ij}\left(x_\alpha, \theta, \frac{2}{n^{1/3}}\right) - E_\theta \psi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right) \right| < \frac{1}{n^{1/3}}. \end{aligned}$$

Since

$$\nu(n) \geqq \left| E_\theta \phi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right) + c_{ij}(\theta) \right|$$

and

$$\nu(n) \geqq \left| c_{ij}(\theta) + E_\theta \psi_{ij}\left(x, \theta, \frac{2}{n^{1/3}}\right) \right|,$$

the validity of (35) implies the validity of (31). Hence $U_n(\theta)$ is a subset of $T_n(\theta)$. From condition (b) of Assumption III and Tshebysheff's inequality it follows that $\lim_{n=\infty} P[U_n(\theta) \mid \theta] = 1$ uniformly in θ. Hence $\lim_{n=\infty} P[T_n(\theta) \mid \theta] = 1$ uniformly in θ. Thus, as can easily be seen,

$$(36)\qquad \lim_{n=\infty} P[R_n(\theta, \lambda_n, q) \mid \theta] = 1$$

for any bounded sequence $\{\lambda_n\}$ uniformly in θ and q.

Let $\dot{J}_{ns}(W_n, q)$ be the intersection of the regions W_n and $I_{ns}(q)$. Furthermore let $\dot{J}_n(W_n, \theta, \lambda, q)$ be equal to $\sum_s \dot{J}_{ns}(W_n, q)$ where the summation is to be taken over all values of s for which $n^{1/2} \mid \theta - \theta_{ns}(q) \mid \leqq \lambda$. Then the common part of $\sum_{s=1}^{\infty} I_{ns}(q)$ and $W_n(\theta, \lambda - 1/q)$ is a subset of $\dot{J}_n(W_n, \theta, \lambda, q)$, and the last is a subset of the common part of $\sum_{s=1}^{\infty} I_{ns}(q)$ and $W_n(\theta, \lambda + 1/q)$. Hence, since $P[\sum_{s=1}^{\infty} I_{ns}(q) \mid \theta] = 1$, we have

$$(37)\qquad \begin{aligned} &\left| P[W_n(\theta, \lambda) \mid \theta] - P[\dot{J}_n(W_n, \theta, \lambda, q) \mid \theta] \right| \\ &\qquad \leqq P[W_n(\theta, \lambda + 1/q) \mid \theta] - P[W_n(\theta, \lambda - 1/q) \mid \theta] \\ &\qquad \leqq P[\lambda - 1/q \leqq n^{1/2} \mid \dot{\theta}_n - \theta \mid \leqq \lambda + 1/q \mid \theta]. \end{aligned}$$

From (37) and (29) it follows that for any sequence $\{q_n\}$ for which $\lim q_n = \infty$, we have

$$(38)\qquad \lim_{n=\infty} \left| P[W_n(\theta, \lambda) \mid \theta] - P[\dot{J}_n(W_n, \theta, \lambda, q_n) \mid \theta] \right| = 0$$

uniformly in θ, W_n and λ.

Since the common part of the regions $\dot{J}_n(W_n, \theta, \lambda, q)$ and $R_n(\theta, \lambda, q)$ is contained as a subset in $J_n(W_n, \theta, \lambda, q)$ and since the last is a subset of $\dot{J}_n(W_n, \theta, \lambda, q)$, we obtain from (38) and (36) that for any bounded sequence

$\{\lambda_n\}$ and for any sequence $\{q_n\}$ for which $\lim_{n=\infty} q_n = \infty$ we have

$$\lim_{n=\infty} \{P[W_n(\theta, \lambda_n) \mid \theta] - P[J_n(W_n, \theta, \lambda_n, q_n) \mid \theta]\} = 0 \tag{39}$$

uniformly in θ and W_n.

Since the common part of the regions $J_n^*(W_n, \theta, \lambda, q)$ and $R_n(\theta, \lambda, q)$ is a subset of $\bar{J}_n^*(W_n, \theta, \lambda, q)$ and the last is a subset of $J_n^*(W_n, \theta, \lambda, q)$ we obtain from (36) that for any bounded sequence $\{\lambda_n\}$ and for any sequence $\{q_n\}$ for which $\lim_{n=\infty} q_n = \infty$, we have

$$\lim_{n=\infty} \{P[\bar{J}_n^*(W_n, \theta, \lambda_n, q_n) \mid \theta] - P[J_n^*(W_n, \theta, \lambda_n, q_n) \mid \theta]\} = 0 \tag{40}$$

uniformly in θ and W_n.

Now we shall evaluate the limit values of $P[J_n(W_n, \theta, \lambda, q) \mid \theta]$ and $P[\bar{J}_n^*(W_n, \theta, \lambda, q) \mid \theta]$. Denote by $A_n(\lambda, q)$ the domain in the space of the variables θ, W_n, θ' and E_n defined as follows: θ and W_n can take arbitrary values, θ' is restricted to values for which $|\theta' - \theta| \leqq \lambda/n^{1/2}$; and for any θ and W_n, E_n is restricted to points which lie in the sum of the sets $J_n(W_n, \theta, \lambda, q)$ and $\bar{J}_n^*(W_n, \theta, \lambda, q)$. Denote furthermore by $\rho_n(\theta', \theta, \dot{\theta}_n)$ the function

$$\rho_n(\theta', \theta, \dot{\theta}_n) = -\frac{1}{2} \sum_j \sum_i n(\theta'^i - \dot{\theta}_n^i)(\theta'^j - \dot{\theta}_n^j) c_{ij}(\theta). \tag{41}$$

Consider the Taylor expansion

$$\begin{aligned} &\sum_\alpha \log f(x_\alpha, \theta') \\ &= \sum_\alpha \log f(x_\alpha, \dot{\theta}_n) + \frac{n}{2} \sum_j \sum_i (\theta'^i - \dot{\theta}_n^i)(\theta'^j - \dot{\theta}_n^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i \partial\theta^j} \end{aligned} \tag{42}$$

where $\bar{\theta}$ lies in the interval $[\theta', \dot{\theta}_n]$.

Since in the domain $A_n(\lambda, q)$ any point E_n lies in the sum of the sets $J_n(W_n, \theta, \lambda, q)$ and $\bar{J}_n^*(W_n, \theta, \lambda, q)$, it follows from the definitions of these sets that E_n lies in the set $\sum_s Q_n[\theta_{ns}(q)]$ where the summation is to be taken over all values of s for which $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$ (the set $Q_n(\theta)$ is defined in (22)). Hence for any E_n in the domain $A_n(\lambda, q)$ we have

$$\begin{aligned} &\left| \frac{1}{n} \sum_\alpha \phi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) + c_{ij}[\theta_{ns}(q)] \right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n), \\ &\left| \frac{1}{n} \sum_\alpha \psi_{ij}\left(x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right) + c_{ij}[\theta_{ns}(q)] \right| \leqq \frac{1}{n^{1/3}} + \nu(n) + \bar{\nu}(n) \end{aligned} \tag{43}$$

for that value of s for which E_n lies in $I_{ns}(q)$. In all that follows, with any

point of the domain $A_n(\lambda, q)$ we shall associate the integer s for which E_n lies in $I_{ns}(q)$. Since $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ in the domain $A_n(\lambda, q)$, we have in the domain $A_n(\lambda, q)$

$$\begin{aligned}\frac{1}{n}\sum_\alpha \psi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right] &\leqq \frac{1}{n}\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \\ &\leqq \frac{1}{n}\sum_\alpha \phi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right], \\ \frac{1}{n}\sum_\alpha \psi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right] &\leqq \frac{1}{n}\sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \\ &\leqq \frac{1}{n}\sum_\alpha \phi_{ij}\left[x_\alpha, \theta_{ns}(q), \frac{1}{n^{1/3}}\right]\end{aligned} \tag{44}$$

for almost all values of n. From the definition of $\bar{\nu}(n)$ and from the validity of the inequality $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ in $A_n(\lambda, q)$ we find that in the domain $A_n(\lambda, q)$

$$|c_{ij}[\theta_{ns}(q)] - c_{ij}(\theta)| \leqq \bar{\nu}(n) \tag{45}$$

for almost all values of n. From (43), (44) and (45) it follows that in the domain $A_n(\lambda, q)$

$$\begin{aligned}\left|\frac{1}{n}\sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) + c_{ij}(\theta)\right| &\leqq \frac{1}{n^{1/3}} + \nu(n) + 2\bar{\nu}(n), \\ \left|\frac{1}{n}\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) + c_{ij}(\theta)\right| &\leqq \frac{1}{n^{1/3}} + \nu(n) + 2\bar{\nu}(n)\end{aligned} \tag{46}$$

for almost all values of n.

Since $n^{1/2}|\theta-\theta_{ns}(q)| \leqq \lambda$ in $A_n(\lambda, q)$, we have

$$n^{1/2}|\theta - \dot{\theta}_n| \leqq \lambda + 1/q \quad \text{in} \quad A_n(\lambda, q) \tag{47}$$

and therefore

$$n^{1/2}|\theta' - \dot{\theta}_n| \leqq 2\lambda + 1/q \quad \text{in} \quad A_n(\lambda, q). \tag{48}$$

Since $\bar{\theta}$ lies in the interval $[\theta', \dot{\theta}_n]$, from (48) we obtain

$$n^{1/2}|\bar{\theta} - \dot{\theta}_n| \leqq 2\lambda + 1/q \quad \text{in} \quad A_n(\lambda, q). \tag{49}$$

From (47) and (49) it follows that in $A_n(\lambda, q)$

$$\sum_\alpha \psi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \leqq \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta})}{\partial\theta^i\partial\theta^j} \leqq \sum_\alpha \phi_{ij}\left(x_\alpha, \theta, \frac{1}{2n^{1/3}}\right) \tag{50}$$

for almost all n. Because of (23), (46) and (50) we obtain from (41) and (42)

$$\lim_{n=\infty}\left\{\text{l.u.b.}_{A_n(\lambda,q)}\left|\sum_\alpha \log f(x_\alpha,\theta')-\sum_\alpha \log f(x_\alpha,\dot\theta_n)-\rho_n(\theta',\theta,\dot\theta_n)\right|\right\}=0, \tag{51}$$

and

$$\lim_{n=\infty}\left\{\text{l.u.b.}_{A_n(\lambda,q)}\left|\left[\sum_\alpha \log f(x_\alpha,\theta)-\sum_\alpha \log f(x_\alpha,\theta')\right]-[\rho_n(\theta,\theta,\dot\theta_n)-\rho_n(\theta',\theta,\dot\theta_n)]\right|\right\}=0. \tag{52}$$

Denote $P[J_{ns}(W_n, q)\,|\,\theta_{ns}(q)]$ by $P_{ns}(W_n, q)$ and $P[\bar J^*_{ns}(W_n, q)\,|\,\theta_{ns}(q)]$ by $\bar P^*_{ns}(W_n, q)$. Substituting $\theta_{ns}(q)$ for θ' we easily obtain from (52)

$$\lim_{n=\infty}\left\{\text{l.u.b.}_{\theta,W_n}\left|P[J_n(W_n,\theta,\lambda,q)\,|\,\theta]-\sum_s P_{ns}(W_n,q)\exp\left(\rho_n[\theta,\theta,\theta^*_{ns}(q)]-\rho_n[\theta_{ns}(q),\theta,\theta^*_{ns}(q)]\right)\right|\right\}=0 \tag{53}$$

where the summation with respect to s is to be taken over all values for which $n^{1/2}|\theta-\theta_{ns}(q)|\leqq\lambda$ and $\theta^*_{ns}(q)$ is a parameter point for which

$$n^{1/2}\,|\,\theta_{ns}(q)-\theta^*_{ns}(q)\,|\leqq 1/q. \tag{54}$$

Since $c_{ij}(\theta)$ is a uniformly continuous and bounded function of θ, it follows from (41) and (54) that

$$\begin{aligned}|\,\rho(\theta,\theta,\theta^*_{ns}(q))-\rho[\theta,\theta,\theta_{ns}(q)]\,| &= \phi_1[\theta,\theta^*_{ns}(q),\theta_{ns}(q)]/q,\\ |\,\rho[\theta_{ns}(q),\theta,\theta^*_{ns}(q)]-\rho[\theta_{ns}(q),\theta,\theta_{ns}(q)]\,| &= \phi_2[\theta,\theta^*_{ns}(q),\theta_{ns}(q)]/q,\end{aligned} \tag{55}$$

where $\phi_1[\theta, \theta^*_{ns}(q), \theta_{ns}(q)]$ and $\phi_2[\theta, \theta^*_{ns}(q), \theta_{ns}(q)]$ are bounded functions of θ, $\theta^*_{ns}(q)$ and $\theta_{ns}(q)$ in the domain $n^{1/2}|\theta-\theta_{ns}(q)|\leqq\lambda$. From (53) and (55) it follows that

$$\limsup_{n=\infty}\left\{\text{l.u.b.}_{\theta,W_n}\left|P[J_n(W_n,\theta,\lambda,q)\,|\,\theta]-\sum_s P_{ns}(W_n,q)\exp\left(\rho_n[\theta,\theta,\theta_{ns}(q)]-\rho_n[\theta_{ns}(q),\theta,\theta_{ns}(q)]\right)\right|\right\}=\epsilon(\lambda,q) \tag{56}$$

where

$$\lim_{q=\infty}\epsilon(\lambda,q)=0 \tag{57}$$

uniformly in λ over any finite positive interval. Similarly we obtain

$$\limsup_{n=\infty} \left\{ \text{l.u.b.}_{\theta, W_n} \left| P[\bar{J}_n^*(W_n, \theta, \lambda, q) \mid \theta] - \sum_s \bar{P}_{ns}^*(W_n, q) \exp\left(\rho_n[\theta, \theta, \theta_{ns}(q)] - \rho_n[\theta_{ns}(q), \theta, \theta_{ns}(q)]\right) \right| \right\} = \eta(\lambda, q) \tag{58}$$

where

$$\lim_{q=\infty} \eta(\lambda, q) = 0 \tag{59}$$

uniformly in λ over any finite positive interval. In the formulas (56) and (58) the summation with respect to s is to be taken for all values for which $|\theta - \theta_{ns}(q)| \leqq \lambda/n^{1/2}$. The expression $\rho_n[\theta_{ns}(q), \theta, \theta_{ns}(q)]$ is obviously equal to zero, hence (56) and (58) can be simplified by substituting zero for this expression. Denote $P[J_{ns}^*(W_n, q) \mid \theta_{ns}(q)]$ by $P_{ns}^*(W_n, q)$. Because of Proposition I we have

$$\lim_{n=\infty} \{\mathfrak{P}[J_{ns}^*(W_n, q) \mid \theta_{ns}(q)] - P_{ns}^*(W_n, q)\} = 0$$

uniformly in s and W_n. Hence we obtain from condition (b) of (24)

$$\lim_{n=\infty} \{P_{ns}(W_n, q) - P_{ns}^*(W_n, q)\} = 0 \tag{60}$$

uniformly in s and W_n. Since $\bar{J}_{ns}^*(W_n, q)$ is the intersection of $Q_n[\theta_{ns}(q)]$ with $J_{ns}^*(W_n, q)$ and since

$$\lim_{n=\infty} P[Q_n[\theta_{ns}(q)] \mid \theta_{ns}(q)] = 1$$

uniformly in s, we have

$$\lim_{n=\infty} \{\bar{P}_{ns}^*(W_n, q) - P_{ns}^*(W_n, q)\} = 0 \tag{61}$$

uniformly in s and W_n. Since for any given λ and q the number of different values of s satisfying the inequality $n^{1/2}|\theta - \theta_{ns}(q)| \leqq \lambda$ is a bounded function of θ, from (56), (58), (60) and (61) we obtain

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} | P[J_n(W_n, \theta, \lambda, q) \mid \theta] - P[\bar{J}_n^*(W_n, \theta, \lambda, q) \mid \theta]|\} = \zeta(\lambda, q) \leqq \epsilon(\lambda, q) + \eta(\lambda, q). \tag{62}$$

Hence

$$\lim_{q=\infty} \zeta(\lambda, q) = 0 \tag{63}$$

uniformly in λ over any finite positive interval.

For any positive λ' the sets $J_n(W_n, \theta, \lambda+\lambda', q)-J_n(W_n, \theta, \lambda, q)$, $\bar{J}_n^*(W_n, \theta, \lambda+\lambda', q)-\bar{J}_n^*(W_n, \theta, \lambda, q)$, $J_n^*(W_n, \theta, \lambda+\lambda', q)-J^*(W_n, \theta, \lambda, q)$, $W_n^*(W_n, \theta, \lambda+\lambda', q)-W_n^*(W_n, \theta, \lambda, q)$ and $W_n(\theta, \lambda+\lambda')-W_n(\theta, \lambda)$ are subsets of the set defined by the inequality

$$\lambda - 1/q \leqq n^{1/2}\left|\theta - \dot{\theta}_n\right| \leqq \lambda + \lambda' + 1/q.$$

Since for any sequence $\{\lambda_n\}$ for which $\lim_{n=\infty} \lambda_n = \infty$ we have

$$\lim_{n=\infty} P[\lambda_n - 1/q \leqq n^{1/2}\left|\theta - \dot{\theta}_n\right| \leqq \lambda_n + \lambda' + 1/q \,|\, \theta] = 0$$

uniformly in θ, q and λ', (39) and (40) hold for any arbitrary sequence $\{\lambda_n\}$ and (63) holds uniformly in λ where λ can take any positive value. Thus from (30), (39), (40), (62) and (63) we obtain

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} \left| P[W_n(\theta, \lambda) \,|\, \theta] - P[W_n^*(W_n, \theta, \lambda, q) \,|\, \theta]\right| \} = \epsilon_3(\lambda, q), \tag{64}$$

where

$$\lim_{q=\infty} \epsilon_3(\lambda, q) = 0 \tag{65}$$

uniformly in λ. Let $\{q_i\}$ $(i=1, 2, \cdots, \text{ad inf.})$ be a sequence of positive integers such that $\lim_{i=\infty} q_i = +\infty$. Furthermore let $\{\eta_i\}$ be a sequence of positive numbers such that $\lim_{i=\infty} \eta_i = 0$. We define $W_n^*(W_n)$ as follows:

$$W_n^*(W_n) = W_n^*(W_n, q_{i+1}) \quad \text{for} \quad n_i < n \leqq n_{i+1} \quad (i = 0, 1, \cdots, \text{ad inf.}). \tag{66}$$

The sequence $\{n_i\}$ $(i=0, 1, 2, \cdots, \text{ad inf.})$ of integers is chosen as follows: Denote by $F_n(\lambda, q)$ the expression

$$\text{l.u.b.}_{\theta, W_n} \left| P[W_n(\theta, \lambda) \,|\, \theta] - P[W_n^*(W_n, \theta, \lambda, q) \,|\, \theta]\right|.$$

The integer n_0 is put equal to 0 and n_i is chosen such that

$$\begin{gathered} n_i > n_{i-1}, \\ F_n(\lambda_i, q_{i+1}) < \epsilon_3(\lambda_i, q_{i+1}) + \eta_i \end{gathered} \tag{67}$$

for all $n > n_i$, and where $\{\lambda_i\}$ denotes a sequence of numbers such that $\lim_{i=\infty} \lambda_i = +\infty$.

Let $\lambda_n' = \lambda_i$, $\eta_n' = \eta_i$, and $q_n' = q_{i+1}$ for $n_i < n \leqq n_{i+1}$ $(i=0, 1, 2, \cdots, \text{ad inf.})$. Then from (64), (65) and (67) we obtain

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} \left| P[W_n(\theta, \lambda_n') \,|\, \theta] - P[W_n^*(W_n, \theta, \lambda_n', q_n') \,|\, \theta]\right| \} = 0. \tag{68}$$

Denote by $W_n^*(W_n, \theta, \lambda)$ the intersection of $W_n^*(W_n)$ and the set defined by the inequality $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$. Since $W_n^*(W_n, \theta, \lambda, q)$ is the intersection of $W_n^*(W_n, q)$ and the region $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$, it follows from (66) that

$$W_n^*(W_n, \theta, \lambda_n', q_n') = W_n^*(W_n, \theta, \lambda_n'). \tag{69}$$

Equation (20) follows from (68) and (69). Hence Lemma 1 is proved.

We shall say that a region V_n^* lies in the space of the maximum likelihood estimates if it has the following property: If E_n is an element of V_n^* then also all those points E_n' for which $\hat{\theta}_n(E_n') = \hat{\theta}_n(E_n)$ are elements of V_n^*. In all the following considerations the symbol * as a superscript in the notation of a region will indicate that the region lies in the space of the maximum likelihood estimates, except if a statement to the contrary is explicitly made. For any region V_n^* we shall denote by $\mathfrak{P}(V_n^*|\theta)$ the probability that the sample point will fall in V_n^* calculated under the assumption that $n^{1/2}(\hat{\theta}_n^1 - \theta^1), \cdots, n^{1/2}(\hat{\theta}_n^k - \theta^k)$ have a joint normal distribution with zero means and covariance matrix $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$.

LEMMA 2. *There exists a function $W_n^*(R_n^*)$ defined over all Borel measurable subsets R_n^* such that*

$$\lim_{n=\infty} \{P[R_n^*|\theta] - \mathfrak{P}[W_n^*(R_n^*)|\theta]\} = 0$$

uniformly in θ and R_n^*.

Proof. Since we assumed that the set $J_{ns}^*(W_n, q)$ defined in (24) is an interval in the space of the maximum likelihood estimates, it follows from Proposition I that

$$\lim_{n=\infty} \{P[J_{ns}^*(W_n, q) \mid \theta] - \mathfrak{P}[J_{ns}^*(W_n, q) \mid \theta]\} = 0 \tag{70}$$

uniformly in θ, W_n, and s. Let $W_n^*(W_n, q)$ be the set defined in (25) and let $W_n^*(W_n, \theta, \lambda, q)$ be the intersection of $W_n^*(W_n, q)$ and the region $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$. For given values of λ and q the number of different values of s, for which $J_{ns}^*(W_n, q)$ has at least a point common with the region $n^{1/2}|\theta - \hat{\theta}_n| \leqq \lambda$, is a bounded function of θ. Hence it follows from (70) that

$$\lim_{n=\infty} \{P[W_n^*(W_n, \theta, \lambda, q) \mid \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda, q) \mid \theta]\} = 0 \tag{71}$$

uniformly in θ and W_n. From (64), (65) and (71) we obtain

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta, W_n} |P[W_n(\theta, \lambda) \mid \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda, q) \mid \theta]|\} = \epsilon(\lambda, q) \tag{72}$$

where

$$\lim_{q=\infty} \epsilon(\lambda, q) = 0 \tag{73}$$

uniformly in λ. The set $W_n(\theta, \lambda)$ denotes the intersection of W_n and the region $n^{1/2}|\theta - \dot{\theta}_n| \leqq \lambda$.

Let $\{q_i\}$ $(i=1, 2, \cdots, \text{ad inf.})$ be a sequence of positive integers such that $\lim_{i=\infty} q_i = \infty$. Furthermore let $\{\eta_i\}$ be a sequence of positive numbers such that $\lim_{i=\infty} \eta_i = 0$. We define $W_n^*(W_n)$ as follows

$$W_n^*(W_n) = W_n^*(W_n, q_{i+1}) \quad \text{for} \quad n_i < n \leqq n_{i+1} \quad (i = 0, 1, 2, \cdots, \text{ad inf.}). \tag{74}$$

The sequence $\{n_i\}$ $(i=0, 1, 2, \cdots, \text{ad inf.})$ of integers is chosen as follows: Denote by $F_n(\lambda, q)$ the expression

$$\text{l.u.b.}_{\theta, W_n} \left| P[W_n(\theta, \lambda) \mid \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda, q) \mid \theta] \right|.$$

The integer n_0 is put equal to zero and n_i is chosen so that

$$\begin{gathered} n_i > n_{i-1}, \\ F_n(\lambda_i, q_{i+1}) < \epsilon(\lambda_i, q_{i+1}) + \eta_i \end{gathered} \tag{75}$$

for all $n > n_i$, and $\{\lambda_i\}$ denotes a sequence of numbers such that $\lim \lambda_i = \infty$. Let $\lambda_n' = \lambda_i$, $\eta_n' = \eta_i$ and $q_n' = q_{i+1}$ for $n_i < n \leqq n_{i+1}$ $(i=0, 1, \cdots, \text{ad inf.})$. From (72), (73) and (75) we obtain

$$\lim_{n=\infty} \left\{ \text{l.u.b.}_{\theta, W_n} \left| P[W_n(\theta, \lambda_n') \mid \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda_n', q_n') \mid \theta] \right| \right\} = 0. \tag{76}$$

Denote by $W_n^*(W_n, \theta, \lambda)$ the intersection of $W_n^*(W_n)$ and the region $n^{1/2}|\theta - \dot{\theta}_n| \leqq \lambda$. Because of (74) we obviously have

$$W_n^*(W_n, \theta, \lambda_n') = W_n^*(W_n, \theta, \lambda_n', q_n').$$

Hence from (76) we obtain

$$\lim_{n=\infty} \left\{ \text{l.u.b.}_{\theta, W_n} \left| P[W_n(\theta, \lambda_n') \mid \theta] - \mathfrak{P}[W_n^*(W_n, \theta, \lambda_n') \mid \theta] \right| \right\} = 0. \tag{77}$$

Since $\lim_{n=\infty} \lambda_n' = \infty$, it follows from (77) that

$$\lim_{n=\infty} \left\{ \text{l.u.b.}_{\theta, W_n} \left| P[W_n \mid \theta] - \mathfrak{P}[W_n^*(W_n) \mid \theta] \right| \right\} = 0. \tag{78}$$

The region W_n may be any Borel measurable subset of the rn-dimensional sample space. In particular, W_n may be any Borel measurable subset R_n^* in the space of the maximum likelihood estimates. Hence Lemma 2 follows from (78).

On the basis of Lemmas 1 and 2 we can restrict ourselves in case of large

samples to subsets of the space of the maximum likelihood estimates and we can substitute $\mathfrak{P}[V_n^*|\theta]$ for $P[V_n^*|\theta]$. Hence, if the sample is sufficiently large, the problem of statistical inference concerning the unknown parameter θ can be reduced to the case where the variates involved have a joint normal distribution.

5. **Tests of simple hypotheses which have uniformly best average power over a family of surfaces.** For any value c let K_c denote a surface in the parameter space. For instance K_c may be defined by the equation $\phi(\theta) = c$ where $\phi(\theta)$ denotes some analytic function of θ. Consider furthermore a non-negative function $w(\theta)$ of θ, called a weight function. For any function $\psi(\theta)$ of θ the symbol $\int_{K_c}\psi(\theta)dA$ will denote the surface integral of the function $\psi(\theta)$ over the surface K_c.

DEFINITION I. *A critical region W_n is said to have uniformly best average power with respect to the surfaces K_c and the weight function $w(\theta)$ if for any region Z_n of size equal to that of W_n we have $\int_{K_c}P(W_n|\theta)w(\theta)dA \geqq \int_{K_c}P(Z_n|\theta)w(\theta)dA$ for all values c for which K_c is defined.*

Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution. The mean values $\theta^1, \cdots, \theta^k$ of the variates $y^1, \cdots, y^k$ are unknown, but the covariance matrix $\|\sigma_{ij}\|$ $(i, j = 1, \cdots, k)$ is known and is nonsingular. Suppose that we wish to test the simple hypothesis that $\theta = \theta_0$. Consider the family of ellipsoids given by

$$\sum_j \sum_i \lambda_{ij}[\theta^i - \theta_0^i][\theta^j - \theta_0^j] = c, \tag{79}$$

where $\|\lambda_{ij}\| = \|\sigma_{ij}\|^{-1}$. For any c denote by S_c the ellipsoid given by (79). Consider a nonsingular linear transformation of the parameter space

$$\theta'^i - \theta_0^i = \beta_{i1}(\theta^1 - \theta_0^1) + \cdots + \beta_{ik}(\theta^k - \theta_0^k) \tag{80}$$

such that the family of ellipsoids S_c is transformed into a family of concentric spheres with the center at θ_0. Denote by S_c' the image of S_c. For any point θ and for any positive ρ consider the set $\omega(\theta, \rho)$ consisting of all points θ_1 which lie on the same S_c as θ and for which $|\theta_1 - \theta| \leqq \rho$. Let

$$\xi(\theta) = \lim_{\rho=0} \frac{A[\omega'(\theta, \rho)]}{A[\omega(\theta, \rho)]}, \tag{81}$$

where $\omega'(\theta, \rho)$ is the image of $\omega(\theta, \rho)$ and for any set ω, $A(\omega)$ denotes the area of ω.

PROPOSITION II. *If the variates $y^1, \cdots, y^k$ have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and a known covariance matrix $\|\sigma_{ij}\|$, then for testing the hypothesis $\theta = \theta_0$ on the basis of a single observation on each of the variates $y^1, \cdots, y^k$, the critical region given by the inequality*

$$\sum_j \sum_i \lambda_{ij}(y^i - \theta_0^i)(y^j - \theta_0^j) \geqq d \qquad (\|\lambda_{ij}\| = \|\sigma_{ij}\|^{-1}) \tag{82}$$

has uniformly best average power with respect to the surfaces S_c defined in (79) *and the weight function given in* (81).

Proof. Consider the linear transformation

$$y'^i - \theta_0^i = \beta_{i1}(y^1 - \theta_0^1) + \cdots + \beta_{ik}(y^k - \theta_0^k), \tag{83}$$

where the matrix $\|\beta_{ij}\|$ is the same as in (80). The variates y'^i $(i=1, \cdots, k)$ are normally and independently distributed with mean values $\theta_0^1, \cdots, \theta_0^k$ (under the hypothesis $\theta=\theta_0$) and have a common variance σ^2. We will assume $\sigma^2=1$, since this can always be achieved by multiplying the matrix $\|\beta_{ij}\|$ by a proportionality factor. The critical region W given in (82) will be transformed into the region W' given by

$$(y'^1 - \theta_0^1)^2 + \cdots + (y'^k - \theta_0^k)^2 \geqq d. \tag{84}$$

Because of (81) we obviously have

$$\int_{S_c} P(Z \mid \theta)\xi(\theta)dA = \int_{S_c'} P(Z' \mid \theta')dA, \tag{85}$$

where Z denotes an arbitrary region in the space of $y^1, \cdots, y^k$ and Z' is the image of Z in the space of $y'^1, \cdots, y'^k$. Hence in order to prove Proposition II we have merely to show that

$$\int_{S_c'} P(W' \mid \theta')dA \geqq \int_{S_c'} P(Z' \mid \theta')dA \tag{86}$$

for any region Z' in the space of $y'^1, \cdots, y'^k$ which has a size equal to that of W'.

By a lemma of Neyman and Pearson(4) we see easily that (86) is proved, if we can show that there exists a function $d(c)$ of c such that

$$\int_{S_c'} p(y' \mid \theta')dA/p(y' \mid \theta_0) \geqq d(c) \text{ within } W' \tag{87}$$

and

$$\int_{S_c'} p(y' \mid \theta')dA/p(y' \mid \theta_0) \leqq d(c) \text{ outside } W' \tag{88}$$

for all positive values of c, where $p(y' \mid \theta')$ denotes the joint density function of $y'^1, \cdots, y'^k$ under the assumption that the true means are $\theta'^1, \cdots, \theta'^k$.

(4) J. Neyman and E. S. Pearson, *Contributions to the theory of testing statistical hypotheses*, Statistical Research Memoirs vol. 1 (1936).

If we denote $y'^i - \theta_0^i$ by v^i and $\theta'^i - \theta_0^i$ by θ^{*i}, we have

$$p(y' \mid \theta') = \frac{1}{(2\pi)^{k/2}} \exp\left(-2^{-1}\sum (v^i - \theta^{*i})^2\right)$$

and

$$p(y' \mid \theta_0) = \frac{1}{(2\pi)^{k/2}} \exp\left(-2^{-1}\sum (v^i)^2\right).$$

Hence

$$\begin{aligned}\int_{S_c'} p(y' \mid \theta')dA &= \left(\frac{1}{2\pi}\right)^{k/2} \int_{S_c'} \exp\left(-2^{-1}\sum (v^i - \theta^{*i})^2\right)dA \\ &= \left(\frac{1}{2\pi}\right)^{k/2} \exp\left(-2^{-1}\sum (v^i)^2\right) \\ &\quad \cdot \int_{S_c'} \exp\left(\sum v^i\theta^{*i} - 2^{-1}\sum (\theta^{*i})^2\right)dA \\ &= p(y' \mid \theta_0) \exp\left(-2^{-1}\sum (\theta^{*i})^2\right) \int_{S_c'} \exp\left(\sum v^i\theta^{*i}\right)dA,\end{aligned}$$

since $\sum(\theta^{*i})^2$ is constant on the surface S_c'. Hence (87) and (88) can be written

$$I(v^1, \cdots, v^k) = \int_{S_c'} \exp \sum v^i\theta^{*i}dA \geqq d^*(c) \text{ within } W', \tag{89}$$

$$I(v^1, \cdots, v^k) \leqq d^*(c) \text{ outside } W'. \tag{90}$$

Denote $\left|(\sum(v^i)^2)^{1/2}\right|$ by r_v and $\left|(\sum(\theta^{*i})^2)^{1/2}\right|$ by r^*. On the surface S_c' we have $r^* = c$. Denote by $\alpha(\theta^*)$ the angle $(0 \leqq \alpha \leqq \pi)$ between the vector v and the vector θ^*. Then we have

$$I(v^1, \cdots, v^k) = \int_{S_c'} \exp\left(cr_v \cos\left[\alpha(\theta^*)\right]\right)dA.$$

Because of the symmetry of the sphere, the value of this integral will not be changed if we substitute $\beta(\theta^*)$ for $\alpha(\theta^*)$ where $\beta(\theta^*)$ denotes the angle $(0 \leqq \beta(\theta^*) \leqq \pi)$ between the vector θ^* and an arbitrarily chosen fixed vector u. Hence $I(v^1, \cdots, v^k)$ depends only on r_v, that is $I(v^1, \cdots, v^k) = I(r_v)$. The inequalities (89) and (90) are obviously proved if we can show that $I(r_v)$ is a monotonically increasing function of r_v. We have

$$\frac{dI(r_v)}{dr_v} = \int_{S_c'} c \cos\left[\beta(\theta^*)\right] \exp\left(cr_v \cos\left[\beta(\theta^*)\right]\right)dA. \tag{91}$$

Denote by ω_1 the subset of S_c' in which $0 \leqq \beta(\theta^*) \leqq \pi/2$, and by ω_2 the subset in which $\pi/2 \leqq \beta(\theta^*) \leqq \pi$. Because of the symmetry of the sphere we obvi-

ously have

$$\int_{\omega_2} c \cos [\beta(\theta^*)] \exp (cr_v \cos [\beta(\theta^*)])dA$$

$$= \int_{\omega_1} c \cos [\pi - \beta(\theta^*)] \exp (cr_v \cos [\pi - \beta(\theta^*)])dA \tag{92}$$

$$= - \int_{\omega_1} c \cos [\beta(\theta^*)] \exp (- cr_v \cos [\beta(\theta^*)])dA.$$

Hence

$$\frac{dI(r_v)}{dr_v} = c \int_{\omega_1} \cos [\beta(\theta^*)] \{\exp (cr_v \cos [\beta(\theta^*)]) - \exp (-cr_v \cos [\beta(\theta^*)])\} dA. \tag{93}$$

The right-hand side of (93) is positive. Hence Proposition II is proved.

Now let us turn back to the general problem of r variates $x^1, \cdots, x^r$ whose joint probability density function $f(x^1, \cdots, x^r, \theta^1, \cdots, \theta^k) = f(x, \theta)$ involves k unknown parameters, as considered in the previous sections.

DEFINITION II. *A sequence* $\{W_n\}$ $(n = 1, \cdots,$ ad inf.) *of critical regions of size* α *for testing the simple hypothesis* $\theta = \theta_0$ *is said to have asymptotically best average power with respect to the family of surfaces* K_c *and the weight function* $w(\theta)$ *if for any sequence* $\{Z_n\}$ *for which* $P(Z_n | \theta_0) = \alpha$ *we have*

$$\limsup_{n=\infty} \left\{ \text{l.u.b.}_c \left[\int_{K_c} P(Z_n | \theta) \frac{w(\theta)}{A(K_c)} dA - \int_{K_c} P(W_n | \theta) \frac{w(\theta)}{A(K_c)} dA \right] \right\} \leqq 0,$$

where

$$A(K_c) = \int_{K_c} w(\theta)dA.$$

We shall prove the following theorem.

THEOREM I. *Let* W_n^* *be a critical region for testing* $\theta = \theta_0$ *defined by the inequality*

$$n \sum_j \sum_i (\theta_n^i - \theta_0^i)(\theta_n^j - \theta_0^j) c_{ij}(\theta_n) \geqq d_n,$$

where the real number d_n *is chosen so that* $P(W_n^* | \theta_0) = \alpha$. *Denote by* S_c *the surface in the parameter space defined by the equation*

$$\sum \sum (\theta^i - \theta_0^i)(\theta^j - \theta_0^j) c_{ij}(\theta_0) = c.$$

Furthermore let $\xi(\theta)$ *be the weight function as defined in* (81) *where* $\|c_{ij}(\theta_0)\|$ *is substituted for* $\|\lambda_{ij}\|$. *Then the test* $\{W_n^*\}$ *has asymptotically best average power with respect to the family of surfaces* S_c *and the weight function* $\xi(\theta)$.

Proof. Because of Lemma 1 we can restrict ourselves to subsets of the space of the maximum likelihood estimates. Let us assume that Theorem I is not true. Then it follows from Lemma 2 that a sequence of values $\{c_n\}$ and a sequence of regions $\{Z_n^*\}$ exist such that $P(Z_n^* \mid \theta_0) = \alpha$ and

$$\limsup_{n=\infty} \left\{ \int_{S_{c_n}} \mathfrak{P}(Z_n^* \mid \theta) \zeta_n(\theta) dA - \int_{S_{c_n}} \mathfrak{P}(W_n^* \mid \theta) \zeta_n(\theta) dA \right\} = \delta > 0, \tag{94}$$

where

$$\zeta_n(\theta) = \xi(\theta) \Big/ \int_{S_{c_n}} \xi(\theta) dA.$$

From (94) it follows that there exists a subsequence $\{n'\}$ of the sequence $\{n\}$ such that

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \mathfrak{P}(Z_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \mathfrak{P}(W_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA \right\} = \delta > 0. \tag{95}$$

It is easy to verify that

$$\lim_{n=\infty} \mathfrak{P}(W_{n'}^* \mid \theta_{n'}) = 1$$

if $\theta_{n'}$ is a point of $S_{c_{n'}}$ and if $\lim_{n=\infty} n'c_{n'} = +\infty$. Under the latter condition also

$$\lim_{n=\infty} \int_{S_{c_{n'}}} \mathfrak{P}(W_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA = 1.$$

Thus (95) can hold only if the sequence $\{n'c_{n'}\}$ is bounded. If $\{n'c_{n'}\}$ is bounded, we obviously have for any sequence of regions $\{V_n^*\}$

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \mathfrak{P}(V_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \overline{P}(V_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA \right\} = 0,$$

where $\overline{P}(V_n^* \mid \theta)$ denotes the probability of V_n^* calculated under the assumption that $n^{1/2}(\theta_n^1 - \theta^1), \cdots, n^{1/2}(\theta_n^k - \theta^k)$ have a joint normal distribution with zero means and covariance matrix equal to $\|c_{ij}(\theta_0)\|^{-1}$. Hence from (95) we obtain

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \overline{P}(Z_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA - \int_{S_{c_{n'}}} \overline{P}(W_{n'}^* \mid \theta) \zeta_{n'}(\theta) dA \right\} = \delta > 0. \tag{96}$$

Denote by $\overline{W}_n^*$ the region defined by the inequality

$$n \sum \sum (\theta_n^i - \theta_0^i)(\theta_n^j - \theta_0^j) c_{ij}(\theta_0) \geqq \bar{d}_n,$$

where $\bar{d}_n$ is chosen so that $\overline{P}(\overline{W}_n^* \mid \theta_0) = \alpha$. Furthermore denote by $\overline{Z}_n^*$ the sum of Z_n^* and the region $n^{1/2} \mid \theta_0 - \theta_n \mid \leqq \lambda_n$, where λ_n is chosen so that

$\overline{P}(\overline{Z}_n^* \mid \theta_0) = \alpha$. Since, as can easily be seen,

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \overline{P}(\overline{W}_{n'}^* \mid \theta)\zeta_{n'}(\theta)dA - \int_{S_{c_{n'}}} \overline{P}(W_{n'}^* \mid \theta)\zeta_{n'}(\theta)dA \right\} = 0$$

and

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}} \overline{P}(\overline{Z}_{n'}^* \mid \theta)\zeta_{n'}(\theta)dA - \int_{S_{c_{n'}}} \overline{P}(Z_{n'}^* \mid \theta)\zeta_{n'}(\theta)dA \right\} = 0,$$

we obtain from (96) a contradiction to Proposition II. Hence Theorem I is proved.

6. **Tests of simple hypotheses which have best constant power on a family of surfaces.**

DEFINITION III. *A critical region W_n for testing $\theta = \theta_0$ is said to have uniformly best constant power on the family of surfaces $\{K_c\}$ if the following two conditions are fulfilled:*

(a) $P(W_n \mid \theta_1) = P(W_n \mid \theta_2)$ *for any pair of points θ_1, θ_2 which lie on the same surface K_c.*

(b) $P(W_n \mid \theta) \geqq P(Z_n/\theta)$ *for any Z_n which satisfies condition* (a) *and for which $P(Z_n \mid \theta_0) = P(W_n \mid \theta_0)$.*

From Proposition II we obtain the following:

PROPOSITION III. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and a known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. Then for testing $\theta = \theta_0$, the region defined in* (82) *has uniformly best constant power on the surfaces S_c defined by the equation*

$$\sum\sum (\theta^i - \theta_0^i)(\theta^j - \theta_0^j)\lambda_{ij} = c. \tag{97}$$

Since the critical region defined in (82) satisfies condition (a) of Definition III, Proposition III is an immediate consequence of Proposition II.

DEFINITION IV. *A sequence of critical regions $\{W_n\}$ for testing $\theta = \theta_0$ is said to be of size α and to have asymptotically best constant power on the surfaces K_c if the following three conditions are fulfilled:*

(a) $P(W_n \mid \theta_0) = \alpha$ $(n = 1, 2, \cdots,$ ad inf.).

(b) $\lim_{n=\infty} \{\text{l.u.b.}_c [\text{l.u.b.}_{\theta \in K_c} P(W_n \mid \theta) - \text{g.l.b.}_{\theta \in K_c} P(W_n \mid \theta)]\} = 0$, *where the symbol* $\text{l.u.b.}_{\theta \in K_c}$ *means that the least upper bound is to be taken with respect to θ restricting θ to points of K_c.*

(c) *For any sequence $\{Z_n\}$ which satisfies* (a) *and* (b) *we have*

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta} [P(Z_n \mid \theta) - P(W_n \mid \theta)]\} = 0.$$

It is easy to verify that the sequence $\{W_n^*\}$ defined in Theorem I satis-

fies the conditions (a) and (b) for $K_c = S_c$, where S_c denotes the surface defined in Theorem I. Thus from Theorem I we obtain the following theorem.

THEOREM II. *Let* $\{W_n^*\}$ *and* S_c *be defined as in Theorem* I. *For testing* $\theta = \theta_0$, *the sequence* $\{W_n^*\}$ *has asymptotically best constant power on the surfaces* S_c.

7. **Most stringent tests of simple hypotheses.** Let θ and θ_0 be two parameter points and let α denote a positive number less than 1. We denote by $P_n(\theta, \theta_0, \alpha)$ the least upper bound of $P(W_n \mid \theta)$ with respect to W_n, where W_n is restricted to regions for which $P(W_n \mid \theta_0) = \alpha$. It is clear that if W_n is a critical region of size α for testing $\theta = \theta_0$, its power function can nowhere exceed the value of $P_n(\theta, \theta_0, \alpha)$, that is $P(W_n \mid \theta) \leqq P_n(\theta, \theta_0, \alpha)$ for all values of θ.

DEFINITION V. *A critical region* W_n *is said to be a most stringent test of the hypothesis* $\theta = \theta_0$ *on the level of significance* α *if* $P(W_n \mid \theta_0) = \alpha$ *and if*

$$\text{l.u.b.}_{\theta} \left[P_n(\theta, \theta_0, \alpha) - P(W_n \mid \theta)\right] \leqq \text{l.u.b.}_{\theta} \left[P_n(\theta, \theta_0, \alpha) - P(Z_n \mid \theta)\right]$$

for all regions Z_n *for which* $P(Z_n \mid \theta_0) = \alpha$.

We shall prove the following proposition.

PROPOSITION IV. *Let* $y^1, \cdots, y^k$ *be* k *variates which have a joint normal distribution with unknown mean values* $\theta^1, \cdots, \theta^k$ *and a known covariance matrix* $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. *Then for testing* $\theta = \theta_0$ *the region* W *defined in* (82) *is a most stringent test.*

Proof. We shall assume that Proposition IV does not hold and we shall arrive at a contradiction. If Proposition IV is not true, then there exists a region Z in the space of $y^1, \cdots, y^k$ such that $P(Z \mid \theta_0) = \alpha$ and

$$\text{(98)} \qquad \text{l.u.b.}_{\theta} \left[P(\theta, \theta_0, \alpha) - P(W \mid \theta)\right] > \text{l.u.b.}_{\theta} \left[P(\theta, \theta_0, \alpha) - P(Z \mid \theta)\right].$$

Let S_c be the surface defined by the equation

$$\sum_j \sum_i (\theta^i - \theta_0^i)(\theta^j - \theta_0^j)\lambda_{ij} = c.$$

The functions $P_1(\theta, \theta_0, \alpha)$ and $P(W \mid \theta)$ are constant on the surface S_c. Hence, on account of (98), there exists a value c_0 such that

$$P(Z \mid \theta) > P(W \mid \theta)$$

for all points θ on S_{c_0}. But this is a contradiction to Proposition II. Hence Proposition IV is proved.

DEFINITION VI. *A sequence of critical regions* $\{W_n\}$ *is said to be an asymptotically most stringent test of the hypothesis* $\theta = \theta_0$ *on the level of significance* α

if $P(W_n|\theta_0)=\alpha$ *and if for any* $\{Z_n\}$ *for which* $P(Z_n|\theta_0)=\alpha$ *we have*

$$\limsup_{n=\infty}\{\text{l.u.b.}_{\theta}[P_n(\theta,\theta_0,\alpha)-P(W_n|\theta)]-\text{l.u.b.}_{\theta}[P_n(\theta,\theta_0,\alpha)-P(Z_n|\theta)]\}\leqq 0.$$

We shall prove the following theorem.

THEOREM III. *Let* W_n^* *be the region defined in Theorem* I. *Then the sequence* $\{W_n^*\}$ *is an asymptotically most stringent test of the hypothesis* $\theta=\theta_0$.

Proof. Denote by $\mathfrak{P}_n(\theta,\theta_0,\alpha)$ the least upper bound of $\mathfrak{P}(Z_n^*|\theta)$ with respect to Z_n^*, where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which $\mathfrak{P}(Z_n^*|\theta_0)=\alpha$. Because of Lemma 1 we have

$$\lim_{n=\infty}[P_n(\theta,\theta_0,\alpha)-\mathfrak{P}_n(\theta,\theta_0,\alpha)]=0 \tag{99}$$

uniformly in θ. Denote by $\overline{P}_n(\theta,\theta_0,\alpha)$ the least upper bound of $\overline{P}(Z_n^*|\theta)$ with respect to Z_n^*, where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which $\overline{P}(Z_n^*|\theta_0)=\alpha$. The symbol $\overline{P}(V_n^*|\theta)$ denotes the probability of V_n^* calculated under the assumption that the joint distribution of $n^{1/2}(\hat{\theta}_n^1-\theta^1),\cdots,n^{1/2}(\hat{\theta}_n^k-\theta^k)$ is normal with zero means and covariance matrix $\|c_{ij}(\theta_0)\|^{-1}$. For any positive λ we have

$$\lim_{n=\infty}[\mathfrak{P}_n(\theta,\theta_0,\alpha)-\overline{P}_n(\theta,\theta_0,\alpha)]=0 \tag{100}$$

uniformly in θ in the domain $|n^{1/2}(\theta-\theta_0)|\leqq\lambda$. Since for any sequence $\{\theta_n\}$ for which $\lim|n^{1/2}(\theta_n-\theta_0)|=+\infty$, we have

$$\lim_{n=\infty}\mathfrak{P}_n(\theta_n,\theta_0,\alpha)=\lim_{n=\infty}\overline{P}_n(\theta_n,\theta_0,\alpha)=1,$$

we obtain from (100)

$$\lim_{n=\infty}[\mathfrak{P}_n(\theta,\theta_0,\alpha)-\overline{P}_n(\theta,\theta_0,\alpha)]=0 \tag{101}$$

uniformly in θ. For any c let S_c be the surface defined by

$$\sum_{j=1}^{k}\sum_{i=1}^{k}(\theta^i-\theta_0^i)(\theta^j-\theta_0^j)c_{ij}(\theta_0)=c.$$

Obviously $\overline{P}_n(\theta,\theta_0,\alpha)$ is constant along the surface S_c. From (99) and (101) we obtain

$$\lim_{n=\infty}\{\text{l.u.b.}_{\theta\in S_c}P_n(\theta,\theta_0,\alpha)-\text{g.l.b.}_{\theta\in S_c}P_n(\theta,\theta_0,\alpha)\}=0 \tag{102}$$

uniformly in θ. We shall derive a contradiction from the assumption that Theorem III is not true. If Theorem III is not true, there exists a sequence $\{Z_n\}$ of regions such that $P(Z_n|\theta_0)=\alpha$ and

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta} [P_n(\theta, \theta_0, \alpha) - P(W_n^* | \theta)] - \text{l.u.b.}_{\theta} [P_n(\theta, \theta_0, \alpha) - P(Z_n | \theta)]\} = \delta > 0. \tag{103}$$

On account of (102) and since

$$\lim_{n=\infty} [\text{l.u.b.}_{\theta \in S_c} P(W_n^* | \theta) - \text{g.l.b.}_{\theta \in S_c} P(W_n^* | \theta)] = 0$$

uniformly in c (see Theorem II), we obtain from (103) that there exists a sequence $\{c_n\}$ and a subsequence $\{n'\}$ of $\{n\}$ such that for all points $\theta_{n'}$ of $S_{c_{n'}}$

$$P(Z_{n'} | \theta_{n'}) > P(W_{n'}^* | \theta_{n'}) + \delta/2$$

for all n greater than a certain n_0. But this contradicts Theorem I. Hence Theorem III is proved.

8. **Definitions of "best" tests of composite hypotheses.** In this section we shall extend the definitions given in the previous sections to the case of composite hypotheses. Let ω be a subset of the parameter space and denote by H_ω the hypothesis that the true parameter point is contained in ω. In all that follows the letter θ printed in boldface will indicate that the parameter point lies in ω. For example, the symbol $\text{l.u.b.}_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$ denotes the least upper bound of the function $f(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ where $\boldsymbol{\theta}$ is restricted to points of ω. For any point $\boldsymbol{\theta}$ and for any real value c let $K_c(\boldsymbol{\theta})$ denote a surface in the parameter space. For instance $K_c(\boldsymbol{\theta})$ may be given by r equations in θ

$$\phi_1(\theta, \boldsymbol{\theta}) = \cdots = \phi_r(\theta, \boldsymbol{\theta}) = 0,$$

where $\phi_1(\theta, \boldsymbol{\theta}), \cdots, \phi_r(\theta, \boldsymbol{\theta})$ are some analytic functions of θ and $\boldsymbol{\theta}$.

DEFINITION VII. *A critical region W_n for testing H_ω is said to have uniformly best average power with respect to a family of surfaces $K_c(\boldsymbol{\theta})$ and a weight function $w(\theta)$ if for any Z_n for which*

$$\text{l.u.b.}_{\boldsymbol{\theta}} P(Z_n | \boldsymbol{\theta}) = \text{l.u.b.}_{\boldsymbol{\theta}} P(W_n | \boldsymbol{\theta})$$

we have

$$\int_{K_c(\boldsymbol{\theta})} P(W_n | \theta) w(\theta) dA \geqq \int_{K_c(\boldsymbol{\theta})} P(Z_n | \theta) w(\theta) dA$$

for any $\boldsymbol{\theta}$ and for any c for which $K_c(\boldsymbol{\theta})$ is defined.

DEFINITION VIII. *A sequence $\{W_n\}$ ($n = 1, 2, \cdots$, ad inf.) of critical regions for testing the hypothesis H_ω is said to have asymptotically best average power with respect to a family of surfaces $K_c(\boldsymbol{\theta})$ and a weight function $w(\theta)$ if the following two conditions are fulfilled:*

(a) *There exists a fixed* α *such that*

$$\text{l.u.b.}_{\theta}\ P(W_n \mid \theta) = \alpha \qquad (n = 1, 2, \cdots, \text{ad inf.})$$

(b) *For any sequence* $\{Z_n\}$ *for which* $\text{l.u.b.}_{\theta}\ P(Z_n \mid \theta) = \alpha$, *we have*

$$\limsup_{n=\infty} \left\{ \text{l.u.b.}_{c,\theta} \left[\int_{K_c(\theta)} P(Z_n \mid \theta) \frac{w(\theta)}{A[K_c(\theta)]} dA - \int_{K_c(\theta)} P(W_n \mid \theta) \frac{w(\theta)}{A[K_c(\theta)]} dA \right] \right\} \leqq 0,$$

where

$$A[K_c(\theta)] = \int_{K_c(\theta)} w(\theta) dA.$$

DEFINITION IX. *A critical region* W_n *for testing* H_ω *is said to have uniformly best constant power on the family of surfaces* $K_c(\theta)$ *if the following two conditions are fulfilled*:

(a) $P(W_n \mid \theta') = P(W_n \mid \theta'')$ *for all pairs of points* θ' *and* θ'' *which lie on the same surface* $K_c(\theta)$.

(b) $P(W_n \mid \theta) \geqq P(Z_n \mid \theta)$ *for any* θ *not in* ω *and for any* Z_n *which satisfies* (a) *and the condition*

$$\text{l.u.b.}_{\theta}\ P(Z_n \mid \theta) = \text{l.u.b.}_{\theta}\ P(W_n \mid \theta).$$

DEFINITION X. *A sequence of critical regions* $\{W_n\}$ *for testing* H_ω *is said to have asymptotically best constant power on the surfaces* $K_c(\theta)$ *if the following three conditions are fulfilled*:

(a) $\text{l.u.b.}_{\theta}\ P(W_n \mid \theta) = \alpha$ $(n = 1, 2, \cdots, \text{ad inf.})$.

(b) $\lim_{n=\infty} \{\text{l.u.b.}_{c,\theta} [\text{l.u.b.}_{\theta \in K_c(\theta)}\ P(W_n \mid \theta) - \text{g.l.b.}_{\theta \in K_c(\theta)}\ P(W_n \mid \theta)]\} = 0$.

(c) *For any sequence* $\{Z_n\}$ *which satisfies* (a) *and* (b) *we have*

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta \in \bar{\omega}} [P(Z_n \mid \theta) - P(W_n \mid \theta)]\} = 0,$$

where $\bar{\omega}$ *is the complement of* ω.

DEFINITION XI. *Denote by* $P_n(\theta, \omega, \alpha)$ *the least upper bound of* $P(Z_n \mid \theta)$ *with respect to* Z_n *subject to the condition* $\text{l.u.b.}_{\theta}\ P(Z_n \mid \theta) = \alpha$. *A critical region* W_n *is said to be a most stringent test of the hypothesis* H_ω *if for some positive* α

$$\text{l.u.b.}_{\theta}\ P(W_n \mid \theta) = \alpha$$

and

$$\text{l.u.b.}_{\theta} [P_n(\theta, \omega, \alpha) - P(W_n \mid \theta)] \leqq \text{l.u.b.}_{\theta} [P_n(\theta, \omega, \alpha) - P(Z_n \mid \theta)]$$

for all regions Z_n *for which* $\text{l.u.b.}_{\theta}\ P(Z_n \mid \theta) = \alpha$.

DEFINITION XII. *A sequence of critical regions* $\{W_n\}$ *is said to be an asymptotically most stringent test of the hypothesis* H_ω *if the following two conditions are fulfilled*:

(a) *There exists a positive* α *such that*

$$\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha \qquad (n = 1, 2, \cdots, \text{ad inf.}).$$

(b) *For any sequence* $\{Z_n\}$ *which satisfies* (a) *for the same* α *we have*

$$\limsup_{n=\infty} \{\text{l.u.b.}_{\theta} [P_n(\theta, \omega, \alpha) - P(W_n \mid \theta)] - \text{l.u.b.}_{\theta} [P_n(\theta, \omega, \alpha) - P(Z_n \mid \theta)]\} \leqq 0.$$

In Definitions VII–XII we have formulated the condition

$$\text{l.u.b.}_{\theta}\, P(W_n \mid \theta) = \alpha.$$

The question can be raised whether, in place of this condition, the requirement that

$$(*) \qquad P(W_n \mid \theta) = \alpha$$

for all points θ should be made; or whether the weaker condition that

$$(**) \qquad \lim_{n=\infty} P(W_n \mid \theta) = \alpha$$

uniformly in θ should be required. Condition (*) has the serious drawback that regions satisfying it do not always exist. Even in cases where (*) can be fulfilled, it imposes too strong a restriction on the possible choice of W_n, which does not seem to be quite justified. It is conceivable that in some cases a region W_n' may exist which does not satisfy (*) but has such an advantageous power function that we prefer it to any region W_n which satisfies(*).

As to the condition (**), we shall see that it is satisfied for the sequence $\{W_n\}$ which is shown in this paper to be asymptotically best according to all three definitions VIII, X and XII. Hence the same sequence $\{W_n\}$ remains asymptotically best if we replace the condition l.u.b.$_\theta$ $P(W_n \mid \theta) = \alpha$ by (**) in the definitions VIII, X and XII.

In the following §§9–11 we shall discuss a linear hypothesis of the following type: $\theta^1 = \theta_0^1, \cdots, \theta^r = \theta_0^r$ $(r < k)$, where $\theta_0^1, \cdots, \theta_0^r$ are some specified values. That is to say, the set ω is the set of all points θ for which the above equations hold. In §12 the general composite hypothesis will be discussed.

9. **Tests of linear composite hypotheses which have uniformly best average power over a family of surfaces.** Let H_ω be the hypothesis that $\theta^1 = \theta_n^1, \cdots, \theta^r = \theta_0^r$ $(r < k)$. We shall introduce the following notation: For any parameter point $\theta = (\theta^1, \cdots, \theta^k)$ the symbol ${}_1\theta$ will denote the vector in the r-dimensional space with the components $\theta^1, \cdots, \theta^r$, and ${}_2\theta$ will denote the vector in the $k-r$ dimensional space with the components $\theta^{r+1}, \cdots, \theta^k$. For any function $\psi(\theta)$ of θ we shall use the synonymous nota-

tion $\psi({}_1\theta, {}_2\theta)$. For instance $P(W|{}_1\theta, {}_2\theta)$ is synonymous with $P(W|\theta)$.

Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$ $(i, j = 1, \cdots, k)$, which is nonsingular. Denote by W the region in the space of $y^1, \cdots, y^k$ given by the inequality

$$\sum_{q=1}^{r}\sum_{p=1}^{r}\bar{\lambda}_{pq}(y^p - \theta_0^p)(y^q - \theta_0^q) \geqq d, \tag{104}$$

where $\|\bar{\lambda}_{pq}\| = \|\sigma_{pq}\|^{-1}$ $(p, q = 1, \cdots, r)$. Consider the nonsingular linear transformation of the variates $y^1, \cdots, y^k$ given by the equations

$$\begin{aligned} y'^p - \theta_0^p &= \beta_{p1}(y^1 - \theta_0^1) + \cdots + \beta_{pr}(y^r - \theta_0^r) && (p = 1, \cdots, r), \\ y'^t &= \gamma_{t1}y^1 + \cdots + \gamma_{tk}y^k && (t = r+1, \cdots, k), \end{aligned} \tag{105}$$

such that $y'^1, \cdots, y'^k$ are independently distributed with unit variances. Denote by $S_c(\theta)$ the surface given by the equations

$$\begin{aligned} &\sum_{q=1}^{r}\sum_{p=1}^{r}\bar{\lambda}_{pq}(\theta^p - \theta_0^p)(\theta^q - \theta_0^q) = c, \\ &\gamma_{t1}\theta^1 + \cdots + \gamma_{tk}\theta^k = \sum_{i=1}^{k}\gamma_{ti}\theta^i \qquad (t = r+1, \cdots, k). \end{aligned} \tag{106}$$

Consider the transformation of the parameter space given by

$$\begin{aligned} \theta'^p - \theta_0^p &= \beta_{p1}(\theta^1 - \theta_0^1) + \cdots + \beta_{pr}(\theta^r - \theta_0^r) && (p = 1, \cdots, r), \\ \theta'^t &= \gamma_{t1}\theta^1 + \cdots + \gamma_{tk}\theta^k && (t = r+1, \cdots, k), \end{aligned} \tag{107}$$

where the coefficients β_{pq} and γ_{tj} are the same as in (105). The transformation (107) transforms the surface $S_c(\theta)$ into a sphere $S'_c(\theta)$ given by

$$\sum_{p=1}^{r}(\theta'^p - \theta_0^p)^2 = c, \qquad \theta'^t = \sum_{i=1}^{k}\gamma_{ti}\theta^i = \theta'^t. \tag{108}$$

For any point θ and for any positive ρ consider the set $\omega(\theta, \rho)$ consisting of all points $^*\theta$ which lie on the same $S_c(\theta)$ as θ, and for which $|{}^*\theta - \theta| \leqq \rho$. Let

$$\xi(\theta) = \lim_{\rho=0}\frac{A[\omega'(\theta, \rho)]}{A[\omega(\theta, \rho)]}, \tag{109}$$

where $\omega'(\theta, \rho)$ is the image of $\omega(\theta, \rho)$ (by transformation (107)) and, for any set ω, $A(\omega)$ denotes the $(r-1)$-dimensional area of ω. We shall prove the following proposition.

PROPOSITION V. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. For testing the hypothesis ${}_1\theta = {}_1\theta_0$ on the basis of a single*

observation on each of the variates $y^1, \cdots, y^k$, *the critical region* W *given in* (104) *has uniformly best average power with respect to the family of surfaces* $S_c(\theta)$ *defined in* (106) *and the weight function* $\xi(\theta)$ *given in* (109).

Proof. Because of (109) we have

$$\int_{S_c(\theta)} P(Z \mid \theta)\xi(\theta)dA = \int_{S_c'(\theta)} P(Z' \mid \theta')dA, \tag{110}$$

where Z denotes an arbitrary region in the space of $y^1, \cdots, y^k$ and Z' is the image of Z by transformation (105). The region W is transformed into W' given by

$$(y'^1 - \theta_0^1)^2 + \cdots + (y'^r - \theta_0^r)^2 \geqq d. \tag{111}$$

In order to prove Proposition V we have merely to show that

$$\int_{S_c'(\theta)} P(W' \mid \theta')dA \geqq \int_{S_c'(\theta)} P(Z' \mid \theta')dA \tag{112}$$

for any $c>0$, for any θ, and for any region Z' in the space of $y'^1, \cdots, y'^k$ for which l.u.b.$_{{}_2\theta'} P(Z' \mid {}_1\theta_0, {}_2\theta') = $l.u.b.$_{{}_2\theta'} P(W' \mid {}_1\theta_0, {}_2\theta')$. For any point θ' of $S_c'(\theta)$ we have ${}_2\theta' = {}_2\theta'$. By a lemma of Neyman and Pearson, (112) is proved if we can show that there exists a function $d(c)$ such that

$$\left.\begin{aligned} &\frac{\int_{S_c'(\theta)} p(y'^1, \cdots, y'^k \mid {}_1\theta', {}_2\theta')dA}{p(y'^1, \cdots, y'^k \mid {}_1\theta_0, {}_2\theta')} \\ &\qquad \geqq d(c) \text{ within } W', \\ &\qquad \leqq d(c) \text{ outside } W' \end{aligned}\right\} \tag{113}$$

for all values of c and θ where $p(y'^1, \cdots, y'^k \mid \theta')$ denotes the joint probability density of $y'^1, \cdots, y'^k$ under the assumption that θ' is the true parameter point. Obviously

$$\frac{p(y'^1, \cdots, y'^k \mid {}_1\theta', {}_2\theta')}{p(y'^1, \cdots, y'^k \mid {}_1\theta_0, {}_2\theta')} = \frac{p(y'^1, \cdots, y'^r \mid {}_1\theta')}{p(y'^1, \cdots, y'^r \mid {}_1\theta_0)}.$$

Hence (113) is equivalent to

$$\left.\begin{aligned} &\frac{\int_{S_c'(\theta)} p(y'^1, \cdots, y'^r \mid {}_1\theta')dA}{p(y'^1, \cdots, y'^r \mid {}_1\theta_0)} \\ &\qquad \geqq d(c) \text{ within } W', \\ &\qquad \leqq d(c) \text{ outside } W'. \end{aligned}\right\} \tag{114}$$

The proof of (114) is omitted, since it is the same as that of the inequalities (87) and (88). Hence Proposition V is proved.

Let the critical region W_n^* be defined by the following inequality

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}(\dot{\theta}_n^p - \theta_0^p)(\dot{\theta}_n^q - \theta_0^q)\bar{c}_{pq}(\dot{\theta}_n) \geqq d_n, \tag{115}$$

where $\|\bar{c}_{pq}(\theta)\| = \|\sigma_{pq}(\theta)\|^{-1}$ $(p, q=1, \cdots, r)$ and $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$ $(i, j$

$=1, \cdots, k$). The constant d_n is chosen so that

$$\text{l.u.b.}_{\theta} P(W^* \mid \theta) = \alpha.$$

Let $Z_n^i(\theta) = n^{1/2}(\theta_n^i - \theta^i)$ $(i=1, \cdots, k)$ and consider the nonsingular linear transformation

$$\begin{aligned} \overline{Z}_n^p(\theta) &= \beta_{p1}(\theta)Z_n^1(\theta) + \cdots + \beta_{pr}(\theta)Z_n^r(\theta) \qquad (p = 1, \cdots, r), \\ \overline{Z}_n^t(\theta) &= \gamma_{t1}(\theta)Z_n^1(\theta) + \cdots + \gamma_{tk}(\theta)Z_n^k(\theta) \qquad (t = r+1, \cdots, k), \end{aligned} \tag{116}$$

such that $\overline{Z}_n^1(\theta), \cdots, \overline{Z}_n^k(\theta)$ would be independently distributed with unit variances if the covariance matrix of $Z_m^1(\theta), \cdots, Z_n^k(\theta)$ were given by $\|\sigma_{ij}(\theta)\|$. Denote by $S_c(\theta)$ the surface defined by the equations

$$\begin{aligned} \sum_{q=1}^{r}\sum_{p=1}^{r}(\theta^p - \theta_0^p)(\theta^q - \theta_0^q)\bar{c}_{pq}(\theta) &= c, \\ \gamma_{t1}(\theta)\theta^1 + \cdots + \gamma_{tk}(\theta)\theta^k &= \sum \gamma_{ti}(\theta)\theta^i \quad (t = r+1, \cdots, k). \end{aligned} \tag{117}$$

For any positive δ denote by S_δ the set of all points θ for which $|\theta^p - \theta_0^p| \leqq \delta$ $(p=1, \cdots, r)$. We shall prove the existence of a positive δ such that for any point θ in S_δ there exists exactly one surface $S_c(\theta)$, that is exactly one value of c and exactly one point θ, such that θ lies on the surface $S_c(\theta)$. This statement is obviously proved if we show that for any point θ in S_δ the set of $k-r$ equations

$$\gamma_{t1}(\theta)(\theta^1 - \theta^1) + \cdots + \gamma_{tk}(\theta)(\theta^k - \theta^k) = 0 \qquad (t = r+1, \cdots, k) \tag{118}$$

has a unique solution in the unknowns $\theta^{r+1}, \cdots, \theta^k$. From the definition of the quantities $\beta_{pq}(\theta)$ and $\gamma_{ti}(\theta)$ it follows that

$$A(\theta)\|\sigma_{ij}(\theta)\|\overline{A}(\theta) = I, \tag{119}$$

where $A(\theta)$ denotes the matrix

$$\left\| \begin{array}{ccccccc} \beta_{11}(\theta) & \beta_{12}(\theta) & \cdots & \beta_{1r}(\theta) & 0 & \cdots & 0 \\ \beta_{12}(\theta) & \beta_{22}(\theta) & \cdots & \beta_{2r}(\theta) & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \beta_{r1}(\theta) & \beta_{r2}(\theta) & \cdots & \beta_{rr}(\theta) & 0 & \cdots & 0 \\ \gamma_{r+1\,1}(\theta) & \gamma_{r+1\,2}(\theta) & \cdots & \gamma_{r+1\,r}(\theta) & \gamma_{r+1\,r+1}(\theta) & \cdots & \gamma_{r+1\,k}(\theta) \\ \cdot & \cdot & \cdots & \cdot & \cdot & \cdots & \cdot \\ \gamma_{k1}(\theta) & \gamma_{k2}(\theta) & \cdots & \gamma_{kr}(\theta) & \gamma_{k\,r+1}(\theta) & \cdots & \gamma_{kk}(\theta) \end{array} \right\|, \tag{120}$$

$\overline{A}(\theta)$ is the transposed of $A(\theta)$ and I denotes the unit matrix. Since $\sigma_{ij}(\theta)$ is a continuous and bounded function of θ and since the determinant $|\sigma_{ij}(\theta)|$ has a positive lower bound, we find that $\beta_{pq}(\theta)$ and $\gamma_{ti}(\theta)$ are continuous and

bounded functions of θ and that the absolute value of the determinant $A(\theta)$ has a positive lower bound. Hence also the absolute value of the determinant

$$\gamma(\theta) = \begin{vmatrix} \gamma_{r+1\,r+1}(\theta) & \cdots & \gamma_{r+1\,k}(\theta) \\ \cdot & \cdots & \cdot \\ \gamma_{k\,r+1}(\theta) & \cdots & \gamma_{kk}(\theta) \end{vmatrix} \tag{121}$$

has a positive lower bound.

Let $\theta = \theta^*$ where θ^* denotes an arbitrary point of ω. Then, since the determinant in (121) has a positive lower bound, the equations (118) have a unique solution in the unknowns $\theta^{r+1}, \cdots, \theta^k$, namely the solution $\theta = \theta^*$. Furthermore we see that the Jacobian of the equations (118), taken at the point $\theta = \theta^*$, is equal to $\gamma(\theta^*)$. Since the absolute value of $\gamma(\theta^*)$ has a positive lower bound, there exists a positive δ such that the equations (118) have a unique solution in θ if $|\theta - \theta^*| \leqq \delta$. This proves the existence of a positive δ such that for any point θ in S_δ there exists exactly one surface $S_c(\theta)$ such that θ lies on $S_c(\theta)$.

Since for the critical region W_n^* defined in (115) we obviously have $\lim_{n=\infty} P(W_n^* | \theta) = 1$ uniformly over the domain $|\theta^p - \theta_0^p| \geqq \delta$, we shall restrict ourselves to the consideration of points θ for which $|\theta^p - \theta_0^p| \leqq \delta$ $(p = 1, \cdots, r)$.

Consider the transformation of the parameter space given by

$$\begin{aligned} \theta'^p &= \theta_0^p = \beta_{p1}(\theta)(\theta^1 - \theta_0^1) + \cdots + \beta_{pr}(\theta)(\theta^r - \theta_0^r) \qquad (p = 1, \cdots, r), \\ \theta'^t &= \gamma_{t1}(\theta)\theta^1 + \cdots + \gamma_{tk}(\theta)\theta^k \qquad (t = r+1, \cdots, k), \end{aligned} \tag{122}$$

where θ denotes the point for which θ lies on $S_c(\theta)$. The transformation (122) transforms $S_c(\theta)$ into the sphere $S_c'(\theta)$ given by

$$\sum_{p=1}^{r} (\theta'^p - \theta_0^p)^2 = c; \quad \theta'^t = \sum \gamma_{ti}(\theta)\theta^i = \theta'^t \qquad (t = r+1, \cdots, k). \tag{123}$$

We define a weight function $\xi(\theta)$ as follows:

$$\xi(\theta) = \lim_{\rho=0} \frac{A[\omega'(\theta, \rho)]}{A[\omega(\theta, \rho)]}, \tag{124}$$

where the symbols on the right-hand side have the same meaning as in (109).

THEOREM IV. *Let the critical region W_n^* for testing ${}_1\theta = {}_1\theta_0$ be the region defined in* (115). *Furthermore, let $S_c(\theta)$ be the surface defined in* (117) *and let $\xi(\theta)$ be the weight function defined in* (124). *Then $\{W_n^*\}$ has asymptotically best average power with respect to the family of surfaces $S_c(\theta)$ and the weight function $\xi(\theta)$.*

Proof. Because of Lemma 1, we can restrict ourselves to subsets of the space of the maximum likelihood estimates. Because of Lemma 2, Theo-

rem IV is proved if we show that for any $\{Z_n^*\}$ for which l.u.b.$_\theta$ $P(Z_n^* | \theta) = \alpha$ we have

$$\limsup_{n=\infty} \left\{ \text{l.u.b.}_{c,\theta} \left[\int_{S_c(\theta)} \mathfrak{P}(Z_n^* | \theta)\zeta_n(\theta)dA - \int_{S_c(\theta)} \mathfrak{P}(W_n^* | \theta)\zeta_n(\theta)dA \right] \right\} \leqq 0,$$

where

$$\zeta_n(\theta) = \xi(\theta) \Big/ \int_{S_c(\theta)} \xi(\theta)dA.$$

If Theorem IV were not true, there would exist a sequence $\{c_n\}$, a sequence $\{\theta_n\}$, a sequence $\{Z_n^*\}$, and a subsequence $\{n'\}$ of $\{n\}$ such that

$$\text{l.u.b.}_{\theta} P(Z_n^* | \theta) = \alpha,$$

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}(\theta_{n'})} \mathfrak{P}(Z_{n'}^* | \theta)\zeta_{n'}(\theta)dA - \int_{S_{c_{n'}}(\theta_{n'})} \mathfrak{P}(W_{n'}^* | \theta)\zeta_{n'}(\theta)dA \right\} = \delta > 0. \tag{125}$$

It is easy to verify that for any sequence $\{c_n\}$ for which $\lim nc_n = +\infty$ we have $\int_{S_{c_n}(\theta)} \mathfrak{P}(W_n^* | \theta)\zeta_n(\theta)dA = 1$ uniformly in θ. Hence (125) can hold only if the sequence $\{n'c_{n'}\}$ is bounded. If $\{n'c_{n'}\}$ is bounded, for any sequence of regions $\{V_n^*\}$ we obviously have

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}(\theta_{n'})} \mathfrak{P}(V_{n'}^* | \theta)\zeta_{n'}(\theta)dA - \int_{S_{c_{n'}}(\theta_{n'})} P_{\theta_{n'}}(V_{n'}^* | \theta)\zeta_{n'}(\theta)dA \right\} = 0, \tag{126}$$

where $P_\theta(V_n^* | \theta)$ denotes the probability of V_n^* calculated under the assumption that $n^{1/2}(\theta_n^1 - \theta^1), \cdots, n^{1/2}(\theta_n^k - \theta^k)$ have a joint normal distribution with zero means and covariance matrix $\sigma_{ij}(\theta)$. Let $W_n^*(\theta)$ be the region defined by

$$n \sum_{q=1}^{r} \sum_{p=1}^{r} (\theta_n^p - \theta_0^p)(\theta_n^q - \theta_0^q)\bar{c}_{pq}(\theta) \geqq d_n.$$

It is clear that if $\{n'c_{n'}\}$ is bounded, we have

$$\lim_{n=\infty} \{ P_\theta(W_{n'}^* | \theta) - P_\theta[W_{n'}^*(\theta) | \theta] \} = 0 \tag{127}$$

uniformly in θ and θ over the domain in which θ is a point of $S_{c_{n'}}(\theta)$. From (125), (126) and (127) we obtain

$$\lim_{n=\infty} \left\{ \int_{S_{c_{n'}}(\theta_{n'})} P_{\theta_{n'}}(Z_{n'}^* | \theta)\zeta_{n'}(\theta)dA - \int_{S_{c_{n'}}(\theta_{n'})} P_{\theta_{n'}}[W_{n'}^*(\theta_{n'}) | \theta]\zeta_{n'}(\theta)dA \right\} = \delta > 0. \tag{128}$$

The surface $S_c(\theta)$ defined in (106) is identical with the surface $S_c(\theta)$ defined

in (117) if in (106) we substitute $c_{ij}(\theta)$ for λ_{ij}, that is we substitute $\bar{c}_{pq}(\theta)$ for $\bar{\lambda}_{pq}$. Similarly, if we substitute $c_{ij}(\theta)$ for λ_{ij} for any point of the surface $S_c(\theta)$, the value of the weight function $\xi(\theta)$ defined in (109) is the same as the value of $\xi(\theta)$ defined in (124). Hence, since

$$\lim_{n=\infty} [\text{l.u.b.}_{\theta} P_{\theta}(Z_{n'}^* \mid \theta)] = \alpha,$$

equation (128) is in contradiction to Proposition V. Thus Theorem IV is proved.

10. **Tests of linear composite hypotheses which have best constant power on a family of surfaces.** The critical region W defined in (104) satisfies condition (a) of Definition IX if $K_c(\theta)$ is equal to $S_c(\theta)$ given in (106). Hence from Proposition V we obtain the following proposition.

PROPOSITION VI. *The region W given in* (104) *has uniformly best constant power on the surfaces $S_c(\theta)$ defined in* (106).

If W_n^* is the region defined in (115) and if $K_c(\theta)$ is equal to the surface $S_c(\theta)$ defined in (117), then $\{W_n^*\}$ satisfies conditions (a) and (b) of Definition X. Hence, from Theorem IV we easily obtain the following theorem.

THEOREM V. *Let W_n^* be the region defined in* (115) *and let $S_c(\theta)$ be the surface defined in* (117), *then for testing ${}_1\theta = {}_1\theta_0$, $\{W_n^*\}$ has asymptotically best constant power along the surfaces $S_c(\theta)$.*

11. **Most stringent tests of linear composite hypotheses.** We shall prove the following proposition.

PROPOSITION VII. *Let $y^1, \cdots, y^k$ be k variates which have a joint normal distribution with unknown mean values $\theta^1, \cdots, \theta^k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. For testing the hypothesis ${}_1\theta = {}_1\theta_0$ on the basis of a single observation on each of the variates $y^1, \cdots, y^k$, the region W given in* (104) *is a most stringent test.*

Proof. First we shall show that $P(\theta, \omega, \alpha)$ is constant along $S_c(\theta)$ where $S_c(\theta)$ is defined in (106). Consider a linear transformation of $y^1, \cdots, y^k$ as defined in (105). Then the transformed variates $y'^1, \cdots, y'^k$ are independently distributed with unit variances. Denote by θ' the image of θ obtained by the transformation (107) and let $P'(\theta', \omega, \alpha)$ be equal to l.u.b. $P(Z' \mid \theta')$ with respect to Z', where Z' may be any region in the space of $y'^1, \cdots, y'^k$ subject to the condition that

$$\text{l.u.b.}_{{}_2\theta'} P(Z' \mid {}_1\theta_0, {}_2\theta') = \alpha.$$

Obviously $P'(\theta', \omega, \alpha) = P(\theta, \omega, \alpha)$.

Hence we have merely to show that $P'(\theta', \omega, \alpha)$ is constant along $S_c'(\boldsymbol{\theta})$ where $S_c'(\boldsymbol{\theta})$ is the image of $S_c(\boldsymbol{\theta})$ and is given by

$$\sum_{p=1}^{r} (\theta'^p - \theta_0^p)^2 = c; \qquad \theta'^t = \boldsymbol{\theta}'^t \qquad (t = r+1, \cdots, k).$$

Let $P^*(\theta', {}_1\theta_0, \alpha)$ be equal to the least upper bound of $P(Z' \mid \theta')$ with respect to all regions Z' in the space of $y'^1, \cdots, y'^k$ for which $P(Z' \mid {}_1\theta_0, {}_2\theta') = \alpha$. Obviously $P^*(\theta', {}_1\theta_0, \alpha) \geqq P'(\theta', \omega, \alpha)$. It is easy to verify that the region V' for which $P(V' \mid {}_1\theta', {}_2\theta') = P^*(\theta', {}_1\theta_0, \alpha)$ is a subset in the space of $y'^1, \cdots, y'^r$. Hence

$$P(V' \mid {}_1\theta_0, {}_2\theta_1') = P(V' \mid {}_1\theta_0, {}_2\theta_2')$$

for any pair of points θ_1' and θ_2', and therefore

$$P^*(\theta', {}_1\theta_0, \alpha) = P'(\theta', \omega, \alpha) = P(\theta, \omega, \alpha).$$

Since $P^*(\theta', {}_1\theta_0, \alpha)$ is constant along $S_c(\boldsymbol{\theta})$, our statement is proved. From this and Proposition V, Proposition VII easily follows.

THEOREM VI. *Let W_n^* be the region defined in* (115). *Then $\{W_n^*\}$ is an asymptotically most stringent test of the hypothesis ${}_1\theta = {}_1\theta_0$.*

Proof. Denote by $\mathfrak{P}_n(\theta, \omega, \alpha)$ the least upper bound of $\mathfrak{P}(Z_n^* \mid \theta)$ with respect to Z_n^*, where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which

$$\text{l.u.b.}_{\boldsymbol{\theta}} \mathfrak{P}(Z_n^* \mid \boldsymbol{\theta}) = \alpha.$$

On account of Lemmas 1 and 2 we have

$$\lim_{n=\infty} \{P_n(\theta, \omega, \alpha) - \mathfrak{P}_n(\theta, \omega, \alpha)\} = 0 \tag{129}$$

uniformly in θ.

Denote by $\overline{P}_n(\theta, {}_1\theta_0, \alpha)$ the least upper bound of $\overline{P}_n(Z_n^* \mid \theta)$ with respect to Z_n^* where Z_n^* is restricted to regions in the space of the maximum likelihood estimates for which

$$\text{l.u.b.}_{{}_2\theta} \overline{P}(Z_n^* \mid {}_1\theta_0, {}_2\theta) = \alpha.$$

The symbol $\overline{P}(V_n^* \mid \theta)$ denotes the probability of V_n^* calculated under the assumption that the joint distribution of $n^{1/2}(\hat{\theta}_n^1 - \theta^1), \cdots, n^{1/2}(\hat{\theta}_n^k - \theta^k)$ is normal with zero means and covariance matrix $\|c_{ij}(\boldsymbol{\theta})\|^{-1}$, where $\boldsymbol{\theta}$ denotes that point for which θ lies on the surface $S_c(\boldsymbol{\theta})$ defined in (117). It can be shown that for any positive λ we have

$$\lim_{n=\infty} \{\mathfrak{P}_n(\theta, \omega, \alpha) - \overline{P}_n(\theta, {}_1\theta_0, \alpha)\} = 0 \tag{130}$$

uniformly in θ in the domain $|{}_1\theta - {}_1\theta_0| \leqq \lambda/n^{1/2}$. Since for any sequence $\{\theta_n\}$ for which

$$\lim_{n=\infty} n^{1/2}\left(\sum_{p=1}^{r}(\theta^p - \theta_0^p)^2\right)^{1/2} = +\infty,$$

we have

$$\lim_{n=\infty} \mathfrak{P}(\theta_n, \omega, \alpha) = \lim_{n=\infty} \overline{P}_n(\theta_n, {}_1\theta_0, \alpha) = 1,$$

we obtain from (130)

$$\lim_{n=\infty} \{\mathfrak{P}_n(\theta, \omega, \alpha) - \overline{P}_n(\theta, {}_1\theta_0, \alpha)\} = 0 \tag{131}$$

uniformly in θ. The function $\overline{P}_n(\theta, {}_1\theta_0, \alpha)$ is constant along the surface $S_c(\theta)$ defined in (117). This can be proved in the same way as the constancy of $P(\theta, \omega, \alpha)$ on $S_c(\theta)$ defined in (106). Hence from (129) and (131) we obtain

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta \in S_c(\theta)} P_n(\theta, \omega, \alpha) - \text{g.l.b.}_{\theta \in S_c(\theta)} P_n(\theta, \omega, \alpha)\} = 0 \tag{132}$$

uniformly in θ and c. According to Theorem V we have

$$\lim_{n=\infty} \{\text{l.u.b.}_{\theta \in S_c(\theta)} P(W_n^* | \theta) - \text{g.l.b.}_{\theta \in S_c(\theta)} P(W_n^* | \theta)\} = 0 \tag{133}$$

uniformly in c and θ. Theorem VI follows from (132), (133) and Theorem IV.

12. **The general composite hypothesis.** In §§9–11 we have considered the linear composite hypothesis ${}_1\theta = {}_1\theta_0$. Now we shall discuss a general composite hypothesis H_ω where ω denotes a subset of the parameter space given by r equations

$$\xi^1(\theta) = \xi^2(\theta) = \cdots = \xi^r(\theta) = 0 \qquad (r < k), \tag{134}$$

that is, ω is the set of all points θ which satisfy equations (134). We make the following assumption.

ASSUMPTION VI. *There exist $k-r$ functions $\xi^{r+1}(\theta), \cdots, \xi^k(\theta)$ such that the following three conditions are fulfilled:*

(a) *The transformation which transforms the point θ into the point ξ with the coordinates $\xi^1(\theta), \cdots, \xi^k(\theta)$ is a topological transformation of Ω into itself.*

(b) *The first and second order partial derivatives of $\xi^1(\theta), \cdots, \xi^k(\theta)$ are uniformly continuous and bounded functions of θ.*

(c) *The greatest lower bound of the absolute value of the Jacobian $\partial(\xi^1, \cdots, \xi^k)/\partial(\theta^1, \cdots, \theta^k)$ is positive.*

Let $\xi = (\xi^1, \cdots, \xi^k)$ denote a variable point of the parameter space Ω.

Since according to Assumption VI the transformation

$$\xi^i = \xi^i(\theta) \qquad (i = 1, \cdots, k) \tag{135}$$

is topological, we can solve the equations (135) and we obtain

$$\theta^i = \theta^i(\xi^1, \cdots, \xi^k) \qquad (i = 1, \cdots, k). \tag{136}$$

From conditions (b) and (c) of Assumption VI it follows that the first and second order derivatives of $\theta^i(\xi^1, \cdots, \xi^k)$ are uniformly continuous and bounded functions of ξ and the absolute value of the Jacobian $\partial(\theta^1, \cdots, \theta^k)/\partial(\xi^1, \cdots, \xi^k)$ has a positive lower bound.

Let $f^*(x, \xi)$ be the probability density function we obtain from the probability density function $f(x, \theta)$ of x by substituting the right-hand side of (136) for θ^i. Hence $f^*(x, \xi)$ is the probability density function of x in the transformed parameter space. It is clear that the maximum likelihood estimate of ξ^i is equal to $\hat{\xi}_n^i = \xi^i(\hat{\theta}_n^1, \cdots, \hat{\theta}_n^k)$, where $\hat{\theta}_n$ is the maximum likelihood estimate of θ.

Denote by I*, II*, $\cdots$, V* the assumptions which we obtain from Assumptions I–V respectively by substituting $f^*(x, \xi)$ for $f(x, \theta)$, ξ for θ and $\hat{\xi}_n$ for $\hat{\theta}_n$. We shall show that Assumptions I*–V* can be derived from Assumptions I–VI.

Assumption I* is an immediate consequence of Assumptions I and VI. Since according to Assumption VI the first derivatives of $\xi^i(\theta)$ are continuous and bounded functions of θ, the transformation (135) is uniformly continuous. Hence, for each positive ϵ^* there exists a positive ϵ such that the inequality $|\hat{\theta}_n - \theta| \leqq \epsilon$ implies the inequality $|\hat{\xi}_n - \xi| \leqq \epsilon^*$. From this and Assumption II we obtain Assumption II*.

Denote by $\nu_i(x, \theta_1, \delta)$ the least upper bound, and by $\mu_i(x, \theta_1, \delta)$ the greatest lower bound of $\partial \log f(x, \theta)/\partial\theta^i$ in the interval $\theta_1 - \delta \leqq \theta \leqq \theta_1 + \delta$. Using the Taylor expansion we obtain

$$\frac{\partial \log f(x, \theta_1^*)}{\partial \theta^i} = \frac{\partial \log f(x, \theta_1)}{\partial \theta_1} + \sum_j (\theta_1^{*j} - \theta_1^j) \frac{\partial^2 \log f(x, \bar{\theta}_1)}{\partial\theta^i \partial\theta^j}, \tag{137}$$

where $\bar{\theta}_1$ lies in the interval $[\theta_1, \theta_1^*]$. From (137) it follows that

$$\left| \frac{\partial \log f(x, \theta_1^*)}{\partial \theta^i} - \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leqq 2\delta \sum_j [\,|\phi_{ij}(x, \theta, \delta)| + |\psi_{ij}(x, \theta, \delta)|\,] \tag{138}$$

for any θ and δ for which

$$\theta - \delta \leqq \theta_1 \leqq \theta + \delta \quad \text{and} \quad \theta - \delta \leqq \theta_1^* \leqq \theta + \delta.$$

From (138) we obtain for any positive δ

$$|\nu_i(x, \theta, \delta) - \mu_i(x, \theta, \delta)| \leqq 2\delta \sum_j [\,|\phi_{ij}(x, \theta, \delta)| + |\psi_{ij}(x, \theta, \delta)|\,]. \tag{139}$$

Let $\{\theta_{1n}\}$ and $\{\theta_{2n}\}$ $(n=1, 2, \cdots, \text{ad inf.})$ be two sequences of parameter points such that

$$\lim_{n=\infty} (\theta_{1n} - \theta_{2n}) = 0. \tag{140}$$

According to Assumption III the expectations $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded functions of θ_1, θ_2 and δ in the domain D_ϵ defined in Assumption III. Hence also the expectations $E_{\theta_1}|\phi_{ij}(x, \theta_2, \delta)|$ and $E_{\theta_1}|\psi_{ij}(x, \theta_2, \delta)|$ are bounded functions of θ_1, θ_2 and δ in the domain D_ϵ. From this and relations (139) and (140) it follows that for any sequence $\{\delta_n\}$ of positive numbers for which $\lim_{n=\infty} \delta_n = 0$ we have

$$\lim_{n=\infty} \{E_{\theta_{1n}}\nu_i(x, \theta_{2n}, \delta_n) - E_{\theta_{1n}}\mu_i(x, \theta_{2n}, \delta_n)\} = 0. \tag{141}$$

Since

$$\mu_i(x, \theta_{2n}, \delta_n) \leqq \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i} \leqq \nu_i(x, \theta_{2n}, \delta_n)$$

it follows from (141) that

$$\begin{aligned} &\lim_{n=\infty} \left\{E_{\theta_{1n}}\nu_i(x, \theta_{2n}, \delta_n) - E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i}\right\} = 0, \\ &\lim_{n=\infty} \left\{E_{\theta_{1n}}\mu_i(x, \theta_{2n}, \delta_n) - E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i}\right\} = 0. \end{aligned} \tag{142}$$

Using the Taylor expansion we have

$$\frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i} = \frac{\partial \log f(x, \theta_{1n})}{\partial \theta^i} + \sum_j (\theta^j_{2n} - \theta^j_{1n}) \frac{\partial^2 \log f(x, \bar{\theta}_n)}{\partial \theta^i \partial \theta^j}, \tag{143}$$

where $\bar{\theta}_n$ lies in the interval $[\theta_{1n}, \theta_{2n}]$. Since the expectations $E_{\theta_1}|\phi_{ij}(x, \theta_2, \delta)|$ and $E_{\theta_1}|\psi_{ij}(x, \theta_2, \delta)|$ are bounded functions of θ_1, θ_2, and δ in the domain D_ϵ, we obtain

$$\lim_{n=\infty} E_{\theta_{1n}} \left\{\sum_j (\theta^j_{2n} - \theta^j_{1n}) \frac{\partial^2 \log f(x, \bar{\theta}_n)}{\partial \theta^i \partial \theta^j}\right\} = 0.$$

Hence it follows from (143) and Assumption IV that

$$\lim_{n=\infty} E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta^i} = \lim_{n=\infty} E_{\theta_{1n}} \frac{\partial \log f(x, \theta_{1n})}{\partial \theta^i} = 0. \tag{144}$$

We obtain from (142) and (144)

$$\lim E_{\theta_{1n}}\nu_i(x, \theta_{2n}, \delta_n) = \lim E_{\theta_{1n}}\mu_i(x, \theta_{2n}, \delta_n) = 0. \tag{145}$$

Denote by $\phi^*_{ij}(x, \xi_1, \delta^*)$ the least upper bound, and by $\psi^*_{ij}(x, \xi, \delta^*)$ the great-

est lower bound of $\partial^2 \log f^*(x, \xi)/\partial\xi^i\partial\xi^j$ in the interval $\xi_1-\delta^* \leqq \xi \leqq \xi_1+\delta^*$. We have

$$\frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i\partial\xi^j} = \sum_l \sum_m \frac{\partial^2 \log f(x, \theta)}{\partial\theta^l\partial\theta^m} \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} + \sum_l \frac{\partial \log f(x, \theta)}{\partial\theta^l} \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j}. \tag{146}$$

Since

$$E_\theta\left(\frac{\partial \log f(x, \theta)}{\partial\theta}\right) = 0,$$

we obtain from (146)

$$E_\xi \frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i\partial\xi^j} = \sum_l \sum_m E_\theta\left[\frac{\partial^2 \log f(x, \theta)}{\partial\theta^l\partial\theta^m} \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j}\right]. \tag{147}$$

Hence the determinant

$$\left| E_\xi \frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i\partial\xi^j} \right| = \left| E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial\theta^i\partial\theta^j} \right| \left(\frac{\partial(\theta^1, \cdots, \theta^k)}{\partial(\xi^1, \cdots, \xi^k)}\right)^2. \tag{148}$$

Since the determinant $\left| -E_\theta\partial^2 \log f(x, \theta)/\partial\theta^i\partial\theta^j \right|$ has a positive lower bound, it follows from (148) and Assumption VI that

$$\left| - E_\xi\partial^2 \log f^*(x, \xi)/\partial\xi^i\partial\xi^j \right| \tag{149}$$

has a positive lower bound.

For any positive δ^* let $\delta(\delta^*)$ be the smallest positive number such that for any two points ξ_1 and ξ_2 for which $|\xi_1-\xi_2| \leqq \delta^*$ we have $|\theta_1-\theta_2| \leqq \delta(\delta^*)$ where θ_1 and θ_2 are the image points of ξ_1 and ξ_2 by transformation (136). From (146) we obtain

$$\begin{aligned} \phi^*_{ij}(x, \xi, \delta^*) &\leqq \sum_l \sum_m \bar{\phi}_{lm}[x, \theta, \delta(\delta^*)] \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} + \sum_l \bar{\nu}_l[x, \theta, \delta(\delta^*)] \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j}, \\ \psi^*_{ij}(x, \xi, \delta^*) &\geqq \sum_l \sum_m \bar{\psi}_{lm}[x, \theta, \delta(\delta^*)] \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} + \sum_l \bar{\mu}_l[x, \theta, \delta(\delta^*)] \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j}, \end{aligned} \tag{150}$$

where θ is the image point of ξ by transformation (136), and the derivatives $\partial\theta^l/\partial\xi^i$, $\partial^2\theta^l/\partial\xi^i\partial\xi^j$ are taken at some points in the interval $[\xi-\delta^*, \xi+\delta^*]$, and the functions $\bar{\phi}_{lm}(x, \theta, \delta)$, $\bar{\psi}_{lm}(x, \theta, \delta)$, $\bar{\mu}_l(x, \theta, \delta)$ and $\bar{\nu}_l(x, \theta, \delta)$ satisfy the inequalities

$$\psi_{lm}(x, \theta, \delta) \leqq \bar{\phi}_{lm}(x, \theta, \delta) \leqq \phi_{lm}(x, \theta, \delta);$$

$$\psi_{lm}(x, \theta, \delta) \leqq \bar{\psi}_{lm}(x, \theta, \delta) \leqq \phi_{lm}(x, \theta, \delta);$$

$$\mu_l(x, \theta, \delta) \leqq \bar{\nu}_l(x, \theta, \delta) \leqq \nu_l(x, \theta, \delta);$$

$$\mu_l(x, \theta, \delta) \leqq \bar{\mu}_l(x, \theta, \delta) \leqq \nu_l(x, \theta, \delta).$$

Let $\{\xi_{1n}\}$ and $\{\xi_{2n}\}$ be two sequences of parameter points such that $\lim \xi_{1n} = \lim \xi_{2n} = \xi$. Let θ_{1n} be the image point of ξ_{1n}, θ_{2n} the image point of ξ_{2n}, and θ the image point of ξ (by transformation (136)). Let furthermore $\{\delta_n^*\}$ be a sequence of positive numbers such that $\lim \delta_n^* = 0$. Then we obviously have $\lim \delta(\delta_n^*) = 0$ and therefore using (145) and Assumption III we obtain

$$\begin{aligned}
&\lim_{n=\infty} E_{\theta_{1n}} \phi_{lm}[x, \theta_{2n}, \delta(\delta_n^*)] = E_\theta \partial^2 \log f(x, \theta)/\partial\theta^l \partial\theta^m, \\
&\lim_{n=\infty} E_{\theta_{1n}} \psi_{lm}[x, \theta_{2n}, \delta(\delta_n^*)] = E_\theta \partial^2 \log f(x, \theta)/\partial\theta^l \partial\theta^m, \\
&\lim_{n=\infty} E_{\theta_{1n}} \nu_l \quad [x, \theta_{2n}, \delta(\delta_n^*)] = 0, \\
&\lim_{n=\infty} E_{\theta_{1n}} \mu_l \quad [x, \theta_{2n}, \delta(\delta_n^*)] = 0
\end{aligned} \tag{151}$$

uniformly in θ. From (150) and (151) and the uniform continuity of the derivatives $\partial\theta^l/\partial\xi^i$ and $\partial^2\theta^l/\partial\xi^i\partial\xi^j$ we obtain

$$\lim_{n=\infty} E_{\xi_{1n}} \phi_{ij}^*(x, \xi_{2n}, \delta_n^*) = \lim_{n=\infty} E_{\xi_{1n}} \psi_{ij}^*(x, \xi_{2n}, \delta_n^*) = E_\xi \frac{\partial^2 \log f^*(x, \xi)}{\partial\xi^i \partial\xi^j} \tag{152}$$

uniformly in ξ.

Because of (150) we have both

$$\begin{aligned}
&|\phi_{ij}^*(x, \xi, \delta^*)| \quad \text{and} \quad |\psi_{ij}^*(x, \xi, \delta^*)| \\
&\leqq \sum_l \sum_m \Big\{ [|\psi_{lm}[x, \theta, \delta(\delta^*)]| \\
&\qquad + |\phi_{lm}[x, \theta, \delta(\delta^*)]|] \Big[\text{l.u.b.}_\xi \left| \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j} \right| \Big] \Big\} \\
&\quad + \sum_l \Big\{ [|\nu_l[x, \theta, \delta(\delta^*)]| + |\mu_l[x, \theta, \delta(\delta^*)]|] \Big[\text{l.u.b.}_\xi \left| \frac{\partial^2\theta^l}{\partial\xi^i\partial\xi^j} \right| \Big] \Big\}
\end{aligned} \tag{153}$$

where the least upper bound with respect to ξ is to be taken over the interval $[\xi - \delta^*, \xi + \delta^*]$.

We shall show that $E_{\theta_1}[\nu_i(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\mu_i(x, \theta_2, \delta)]^2$ are bounded func-

tions of θ_1, θ_2 and δ in the domain D_ϵ for sufficiently small ϵ. Our statement is proved if we show that $E_{\theta_1}[\partial \log f(x, \theta_1)/\partial\theta^i]^2$, $E_{\theta_1}[\nu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ and $E_{\theta_1}[\mu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ are bounded in D_ϵ. The first of these expressions is bounded because of Assumption V. From (138) it follows that

$$\left| \nu_i(x, \theta_2, \delta) - \frac{\partial \log f(x, \theta_1)}{\partial\theta^i} \right| \leqq 2\delta' \sum_j [\,|\phi_{ij}(x, \theta_2, \delta')| + |\psi_{ij}(x, \theta_2, \delta')|\,]$$

and

$$\left| \mu_i(x, \theta_2, \delta) - \frac{\partial \log f(x, \theta_1)}{\partial\theta^i} \right| \leqq 2\delta' \sum_j [\,|\phi_{ij}(x, \theta_2, \delta')| + |\psi_{ij}(x, \theta_2, \delta')|\,]$$

where $\delta' = \delta + \sum_j |\theta_1^j - \theta_2^j|$. From the above inequalities and the fact that $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded in D_ϵ it follows that for sufficiently small ϵ the expressions $E_{\theta_1}[\nu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ and $E_{\theta_1}[\mu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial\theta^i]^2$ are bounded in D_ϵ. Hence our statement is proved.

Since the derivatives $\partial\theta^l/\partial\xi^i$ and $\partial^2\theta^l/\partial\xi^i\partial\xi^j$ are bounded functions of ξ, and since $E_{\theta_1}[\phi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta_1}[\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded functions of θ_1, θ_2 and δ in the domain D_ϵ, it follows from (153) that there exists a positive ϵ^* such that $E_{\xi_1}[\phi_{ij}^*(x, \xi_2, \delta^*)]^2$ and $E_{\xi_1}[\psi_{ij}^*(x, \xi_2, \delta^*)]^2$ are bounded functions of ξ_1, ξ_2 and δ^* in the domain defined by $|\xi_1 - \xi_2| \leqq \epsilon^*$ and $|\delta^*| \leqq \epsilon^*$. Assumption III* follows from the latter statement and the relations (152) and (149).

Assumption IV* is an immediate consequence of Assumption IV.

We have

$$\frac{\partial \log f^*(x, \xi)}{\partial\xi^i} = \sum_j \frac{\partial \log f(x, \theta)}{\partial\theta^j} \frac{\partial\theta^j}{\partial\xi^i}. \tag{154}$$

For any points x and θ denote the maximum of the k expressions $|\partial \log f(x, \theta)/\partial\theta^1|, \cdots, |\partial \log f(x, \theta)/\partial\theta^k|$ by $\rho(x, \theta)$. From Assumption V it follows easily that

$$E_\theta[\rho(x, \theta)]^{2+\eta} \tag{155}$$

is a bounded function of θ. Since the derivatives $\partial\theta^j/\partial\xi^i$ are bounded functions of ξ, Assumption V* follows from (154) and (155).

Denote by $c_{ij}^*(\theta)$ the function of θ we obtain from $-E_\xi \partial^2 \log f^*(x, \xi)/\partial\xi^i\partial\xi^j$ by substituting $\xi(\theta)$ for ξ. Then we obtain from (147)

$$c_{ij}^*(\theta) = \sum_l \sum_m c_{lm}(\theta) \frac{\partial\theta^l}{\partial\xi^i} \frac{\partial\theta^m}{\partial\xi^j}. \tag{156}$$

Denote by A the matrix $\|\partial\theta^i/\partial\xi^j\|$ $(i, j = 1, \cdots, k)$ and let $\overline{A}$ be the trans-

posed of A. Then equation (156) can be written as

$$\|c^*_{ij}(\theta)\| = \overline{A}\|c_{ij}(\theta)\|A. \tag{157}$$

Let $\|\sigma^*_{ij}(\theta)\|$ be the inverse of the matrix $\|c^*_{ij}(\theta)\|$. Futhermore let B be the matrix $\|\partial\xi^i/\partial\theta^j\|$ $(i, j=1, \cdots, k)$ and denote by $\overline{B}$ the transposed of B. Since $B=A^{-1}$, we obtain from (157)

$$\|\sigma^*_{ij}(\theta)\| = B\|\sigma_{ij}(\theta)\|\overline{B}, \tag{158}$$

where $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$. Equation (158) can be written as

$$\sigma^*_{ij}(\theta) = \sum_l \sum_m \frac{\partial\xi^i}{\partial\theta^l}\frac{\partial\xi^j}{\partial\theta^m}\sigma_{lm}(\theta). \tag{159}$$

Let

$$\|\bar{c}^*_{pq}(\theta)\| = \|\sigma^*_{pq}(\theta)\|^{-1} \qquad (p, q = 1, \cdots, r), \tag{160}$$

Denote by W_n^* the critical region defined by the inequality

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\dot{\theta}_n)\xi^q(\dot{\theta}_n)\bar{c}^*_{pq}(\dot{\theta}_n) \geqq d_n, \tag{161}$$

where the constant d_n is chosen so that

$$\text{l.u.b.}_{\theta}\ P(W^* \mid \theta) = \alpha.$$

The point θ is restricted to points of the set ω defined by equations (134).

For each positive c and for each point θ of ω we define the surface $S_c(\theta)$ by the equations

$$\begin{aligned}
&\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\theta)\xi^q(\theta)\bar{c}^*_{pq}(\theta) = c,\\
&\sum_{j=1}^{k}\gamma_{tj}(\theta)\xi^j(\theta) = \sum_{j=1}^{k}\gamma_{tj}(\theta)\xi^j(\theta) \qquad (t = r+1, \cdots, k),
\end{aligned} \tag{162}$$

where the coefficients $\gamma_{tj}(\theta)$ satisfy the following condition: There exists a matrix $\|\beta_{pq}(\theta)\|$ $(p, q=1, \cdots, r)$ such that if we form the matrix $A(\theta)$ given in (120) then

$$A(\theta)\|\sigma^*_{ij}(\theta)\|\overline{A}(\theta) = I, \tag{163}$$

where $\overline{A}(\theta)$ denotes the transposed of $A(\theta)$ and I denotes the unit matrix.

Consider the transformation of the parameter space given by

$$\begin{aligned}
\theta'^p &= \beta_{p1}(\theta)\xi^1(\theta) + \cdots + \beta_{pr}(\theta)\xi^r(\theta) \qquad (p = 1, \cdots, r),\\
\theta'^t &= \gamma_{t1}(\theta)\xi^1(\theta) + \cdots + \gamma_{tk}(\theta)\xi^k(\theta) \qquad (t = r+1, \cdots, k),
\end{aligned} \tag{164}$$

where θ denotes the point of ω for which θ lies on $S_c(\theta)$ for some value of c. The matrix $\|\beta_{pq}(\theta)\|$ is chosen so that (163) is fulfilled. The transformation (164) transforms $S_c(\theta)$ into the sphere $S_c'(\theta)$ given by

$$\sum_{p=1}^{r} (\theta'^p)^2 = c, \qquad \theta'^t = \sum_{j=1}^{k} \gamma_{tj}(\theta)\xi^j(\theta).$$

We define a weight function $\eta(\theta)$ as follows:

$$\eta(\theta) = \lim_{\rho=0} A[\omega'(\theta, \rho)]/A[\omega(\theta, \rho)], \tag{165}$$

where the symbols on the right-hand side of (165) are defined as in (109).

Since Assumptions I*–V* are fulfilled if Assumption VI holds, we obtain from Theorems IV, V and VI the following theorem.

THEOREM VII. *Let W_n^* be the region defined in* (161) *and let $S_c(\theta)$ be the surface defined in* (162). *Furthermore let $\eta(\theta)$ be the weight function defined in* (165). *If Assumption* VI *holds, then for testing the hypothesis $\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$ the sequence $\{W_n^*\}$*

(a) *has asymptotically best average power with respect to the surfaces $S_c(\theta)$ and the weight function $\eta(\theta)$;*

(b) *has asymptotically best constant power on the surfaces $S_c(\theta)$;*

(c) *is an asymptotically most stringent test.*

13. **Optimum properties of the likelihood ratio test.** For testing a composite hypothesis H_ω, Neyman and Pearson introduced a statistic(5), called likelihood ratio, defined as follows: The density function in the sample space is given by $\prod_{\alpha=1}^{n} f(x_\alpha, \theta)$. Denote by $P(x_1, \cdots, x_n)$ the maximum of this function with respect to $\theta^1, \cdots, \theta^k$, and let $P_\omega(x_1, \cdots, x_n)$ be the conditional maximum with respect to $\theta^1, \cdots, \theta^k$, subject to the condition that θ must be a point of ω. Then the likelihood ratio for testing the hypothesis H_ω is given by $\lambda_n(\omega, E_n) = P_\omega(x_1, \cdots, x_n)/P(x_1, \cdots, x_n)$. It is obvious that the value of $\lambda_n(\omega, E_n)$ always lies between 0 and 1. Neyman and Pearson recommend the use of the left tail as critical region, that is the hypothesis H_ω is rejected if the value of $\lambda_n(\omega, E_n)$ is less than a certain constant $\bar{\lambda}_n(\omega)$. Denote the region $\lambda_n(\omega, E_n) < \bar{\lambda}_n(\omega)$ by $L_n(\omega)$. In all that follows we choose the constant $\bar{\lambda}_n(\omega)$ so that

$$\text{l.u.b.}_{\theta} P[L_n(\omega) \mid \theta] = \alpha.$$

We shall prove that there exists a finite value B such that $-2 \log \bar{\lambda}_n(\omega) < B$ for all n and for all ω. Consider the Taylor expansion

(5) See in this connection J. Neyman and E. S. Pearson, *On the use and interpretation of certain test criteria for purposes of statistical inference*, Biometrika vol. 20A (1928).

$$\sum_\alpha \log f(x_\alpha, \theta) = \sum_\alpha \log f(x_\alpha, \dot{\theta}_n) + \frac{1}{2}\sum\sum n(\dot{\theta}_n^i - \theta^i)(\dot{\theta}_n^j - \theta^j) \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i\partial\theta^j} \frac{1}{n}$$

where $\bar{\theta}_n$ lies in the interval $[\dot{\theta}_n, \theta]$. Since

$$\sum_\alpha \log f(x_\alpha, \dot{\theta}_n) - \sum_\alpha \log f(x_\alpha, \theta) \geqq -\log \lambda_n(\omega, E_n),$$

we have

$$-2\log\lambda_n(\omega, E_n) \leqq -\sum\sum n(\dot{\theta}_n^i - \theta^i)(\dot{\theta}_n^j - \theta^j) \frac{1}{n}\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i\partial\theta^j}.$$

Since $n^{-1}\sum_\alpha \partial^2 \log f(x_\alpha, \bar{\theta}_n)/\partial\theta^i\partial\theta^j$ converges stochastically to $-c_{ij}(\theta)$ under the assumption that θ is the true parameter point, it follows easily from Proposition I that for any $\epsilon > 0$ there exists a positive value $A(\epsilon)$ such that for any ω

$$\limsup_{n=\infty}\left\{\text{l.u.b.}_\theta P\left[-\sum\sum(\dot{\theta}_n^i - \theta^i)(\dot{\theta}_n^j - \theta^j)\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i\partial\theta^j} \geqq A(\epsilon) \,\middle|\, \theta\right]\right\} \leqq \epsilon$$

and $\lim_{\epsilon=0} A(\epsilon) = +\infty$. Hence

$$\limsup_{n=\infty}\{\text{l.u.b.}_{\theta,\omega} P[-2\log\lambda_n(\omega, E_n) \geqq A(\epsilon) \mid \theta]\} \leqq \epsilon. \tag{166}$$

This proves the existence of a finite number B with the required property.

For any subsets Γ and Γ' of the parameter space we denote by $\delta(\Gamma, \Gamma')$ the greatest lower bound of the distance between θ and θ' where θ is restricted to points of Γ and θ' is restricted to points of Γ'. We shall call $\delta(\Gamma, \Gamma')$ the distance of the sets Γ and Γ'.

Let $\{\theta_n\}$ be a sequence of parameter points such that $\lim \delta(\theta_n, \omega) = 0$ and $\lim_{n=\infty} n^{1/2}\delta(\theta_n, \omega) = +\infty$. We shall prove that there exists a positive ν_0 such that for any constant A

$$\lim_{n=\infty} P\left[\sum_\alpha \log f(x_\alpha, \dot{\theta}_n) - \sum_\alpha \log f(x_\alpha, \theta^*) > A \,\middle|\, \theta_n\right] = 1, \tag{167}$$

where θ^* denotes a point of ω for which $|\theta^* - \theta_n| \leqq \nu_0$ and $\sum_\alpha \log f(x_\alpha, \theta^*) \geqq \sum_\alpha \log f(x_\alpha, \theta)$ for all θ in the domain $|\theta - \theta_n| \leqq \nu_0$. Consider the Taylor expansion

$$\sum_\alpha \log f(x_\alpha, \theta^*) - \sum_\alpha \log f(x_\alpha, \dot{\theta}_n) = \frac{n}{2}\sum\sum(\dot{\theta}_n^i - \theta^{*i})(\dot{\theta}_n^j - \theta^{*j})\frac{1}{n}\sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i\partial\theta^j}, \tag{168}$$

where $\bar{\theta}_n$ lies in the interval $[\hat{\theta}_n, \theta^*]$. Because of conditions (a) and (b) of Assumption III for any $\epsilon > 0$ the positive number ν_0 can be chosen so that

$$\lim_{n=\infty} P\left[\left| c_{ij}(\theta^*) + \frac{1}{n}\sum_{\alpha} \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j}\right| < \epsilon \,\middle|\, \theta_n\right] = 1. \tag{169}$$

Since $\|c_{ij}(\theta)\|$ is positive definite and the determinant $|c_{ij}(\theta)|$ has a positive lower bound, we obtain from Proposition I, (168) and (169) that for some positive, ν_0, (167) holds. Hence our statement is proved.

We say that the likelihood ratio test is uniformly consistent if for any positive ν

$$\lim_{n=\infty} P[L_n(\omega) \mid \theta] = 1$$

uniformly in ω and θ over the domain $\delta(\theta, \omega) \geqq \nu$. We postulate the following assumption.

ASSUMPTION VII. *The likelihood ratio test is uniformly consistent.*

This assumption together with the uniform consistency of the maximum likelihood estimate $\hat{\theta}_n$ will be proved in a forthcoming paper on the basis of some weak assumptions on the density function $f(x, \theta)$.

Let $\omega_\nu(\theta)$ be the intersection of ω with the set of all points θ' for which $\sum_{i=1}^{k} |\theta^i - \theta'^i| \geqq \nu$. From Assumption VII it follows that for any positive ν

$$\lim_{n=\infty} P\{-2 \log \lambda_n[\omega_\nu(\theta), E_n] \geqq -2 \log \bar{\lambda}_n[\omega_\nu(\theta)] \mid \theta\} = 1 \tag{170}$$

uniformly in θ.

Let $\{\theta_n\}$ be a sequence of parameter points such that $\lim_{n=\infty} \delta(\theta_n, \omega) = 0$ and $\lim_{n=\infty} n^{1/2}\delta(\theta_n, \omega) = +\infty$. Denote by ω_n the set of all points θ for which $|\theta - \theta_n| \leqq \nu_0$. Since $-2 \log \bar{\lambda}_n[\omega_\nu(\theta)]$ has a finite upper bound it follows from (167) that for a sufficiently small ν_0

$$\lim_{n=\infty} P\{-2 \log \lambda_n[\omega_n, E_n] \geqq -2 \log \bar{\lambda}_n[\omega_{\nu_0}(\theta_n)] \mid \theta_n\} = 1. \tag{171}$$

Obviously

$$-2 \log \lambda_n(\omega, E_n) = \text{minimum}\{-2 \log \lambda_n[\omega_{\nu_0}(\theta_n), E_n], -2 \log \lambda_n(\omega_n, E_n)\}.$$

From (170) and (171) we obtain

$$\lim_{n=\infty} P\{-2 \log \lambda_n(\omega, E_n) \geqq -2 \log \bar{\lambda}_n[\omega_{\nu_0}(\theta_n)] \mid \theta_n\} = 1. \tag{172}$$

Since $-2 \log \lambda_n[\omega_{\nu_0}(\theta_n), E_n] \geqq -2 \log \lambda_n(\omega, E_n)$, we have $-2 \log \bar{\lambda}_n[\omega_{\nu_0}(\theta_n)] \geqq -2 \log \bar{\lambda}_n(\omega)$. Hence from (172) we obtain

$$\lim_{n=\infty} P[L_n(\omega) \mid \theta_n] = 1. \tag{173}$$

From Assumption VII it follows that (173) holds for any sequence $\{\theta_n\}$ for which $\lim n^{1/2}\delta(\theta_n, \omega) = +\infty$.

Now let us consider the case where ω is the set given by ${}_1\theta = {}_1\theta_0$. Let $\{\theta_n\}$ be a sequence of parameter points for which the sequence $n^{1/2}\delta(\theta_n, \omega)$ is bounded. Denote by $\dot{\theta}_n(\omega)$ the parameter point in ω for which

$$\prod_{\alpha=1}^{n} f[x_\alpha, \dot{\theta}_n(\omega)] = P_\omega(x_1, \cdots, x_n).$$

Let θ_n be the point for which $\delta(\theta_n, \omega) = \delta(\theta_n, \theta_n)$. Denote by $T_n(\nu)$ the region given by the inequalities

$$\begin{aligned} |\dot{\theta}_n - \theta_n| &< \nu, \\ |\dot{\theta}_n(\omega) - \theta_n| &< \nu. \end{aligned} \tag{174}$$

We shall prove that

$$\lim P[T_n(\nu) \mid \theta_n] = 1 \qquad \text{for any } \nu > 0. \tag{175}$$

Consider the Taylor expansion

$$\begin{aligned} \sum_\alpha \log f(x_\alpha, \theta_n) = \sum_\alpha \log f(x_\alpha, \dot{\theta}_n) \\ + \frac{n}{2} \sum\sum (\dot{\theta}_n^i - \theta_n^i)(\dot{\theta}_n^j - \theta_n^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \bar{\theta}_n)}{\partial\theta^i \partial\theta^j}. \end{aligned} \tag{176}$$

Since $[n^{-1}\sum_\alpha \partial^2 \log f(x_\alpha, \bar{\theta}_n)/\partial\theta^i\partial\theta^j + c_{ij}(\theta_n)]$ converges stochastically to zero, and since $n^{1/2}|\theta_n^i - \theta_n^i|$ is bounded, we easily obtain from Proposition I that for any $\epsilon > 0$ there exists a constant B_ϵ such that

$$\limsup_{n=\infty} P[\sum \log f(x_\alpha, \dot{\theta}_n) - \sum \log f(x_\alpha, \theta_n) \geqq B_\epsilon \mid \theta_n] = \epsilon.$$

Since $-2 \log \lambda_n(\omega, E_n) \leqq 2\sum \log f(x_\alpha, \dot{\theta}_n) - 2\sum \log f(x_\alpha, \theta_n)$ we have

$$\limsup_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) \geqq 2B_\epsilon \mid \theta_n] \leqq \epsilon. \tag{177}$$

Denote by ω_n the subset of ω in which $|\theta - \theta_n| \geqq \nu$. From Assumption VII it follows that

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega_n, E_n) \geqq -2 \log \bar{\lambda}_n(\omega_n) \mid \theta_n] = 1. \tag{178}$$

Since $-2 \log \bar{\lambda}_n(\omega) \leqq -2 \log \bar{\lambda}_n(\omega_n)$ we have

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega_n, E_n) \geqq -2 \log \bar{\lambda}_n(\omega) \mid \theta_n] = 1. \tag{179}$$

For any given constant B there exists a positive $\alpha < 1$ such that $-2 \log \bar{\lambda}_n(\omega)$

$\geqq B$ if l.u.b.$_{\theta}$ $P[L_n(\omega)|\theta]=\alpha$. Hence from (179) we obtain

$$\lim_{n=\infty} P[-2\log\lambda_n(\omega_n, E_n) \geqq B \mid \theta_n] = 1 \tag{180}$$

for any constant B. From (180) it follows that

$$\lim_{n=\infty} \{P[-2\log\lambda_n(\omega, E_n) \geqq B \mid \theta_n] - P[|\dot{\theta}_n(\omega) - \theta_n| \geqq \nu \mid \theta_n]\} \geqq 0 \tag{181}$$

for arbitrary values of B. From (181) and (177) we obtain

$$\lim_{n=\infty} P[|\dot{\theta}_n(\omega) - \theta_n| \geqq \nu \mid \theta_n] = 0.$$

Hence (175) is proved.

Consider the Taylor expansion

$$\begin{aligned}\sum_{\alpha} \log f(x_\alpha, \theta_n) &= \sum_{\alpha} \log f[x_\alpha, \dot{\theta}_n(\omega)] \\ &+ \frac{n}{2}\sum_{r+1}^{k}\sum_{r+1}^{k}[\dot{\theta}_n^i(\omega) - \theta_n^i][\dot{\theta}_n^j(\omega) - \theta_n^j]\,\frac{1}{n}\sum_{\alpha}\frac{\partial^2 \log f(x_\alpha, \ddot{\theta}_n)}{\partial\theta^i\partial\theta^j}\end{aligned} \tag{182}$$

where $\ddot{\theta}_n$ lies in the interval $[\theta_n, \dot{\theta}_n(\omega)]$. In the following arguments we shall use the following lemma: *Let* $\|\lambda_{ij}\|$ $(i, j=1, \cdots, k)$ *be a definite matrix and for each integer* s *let* λ_{ij}^s *be a real number such that* $\lim_{s=\infty}\lambda_{ij}^s=\lambda_{ij}$. *Then*

$$\lim_{s=\infty}\left(\sum\sum \nu_i\nu_j\lambda_{ij}/\sum\sum \nu_i\nu_j\lambda_{ij}^s\right) = 1$$

uniformly in $\nu_1, \cdots, \nu_k$. From (175), the Taylor expansions (176) and (182), and the above lemma it follows that for any $\epsilon>0$

$$\begin{aligned}\lim_{n=\infty} P[(1+\epsilon)q_n - (1-\epsilon)\bar{q}_n \geqq -2\log\lambda_n(\omega, E_n) \\ \geqq (1-\epsilon)q_n - (1+\epsilon)\bar{q}_n \mid \theta_n] = 1,\end{aligned} \tag{183}$$

where

$$q_n = \sum_{1}^{k}\sum_{1}^{k} n(\dot{\theta}_n^i - \theta_n^i)(\dot{\theta}_n^j - \theta_n^j)c_{ij}(\theta_n) \tag{184}$$

and

$$\bar{q}_n = \sum_{r+1}^{k}\sum_{r+1}^{k} n[\dot{\theta}_n^i(\omega) - \theta_n^i][\dot{\theta}_n^j(\omega) - \theta_n^j]c_{ij}(\theta_n). \tag{185}$$

Since $-2\log\lambda_n(\omega, E_n)\geqq 0$, we obtain from (183)

$$\lim_{n=\infty} P[(1+\epsilon)q_n - (1-\epsilon)\bar{q}_n \geqq 0 \mid \theta_n] = 1. \tag{186}$$

Since the sequence $\{n^{1/2}\delta(\theta_n, \theta_n)\}$ is bounded, the expression q_n is bounded in the probability sense, that is for each positive ρ there exists a positive value A_ρ such that

$$\lim_{n=\infty} P(q_n > A_\rho \mid \theta_n) \leqq \rho.$$

From (186) it follows that $\bar{q}_n$ is also bounded in the probability sense. Hence because of (183) we have

$$\lim_{n=\infty} P[q_n - \bar{q}_n + \epsilon \geqq -2 \log \lambda_n(\omega, E_n) \geqq q_n - \bar{q}_n - \epsilon \mid \theta_n] = 1 \tag{187}$$

for any $\epsilon > 0$. From the Taylor expansions

$$\begin{aligned} y_n^i(\theta_n) &= \frac{1}{n^{1/2}} \sum_\alpha \frac{\partial \log f(x_\alpha, \theta_n)}{\partial \theta^i} \\ &= -\sum_{j=1}^{k} n^{1/2}(\theta_n^j - \theta_n^j) \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta_n^*)}{\partial \theta^i \partial \theta^j} \qquad (i = 1, \cdots, k) \end{aligned} \tag{188}$$

and

$$y_n^i(\theta_n) = -\sum_{j=r+1}^{k} n^{1/2}[\theta_n^j(\omega) - \theta_n^j] \frac{1}{n} \sum_\alpha \frac{\partial^2 \log f(x_\alpha, \theta_n^{**})}{\partial \theta^i \partial \theta^j} \tag{189}$$

$$(i = r + 1, \cdots, k)$$

we obtain

$$y_n^i(\theta_n) = +\sum_{j=1}^{k} n^{1/2}(\theta_n^j - \theta_n^j)[c_{ij}(\theta_n) + \epsilon_{ijn}(E_n)], \tag{190}$$

$$n^{1/2}(\theta_n^i - \theta_n^i) = \sum_{j=1}^{k} [\sigma_{ij}(\theta_n) + \eta_{ijn}(E_n)] y_n^j(\theta_n) \qquad (i = 1, \cdots, k), \tag{191}$$

and

$$n^{1/2}[\theta_n^i(\omega) - \theta_n^i] = \sum_{j=r+1}^{k} [\bar{\sigma}_{ij}(\theta_n) + \zeta_{ijn}(E_n)] y_n^j(\theta_n) \quad (i = r + 1, \cdots, k), \tag{192}$$

where

$$\begin{aligned} \|\sigma_{ij}(\theta_n)\| &= \|c_{ij}(\theta_n)\|^{-1} \qquad (i, j = 1, \cdots, k), \\ \|\bar{\sigma}_{ij}(\theta)\| &= \|c_{ij}(\theta)\|^{-1} \qquad (i, j = r + 1, \cdots, k), \end{aligned}$$

and for any positive ν we have

$$\begin{aligned} \lim_{n=\infty} P[\,|\epsilon_{ijn}(E_n)| < \nu \mid \theta_n] &= \lim P[\,|\eta_{ijn}(E_n)| < \nu \mid \theta_n] \\ &= \lim_{n=\infty} P[\,|\zeta_{ijn}(E_n)| < \nu \mid \theta_n] = 1. \end{aligned}$$

Hence

$$\sum_{r+1}^{k}\sum_{r+1}^{k} n[\theta_n^{i}(\omega) - \theta_n^{i}][\theta_n^{j}(\omega) - \theta_n^{j}]c_{ij}(\theta_n)$$

$$= \sum_{r+1}^{k}\sum_{r+1}^{k}\sum_{r+1}^{k}\sum_{r+1}^{k} [c_{ij}(\theta_n)\bar{\sigma}_{il}(\theta_n)\bar{\sigma}_{jm}(\theta_n) + \rho_{ijlmn}(E_n)]y_n^{l}(\theta_n)y_n^{m}(\theta_n)$$

$$= \sum_{r+1}^{k}\sum_{r+1}^{k} [\bar{\sigma}_{ij}(\theta_n) + \rho'_{ijn}(E_n)]y_n^{i}(\theta_n)y_n^{j}(\theta_n)$$

$$= \sum_{j=r+1}^{k}\sum_{i=r+1}^{k}\sum_{m=1}^{k}\sum_{l=1}^{k} [\bar{\sigma}_{ij}(\theta_n)c_{il}(\theta_n)c_{jm}(\theta_n) + \rho''_{ijmln}(E_n)]n(\theta_n^{l} - \theta_n^{l})(\theta_n^{m} - \theta_n^{m}),$$

where

$$\lim_{n=\infty} P[|\rho''_{ijmln}(E_n)| < \nu | \theta_n] = 1$$

for any positive ν. If at least one of the integers l and m is greater than r, we have

$$\sum_{j=r+1}^{k}\sum_{i=r+1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{il}(\theta_n)c_{jm}(\theta_n) = c_{ml}(\theta_n).$$

Hence

$$\sum_{m=1}^{k}\sum_{l=1}^{k} v_m v_l c_{lm}(\theta_n) - \sum_{j=r+1}^{k}\sum_{i=r+1}^{k}\sum_{l=1}^{k}\sum_{m=1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{im}(\theta_n)c_{jm}(\theta_n)v_l v_m$$

$$= \sum_{m=1}^{r}\sum_{l=1}^{r} v_l v_m \left[c_{lm}(\theta_n) - \sum_{j=r+1}^{k}\sum_{i=r+1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{im}(\theta_n)c_{jl}(\theta_n) \right].$$

The coefficient

$$A_{lm}(\theta_n) = c_{lm}(\theta_n) - \sum_{r+1}^{k}\sum_{r+1}^{k} \bar{\sigma}_{ij}(\theta_n)c_{im}(\theta_n)c_{jl}(\theta_n)$$

can be written as the following ratio:

(193)
$$A_{lm}(\theta_n) = \frac{\begin{vmatrix} c_{lm}(\theta_n) & c_{l\,r+1}(\theta_n) & \cdots & c_{lk}(\theta_n) \\ c_{m\,r+1}(\theta_n) & c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdot & \cdots & \cdot \\ c_{mk}(\theta_n) & c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix}}{\begin{vmatrix} c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdots & \cdot \\ c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix}}.$$

It is known that if A' is the adjoint of any determinant A, and M and M'

are corresponding m-rowed minors of A and A' respectively, then M' is equal to the product of A^{m-1} by the algebraic complement of M.

Let A be the k-rowed determinant $|\sigma_{ij}(\theta_n)|$ $(i, j=1, \cdots, k)$ and let M be the $(k-r+1)$-rowed minors

$$M = \begin{vmatrix} \sigma_{lm}(\theta_n) & \sigma_{l\,r+1}(\theta_n) & \cdots & \sigma_{lk}(\theta_n) \\ \sigma_{r+1\,m}(\theta_n) & \sigma_{r+1\,r+1}(\theta_n) & \cdots & \sigma_{r+1\,k}(\theta_n) \\ \cdot & \cdot & \cdots & \cdot \\ \sigma_{km}(\theta_n) & \sigma_{k\,r+1}(\theta_n) & \cdots & \sigma_{kk}(\theta_n) \end{vmatrix}. \tag{194}$$

Then we have

$$M' = A^{k-r+1} \begin{vmatrix} c_{lm}(\theta_n) & c_{l\,r+1}(\theta_n) & \cdots & c_{lk}(\theta_n) \\ c_{m\,r+1}(\theta_n) & c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdot & \cdots & \cdot \\ c_{mk}(\theta_n) & c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix} = A^{k-r}\overline{M}, \tag{195}$$

where $\overline{M}$ denotes the algebraic complement of M. Let M_1 be the $(k-r)$-rowed minor $|\sigma_{ij}(\theta)|$ $(i, j=r+1, \cdots, k)$. Then we obtain

$$M_1' = A^{k-r} \begin{vmatrix} c_{r+1\,r+1}(\theta_n) & \cdots & c_{r+1\,k}(\theta_n) \\ \cdot & \cdots & \cdot \\ c_{k\,r+1}(\theta_n) & \cdots & c_{kk}(\theta_n) \end{vmatrix} = A^{k-r-1}\overline{M}_1, \tag{196}$$

where $\overline{M}_1$ denotes the algebraic complement of M_1.

From (193), (195) and (196) we obtain

$$A_{lm}(\theta_n) = \overline{M}/\overline{M}_1 = \bar{c}_{lm}(\theta_n) \qquad (l, m = 1, \cdots, r),$$

where $\|\bar{c}_{lm}(\theta_n)\| = \|\sigma_{lm}(\theta_n)\|^{-1}$ $(l, m=1, \cdots, r)$. Hence

$$q_n - \bar{q}_n = n\sum_{q=1}^{r}\sum_{p=1}^{r}(\dot{\theta}_n^p - \theta_n^p)(\dot{\theta}_n^q - \theta_n^q)[\bar{c}_{pq}(\theta_n) + \eta_{pqn}(E_n)] + \rho_n(E_n),$$

where for any positive ν

$$\lim_{n=\infty} P[\,|\eta_{pqn}(E_n)| < \nu \,|\, \theta_n] = \lim_{n=\infty} P[\,|\rho_n(E_n)| < \nu \,|\, \theta_n] = 1.$$

From (187) it follows that for any positive ϵ

$$\lim_{n=\infty} P\Big\{-2\log\lambda_n(\omega, E_n) - \epsilon \leqq \sum_{1}^{r}\sum_{1}^{r} n(\dot{\theta}_n^p - \theta_n^p)(\dot{\theta}_n^q - \theta_n^q) \cdot [\bar{c}_{pq}(\theta_n) + \eta_{pqn}(E_n)] \leqq -2\log\lambda_n(\omega, E_n) + \epsilon \,\Big|\, \theta_n\Big\} = 1. \tag{197}$$

Since $\theta_n^p = \theta_0^p$ $(p = 1, \cdots, r)$ from (197) we obtain because of the boundedness of the sequence $\{n^{1/2}(\theta_n - \theta_n)\}$

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leqq \sum\sum n(\dot{\theta}_n^p - \theta_0^p)(\dot{\theta}_n^q - \theta_0^q)\bar{c}_{pq}(\theta_n) \leqq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1.$$

The above equation remains valid if we substitute $\dot{\theta}_n$ for θ_n, that is

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leqq n\sum\sum (\dot{\theta}_n^p - \theta_0^p)(\dot{\theta}_n^q - \theta_0^q)\bar{c}_{pq}(\dot{\theta}_n) \leqq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1. \tag{198}$$

Let W_n^* be the critical region defined in (115). Since (173) holds for any sequence $\{\theta_n\}$ for which $\lim n^{1/2}\delta(\theta_n, \omega) = \infty$, we obtain from (198)

$$\lim_{n=\infty} \{P(W_n^* \mid \theta) - P[L_n(\omega) \mid \theta]\} = 0 \tag{199}$$

uniformly in θ.

Now we consider the general case where ω is given by r equations

$$\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$$

such that Assumption VI is satisfied. As we have seen in §12, the whole theory remains valid if we replace the parameters $\theta^1, \cdots, \theta^k$ by the new parameters

$$\xi^1 = \xi^1(\theta), \cdots, \xi^k = \xi^k(\theta),$$

where the functions $\xi^1(\theta), \cdots, \xi^k(\theta)$ satisfy Assumption VI. Hence from (199) it follows that

$$\lim_{n=\infty} \{P(W_n^* \mid \theta) - P[L_n(\omega) \mid \theta]\} = 0 \tag{200}$$

uniformly in θ, where W_n^* denotes the region defined in (161).

From (200) and Theorem VII we obtain the following theorem.

THEOREM VIII. *Let $S_c(\theta)$ be the surface defined in* (162) *and $\eta(\theta)$ be the weight function defined in* (165). *If Assumption* VI *holds, then for testing the hypothesis $\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$ the likelihood ratio test*

(a) *has asymptotically best average power with respect to the surfaces $S_c(\theta)$ and weight function $\eta(\theta)$;*

(b) *has asymptotically best constant power on the surfaces $S_c(\theta)$;*

(c) *is an asymptotically most stringent test.*

14. **Large sample distribution of the likelihood ratio.** S. S. Wilks[6] has derived the large sample distribution of the likelihood ratio $\lambda_n(\omega, E_n)$ if ω is

[6] S. S. Wilks, *Distribution of the likelihood ratio in large samples*, Ann. Math. Statist. 1938.

a linear subspace of the parameter space and if the hypothesis to be tested is true. Here we derive the large sample distribution of $\lambda_n(\omega, E_n)$ for any set ω satisfying Assumption VI in both cases, when the hypothesis to be tested is true, and when it is not true.

Let $u_1, \cdots, u_r$ be r independently and normally distributed variates with unit variances. Denote the expected value of u_p by μ_p. The distribution of the statistic

$$U^2 = u_1^2 + \cdots + u_r^2$$

is known(7). The only parameter involved in this distribution is $\lambda^2 = \mu_1^2 + \cdots + \mu_r^2$. Let us denote the cumulative distribution of U^2 by $F_r(\lambda^2, t)$, that is,

$$P[(U^2 < t)] = F_r(\lambda^2, t) \qquad (\lambda^2 = \mu_1^2 + \cdots + \mu_r^2). \tag{201}$$

Obviously $F_r(0, t)$ is the χ^2-distribution with r degrees of freedom.

Let $v_1, \cdots, v_r$ be r variates which have a joint normal distribution. Denote by μ_p the mean value of v_p and by σ_{pq} the covariance between v_p and v_q. Consider the statistic

$$V^2 = \sum_{q=1}^{r} \sum_{p=1}^{r} \lambda_{pq} v_p v_q, \tag{202}$$

where $\|\lambda_{pq}\| = \|\sigma_{pq}\|^{-1}$. It is easy to verify that the distribution of V^2 is given by

$$P(V^2 < t) = F_r(\lambda^2, t), \tag{203}$$

where

$$\lambda^2 = \sum \sum \lambda_{pq} \mu_p \mu_q. \tag{204}$$

We will now derive the limit distribution of the expression on the left-hand side of (161), that is of the statistic

$$Q_n = n \sum_{q=1}^{r} \sum_{p=1}^{r} \xi^p(\dot{\theta}_n) \xi^q(\dot{\theta}_n) \bar{c}^*_{pq}(\dot{\theta}_n). \tag{205}$$

The joint distribution of the variates $n^{1/2}[\xi^1(\dot{\theta}_n) - \xi^1(\theta)], \cdots, n^{1/2}[\xi^r(\dot{\theta}_n) - \xi^r(\theta)]$ converges with $n \to \infty$ uniformly towards the cumulative normal distribution with zero means and covariance matrix $\|\sigma^*_{pq}(\theta)\| = \|\bar{c}^*_{pq}(\theta)\|^{-1}$. Since $\dot{\theta}_n$ is a uniformly consistent estimate of θ and since $c_{ij}(\theta)$ is a uniformly continuous function of θ, the statistic

$$\overline{Q}_n = n \sum \sum \xi^p(\dot{\theta}_n) \xi^q(\dot{\theta}_n) \bar{c}^*_{pq}(\theta) \tag{206}$$

---

(7) See for instance P. C. Tang, *The power function of the analysis of variance tests*, Statistical Research Memoirs vol. 2 (1938).

has the same limit distribution as Q_n, that is,

$$\lim_{n=\infty} \{P[\overline{Q}_n < t \mid \theta] - P[Q_n < t \mid \theta]\} = 0 \tag{207}$$

uniformly in θ and t.

It is easy to see that

$$\lim_{n=\infty} \{P[\overline{Q}_n < t \mid \theta] - F_r[\lambda_n^2(\theta), t] = 0 \tag{208}$$

uniformly in θ and t, where

$$\lambda_n^2(\theta) = n \sum \sum \xi^p(\theta)\xi^q(\theta)\bar{c}_{pq}^*(\theta). \tag{209}$$

Hence, because of (207) we have

$$\lim_{n=\infty} \{P[Q_n < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0 \tag{210}$$

uniformly in θ and t.

Let $\{\theta_n\}$ $(n=1, 2, \cdots, \text{ad inf.})$ be a sequence of parameter points for which $n^{1/2}\delta(\theta_n, \omega)$ is bounded. Then we obtain from (198)

$$\lim_{n=\infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leqq Q_n \leqq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1 \tag{211}$$

for any positive ϵ. From (210) and (211) it follows that

$$\lim_{n=\infty} \{P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta_n] - F_r[\lambda_n^2(\theta_n), t]\} = 0 \tag{212}$$

uniformly in t. Since (173) holds for any sequence $\{\theta_n\}$ for which $\lim n^{1/2}\delta(\theta_n, \omega) = +\infty$, we obtain from (212)

$$\lim_{n=\infty} \{P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0 \tag{213}$$

uniformly in θ and t. Hence we have proved the following theorem.

THEOREM IX. *Let $F_r(\lambda^2, t)$ be the distribution function defined in* (201) *and let $\lambda_n(\omega, E_n)$ be the likelihood ratio statistic for testing the hypothesis $\xi^1(\theta) = \cdots = \xi^r(\theta) = 0$. Let furthermore $\lambda_n^2(\theta)$ be the expression defined in* (209). *Then, if Assumption* VI *holds, we have*

$$\lim_{n=\infty} \{P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0$$

uniformly in t and θ. If the hypothesis to be tested is true, that is if θ is a point of ω, $\lambda_n^2(\theta) = 0$ and therefore the limit distribution of $-2 \log \lambda_n(\omega, E_n)$ is the χ^2-distribution with r degrees of freedom.

15. **Summary.** Let $f(x^1, \cdots, x^m, \theta^1, \cdots, \theta^k)$ be the joint probability density function of the variates $x^1, \cdots, x^m$ involving k unknown parameters $\theta^1, \cdots, \theta^k$. Any set of values $\theta^1, \cdots, \theta^k$ can be represented by a point θ of the k-dimensional Cartesian space with the coordinates $\theta^1, \cdots, \theta^k$. Let ω be the subset of the parameter space defined by the equations

$$\xi^1(\theta) = \xi^2(\theta) = \cdots = \xi^r(\theta) = 0 \qquad (r \leqq k),$$

that is, ω is the set of all points θ for which the above equations are fulfilled. Denote by H_ω the hypothesis that the true parameter point θ is an element of ω. In this paper the question of an appropriate test of the hypothesis H_ω is discussed when the number of observations is large.

The following notations have been introduced. The point $\dot{\theta}_n$ denotes the point with the coordinates $\dot{\theta}_n^1, \cdots, \dot{\theta}_n^k$ where $\dot{\theta}_n^i$ is the maximum likelihood estimate of θ^i based on n independent observations on $x^1, \cdots, x^m$. The expected value of $-\partial^2 \log f(x^1, \cdots, x^m, \theta^1, \cdots, \theta^k)\partial\theta^i\partial\theta^j$ is denoted by $c_{ij}(\theta)$ and $\|\sigma_{ij}(\theta)\| = \|c_{ij}(\theta)\|^{-1}$. Furthermore $\sum_{l=1}^{k}\sum_{m=1}^{k}(\partial\xi^p/\partial\theta^l)(\partial\xi^q/\partial\theta^m)\sigma_{lm}(\theta)$ $(p, q = 1, \cdots, r)$ is denoted by $\sigma_{pq}^*(\theta)$ and

$$\|\bar{c}_{pq}^*(\theta)\| = \|\sigma_{pq}^*(\theta)\|^{-1} \qquad (p, q = 1, \cdots, r).$$

The region W_n^* denotes the critical region defined by the inequality

$$n\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\dot{\theta}_n)\xi^q(\dot{\theta}_n)\bar{c}_{pq}^*(\dot{\theta}_n) \geqq d_n,$$

where n is the number of independent observations on $x^1, \cdots, x^m$, and the constant d_n is chosen so that the least upper bound of the probability that the sample point falls within W_n^*, calculated under the restriction that the true parameter point lies in ω, is equal to a given positive $\alpha < 1$.

Let λ_n be the likelihood ratio statistic for testing H_ω and let L_n be the critical region defined by the inequality

$$\lambda_n \leqq \bar{\lambda}_n,$$

where the constant $\bar{\lambda}_n$ is chosen so that the least upper bound of the probability that the sample point falls within L_n, calculated under the restriction that the true parameter point lies in ω, is equal to α.

Under certain assumptions on $f(x^1, \cdots, x^m, \theta^1, \cdots, \theta^k)$ and the functions $\xi^1(\theta), \cdots, \xi^r(\theta)$ the following results have been obtained:

I. For testing the hypothesis H_ω the critical regions W_n^* and L_n both: (1) have asymptotically best average power over a family of surfaces defined in (162); (2) have asymptotically best constant power along the surfaces defined in (162); (3) are asymptotically most stringent tests. The exact definitions of these notions are given in Definitions VIII, X and XII, respectively.

II. The statistics $-2 \log \lambda_n$ and $n\sum_{q=1}^{r}\sum_{p=1}^{r}\xi^p(\dot{\theta}_n)\xi^q(\dot{\theta}_n)\bar{c}_{pq}^*(\dot{\theta}_n)$ have the

same limit distribution. The limit distribution of $-2 \log \lambda_n$ is the χ^2-distribution with r degrees of freedom if the hypothesis to be tested is true. If the true parameter point θ_n is not an element of ω, the distribution of $-2 \log \lambda_n$ approaches the distribution of a sum of non-central squares

$$U^2 = u_1^2 + \cdots + u_r^2,$$

where the variates $u_1, \cdots, u_r$ are independently and normally distributed with unit variances and

$$\sum_{p=1}^{r} (Eu_p)^2 = n \sum \sum \xi^p(\theta_n)\xi^q(\theta_n)\bar{c}^*_{pq}(\theta_n).$$

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.