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

BY

ABRAHAM WALD

Table of contents

1. Introduction ......................................................... 426
2. Assumptions on the density function \( f(x, \theta) \) .......................... 428
3. The joint limit distribution of \( \hat{\theta}_n \) ................................... 429
4. Reduction of the general problem to the case of a multivariate normal distribution 433
5. Tests of simple hypotheses which have uniformly best average power over a family of surfaces ..................................................... 445
6. Tests of simple hypotheses which have best constant power on a family of surfaces ... 450
7. Most stringent tests of simple hypotheses .................................. 451
8. Definitions of “best” tests of composite hypotheses .......................... 453
9. Tests of linear composite hypotheses which have uniformly best average power over a family of surfaces ..................................................... 455
10. Tests of linear composite hypotheses which have best constant power on a family of surfaces ..................................................... 461
11. Most stringent tests of linear composite hypotheses ....................... 461
12. The general composite hypothesis ..................................... 463
13. Optimum properties of the likelihood ratio test ........................... 470
14. Large sample distribution of the likelihood ratio .......................... 478
15. Summary .................................................................. 481

1. Introduction. In this paper we shall deal with the following general problem: Let \( f(x_1, x_2, \ldots, x_r, \theta_1, \ldots, \theta_k) \) be the joint probability density function of the variates (chance variables) \( x_1, \ldots, x_r \) involving \( k \) unknown parameters \( \theta_1, \ldots, \theta_k \). Any set of \( k \) values \( \theta_1, \ldots, \theta_k \) can be represented by a point \( \theta \) in the \( k \)-dimensional Cartesian space with the coordinates \( \theta_1, \ldots, \theta_k \). We shall denote the set of all possible parameter points by \( \Omega \). The set \( \Omega \) is called parameter space. The parameter space \( \Omega \) may be the whole \( k \)-dimensional Cartesian space, or a subset of it. For any subset \( \omega \) of \( \Omega \), we shall denote by \( H_\omega \) the hypothesis that the parameter point lies in \( \omega \). If \( \omega \) consists of a single point, \( H_\omega \) is called a simple hypothesis, otherwise \( H_\omega \) is called a composite hypothesis. In this paper we shall discuss the question of an appropriate test of the hypothesis \( H_\omega \) based on a large number of independent observations on \( x_1, \ldots, x_r \).

For simplicity we shall introduce the following notations: The letter \( \theta \) or \( \theta_i \) for any subscript \( i \) will denote a point in the parameter space \( \Omega \). The letter \( x \)
will denote the random vector with the components \(x_1, \ldots, x_r\), and \(x_a\) will denote the random vector with the components \(x_{a1}, \ldots, x_{ar}\) where \(x_{ai}\) \((i = 1, 2, \ldots, r)\) denotes the \(a\)th observation on \(x^i\). In general, the components of a vector \(v\) in the \(s\)-dimensional space will be denoted by \(v_1, \ldots, v_s\), 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 \(\theta\) 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 \(\theta\) 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 \(\theta_n, \ldots, \theta^k\) of \(\theta^1, \ldots, \theta^k\) we mean values of \(\theta^1, \ldots, \theta^k\) for which \(\prod_{i=1}^{n} f(x_{ai}^1, \ldots, x_{ai}^r, \theta^1, \ldots, \theta^k)\) becomes a maximum. The subscript \(n\) in the symbol \(\theta_n\) will indicate that the maximum likelihood estimate is based on \(n\) independent observations on \(x_1, \ldots, x_r\). A region \(W_n\) in the \(rn\)-dimensional sample space is called a critical region for testing the hypothesis \(H_n\) if we decide to reject \(H_n\) when and only when the observed sample point falls within \(W_n\). For any \(\theta\) not in \(\omega\) the value of \(P(W_n | \theta)\) is called the power of the critical region \(W_n\) with respect to the alternative hypothesis \(\theta\). The least upper bound of \(P(W_n | \theta)\) with respect to \(\theta\), restricting \(\theta\) to \(\omega\), 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 \(\theta\) 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 \(n^{1/2}(\theta_n - \theta_0) \geq A_n\) 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 \(\theta_n\) denotes the vector with the components \(\theta_n^1, \ldots, \theta_n^k\) where \(\theta_n^i\) is the maximum likelihood estimate of \(\theta^i\), and if \(A\) is a real number, then the inequality
\[
n^{1/2}(\theta_n - \theta) < A
\]

denotes the set of inequalities
\[ n^{1/2}(\theta_n^i - \theta^i) < A \quad (i = 1, \ldots, k). \]
Or, if \( t \) is a vector with the components \( t^1, \ldots, t^k \), then
\[ n^{1/2}(\theta_n - \theta) < t \]
denotes the set of inequalities
\[ n^{1/2}(\theta_n^i - \theta^i) < t^i \quad (i = 1, \ldots, k). \]

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 \( E\psi(x) \) the expected value of \( \psi(x) \) under the assumption that \( \theta \) is the true parameter point, that is
\[ E\psi(x) = \int_{-\infty}^{+\infty} \psi(x) f(x, \theta) dx. \]

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

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 \( \theta_n = (\theta_{n1}, \ldots, \theta_{nk}) \) exists and the second order partial derivatives \( \partial^2 f(x, \theta)/\partial \theta^i \partial \theta^j \) (\( \alpha = 1, \ldots, n; i, j = 1, \ldots, k \)) are continuous functions of \( \theta \). It is assumed that
\[ \lim_{n \to \infty} P(D_n | \theta) = 1 \quad \text{uniformly in } \theta. \]

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 \( \theta_n \) as a single-valued function of \( E_n \) defined for all points of \( D_n \).

**Assumption II.** For any positive \( \epsilon \)
\[ \lim_{n \to \infty} P \left[ \left| \theta_n - \theta \right| < \epsilon \mid \theta \right] = 1 \]
uniformly in \( \theta \), where \( \theta_n \) denotes the vector with the components \( \theta^1_n, \ldots, \theta^k_n \) and \( \theta^i_n \) is the maximum likelihood estimate of \( \theta^i \).

Assumption II is somewhat more than consistency of the maximum likelihood estimate \( \hat{\theta}_n \). In fact consistency means only that for any positive \( \epsilon \)
\[ \lim_{n \to \infty} P \left[ \left| \hat{\theta}_n - \theta \right| < \epsilon \mid \theta \right] = 1, \]
without requiring that the convergence be uniform in \( \theta \). If \( \hat{\theta}_n \) satisfies Assump-
tion II we shall say that $\theta_n$ is a uniformly consistent estimate of $\theta$. A rigorous proof of the consistency of $\theta_n$ (under certain restrictions on $f(x, \theta)$) was given by J. L. Doob\(^{(3)}\). The uniform consistency of $\theta_n$ 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 $\lim_{n \to \infty} \theta_{1n} = \lim_{n \to \infty} \theta_{2n} = \theta$ and $\lim \delta_n = 0$ we have

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

uniformly in $\theta$.

(b) There exists a positive $\epsilon$ such that the expectations $E_{\theta} [\psi_{ij}(x, \theta_2, \delta)]^2$ and $E_{\theta} [\psi_{ij}(x, \theta_2, \delta)]^2$ are bounded functions of $\theta_1$, $\theta_2$ and $\delta$ in the domain $D$, defined by the inequalities $|\theta_1 - \theta_2| \leq \epsilon$ and $|\delta| \leq \epsilon$.

(c) The greatest lower bound with respect to $\theta$ 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.** $\int_{-\infty}^{+\infty} \frac{\partial f(x, \theta)}{\partial \theta^i} dx = \int_{-\infty}^{+\infty} \frac{\partial^2 f(x, \theta)}{\partial \theta^i \partial \theta^j} dx = 0$.

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

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

identically in $\theta$. 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 $\eta$ such that

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

are bounded functions of $\theta$.

3. **The joint limit distribution of $\theta_n$.** Denote $n^{1/2}(\theta_n - \theta^i)$ by $z_n^i(\theta^i)$ and let $z_n(\theta)$ be the vector with the components $z_n^i(\theta^i), \cdots, z_n^k(\theta^k)$. 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.

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

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

where

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

Proof. Because of Assumption IV, we have

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

From

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

we obtain because of Assumption IV

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

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

\[
\sum_a \frac{\partial \log f(x_a, \theta)}{\partial \theta^i} = \sum_a \frac{\partial \log f(x_a, \theta)}{\partial \theta^i} + \sum_i (\theta_i - \bar{\theta}) \left[ \sum_a \frac{\partial^2 \log f(x_a, \bar{\theta})}{\partial \theta^i \partial \theta^j} \right],
\]

where \( \bar{\theta} \) lies on the segment connecting the points \( \theta \) and \( \theta_i \). Denote \( n^{-1/2} \sum \partial \log f(x_a, \theta) / \partial \theta^i \) by \( y_n(\theta) \) and let \( y_n(\theta) \) be the vector with the components \( y_n^1(\theta), \ldots, y_n^k(\theta) \). Substituting \( \theta_i \) for \( \theta_i \) in (3), the left-hand side of (3) becomes equal to zero and we obtain

\[
y_n^i(\theta) + \sum_i \left\{ n^{1/2} (\theta_i - \bar{\theta}) \right\} \frac{1}{n} \left[ \sum_a \frac{\partial^2 \log f(x_a, \bar{\theta})}{\partial \theta^i \partial \theta^j} \right] = 0
\]

or

\[
y_n^i(\theta) + \sum_i z_n^i(\theta) \frac{1}{n} \left[ \sum_a \frac{\partial^2 \log f(x_a, \bar{\theta})}{\partial \theta^i \partial \theta^j} \right] = 0.
\]
Let $\nu$ be an arbitrary positive number and let $Q_n(\theta)$ be the subset of $D_n$ for which the inequality

$$
\left| \frac{1}{n} \sum_a \frac{\partial^2 \log f(x_a, \bar{\theta})}{\partial \theta_i \partial \theta^i} + c_i(\theta) \right| < \nu
$$

holds. We will prove that

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

uniformly in $\theta$.

Let $\tau_0$ be a positive number such that

$$
\left| \psi_i(x, \theta, \tau_0) \right| < \frac{\nu}{2}
$$

for all values of $\theta$. Because of condition (a) of Assumption III, such a $\tau_0$ certainly exists. Denote by $R_n(\theta)$ the subset of $D_n$ consisting of all points $E_n$ for which the inequality

$$
\left| \theta_n - \theta \right| \leq \tau_0
$$

holds. Because of Assumption II

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

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

$$
\left| \bar{\theta} - \theta^i \right| \leq \tau_0 \quad (i = 1, \ldots, k).
$$

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

$$
\sum_a \psi_i(x_a, \theta, \tau_0) \leq \sum_a \frac{\partial^2 \log f(x_a, \bar{\theta})}{\partial \theta_i \partial \theta^i} \leq \sum_a \phi_i(x_a, \theta, \tau_0)
$$

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

$$
\left| \frac{1}{n} \sum_a \phi_i(x_a, \theta, \tau_0) - E\phi_i(x, \theta, \tau_0) \right| < \frac{\nu}{2}
$$

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

$$
\left| \frac{1}{n} \sum_a \psi_i(x_a, \theta, \tau_0) - E\psi_i(x, \theta, \tau_0) \right| < \frac{\nu}{2}.
$$

It follows from (b) of Assumption III and Tshebysheff’s inequality that
\[ \lim_{n \to \infty} P\left[ S_n(\theta) \mid \theta \right] = \lim_{n \to \infty} P\left[ T_n(\theta) \mid \theta \right] = 1 \]

uniformly in \( \theta \). 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)

\[ \left| \frac{1}{n} \sum_{a} \phi_{ij}(x_a, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial \theta^i \partial \theta^j} \right| < \nu, \]

\[ \left| \frac{1}{n} \sum_{a} \psi_{ij}(x_a, \theta, \tau_0) - E_\theta \frac{\partial^2 \log f(x, \theta)}{\partial \theta^i \partial \theta^j} \right| < \nu. \]

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

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

\[ z^i_n(\theta) = \sum_j \gamma^i_n(\theta) \left[ \sigma_{ij}(\theta) + \nu \epsilon_{iun}(E_n, \theta, \nu) \right], \]

where \( \epsilon_{iun}(E_n, \theta, \nu) \) is a bounded function of \( E_n, \theta \) and \( \nu \), provided that, for each \( \theta \), \( E_n \) is restricted to points of \( Q_n(\theta) \) and \( \nu \) is less than a certain positive number \( \nu_0 \).

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

\[ \tilde{z}_n(\theta) = \sum_i \gamma^i_n(\theta) \sigma_{ij}(\theta) \]

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

\[ \lim_{n \to \infty} \left\{ P\left[ \tilde{z}_n(\theta) < t \mid \theta \right] - P\left[ z_n(\theta) < t \mid \theta \right] \right\} = 0 \]

uniformly in \( t \) and \( \theta \).

Denote \( \sum_i \gamma^i_n(\theta) \sigma_{ij}(\theta) \) by \( \tilde{z}^i_n(\theta) \) and let \( \hat{z}_n(\theta) \) be the vector with the components \( \hat{z}^1_n(\theta), \ldots, \hat{z}^k_n(\theta) \). 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 \( \theta \), towards the \( k \)-variate cumulative normal distribution with zero means and covariance matrix \( [e_{ij}(\theta)] \). From this we easily obtain that \( P[\hat{z}_n(\theta) < t \mid \theta] \) converges with \( n \to \infty \), uniformly in \( t \) and \( \theta \), towards the cumulative joint normal distribution with zero means and covariance matrix \( [\sigma_{ij}(\theta)] \). Since \( \nu \) can be chosen arbitrarily small, we obviously have because of (17)

\[ \lim_{n \to \infty} \left\{ P[\hat{z}_n(\theta) < t \mid \theta] - P[\tilde{z}_n(\theta) < t \mid \theta] \right\} = 0 \]

uniformly in \( t \) and \( \theta \). 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 \( n \)-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 \( n \)-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 \( b_n(E'_n) = b_n(E_n) \).

(b) \( \lim_{n \to \infty} \{ P[W^*_n(W_n) \mid \theta] - P[W_n \mid \theta] \} = 0 \) uniformly in \( \theta \) and \( W_n \).

**Proof.** Let \( \lambda \) 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 \( \left| n^{1/2} (\theta - \theta) \right| \leq \lambda \). Similarly let \( W^*_n(W_n, \theta, \lambda) \) be the intersection of \( W^*_n(W_n) \) and the region \( \left| n^{1/2} (\theta - \theta) \right| \leq \lambda \).

For any function \( \Phi(v) \) we will denote by \( \text{g.l.b.} \Phi(v) \) and \( \text{l.u.b.} \Phi(v) \) 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 \( \lim_{n \to \infty} \lambda_n = \infty \)

\[
\lim_{n \to \infty} \{ \text{g.l.b.} \left[ P\left( \left| n^{1/2} (\theta_n - \theta) \right| \leq \lambda_n \mid \theta \right) \right] \} = 1,
\]

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

\[
\lim_{n \to \infty} \{ P[W_n(\theta, \lambda_n) \mid \theta] - P[W^*_n(W_n, \theta, \lambda_n) \mid \theta] \} = 0
\]

uniformly in \( W_n \) and \( \theta \).

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

\[
\frac{r_1 - 1/2}{qn^{1/2}} < \theta_n^1 < \frac{r_1 + 1/2}{qn^{1/2}}, \ldots, \frac{r_k - 1/2}{qn^{1/2}} < \theta_n^k < \frac{r_k + 1/2}{qn^{1/2}}.
\]

Furthermore denote by \( \theta_n^{(r_1, \ldots, r_k, q)} \) the parameter point with the coordinates \( r_1/qn^{1/2}, \ldots, r_k/qn^{1/2} \). We order the system of all sets of \( k \) integers \( (r_1, \ldots, r_k) \) in a sequence and we shall denote by \( I_n(q) \) the interval \( I_n(r_1, \ldots, r_k, q) \) where \( (r_1, \ldots, r_k) \) is the \( s \)-th element in the ordered sequence \( (s = 1, 2, \ldots, \text{ad inf.}) \). Similarly \( \theta_n^{(q)} \) denotes the parameter point \( \theta_n^{(r_1, \ldots, r_k, q)} \) where \( (r_1, \ldots, r_k) \) is the \( s \)-th element in the ordered sequence.
Let $J_n(W_n, q)$ be the common part of the three regions $I_{n}(q)$, $W_n$ and $Q_n[\theta_n(q)]$ where for any $\theta$, $Q_n(\theta)$ is defined by the inequalities
\begin{align*}
&\left| \frac{1}{n} \sum_a \phi_{ij}(x_a, \theta, \frac{1}{n^{1/2}}) + c_{ij}(\theta) \right| \leq \frac{1}{n^{1/2}} + v(n) + v(n), \\
&\left| \frac{1}{n} \sum_a \psi_{ij}(x_a, \theta, \frac{1}{n^{1/2}}) + c_{ij}(\theta) \right| \leq \frac{1}{n^{1/2}} + v(n) + v(n).
\end{align*}
(22)

The expression $v(n)$ is equal to l.u.b. $[E_{\theta} \phi_{ij}(x, \theta, 2/n^{1/2}) - E_{\theta} \phi_{ij}(x, \theta, 2/n^{1/2})]$ and the expression $v(n)$ is defined as follows: Denote by $c_{ij}(\theta, n)$ the least upper bound of $|c_{ij}(\theta) - c_{ij}(\theta)|$ with respect to $\theta$ where $\theta$ can take only values in the interval $[\theta - 1/n^{1/2}, \theta + 1/n^{1/2}]$. Then $v(n)$ is defined as the least upper bound of $c_{ij}(\theta, n)$ with respect to $\theta, i$ and $j$. Because of condition (a) of Assumption III, we obtain
\begin{equation}
\lim_{n \to \infty} v(n) = \lim_{n \to \infty} v(n) = 0.
\end{equation}
(23)

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

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

(24b) $\lim_{n \to \infty} \{ L.U.B. \{ P[J_n(W_n, q) | \theta_n(q)] - L.U.B. \{ I_n(q) | \theta_n(q)\} \} = 0$ where $L.U.B. \{ V_n \} \theta$ denotes the probability of $V_n$ calculated on the basis that the joint distribution of $n^{1/2}(\theta_n - \theta)$, $\cdot \cdot \cdot , n^{1/2}(\theta_n - \theta)$ is normal with zero means and covariance matrix $||\sigma_{ij}(\theta)|| = ||c_{ij}(\theta)||^{-1}$.

The existence of such a set $J^*_n(W_n, q)$ can be proved as follows: Obviously there exists a subset $J^*_n(W_n, q)$ of $I_{n}(q)$ such that (24a) is fulfilled and
\begin{equation}
L.U.B. \{ P[J_n(W_n, q) | \theta_n(q)] \} = \min \{ P[J_n(W_n, q) | \theta_n(q)] \}, L.U.B. \{ I_n(q) | \theta_n(q)\} \}.
\end{equation}

Since $J_n(W_n, q)$ is a subset of $I_{n}(q)$ and since $\lim_{n \to \infty} \{ P[J_n(q) | \theta_n(q)] \} - L.U.B. \{ I_n(q) | \theta_n(q)\} \} = 0$ uniformly in $s$, the above defined subset $J^*_n(W_n, q)$ satisfies also the condition (24b). We define
\begin{equation}
W^*_n(q) = \sum_{s=1}^{\infty} J^*_n(W_n, q).
\end{equation}
(25)

Furthermore we define the regions $J_n(W_n, \theta, \lambda, q)$ and $J^*_n(W_n, \theta, \lambda, q)$ as follows:
\begin{align*}
&J_n(W_n, \theta, \lambda, q) = \sum_s J_n(W_n, q), \\
&J^*_n(W_n, \theta, \lambda, q) = \sum_s J^*_n(W_n, q).
\end{align*}
(26)
where the summation is to be taken over all values of \( s \) for which \( |\theta - \theta_{ns}(q)| \leq \lambda/n^{1/2} \).

Let \( T_n^*(W_n, q) \) be the intersection of \( J_n^*(W_n, q) \) and \( Q_n[\theta_{ns}(q)] \), and let \( J_n^*(W_n, \theta, \lambda, q) = \sum_n T_n^*(W_n, q) \) where the summation is to be taken over all values of \( s \) for which \( |\theta - \theta_{ns}(q)| \leq \lambda/n^{1/2} \). Furthermore let \( W_n^*(W_n, \theta, \lambda, q) \) be the intersection of \( W_n^*(W_n, q) \) with the region \(|n^{1/2}(\theta_n - \theta)| \leq \lambda\).

If for a value of \( s \) we have \( n^{1/2}|\theta - \theta_{ns}(q)| \leq \lambda \) then for all points of \( I_n(q) \) we have \( n^{1/2}|\theta - \theta_{n}| \leq \lambda + 1/q \). If there exists a point in \( I_n(q) \) for which \( n^{1/2}|\theta - \theta_{n}| \leq \lambda - 1/q \) then \( n^{1/2}|\theta - \theta_{ns}(q)| \leq \lambda \). 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

\[
\left| P\left[ W_n^*(W_n, \theta, \lambda, q) \mid \theta \right] - P\left[ J_n^*(W_n, \theta, \lambda, q) \mid \theta \right] \right| 
\leq P\left[ W_n^*(W_n, \theta, \lambda + 1/q, q) \mid \theta \right] - P\left[ W_n^*(W_n, \theta, \lambda - 1/q, q) \mid \theta \right] 
\leq P\left[ |\lambda - 1/q| \leq n^{1/2} \mid \theta_n - \theta \right| \leq |\lambda + 1/q| \mid \theta \right].
\]

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 \( \theta \). Hence also \( E_\theta [\phi_{ij}(x, \theta, \delta)] \) and \( E_\theta [\psi_{ij}(x, \theta, \delta)] \) are bounded functions of \( \theta \). Substituting \( \delta = 0 \) we find that \( c_{ij}(\theta) \) is a bounded function of \( \theta \). 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

\[
\lim_{q \to 0} \left\{ \text{l.u.b.} \ P\left[ |\lambda - 1/q| \leq n^{1/2} \mid \theta_n - \theta \right| \leq |\lambda + 1/q| \mid \theta \right] \right\} = 0
\]

uniformly in \( \lambda \). From (28) and (29) we obtain

\[
\lim_{n \to \infty} \sup_{\theta, \theta_n} \left\{ \text{l.u.b.} \ P\left[ W_n^*(W_n, \theta, \lambda, q) \mid \theta \right] - P\left[ J_n^*(W_n, \theta, \lambda, q) \mid \theta \right] \right\} = \varepsilon_1(\lambda, q)
\]

where \( \lim_{q \to 0} \varepsilon_1(\lambda, q) = 0 \) uniformly in \( \lambda \).

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 \( |\theta - \theta_{ns}(q)| \leq \lambda/n^{1/2} \). 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

\[
\frac{1}{n} \sum_a \phi_{ij} \left( x_a, \theta, \frac{2}{n^{1/3}} \right) + c_{ij}(\theta) \leq \frac{1}{n^{1/3}} + \nu(n),
\]

\[
\frac{1}{n} \sum_a \psi_{ij} \left( x_a, \theta, \frac{2}{n^{1/3}} \right) + c_{ij}(\theta) \leq \frac{1}{n^{1/3}} + \nu(n).
\]

In fact, from (31) it follows that
\begin{equation}
\begin{aligned}
\phi_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leq \frac{1}{n} \sum_a \phi_{ij}\left( x_a, \theta_1, \frac{2}{n^{1/3}} \right) \leq \phi_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n), \\
\psi_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leq \frac{1}{n} \sum_a \psi_{ij}\left( x_a, \theta_1, \frac{2}{n^{1/3}} \right) \leq \psi_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n).
\end{aligned}
\end{equation}

Since

\begin{equation}
\begin{aligned}
\frac{1}{n} \sum_a \psi_{ij}\left( x_a, \theta_1, \frac{2}{n^{1/3}} \right) &\leq \frac{1}{n} \sum_a \phi_{ij}\left( x_a, \theta_n(q), \frac{1}{n^{1/3}} \right) \\
&\leq \frac{1}{n} \sum_a \phi_{ij}\left( x_a, \theta, \frac{2}{n^{1/3}} \right)
\end{aligned}
\end{equation}

and

\begin{equation}
\begin{aligned}
\frac{1}{n} \sum_a \psi_{ij}\left( x_a, \theta_1, \frac{2}{n^{1/3}} \right) &\leq \frac{1}{n} \sum_a \psi_{ij}\left( x_a, \theta_n(q), \frac{1}{n^{1/3}} \right) \\
&\leq \frac{1}{n} \sum_a \phi_{ij}\left( x_a, \theta, \frac{2}{n^{1/3}} \right)
\end{aligned}
\end{equation}

for almost all \( n \) and for all \( s \) for which \( n^{1/3} | \theta - \theta_n(q) | \leq \lambda \), we obtain from (32)

\begin{equation}
\begin{aligned}
\phi_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leq \frac{1}{n} \sum_a \phi_{ij}\left( x_a, \theta_n(q), \frac{1}{n^{1/3}} \right) \leq \phi_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n), \\
\psi_{ij}(\theta) - \frac{1}{n^{1/3}} - \nu(n) &\leq \frac{1}{n} \sum_a \psi_{ij}\left( x_a, \theta_n(q), \frac{1}{n^{1/3}} \right) \leq \psi_{ij}(\theta) + \frac{1}{n^{1/3}} + \nu(n),
\end{aligned}
\end{equation}

for almost all \( n \) and for all \( s \) for which \( n^{1/3} | \theta - \theta_n(q) | \leq \lambda \). Since \( | \phi_{ij} - \psi_{ij} | \) \( \leq \nu(n) \) for almost all \( n \) and for all \( s \) for which \( n^{1/3} | \theta - \theta_n(q) | \leq \lambda \), we obtain from (33)

\begin{equation}
\begin{aligned}
\phi_{ij}[\theta_n(q)] - \frac{1}{n^{1/3}} - \nu(n) - \nu(n) &\leq \frac{1}{n} \sum_a \phi_{ij}\left( x_a, \theta_n(q), \frac{1}{n^{1/3}} \right) \\
&\leq \phi_{ij}[\theta_n(q)] + \frac{1}{n^{1/3}} + \nu(n) + \nu(n),
\end{aligned}
\end{equation}

(34)

\begin{equation}
\begin{aligned}
\phi_{ij}[\theta_n(q)] - \frac{1}{n^{1/3}} - \nu(n) - \nu(n) &\leq \frac{1}{n} \sum_a \psi_{ij}\left( x_a, \theta_n(q), \frac{1}{n^{1/3}} \right) \\
&\leq \psi_{ij}[\theta_n(q)] + \frac{1}{n^{1/3}} + \nu(n) + \nu(n)
\end{aligned}
\end{equation}

for almost all \( n \) and for all \( s \) for which \( n^{1/3} | \theta - \theta_n(q) | \leq \lambda \). The inequalities (34) are equivalent to the inequalities (22) if in (22) \( \theta_n(q) \) is substituted for \( \theta \). Hence our statement about the region \( R_n(\theta, \lambda, q) \) is proved.
Consider the region $U_n(\theta)$ defined by the inequalities

$$
| \frac{1}{n} \sum_a \phi_{ij} \left( x_a, \theta, \frac{2}{n^{1/3}} \right) - E \phi_{ij} \left( x, \theta, \frac{2}{n^{1/3}} \right) | < \frac{1}{n^{1/3}},
$$

and

$$
| \frac{1}{n} \sum_a \psi_{ij} \left( x_a, \theta, \frac{2}{n^{1/3}} \right) - E \psi_{ij} \left( x, \theta, \frac{2}{n^{1/3}} \right) | < \frac{1}{n^{1/3}}.
$$

Since

$$
\nu(n) \geq \left| E \phi_{ij} \left( x, \theta, \frac{2}{n^{1/3}} \right) + \psi_{ij} \left( x, \theta, \frac{2}{n^{1/3}} \right) \right|
$$

and

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

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 Tshebycheff's inequality it follows that $\lim_{\theta \to \infty} P \left[ U_n(\theta) \mid \theta \right] = 1$ uniformly in $\theta$. Hence $\lim_{\theta \to \infty} P \left[ T_n(\theta) \mid \theta \right] = 1$ uniformly in $\theta$. Thus, as can easily be seen,

$$
(36) \quad \lim_{\theta \to \infty} P \left[ R_n(\theta, \lambda_n, q) \mid \theta \right] = 1
$$

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

Let $J_n(W_n, q)$ be the intersection of the regions $W_n$ and $J_n(q)$. Furthermore let $J_n(W_n, \theta, \lambda, q)$ be equal to $\sum s J_n(s)$ where the summation is to be taken over all values of $s$ for which $n^{1/2} | \theta - \theta_n(q) | \leq \lambda$. Then the common part of $\sum_{s=1}^{n} J_n(s)$ and $W_n(\theta, \lambda - 1/q)$ is a subset of $J_n(W_n, \theta, \lambda, q)$, and the last is a subset of the common part of $\sum_{s=1}^{n} J_n(s)$ and $W_n(\theta, \lambda + 1/q)$. Hence, since $P \left[ J_n(W_n, \theta, \lambda, q) \mid \theta \right] = 1$, we have

$$
| P \left[ W_n(\theta, \lambda) \mid \theta \right] - P \left[ J_n(W_n, \theta, \lambda, q) \mid \theta \right] |
$$

\(\leq P \left[ W_n(\theta, \lambda + 1/q) \mid \theta \right] - P \left[ W_n(\theta, \lambda - 1/q) \mid \theta \right] \)

\(\leq P \left[ \lambda - 1/q \leq n^{1/2} | \theta - \theta_n(q) | \leq \lambda + 1/q \mid \theta \right].
$$

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

$$
(38) \quad \lim_{n \to \infty} \left| P \left[ W_n(\theta, \lambda) \mid \theta \right] - P \left[ J_n(W_n, \theta, \lambda, q_n) \mid \theta \right] \right| = 0
$$

uniformly in $\theta$, $W_n$ and $\lambda$.

Since the common part of the regions $J_n(W_n, \theta, \lambda, q)$ 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 $J_n(W_n, \theta, \lambda, q)$, we obtain from (38) and (36) that for any bounded sequence
\{\lambda_n\} and for any sequence \{q_n\} for which \(\lim_{n \to \infty} q_n = \infty\) we have
\[
\lim_{n \to \infty} \left\{ P \left[ W_n(\theta, \lambda_n) \mid \theta \right] - P \left[ J_n(W_n, \theta, \lambda_n, q_n) \mid \theta \right] \right\} = 0
\]
uniformly in \(\theta\) 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 \(J_n^*(W_n, \theta, \lambda, q)\) 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 \(\lim_{n \to \infty} q_n = \infty\), we have
\[
\lim_{n \to \infty} \left\{ P \left[ J_n^*(W_n, \theta, \lambda_n, q_n) \mid \theta \right] - P \left[ J_n^*(W_n, \theta, \lambda_n, q_n) \mid \theta \right] \right\} = 0
\]
uniformly in \(\theta\) and \(W_n\).

Now we shall evaluate the limit values of \(P \left[ J_n(W_n, \theta, \lambda, q) \mid \theta \right]\) and \(P \left[ J_n^*(W_n, \theta, \lambda, q) \mid \theta \right]\). Denote by \(A_n(\lambda, q)\) the domain in the space of the variables \(\theta, W_n, \theta'\) and \(E_n\) defined as follows: \(\theta\) and \(W_n\) can take arbitrary values, \(\theta'\) is restricted to values for which \(|\theta' - \theta| \leq 1/n^{1/2}\); and for any \(\theta\) 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 \(J_n^*(W_n, \theta, \lambda, q)\). Denote furthermore by \(\rho_n(\theta', \theta, \theta_n)\) the function
\[
\rho_n(\theta', \theta, \theta_n) = -\frac{1}{2} \sum_i \sum_j n(\theta'^i - \theta'_n)(\theta'^j - \theta'_n) c_{ij}(\theta).
\]
Consider the Taylor expansion
\[
\sum_{\alpha} \log f(x_{\alpha}, \theta') = \sum_{\alpha} \log f(x_{\alpha}, \theta_n) + \frac{n}{2} \sum_i \left( \theta'^i - \theta'_n \right) \left( \theta'^i - \theta'_n \right) \frac{1}{n} \sum_{\alpha} \frac{\partial^2 \log f(x_{\alpha}, \theta)}{\partial \theta^i \partial \theta^i}
\]
where \(\theta\) lies in the interval \([\theta', \theta_n]\).

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 \(J_n^*(W_n, \theta, \lambda, q)\), it follows from the definitions of these sets that \(E_n\) lies in the set \(\sum Q_n([\theta_n(q)]\) where the summation is to be taken over all values of \(s\) for which \(|s| < 1/n^{1/2}\). Hence for any \(E_n\) in the domain \(A_n(\lambda, q)\) we have
\[
\frac{1}{n} \sum_{\alpha} \phi_{ij} \left( x_{\alpha}, \theta_n(q), \frac{1}{n^{1/2}} \right) + c_{ij} \left[ \theta_n(q) \right] \leq \frac{1}{n^{1/2}} + \nu(n) + \nu(n),
\]
\[
\frac{1}{n} \sum_{\alpha} \psi_{ij} \left( x_{\alpha}, \theta_n(q), \frac{1}{n^{1/2}} \right) + c_{ij} \left[ \theta_n(q) \right] \leq \frac{1}{n^{1/2}} + \nu(n) + \nu(n)
\]
for that value of \(s\) for which \(E_n\) lies in \(I_n(q)\). 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 \( I_\nu(q) \). Since \( n^{1/2} | \theta - \theta_n(q) | \leq \lambda \) in the domain \( A_n(\lambda, q) \), we have in the domain \( A_n(\lambda, q) \)

\[
\frac{1}{n} \sum_a \psi_{ij} \left[ x_a, \theta_n(q), \frac{1}{n^{1/3}} \right] \leq \frac{1}{n} \sum_a \psi_{ij} \left( x_a, \theta, \frac{1}{2n^{1/3}} \right)
\]

(44)

\[
\frac{1}{n} \sum_a \psi_{ij} \left[ x_a, \theta_n(q), \frac{1}{n^{1/3}} \right] \leq \frac{1}{n} \sum_a \phi_{ij} \left[ x_a, \theta_n(q), \frac{1}{n^{1/3}} \right],
\]

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

(45)

\[
| c_{ij}[\theta_n(q)] - c_{ij}(\theta) | \leq \nu(n)
\]

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

(46)

\[
\left| \frac{1}{n} \sum_a \phi_{ij} \left( x_a, \theta, \frac{1}{2n^{1/3}} \right) + c_{ij}(\theta) \right| \leq \frac{1}{n^{1/3}} + \nu(n) + 2\nu(n),
\]

(47)

\[
\left| \frac{1}{n} \sum_a \psi_{ij} \left( x_a, \theta, \frac{1}{2n^{1/3}} \right) + c_{ij}(\theta) \right| \leq \frac{1}{n^{1/3}} + \nu(n) + 2\nu(n)
\]

for almost all values of \( n \).

Since \( n^{1/2} | \theta - \theta_n(q) | \leq \lambda \) in \( A_n(\lambda, q) \), we have

(48)

\[
n^{1/2} | \theta - \theta_n | \leq \lambda + 1/q \text{ in } A_n(\lambda, q)
\]

and therefore

(49)

\[
n^{1/2} | \theta' - \theta_n | \leq 2\lambda + 1/q \text{ in } A_n(\lambda, q).
\]

Since \( \tilde{\theta} \) lies in the interval \([\theta', \theta_n] \), from (48) we obtain

(50)

\[
\sum_a \psi_{ij} \left( x_a, \theta, \frac{1}{2n^{1/3}} \right) \leq \sum_a \phi_{ij} \left( x_a, \theta, \frac{1}{2n^{1/3}} \right)
\]
for almost all $n$. Because of (23), (46) and (50) we obtain from (41) and (42)

\begin{equation}
\lim_{n \to \infty} \left\{ \text{l.u.b.} \left| \sum_{a} \log f(x_a, \theta') - \sum_{a} \log f(x_a, \theta) - \rho_n(\theta', \theta, \theta_n) \right| \right\} = 0,
\end{equation}

and

\begin{equation}
\lim_{n \to \infty} \left\{ \text{l.u.b.} \left[ \sum_{a} \log f(x_a, \theta) - \sum_{a} \log f(x_a, \theta') \right] - [\rho_n(\theta, \theta, \theta_n) - \rho_n(\theta', \theta, \theta_n)] \right\} = 0.
\end{equation}

Denote $P[J_n(W_n, q) | \theta_n(q)]$ by $P_n(W_n, q)$ and $P[J_n(W_n, q) | \theta_n(q)]$ by $P_n(W_n, q)$. Substituting $\theta_n(q)$ for $\theta'$ we easily obtain from (52)

\begin{equation}
\lim_{n \to \infty} \left\{ \text{l.u.b.} \left| P[J_n(W_n, \theta, \lambda, q) | \theta] - \sum_{s} P_n(W_n, q) \exp (\rho_n[\theta, \theta, \theta_n(q)] - \rho_n[\theta_n(q), \theta, \theta_n(q)]) \right| \right\} = 0
\end{equation}

where the summation with respect to $s$ is to be taken over all values for which $n^{1/2} | \theta - \theta_n(q) | \leq \eta$ and $\theta_n(q)$ is a parameter point for which

\begin{equation}
n^{1/2} | \theta_n(q) - \theta_n(q) | \leq 1/q.
\end{equation}

Since $c_2(\theta)$ is a uniformly continuous and bounded function of $\theta$, it follows from (41) and (54) that

\begin{equation}
| \rho(\theta, \theta, \theta_n(q)) - \rho[\theta, \theta, \theta_n(q)] | = \frac{\phi_1[\theta, \theta_n(q), \theta_n(q)]}{q},
\end{equation}

\begin{equation}
| \rho[\theta_n(q), \theta, \theta_n(q)] - \rho[\theta_n(q), \theta, \theta_n(q)] | = \frac{\phi_2[\theta, \theta_n(q), \theta_n(q)]}{q},
\end{equation}

where $\phi_1[\theta, \theta_n(q), \theta_n(q)]$ and $\phi_2[\theta, \theta_n(q), \theta_n(q)]$ are bounded functions of $\theta, \theta_n(q)$ and $\theta_n(q)$ in the domain $n^{1/2} | \theta - \theta_n(q) | \leq \eta$. From (53) and (55) it follows that

\begin{equation}
\lim_{n \to \infty} \left\{ \text{l.u.b.} \left| P[J_n(W_n, \theta, \lambda, q) | \theta] - \sum_{s} P_n(W_n, q) \exp (\rho_n[\theta, \theta, \theta_n(q)] - \rho_n[\theta_n(q), \theta, \theta_n(q)]) \right| \right\} = \epsilon(\lambda, q)
\end{equation}

where

\begin{equation}
\lim_{q \to \infty} \epsilon(\lambda, q) = 0
\end{equation}
uniformly in \( \lambda \) over any finite positive interval. Similarly we obtain

\[
\limsup_{n \to \infty} \left\{ \text{l.u.b.} \sup_{\theta, q} \left\{ P[J_n(W_n, \theta, \lambda, q) \mid \theta] - P_n^*(W_n, q) \right\} \right\} = \eta(\lambda, q)
\]

where

\[
\lim_{q \to \infty} \eta(\lambda, q) = 0
\]

uniformly in \( \lambda \) 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 \( |\theta - \theta_n(q)| \leq \lambda/n^{1/2} \). The expression \( \rho_n[\theta_n(q), \theta, \theta_n(q)] \) is obviously equal to zero, hence (56) and (58) can be simplified by substituting zero for this expression. Denote \( P[J_n^*(W_n, q), \theta_n(q)] \) by \( P_n^*(W_n, q) \). Because of Proposition I we have

\[
\lim_{n \to \infty} \{P[J_n^*(W_n, q), \theta_n(q)] - P_n^*(W_n, q)\} = 0
\]

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

\[
\lim_{n \to \infty} \{P_n(W_n, q) - P_n^*(W_n, q)\} = 0
\]

uniformly in \( s \) and \( W_n \). Since \( J_n^*(W_n, q) \) is the intersection of \( Q_n[\theta_n(q)] \) with \( J_n^*(W_n, q) \) and since

\[
\lim_{n \to \infty} P[Q_n[\theta_n(q)] \mid \theta_n(q)] = 1
\]

uniformly in \( s \), we have

\[
\lim_{n \to \infty} \{P_n(W_n, q) - P_n^*(W_n, q)\} = 0
\]

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

\[
\sup_{n \to \infty} \left\{ \text{l.u.b.} \sup_{\theta, q} \left\{ P[J_n(W_n, \theta, \lambda, q) \mid \theta] - P[J_n^*(W_n, \theta, \lambda, q) \mid \theta] \right\} \right\} = \zeta(\lambda, q) \leq \epsilon(\lambda, q) + \eta(\lambda, q).
\]

Hence

\[
\lim_{q \to \infty} \zeta(\lambda, q) = 0
\]
uniformly in \( \lambda \) over any finite positive interval.

For any positive \( \lambda' \) the sets \( J_n(W_n, \theta, \lambda, q' - J_n(W_n, \theta, \lambda, q), J_n*(W_n, \theta, \lambda, q') - J_n*(W_n, \theta, \lambda, q), J_n*(W_n, \theta, \lambda, q') - J_n(W_n, \theta, \lambda, q), W_n*(W_n, \theta, \lambda, q', q') - W_n*(W_n, \theta, \lambda, q) \) and \( W_n(\theta, \lambda + \lambda') - W_n(\theta, \lambda) \) are subsets of the set defined by the inequality

\[
\lambda - 1/q \leq n^{1/2} | \theta - \theta_n | \leq \lambda + \lambda' + 1/q.
\]

Since for any sequence \( \{ \lambda_n \} \) for which \( \lim_{n \to \infty} \lambda_n = \infty \) we have

\[
\lim_{n \to \infty} P[\lambda_n - 1/q \leq n^{1/2} | \theta - \theta_n | \leq \lambda_n + \lambda' + 1/q | \theta] = 0
\]

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

\[
\limsup_{n \to \infty} \{ \text{l.u.b. } P[W_n(\theta, \lambda) | \theta] - P[W_n*(W_n, \theta, \lambda, q) | \theta] | \} = \epsilon_3(\lambda, q),
\]

where

\[
\lim_{q \to \infty} \epsilon_3(\lambda, q) = 0
\]

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

\[
W_n^*(W_n) = W_n^*(W_n, q_{i+1}) \quad \text{for} \quad n_i < n \leq n_{i+1} \quad (i = 0, 1, 2, \ldots, \text{ad inf.}).
\]

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

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

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

\[
n_i > n_{i-1},
\]

\[
F_n(\lambda_i, q_{i+1}) < \epsilon_3(\lambda_i, q_{i+1}) + \eta_i
\]

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

Let \( \lambda'_i = \lambda_i, \eta'_i = \eta_i, \) and \( q'_i = q_{i+1} \) for \( n_i < n \leq n_{i+1} \) \( (i = 0, 1, 2, \ldots, \text{ad inf.}) \).

Then from (64), (65) and (67) we obtain

\[
\lim_{n \to \infty} \{ \text{l.u.b. } P[W_n(\theta, \lambda'_i) | \theta] - P[W_n^*(W_n, \theta, \lambda'_i, q'_i) | \theta] | \} = 0.
\]
Denote by $W_n^*(W_n, \theta, \lambda)$ the intersection of $W_n^*(W_n)$ and the set defined by the inequality $n^{1/2}|\theta - \theta_n| \leq \lambda$. Since $W_n^*(W_n, \theta, \lambda, q)$ is the intersection of $W_n^*(W_n, q)$ and the region $n^{1/2}|\theta - \theta_n| \leq \lambda$, it follows from (66) that

\begin{equation}
W_n^*(W_n, \theta, \lambda', q') = W_n^*(W_n, \theta, \lambda').
\end{equation}

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 $\theta_n(E'_n) = \theta_n(E_n)$ 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 $\Psi(V_n^*)|\theta)$ the probability that the sample point will fall in $V_n^*$ calculated under the assumption that $n^{1/2}|\theta_n - \theta|$, $\ldots$, $n^{1/2}|\theta_n - \theta|^k$ have a joint normal distribution with zero means and covariance matrix $||\sigma_i(\theta)|| = ||\sigma_i(\theta)||^{-1}$.

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

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

uniformly in $\theta$ and $R_n^*$.

**Proof.** Since we assumed that the set $J_n^*(W_n, q)$ defined in (24) is an interval in the space of the maximum likelihood estimates, it follows from Proposition 1 that

\begin{equation}
\lim_{n \to \infty} \{P[J_n^*(W_n, q)|\theta] - \Psi[J_n^*(W_n, q)|\theta]\} = 0
\end{equation}

uniformly in $\theta$, $W_n$, 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 $n^{1/2}|\theta - \theta_n| \leq \lambda$. For given values of $\lambda$ and $q$ the number of different values of $s$, for which $J_n^*(W_n, q)$ has at least a point common with the region $n^{1/2}|\theta - \theta_n| \leq \lambda$, is a bounded function of $\theta$. Hence it follows from (70) that

\begin{equation}
\lim_{n \to \infty} \{P[W_n^*(W_n, \theta, \lambda, q)|\theta] - \Psi[W_n^*(W_n, \theta, \lambda, q)|\theta]\} = 0
\end{equation}

uniformly in $\theta$ and $W_n$. From (64), (65) and (71) we obtain

\begin{equation}
\limsup_{n \to \infty} \{\text{l.u.b.} \{P[W_n(\theta, \lambda)|\theta] - \Psi[W_n(\theta, \lambda, q)|\theta]\}\} = \epsilon(\lambda, q)
\end{equation}

where
uniformly in \( \lambda \). The set \( W_n(\theta, \lambda) \) denotes the intersection of \( W_n \) and the region \( n^{1/2}|\theta - \hat{\theta}_{n}| \leq \lambda \).

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

\[
W_n^*(W_n) = W_n(W_n, q_{i+1}) \quad \text{for} \quad n_i < n \leq n_{i+1} \quad (i = 0, 1, 2, \cdots, \text{ad inf.}) .
\]

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

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

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

\[
n_i > n_{i-1},
\]

\[
F_n(\lambda_i, q_{i+1}) < \epsilon(\lambda_i, q_{i+1}) + \eta_i
\]

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

\[
\lim_{n \to \infty} \{ \text{l.u.b. } | P[W_n(\theta, \lambda_i')] - \mathbb{P}[W_n^*(W_n, \theta, \lambda_i', q_i')] | \} = 0 .
\]

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

\[
W_n^*(W_n, \theta, \lambda_i') = W_n^*(W_n, \theta, \lambda_i', q_i') .
\]

Hence from (76) we obtain

\[
\lim_{n \to \infty} \{ \text{l.u.b. } | P[W_n(\theta, \lambda_i')] - \mathbb{P}[W_n^*(W_n, \theta, \lambda_i')] | \} = 0 .
\]

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

\[
\lim_{n \to \infty} \{ \text{l.u.b. } | P[W_n(\theta, \lambda_i')] - \mathbb{P}[W_n^*(W_n, \theta)] | \} = 0 .
\]

The region \( W_n \) may be any Borel measurable subset of the \( r^n \)-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 \( \Psi[V^*_n|\theta] \) for \( P[V^*_n|\theta] \). Hence, if the sample is sufficiently large, the problem of statistical inference concerning the unknown parameter \( \theta \) 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 \( \theta \). Consider furthermore a non-negative function \( w(\theta) \) of \( \theta \), called a weight function. For any function \( \psi(\theta) \) of \( \theta \) 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 \( \int_{K_c} P(W_n|\theta)w(\theta)dA \geq \int_{K_c} P(Z_n|\theta)w(\theta)dA \) for all values \( c \) for which \( K_c \) is defined.

Let \( y^1, \ldots, y^k \) be \( k \) variates which have a joint normal distribution. The mean values \( \theta^1, \ldots, \theta^k \) of the variates \( y^1, \ldots, y^k \) are unknown, but the covariance matrix \( \|\sigma_{ij}\| (i, j = 1, \ldots, k) \) 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

\[
\sum_{i} \sum_{j} \lambda_{ij} [\theta^i - \theta_0^i] [\theta^j - \theta_0^j] = c,
\]

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

\[
\theta^i - \theta_0^i = \beta_{i1}(\theta^1 - \theta_0^1) + \cdots + \beta_{ik}(\theta^k - \theta_0^k)
\]

such that the family of ellipsoids \( S_c \) is transformed into a family of concentric spheres with the center at \( \theta_0 \). Denote by \( S'_c \) the image of \( S_c \). For any point \( \theta \) and for any positive \( \rho \) consider the set \( \omega(\theta, \rho) \) consisting of all points \( \theta_1 \) which lie on the same \( S_c \) as \( \theta \) and for which \( |\theta_1 - \theta| \leq \rho \). Let

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

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

**Proposition II.** If the variates \( y^1, \ldots, y^k \) have a joint normal distribution with unknown mean values \( \theta^1, \ldots, \theta^k \) 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 \( y^1, \ldots, y^k \), the critical region given by the inequality
has uniformly best average power with respect to the surfaces $S_e$ defined in (79) and the weight function given in (81).

**Proof.** Consider the linear transformation

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

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 $\sigma^2$. 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

$$\sum_{i, j} \lambda_{ij}(y^i - \theta_0)(y^j - \theta_0') \geq d$$

Because of (81) we obviously have

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

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

$$\int_{S_e} P(W' \mid \theta') dA \geq \int_{S_e} P(Z' \mid \theta') dA$$

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

$$\int_{S_e} \rho(y' \mid \theta') dA / \rho(y' \mid \theta_0) \geq d(c)$$

within $W'$

and

$$\int_{S_e} \rho(y' \mid \theta') dA / \rho(y' \mid \theta_0) \leq d(c)$$

outside $W'$ for all positive values of $c$, where $\rho(y' \mid \theta')$ denotes the joint density function of $y^1, \cdots, y^k$ under the assumption that the true means are $\theta_1, \cdots, \theta_k$.

If we denote \( y' - \theta_0 \) by \( v \) and \( \theta' - \theta_0 \) by \( \theta^* \), we have

\[
\rho(y' | \theta') = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}\sum (v^i - \theta^*)^2\right)
\]

and

\[
\rho(y' | \theta_0) = \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}\sum (v^i)^2\right).
\]

Hence

\[
\int_{S'_c} \rho(y' | \theta') \, dA = \left(\frac{1}{2\pi}\right)^{k/2} \int_{S'_c} \exp\left(-\frac{1}{2}\sum (v^i - \theta^*)^2\right) \, dA
\]

\[
\int_{S'_c} \exp\left(\sum v^i \theta^* - \frac{1}{2} \sum (\theta^*)^2\right) \, dA
\]

\[
\rho(y' | \theta_0) \exp\left(-\frac{1}{2}\sum (\theta^*)^2\right) \int_{S'_c} \exp\left(\sum v^i \theta^*\right) \, dA,
\]

since \( \sum (\theta^*)^2 \) is constant on the surface \( S'_c \). Hence (87) and (88) can be written

\[(89) \quad I(v_1, \ldots, v_k) = \int_{S'_c} \exp\left(\sum v^i \theta^* dA \geq d^*(c)\right) \text{ within } W',\]

\[(90) \quad I(v_1, \ldots, v_k) \leq d^*(c) \text{ outside } W'.\]

Denote \( |\sum (v^i)^2|^{1/2} \) by \( r \), and \( |\sum (\theta^*)^2|^{1/2} \) by \( r^* \). On the surface \( S'_c \) we have \( r^* = c \). Denote by \( \alpha(\theta^*) \) the angle \( (0 \leq \alpha \leq \pi) \) between the vector \( v \) and the vector \( \theta^* \). Then we have

\[
I(v_1, \ldots, v_k) = \int_{S'_c} \exp\left(c r \cos \left[\alpha(\theta^*)\right]\right) \, dA.
\]

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 \( (0 \leq \beta(\theta^*) \leq \pi) \) between the vector \( \theta^* \) and an arbitrarily chosen fixed vector \( u \). Hence \( I(v_1, \ldots, v_k) \) depends only on \( r_v \), that is \( I(v_1, \ldots, v_k) = I(r_v) \). 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

\[(91) \quad \frac{dI(r_v)}{dr_v} = \int_{S'_c} c \cos \left[\beta(\theta^*)\right] \exp\left(c r_v \cos \left[\beta(\theta^*)\right]\right) \, dA.
\]

Denote by \( \omega_1 \) the subset of \( S'_c \) in which \( 0 \leq \beta(\theta^*) \leq \pi/2 \), and by \( \omega_2 \) the subset in which \( \pi/2 \leq \beta(\theta^*) \leq \pi \). Because of the symmetry of the sphere we obvi-
ABRAHAM WALD

November

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

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

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

Hence

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

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 \( x_1, \ldots, x_r \) whose joint probability density function \( f(x_1, \ldots, x_r, \theta_1, \ldots, \theta_k) = f(x, \theta) \) involves \( k \) unknown parameters, as considered in the previous sections.

**Definition II.** A sequence \( \{ W_n \} \) \((n=1, \ldots, \infty)\) of critical regions of size \( \alpha \) 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

\[ \limsup \{ \text{l.u.b.} \left[ \int_{K_c} P(Z_n | \theta) \frac{w(\theta)}{A(K_c)} dA - \int_{K_c} P(W_n | \theta) \frac{w(\theta)}{A(K_c)} dA \right] \} \leq 0, \]

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

\[ n \sum_i \sum_j (\theta_0 - \theta_0)(\theta_0 - \theta_0)c_{ij}(\theta_0) \geq d_n, \]

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

\[ \sum_i \sum_j (\theta - \theta_0)(\theta - \theta_0)c_{ij}(\theta_0) = c. \]

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 1 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 \( P(Z_n^* | \theta_o) = \alpha \) and

\[
\text{(94)} \quad \limsup_{n \to \infty} \left\{ \int_{S_n} \Psi(Z_n^* | \theta) \xi_n(\theta) dA - \int_{S_n} \Psi(W_n^* | \theta) \xi_n(\theta) dA \right\} = \delta > 0,
\]

where

\[
\xi_n(\theta) = \xi(\theta) / \int_{S_n} \xi(\theta) dA.
\]

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

\[
\text{(95)} \quad \lim_{n \to \infty} \left\{ \int_{S_{n'}} \Psi(Z_{n'}^* | \theta) \xi_{n'}(\theta) dA - \int_{S_{n'}} \Psi(W_{n'}^* | \theta) \xi_{n'}(\theta) dA \right\} = \delta > 0.
\]

It is easy to verify that

\[
\lim_{n \to \infty} \Psi(W_n^* | \theta_{n'}) = 1
\]

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

\[
\lim_{n \to \infty} \int_{S_{n'}} \Psi(W_n^* | \theta) \xi_{n'}(\theta) dA = 1.
\]

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 \( \{V_n^*\} \)

\[
\lim_{n \to \infty} \left\{ \int_{S_{n'}} \Psi(V_{n'}^* | \theta) \xi_{n'}(\theta) dA - \int_{S_{n'}} \Omega(V_{n'}^* | \theta) \xi_{n'}(\theta) dA \right\} = 0,
\]

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

\[
\text{(96)} \quad \lim_{n \to \infty} \left\{ \int_{S_{n'}} \Omega(Z_{n'}^* | \theta) \xi_{n'}(\theta) dA - \int_{S_{n'}} \Omega(W_{n'}^* | \theta) \xi_{n'}(\theta) dA \right\} = \delta > 0.
\]

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

\[
n \sum \sum (\theta_i^* - \theta_0^*)(\theta_i^* - \theta_0^*)c_{ij}(\theta_0) \geq d_n,
\]

where \( d_n \) is chosen so that \( \Omega(\overline{W}_n^* | \theta_0) = \alpha \). Furthermore denote by \( Z_n^* \) the sum of \( Z_n^* \) and the region \( n^{1/2} |\theta_o - \theta| \leq \lambda_n \), where \( \lambda_n \) is chosen so that
\( P(Z^*_n | \theta_0) = \alpha. \) Since, as can easily be seen,

\[
\lim_{n \to \infty} \left\{ \int_{S_n} P(W^*_n | \theta) \zeta_n(\theta) dA - \int_{S_n} P(W^*_n | \theta_0) \zeta_n(\theta_0) dA \right\} = 0
\]

and

\[
\lim_{n \to \infty} \left\{ \int_{S_n} P(Z^*_n | \theta) \zeta_n(\theta) dA - \int_{S_n} P(Z^*_n | \theta_0) \zeta_n(\theta_0) dA \right\} = 0,
\]

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) \( P(W_n | \theta_1) = P(W_n | \theta_2) \) for any pair of points \( \theta_1, \theta_2 \) which lie on the same surface \( K_c \).

(b) \( P(W_n | \theta) \geq P(Z_n/\theta) \) for any \( Z_n \) which satisfies condition (a) and for which \( P(Z_n | \theta_0) = P(W_n | \theta_0) \).

From Proposition II we obtain the following:

**Proposition III.** Let \( y^1, \ldots, y^k \) be \( k \) variates which have a joint normal distribution with unknown mean values \( \theta^1, \ldots, \theta^k \) and a known covariance matrix \( ||\sigma|| = ||\lambda||^{-1} \). 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

\[
\sum \sum (\theta^i - \theta^*_0)(\theta^j - \theta^*_0)\lambda_{ij} = c.
\]

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 \( \alpha \) and to have asymptotically best constant power on the surfaces \( K_c \) if the following three conditions are fulfilled:

(a) \( P(W_n | \theta_0) = \alpha \) (\( n = 1, 2, \ldots \), ad inf.).

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

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

\[
\lim_{n \to \infty} \left\{ \text{l.u.b.} \left[ P(Z_n | \theta) - P(W_n | \theta) \right] \right\} = 0.
\]

It is easy to verify that the sequence \( \{ W_n^* \} \) defined in Theorem I satis-
fies the conditions (a) and (b) for $K_e = S_e$, where $S_e$ denotes the surface defined in Theorem I. Thus from Theorem I we obtain the following theorem.

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

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

**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 $\alpha$ if $P(W_n | \theta_0) = \alpha$ and if

$$\sup_{\theta} \left[ P_n(\theta, \theta_0, \alpha) - P(W_n | \theta) \right] \leq \sup_{\theta} \left[ P_n(\theta, \theta_0, \alpha) - P(Z_n | \theta) \right]$$

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

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 $P(Z | \theta_0) = \alpha$ and

$$\sup_{\theta} \left[ P(\theta, \theta_0, \alpha) - P(W | \theta) \right] > \sup_{\theta} \left[ P(\theta, \theta_0, \alpha) - P(Z | \theta) \right].$$

Let $S_e$ be the surface defined by the equation

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

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

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

for all points $\theta$ on $S_e$. 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 $\alpha$.
if \( P(W_n | \theta_0) = \alpha \) and if for any \( \{ Z_n \} \) for which \( P(Z_n | \theta_0) = \alpha \) we have
\[
\lim \sup_{n \to \infty} \{ \text{l.u.b.} [P_n(\theta, \theta_0, \alpha) - P(W_n | \theta)] - \text{l.u.b.} [P_n(\theta, \theta_0, \alpha) - P(Z_n | \theta)] \} \leq 0.
\]

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 \( \Psi_n(\theta, \theta_0, \alpha) \) the least upper bound of \( \Psi(Z_n^* | \theta) \) with respect to \( Z_n^* \), where \( Z_n^* \) is restricted to regions in the space of the maximum likelihood estimates for which \( \Psi(Z_n^* | \theta_0) = \alpha \). Because of Lemma 1 we have
\[
\lim_{n \to \infty} \left[ P_n(\theta, \theta_0, \alpha) - \Psi_n(\theta, \theta_0, \alpha) \right] = 0
\]
uniformly in \( \theta \). 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 \( n^{1/2}(\theta_n^* - \theta_0) \), \ldots, \( n^{1/2}(\theta_n^* - \theta_k) \) is normal with zero means and covariance matrix \( \| c_{ij}(\theta_0) \|^{-1} \). For any positive \( \lambda \) we have
\[
\lim_{n \to \infty} \left[ \Psi_n(\theta, \theta_0, \alpha) - \overline{P}_n(\theta, \theta_0, \alpha) \right] = 0
\]
uniformly in \( \theta \) in the domain \( |n^{1/2}(\theta - \theta_0)| \leq \lambda \). Since for any sequence \( \{ \theta_n \} \) for which \( \lim |n^{1/2}(\theta_n - \theta_0)| = +\infty \), we have
\[
\lim_{n \to \infty} \Psi_n(\theta_n, \theta_0, \alpha) = \lim_{n \to \infty} \overline{P}_n(\theta_n, \theta_0, \alpha) = 1,
\]
we obtain from (100)
\[
\lim_{n \to \infty} \left[ \Psi_n(\theta, \theta_0, \alpha) - \overline{P}_n(\theta, \theta_0, \alpha) \right] = 0
\]
uniformly in \( \theta \). For any \( c \) let \( S_c \) be the surface defined by
\[
\sum_{i=1}^{k} \sum_{j=1}^{k} (\theta_i - \theta_0)(\theta_j - \theta_0)c_{ij}(\theta_0) = c.
\]
Obviously \( \overline{P}_n(\theta, \theta_0, \alpha) \) is constant along the surface \( S_c \). From (99) and (101) we obtain
\[
\lim_{n \to \infty} \left\{ \text{l.u.b.} P_n(\theta, \theta_0, \alpha) - \text{g.l.b.} P_n(\theta, \theta_0, \alpha) \right\} = 0
\]
uniformly in \( \theta \). 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


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

(103)

On account of (102) and since

\[ \lim_{\theta \in S_\varepsilon} \{ \text{l.u.b. } P(W_n^* | \theta) - \text{g.l.b. } P(W_n^* | \theta) \} = 0 \]

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 \( \omega \) be a subset of the parameter space and denote by \( H_\omega \) the hypothesis that the true parameter point is contained in \( \omega \). In all that follows the letter \( \theta \) printed in boldface will indicate that the parameter point lies in \( \omega \). For example, the symbol \( \text{l.u.b. } f(\theta) \) denotes the least upper bound of the function \( f(\theta) \) with respect to \( \theta \) where \( \theta \) is restricted to points of \( \omega \). For any point \( \theta \) and for any real value \( c \) let \( K_c(\theta) \) denote a surface in the parameter space. For instance \( K_c(\theta) \) may be given by \( r \) equations in \( \theta \)

\[ \phi_1(\theta, \theta) = \cdots = \phi_r(\theta, \theta) = 0, \]

where \( \phi_1(\theta, \theta), \cdots, \phi_r(\theta, \theta) \) are some analytic functions of \( \theta \) and \( \theta \).

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

\[ \text{l.u.b. } P(Z_n | \theta) = \text{l.u.b. } P(W_n | \theta) \]

we have

\[ \int_{K_c(\theta)} P(W_n | \theta) w(\theta) dA \geq \int_{K_c(\theta)} P(Z_n | \theta) w(\theta) dA \]

for any \( \theta \) and for any \( c \) for which \( K_c(\theta) \) is defined.

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

$$\sup_\theta P(W_n | \theta) = \alpha \quad (n = 1, 2, \ldots, \text{ad inf.})$$

(b) For any sequence $\{Z_n\}$ for which $\sup_\theta P(Z_n | \theta) = \alpha$, we have

$$\limsup_{n \to \infty} \left\{ \sup_\theta \left[ \int_{K_c(\theta)} P(Z_n | \theta) \frac{w(\theta)}{A[K_c(\theta)]} dA \right. \right.$$

$$\left. \left. - \int_{K_c(\theta)} P(W_n | \theta) \frac{w(\theta)}{A[K_c(\theta)]} dA \right]\right\} \leq 0,$$

where

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

**Definition IX.** A critical region $W_n$ for testing $H_0$ is said to have uniformly best constant power on the family of surfaces $K_c(\omega)$ if the following two conditions are fulfilled:

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

(b) $P(W_n | \theta) \geq P(Z_n | \theta)$ for any $\theta$ not in $\omega$ and for any $Z_n$ which satisfies (a) and the condition

$$\sup_\theta P(Z_n | \theta) = \sup_\theta P(W_n | \theta).$$

**Definition X.** A sequence of critical regions $\{W_n\}$ for testing $H_0$ is said to have asymptotically best constant power on the surfaces $K_c(\omega)$ if the following three conditions are fulfilled:

(a) $\sup_\theta P(W_n | \theta) = \alpha (n = 1, 2, \ldots, \text{ad inf.})$.

(b) $\lim_{\omega \to \infty} \left\{ \sup_\theta \left[ \sup_{\omega \in K_c(\theta)} P(W_n | \theta) - \inf_{\theta \in K_c(\theta)} P(W_n | \theta) \right] \right\} = 0$.

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

$$\lim_{\omega \to \infty} \left\{ \sup_{\theta \in \omega} \left[ P(Z_n | \theta) - P(W_n | \theta) \right] \right\} = 0,$$

where $\omega$ is the complement of $\omega$.

**Definition XI.** Denote by $P_n(\theta, \omega, \alpha)$ the least upper bound of $P(Z_n | \theta)$ with respect to $Z_n$ subject to the condition $\sup_\theta P(Z_n | \theta) = \alpha$. A critical region $W_n$ is said to be a most stringent test of the hypothesis $H_0$ if for some positive $\alpha$

$$\sup_\theta P(W_n | \theta) = \alpha$$

and

$$\sup_\theta [P_n(\theta, \omega, \alpha) - P(W_n | \theta)] \leq \sup_\theta [P_n(\theta, \omega, \alpha) - P(Z_n | \theta)]$$

for all regions $Z_n$ for which $\sup_\theta P(Z_n | \theta) = \alpha$. 

---

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Definition XII. A sequence of critical regions \( \{ W_n \} \) is said to be an asymptotically most stringent test of the hypothesis \( H_a \) if the following two conditions are fulfilled:

(a) There exists a positive \( \alpha \) such that

\[
\operatorname{l.u.b.} P(W_n \mid \theta) = \alpha \quad (n = 1, 2, \ldots, \text{ad inf}).
\]

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

\[
\limsup_{n \to \infty} \left\{ \operatorname{l.u.b.} \left[ P_n(\theta, \omega, \alpha) - P(W_n \mid \theta) \right] - \operatorname{l.u.b.} \left[ P_n(\theta, \omega, \alpha) - P(Z_n \mid \theta) \right] \right\} \leq 0.
\]

In Definitions VII–XII we have formulated the condition

\[
\operatorname{l.u.b.} P(W_n \mid \theta) = \alpha.
\]

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

\[(*) \quad P(W_n \mid \theta) = \alpha\]

for all points \( \theta \) should be made; or whether the weaker condition that

\[(**) \quad \lim_{n \to \infty} P(W_n \mid \theta) = \alpha\]

uniformly in \( \theta \) 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 \( \operatorname{l.u.b.} P(W_n \mid \theta) = \alpha \) by (**) in the definitions VIII, X and XII.

In the following §§9–11 we shall discuss a linear hypothesis of the following type: \( \theta^1 = \theta_0, \ldots, \theta^r = \theta_0^* \) \( (r < k) \), where \( \theta_0, \ldots, \theta_0^* \) are some specified values. That is to say, the set \( \omega \) is the set of all points \( \theta \) 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 \( H_a \) be the hypothesis that \( \theta^1 = \theta_0, \ldots, \theta^r = \theta_0^* \) \( (r < k) \). We shall introduce the following notation: For any parameter point \( \theta = (\theta^1, \ldots, \theta^k) \) the symbol \( \theta^0 \) will denote the vector in the \( r \)-dimensional space with the components \( \theta^1, \ldots, \theta^r \), and \( z \theta \) will denote the vector in the \( k-r \) dimensional space with the components \( \theta^{r+1}, \ldots, \theta^k \). For any function \( \psi(\theta) \) of \( \theta \) we shall use the synonymous nota-
tion $\psi(\theta, s)$). For instance $P(W|\theta, s\theta)$ is synonymous with $P(W|\theta)$.

Let $y_1, \ldots, y_k$ be $k$ variates which have a joint normal distribution with unknown mean values $\theta_1, \ldots, \theta_k$ and known covariance matrix $||\sigma_{ij}|| = ||\lambda_{ij}||^{-1}$ ($i, j = 1, \ldots, k$), which is nonsingular. Denote by $W$ the region in the space of $y_1, \ldots, y_k$ given by the inequality

$$\sum_{p=1}^{r} \sum_{q=1}^{r} \lambda_{pq}(y_{p} - \theta_{0}^p)(y_{q} - \theta_{0}^q) \geq d,$$

where $||\lambda_{pq}|| = ||\sigma_{pq}||^{-1}$ ($p, q = 1, \ldots, r$). Consider the nonsingular linear transformation of the variates $y_1, \ldots, y_k$ given by the equations

$$y_{p}' - \theta_{0}^p = \beta_{p1}(y_1 - \theta_0^1) + \cdots + \beta_{pr}(y_r - \theta_0^r) \quad (p = 1, \ldots, r),$$

$$y_{t}' = \gamma_{t1}y_1 + \cdots + \gamma_{tk}y_k \quad (t = r + 1, \ldots, k),$$

such that $y_1', \ldots, y_k'$ are independently distributed with unit variances. Denote by $S_c(\theta)$ the surface given by the equations

$$\sum_{p=1}^{r} \sum_{q=1}^{r} \lambda_{pq}(\theta_{p}' - \theta_{0}^p)(\theta_{q}' - \theta_{0}^q) = c,$$

$$\gamma_{t1}\theta_1^1 + \cdots + \gamma_{tk}\theta_k^k = \sum_{i=1}^{k} \gamma_{ti}\theta_i^i \quad (t = r + 1, \ldots, k).$$

Consider the transformation of the parameter space given by

$$\theta_{p}' - \theta_{0}^p = \beta_{p1}(\theta_{1}' - \theta_0^1) + \cdots + \beta_{pr}(\theta_{r}' - \theta_0^r) \quad (p = 1, \ldots, r),$$

$$\theta_{t}' = \gamma_{t1}\theta_1^1 + \cdots + \gamma_{tk}\theta_k^k \quad (t = r + 1, \ldots, k),$$

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

$$\sum_{p=1}^{r} (\theta_{p}' - \theta_{0}^p)^2 = c, \quad \theta_{t}' = \sum_{i=1}^{k} \gamma_{ti}\theta_i^i = \theta_i'.$$

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

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

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

**Proposition V.** Let $y_1, \ldots, y_k$ be $k$ variates which have a joint normal distribution with unknown mean values $\theta_1, \ldots, \theta_k$ and known covariance matrix $||\sigma_{ij}|| = ||\lambda_{ij}||^{-1}$. For testing the hypothesis $\theta = \theta_0$ on the basis of a single
observation on each of the variates $y_1, \ldots, y_k$, 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

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

where $Z$ denotes an arbitrary region in the space of $y_1, \ldots, y_k$ and $Z'$ is the image of $Z$ by transformation (105). The region $W$ is transformed into $W'$ given by

$$(\theta - \theta_0)^2 + \cdots + (\theta^r - \theta_0)^2 \geq d.$$

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

$$\int_{S_c(\theta)} P(W' | \theta) dA \geq \int_{S_c(\theta)} P(Z' | \theta') dA$$

for any $c > 0$, for any $\theta$, and for any region $Z'$ in the space of $y_1', \ldots, y_k'$ for which l.u.b. $P(Z' | \theta_0, \theta') = \text{l.u.b.} P(W' | \theta_0, \theta')$. For any point $\theta'$ of $S_c(\theta)$ we have $\theta' = \theta_0$. By a lemma of Neyman and Pearson, (112) is proved if we can show that there exists a function $d(c)$ such that

$$\int_{S_c(\theta)} p(y_1', \ldots, y_k' | \theta', \theta_0) dA \geq d(c) \text{ within } W',$$

$$p(y_1', \ldots, y_k' | \theta_0, \theta') \leq d(c) \text{ outside } W'$$

for all values of $c$ and $\theta$ where $p(y_1', \ldots, y_k' | \theta')$ denotes the joint probability density of $y_1', \ldots, y_k'$ under the assumption that $\theta'$ is the true parameter point. Obviously

$$\frac{p(y_1', \ldots, y_k' | \theta', \theta_0)}{p(y_1', \ldots, y_k' | \theta_0, \theta')} = \frac{p(y_1', \ldots, y_k' | \theta_0)}{p(y_1', \ldots, y_k' | \theta_0)}.$$

Hence (113) is equivalent to

$$\int_{S_c(\theta)} p(y_1', \ldots, y_k' | \theta') dA \geq d(c) \text{ within } W',$$

$$p(y_1', \ldots, y_k' | \theta_0) \leq d(c) \text{ outside } W'.$$

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 $W^*$ be defined by the following inequality

$$n \sum_{q=1}^r \sum_{p=1}^r (\theta^p_n - \theta_0^p)(\theta^q_n - \theta_0^q) \varepsilon_{pq}(\theta_n) \geq d_n,$$

where $\|\varepsilon_{pq}(\theta)\| = \|\sigma_{pq}(\theta)\|^{-1}$ ($p, q = 1, \ldots, r$) and $\|\sigma_{ij}(\theta)\| = \|\epsilon_{ij}(\theta)\|^{-1}$ ($i, j$
ABRAHAM WALD

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

\[
Z_n^p(\theta) = \beta_{p1}(\theta)Z_n^1(\theta) + \cdots + \beta_{pr}(\theta)Z_n^r(\theta) \quad (p = 1, \cdots, r),
\]

\[
Z_n^t(\theta) = \gamma_{t1}(\theta)Z_n^1(\theta) + \cdots + \gamma_{tk}(\theta)Z_n^k(\theta) \quad (t = r + 1, \cdots, k),
\]

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

\[
\sum_{p=1}^{r} \sum_{t=1}^{r} (\theta^p - \theta^p_0)(\theta^t - \theta^t_0)e_{pt}(\theta) = c,
\]

\[
\gamma_{t1}(\theta)e^t_1 + \cdots + \gamma_{tk}(\theta)e^t_k = \sum \gamma_{ti}(\theta)e^i_t \quad (t = r + 1, \cdots, k).
\]

For any positive $\delta$ denote by $S_\delta$ the set of all points $\theta$ for which $|\theta^p - \theta^p_0| \leq \delta$ ($p = 1, \cdots, r$). We shall prove the existence of a positive $\delta$ such that for any point $\theta$ in $S_\delta$ there exists exactly one surface $S_\alpha(\theta)$, that is exactly one value of $\delta$ and exactly one point $\theta$, such that $\theta$ lies on the surface $S_\alpha(\theta)$. This statement is obviously proved if we show that for any point $\theta$ in $S_\delta$ the set of $k - r$ equations

\[
\gamma_{t1}(\theta)e^t_1 + \cdots + \gamma_{tk}(\theta)e^t_k = 0 \quad (t = r + 1, \cdots, k)
\]

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

\[
A(\theta)||\sigma_{ij}(\theta)||\bar{A}(\theta) = I,
\]

where $A(\theta)$ denotes the matrix

\[
\begin{bmatrix}
\beta_{11}(\theta) & \beta_{12}(\theta) & \cdots & \beta_{1r}(\theta) & 0 & \cdots & 0 \\
\beta_{21}(\theta) & \beta_{22}(\theta) & \cdots & \beta_{2r}(\theta) & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\
\beta_{r1}(\theta) & \beta_{r2}(\theta) & \cdots & \beta_{rr}(\theta) & 0 & \cdots & 0 \\
\gamma_{r+1\ 1}(\theta) & \gamma_{r+1\ 2}(\theta) & \cdots & \gamma_{r+1\ r}(\theta) & \gamma_{r+1\ r+1}(\theta) & \cdots & \gamma_{r+1\ k}(\theta) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\
\gamma_{k1}(\theta) & \gamma_{k2}(\theta) & \cdots & \gamma_{kr}(\theta) & \gamma_{k\ r+1}(\theta) & \cdots & \gamma_{kk}(\theta)
\end{bmatrix}
\]

$\bar{A}(\theta)$ is the transposed of $A(\theta)$ and $I$ denotes the unit matrix. Since $\sigma_{ij}(\theta)$ is a continuous and bounded function of $\theta$ 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 $\theta$ and that the absolute value of the determinant $A(\theta)$ has a positive lower bound. Hence also the absolute value of the determinant

$$
(121) \quad \gamma(\theta) = \begin{vmatrix}
\gamma_{r+1, r+1}(\theta) & \cdots & \gamma_{r+1, k}(\theta) \\
\vdots & \ddots & \vdots \\
\gamma_{k, r+1}(\theta) & \cdots & \gamma_{k, k}(\theta)
\end{vmatrix}
$$

has a positive lower bound.

Let $\theta = \theta^*$ where $\theta^*$ denotes an arbitrary point of $\omega$. 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 $\delta$ such that the equations (118) have a unique solution in $\theta$ if $|\theta - \theta^*| \leq \delta$. This proves the existence of a positive $\delta$ such that for any point $\theta$ in $S_\delta$ there exists exactly one surface $S_c(\theta)$ such that $\theta$ lies on $S_c(\theta)$.

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

Consider the transformation of the parameter space given by

$$
(122) \quad \theta^p = \theta_0^p + \beta_{p1}(\theta)(\theta^1 - \theta_0^1) + \cdots + \beta_{pr}(\theta)(\theta^r - \theta_0^r) \quad (p = 1, \cdots, r),
\quad \theta^t = \gamma_{t1}(\theta)\theta^1 + \cdots + \gamma_{tk}(\theta)\theta^k \quad (t = r + 1, \cdots, k),
$$

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

$$
(123) \quad \sum_{p=1}^r (\theta^p - \theta_0^p)^2 = c; \quad \theta^t = \sum_{t=1}^k \gamma_{ti}(\theta)\theta^i = \theta^t \quad (t = r + 1, \cdots, k).
$$

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

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

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 $\theta = \theta_0$ 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 \( \text{l.u.b.} \ P(Z_n^*|\theta) = \alpha \) we have

\[
\limsup_{n \to \infty} \left\{ \text{l.u.b.} \left[ \int_{S_n(\theta)} \mathbb{P}(Z_n^*|\theta)\xi_n(\theta)dA - \int_{S_n(\theta)} \mathbb{P}(W_n^*|\theta)\xi_n(\theta)dA \right] \right\} \leq 0,
\]

where

\[
\xi_n(\theta) = \frac{\xi(\theta)}{\int_{S_n(\theta)} \xi(\theta)dA}.
\]

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

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

(125)

\[
\lim_{n \to \infty} \left\{ \int_{S_{n'}(\theta_{n'})} \mathbb{P}(Z_n^*|\theta)\xi_{n'}(\theta)dA - \int_{S_{n'}(\theta_{n'})} \mathbb{P}(W_n^*|\theta)\xi_{n'}(\theta)dA \right\} = \delta > 0.
\]

It is easy to verify that for any sequence \( \{c_n\} \) for which \( \lim n c_n = +\infty \) we have \( \int_{S_n(\theta)} \mathbb{P}(W_n^*|\theta)\xi_n(\theta)dA = 1 \) uniformly in \( \theta \). 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

(126)

\[
\lim_{n \to \infty} \left\{ \int_{S_{n'}(\theta_{n'})} \mathbb{P}(V_n^*|\theta)\xi_{n'}(\theta)dA - \int_{S_{n'}(\theta_{n'})} P_{\theta_{n'}}(V_n^*|\theta)\xi_{n'}(\theta)dA \right\} = 0,
\]

where \( P_{\theta}(V_n^*|\theta) \) denotes the probability of \( V_n^* \) calculated under the assumption that \( n^{1/2}(\theta_n^* - \theta^0), \ldots, n^{1/2}(\theta_n^{r} - \theta^0) \) have a joint normal distribution with zero means and covariance matrix \( \sigma_{ij}(\theta) \). Let \( W_n^*(\theta) \) be the region defined by

\[
n \sum_{p=1}^{r} \sum_{q=1}^{r} (\theta_n^p - \theta^0)^*(\theta_n^q - \theta^0)\xi_{pq}(\theta) \geq d_n.
\]

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

(127)

\[
\lim_{n \to \infty} \left\{ P_{\theta}(W_n^*|\theta) - P_{\theta}[W_n^*(\theta)|\theta] \right\} = 0
\]

uniformly in \( \theta \) and \( \theta \) over the domain in which \( \theta \) is a point of \( S_{c^*}(\theta) \). From (125), (126) and (127) we obtain

(128)

\[
\lim_{n \to \infty} \left\{ \int_{S_{n'}(\theta_{n'})} P_{\theta_{n'}}(Z_n^*|\theta)\xi_{n'}(\theta)dA - \int_{S_{n'}(\theta_{n'})} P_{\theta_{n'}}[W_n^*(\theta_{n'})|\theta]\xi_{n'}(\theta)dA \right\} = \delta > 0.
\]

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 $\lambda_{ii}$, that is we substitute $\xi(\theta)$ for $\lambda_{pq}$. Similarly, if we substitute $c_{ii}(\theta)$ for $\lambda_{ii}$ 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

$$\lim_{n \to \infty} \left\{ \begin{array}{c} \text{l.u.b. } P_0(Z_n^*, \theta) \\ \theta \end{array} \right\} = \alpha,$$

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 $\theta = \theta_0$, $\{W_n^*\}$ 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 $y_1, \ldots, y_k$ be $k$ variates which have a joint normal distribution with unknown mean values $\theta_1, \ldots, \theta_k$ and known covariance matrix $\|\sigma_{ij}\| = \|\lambda_{ij}\|^{-1}$. For testing the hypothesis $\theta = \theta_0$ on the basis of a single observation on each of the variates $y_1, \ldots, y_k$, 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 $y_1, \ldots, y_k$ as defined in (105). Then the transformed variates $y'^1, \ldots, y'^k$ are independently distributed with unit variances. Denote by $\theta'$ the image of $\theta$ obtained by the transformation (107) and let $P'(\theta', \omega, \alpha)$ be equal to l.u.b. $P(Z'|\theta')$ with respect to $Z'$, where $Z'$ may be any region in the space of $y'^1, \ldots, y'^k$ subject to the condition that

$$\text{l.u.b. } P(Z'|y_{\theta_0}, \theta') = \alpha.$$

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

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
Hence we have merely to show that $P'(\theta', \omega, \alpha)$ is constant along $S'_c (\theta)$ where $S'_c (\theta)$ is the image of $S_c (\theta)$ and is given by

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

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

$$P(V'| \theta_0, \theta'_1) = P(V'| \theta_0, \theta'_2)$$

for any pair of points $\theta'_1$ and $\theta'_2$, and therefore

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

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

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

**Proof.** Denote by $B(\theta, \omega, \alpha)$ the least upper bound of $B(Z^*_r | \theta)$ with respect to $Z^*_r$, where $Z^*_r$ is restricted to regions in the space of the maximum likelihood estimates for which

$$\text{l.u.b. } B(Z^*_r | \theta) = \alpha.$$  

On account of Lemmas 1 and 2 we have

$$\lim_{n \to \infty} \{ P_n(\theta, \omega, \alpha) - B_n(\theta, \omega, \alpha) \} = 0$$

uniformly in $\theta$.

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

$$\text{l.u.b. } \overline{B}(Z^*_r | \theta_0, \theta) = \alpha.$$  

The symbol $\overline{B}(V^*_r | \theta)$ denotes the probability of $V^*_r$ calculated under the assumption that the joint distribution of $n^{1/2}(\theta_n^1 - \theta_1^1), \ldots, n^{1/2}(\theta_n^r - \theta^r)$ is normal with zero means and covariance matrix $\|c_{ij}(\theta)\|^{-1}$, where $\theta$ denotes that point for which $\theta$ lies on the surface $S_c (\theta)$ defined in (117). It can be shown that for any positive $\lambda$ we have

$$\lim_{n \to \infty} \{ B_n(\theta, \omega, \alpha) - \overline{B}_n(\theta, \theta_0, \alpha) \} = 0.$$
uniformly in $\theta$ in the domain $|\theta - \theta_0| \leq \lambda / n^{1/2}$. Since for any sequence $\{\theta_n\}$ for which

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

we have

$$\lim_{n \to \infty} \Psi(\theta_n, \omega, \alpha) = \lim_{n \to \infty} \bar{P}_n(\theta_n, \theta_0, \alpha) = 1,$$

we obtain from (130)

$$(131) \quad \lim_{n \to \infty} \{ \Psi_n(\theta, \omega, \alpha) - \bar{P}_n(\theta, \theta_0, \alpha) \} = 0$$

uniformly in $\theta$. The function $\bar{P}_n(\theta, \theta_0, \alpha)$ is constant along the surface $S_\theta(0)$ defined in (117). This can be proved in the same way as the constancy of $P(\theta, \omega, \alpha)$ on $S_\theta(0)$ defined in (106). Hence from (129) and (131) we obtain

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

uniformly in $\theta$ and $\omega$. According to Theorem V we have

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

uniformly in $\omega$ and $\theta$. Theorem VI follows from (132), (133) and Theorem IV.

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

$$(134) \quad \xi^1(\theta) = \xi^2(\theta) = \cdots = \xi^r(\theta) = 0 \quad (r < k),$$

that is, $\omega$ is the set of all points $\theta$ 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 $\theta$ into the point $\xi$ with the coordinates $\xi^1(\theta), \cdots, \xi^k(\theta)$ is a topological transformation of $\Omega$ 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 $\theta$.

(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 $\Omega$. 
Since according to Assumption VI the transformation

\[ \xi^i = \xi^i(\theta) \quad (i = 1, \ldots, k) \]

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

\[ \theta^i = \theta^i(\xi^1, \ldots, \xi^k) \quad (i = 1, \ldots, k). \]

From conditions (b) and (c) of Assumption VI it follows that the first and second order derivatives of \( \theta^i(\xi^1, \ldots, \xi^k) \) are uniformly continuous and bounded functions of \( \xi \) and the absolute value of the Jacobian \( \partial(\theta^1, \ldots, \theta^k)/\partial(\xi^1, \ldots, \xi^k) \) 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 \( \theta^i \). 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 \( \xi^i \) is equal to \( \xi_n^i = \xi^i(\theta_n, \ldots, \theta_n^k) \), where \( \theta_n \) is the maximum likelihood estimate of \( \theta \).

Denote by I*, II*, \ldots, V* the assumptions which we obtain from Assumptions I–V respectively by substituting \( f(x, \xi) \) for \( f(x, \theta) \), \( \xi \) for \( \theta \) and \( \xi_n \) for \( \theta_n \). 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 \( \theta \), the transformation (135) is uniformly continuous. Hence, for each positive \( \varepsilon^* \) there exists a positive \( \varepsilon \) such that the inequality \( \theta_n - \theta \leq \varepsilon \) implies the inequality \( \xi_n - \xi \leq \varepsilon^* \). From this and Assumption II we obtain Assumption II*.

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

\[ \frac{\partial \log f(x, \theta^*)}{\partial \theta^i} = \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} + \sum_j (\theta^*_j - \theta^*_1) \frac{\partial^2 \log f(x, \theta_1)}{\partial \theta^i \partial \theta^j}, \]

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

\[ \left| \frac{\partial \log f(x, \theta^*_1)}{\partial \theta^i} - \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leq 2\delta \sum_i \left[ \left| \phi_{ij}(x, \theta, \delta) \right| + \left| \psi_{ij}(x, \theta, \delta) \right| \right] \]

for any \( \theta \) and \( \delta \) for which

\[ \theta - \delta \leq \theta_1 \leq \theta + \delta \quad \text{and} \quad \theta - \delta \leq \theta_1^* \leq \theta + \delta. \]

From (138) we obtain for any positive \( \delta \)

\[ \left| \nu_i(x, \theta, \delta) - \mu_i(x, \theta, \delta) \right| \leq 2\delta \sum_i \left[ \left| \phi_{ij}(x, \theta, \delta) \right| + \left| \psi_{ij}(x, \theta, \delta) \right| \right]. \]
Let \( \{ \theta_{1n} \} \) and \( \{ \theta_{2n} \} \) \((n = 1, 2, \ldots, \text{ad inf.})\) be two sequences of parameter points such that

\[
\lim_{n \to \infty} (\theta_{1n} - \theta_{2n}) = 0.
\]

According to Assumption III the expectations \( E_{\theta_1} [\phi_i(x, \theta_1, \delta)]^2 \) and \( E_{\theta_1} [\psi_i(x, \theta_1, \delta)]^2 \) are bounded functions of \( \theta_1, \theta_2 \) and \( \delta \) in the domain \( D_\varepsilon \) defined in Assumption III. Hence also the expectations \( E_{\theta_1} [\phi_i(x, \theta_2, \delta)] \) and \( E_{\theta_1} [\psi_i(x, \theta_2, \delta)] \) are bounded functions of \( \theta_1, \theta_2 \) and \( \delta \) in the domain \( D_\varepsilon \).

From this and relations (139) and (140) it follows that for any sequence \( \{ \delta_n \} \) of positive numbers for which \( \lim_{n \to \infty} \delta_n = 0 \) we have

\[
\lim_{n \to \infty} \left\{ E_{\theta_1} \nu_i(x, \theta_{2n}, \delta_n) - E_{\theta_1} \mu_i(x, \theta_{2n}, \delta_n) \right\} = 0.
\]

Since

\[
\mu_i(x, \theta_{2n}, \delta_n) \leq \frac{\partial \log f(x, \theta_{2n})}{\partial \theta_i} \leq \nu_i(x, \theta_{2n}, \delta_n)
\]

it follows from (141) that

\[
\lim_{n \to \infty} \left\{ E_{\theta_1} \nu_i(x, \theta_{2n}, \delta_n) - E_{\theta_1} \mu_i(x, \theta_{2n}, \delta_n) \right\} = 0,
\]

(142)

\[
\lim_{n \to \infty} \left\{ E_{\theta_1} \mu_i(x, \theta_{2n}, \delta_n) - E_{\theta_1} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta_i} \right\} = 0.
\]

Using the Taylor expansion we have

\[
\frac{\partial \log f(x, \theta_{2n})}{\partial \theta_i} = \frac{\partial \log f(x, \theta_{1n})}{\partial \theta_i} + \sum_i (\theta_{2n}^i - \theta_{1n}^i) \frac{\partial^2 \log f(x, \bar{\theta}_n)}{\partial \theta_i^2},
\]

where \( \bar{\theta}_n \) lies in the interval \( [\theta_{1n}, \theta_{2n}] \). Since the expectations \( E_{\theta_1} [\phi_i(x, \theta_2, \delta)] \) and \( E_{\theta_1} [\psi_i(x, \theta_2, \delta)] \) are bounded functions of \( \theta_1, \theta_2 \), and \( \delta \) in the domain \( D_\varepsilon \), we obtain

\[
\lim_{n \to \infty} E_{\theta_1} \left\{ \sum_i (\theta_{2n}^i - \theta_{1n}^i) \frac{\partial^2 \log f(x, \bar{\theta}_n)}{\partial \theta_i^2} \right\} = 0.
\]

Hence it follows from (143) and Assumption IV that

\[
\lim_{n \to \infty} E_{\theta_1} \frac{\partial \log f(x, \theta_{2n})}{\partial \theta_i} = \lim_{n \to \infty} E_{\theta_1} \frac{\partial \log f(x, \theta_{1n})}{\partial \theta_i} = 0.
\]

We obtain from (142) and (144)

\[
\lim_{n \to \infty} E_{\theta_1} \nu_i(x, \theta_{2n}, \delta_n) = \lim_{n \to \infty} E_{\theta_1} \mu_i(x, \theta_{2n}, \delta_n) = 0.
\]

Denote by \( \phi_\ast(x, \xi_1, \delta^\ast) \) the least upper bound, and by \( \psi_\ast(x, \xi, \delta^\ast) \) the great-
est lower bound of $\frac{\partial^2 \log f^*(x, \xi)}{\partial \xi^i \partial \xi^j}$ in the interval $\xi_1 - \delta^* \leq \xi \leq \xi_1 + \delta^*$. We have

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

(146)

Since

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

we obtain from (146)

$$E_v \frac{\partial^2 \log f^*(x, \xi)}{\partial \xi^i \partial \xi^j} = \sum_i \sum_m E_v \left[ \frac{\partial^2 \log f(x, \theta)}{\partial \theta^i \partial \theta^m} \frac{\partial \theta^i}{\partial \xi^i} \frac{\partial \theta^m}{\partial \xi^j} \right].$$

(147)

Hence the determinant

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

(148)

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

$$\left| - E_v \frac{\partial^2 \log f^*(x, \xi)}{\partial \xi^i \partial \xi^j} \right|$$

has a positive lower bound.

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

$$\phi_{ij}(x, \xi, \delta^*) \leq \sum_i \sum_m \phi_{im}(x, \theta, \delta(\delta^*)) \frac{\partial \theta^i}{\partial \xi^i} \frac{\partial \theta^m}{\partial \xi^j} + \sum_i \psi_i(x, \theta, \delta(\delta^*)) \frac{\partial \theta^i}{\partial \xi^i},$$

(150)

$$\psi_{ij}(x, \xi, \delta^*) \geq \sum_i \sum_m \psi_{im}(x, \theta, \delta(\delta^*)) \frac{\partial \theta^i}{\partial \xi^i} \frac{\partial \theta^m}{\partial \xi^j} + \sum_i \mu_i(x, \theta, \delta(\delta^*)) \frac{\partial \theta^i}{\partial \xi^i},$$

where $\theta$ is the image point of $\xi$ by transformation (136), and the derivatives $\frac{\partial \theta^i}{\partial \xi^i}, \frac{\partial \theta^j}{\partial \xi^j}$ are taken at some points in the interval $[\xi - \delta^*, \xi + \delta^*]$, and the functions $\phi_{im}(x, \theta, \delta), \psi_{im}(x, \theta, \delta), \mu_i(x, \theta, \delta)$ and $\psi_i(x, \theta, \delta)$ satisfy the inequalities.
Let \( \{ \xi_1^n \} \) and \( \{ \xi_2^n \} \) be two sequences of parameter points such that \( \lim \xi_1^n = \lim \xi_2^n = \xi \). Let \( \theta_1^n \) be the image point of \( \xi_1^n \), \( \theta_2^n \) the image point of \( \xi_2^n \), and \( \theta \) the image point of \( \xi \) (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

\[
\lim_{n \to \infty} E_{\theta_1^n} \phi_{\text{seq}} [x, \theta_2^n, \delta(\delta_n^*)] = E_{\theta} \partial^2 \log f(x, \theta) / \partial \theta^2 \partial \theta^m,
\]

\[
\lim_{n \to \infty} E_{\theta_2^n} \psi_{\text{seq}} [x, \theta_2^n, \delta(\delta_n^*)] = E_{\theta} \partial^2 \log f(x, \theta) / \partial \theta^2 \partial \theta^m,
\]

\[
\lim_{n \to \infty} E_{\theta_1^n} \mu_{\text{seq}} [x, \theta_2^n, \delta(\delta_n^*)] = 0,
\]

\[
\lim_{n \to \infty} E_{\theta_2^n} \mu_{\text{seq}} [x, \theta_2^n, \delta(\delta_n^*)] = 0
\]

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

\[
\lim_{n \to \infty} E_{\theta_1^n} \phi_{\text{seq}} (x, \xi_2^n, \delta_n^*) = \lim_{n \to \infty} E_{\theta_2^n} \psi_{\text{seq}} (x, \xi_2^n, \delta_n^*) = E_{\theta} \partial^2 \log f^*(x, \xi) / \partial \xi^i \partial \xi^j,
\]

uniformly in \( \xi \).

Because of (150) we have both

\[
| \phi_{\text{seq}} (x, \xi, \delta^*) | \quad \text{and} \quad | \psi_{\text{seq}} (x, \xi, \delta^*) | \leq \sum_{i} \sum_{m} \left\{ \left| \phi_{\text{seq}} [x, \theta, \delta(\delta^*)] \right| \right\}
\]

\[
+ \left\{ | \phi_{\text{seq}} [x, \theta, \delta(\delta^*)] | \left[ \left. \frac{\partial \theta}{\partial \xi^i} \right|_{\xi_1} \right] \right\}
\]

\[
+ \sum_{i} \left\{ \left| \psi_{\text{seq}} [x, \theta, \delta(\delta^*)] \right| + \left| \mu_{\text{seq}} [x, \theta, \delta(\delta^*)] \right| \left[ \left. \frac{\partial \theta}{\partial \xi^i} \right|_{\xi_1} \right] \right\}
\]

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

We shall show that \( E_{\theta_1} [\nu_{\text{seq}} (x, \theta_2, \delta)]^2 \) and \( E_{\theta_2} [\mu_{\text{seq}} (x, \theta_2, \delta)]^2 \) are bounded func-
tions of \( \theta_1, \theta_2 \) and \( \delta \) in the domain \( D \), for sufficiently small \( \epsilon \). Our statement is proved if we show that \( E_\delta [\partial \log f(x, \theta_1)/\partial \theta^1]^2 \), \( E_\delta [\rho(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial \theta^1]^2 \) and \( E_\delta [\mu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial \theta^i]^2 \) are bounded in \( D_\epsilon \). The first of these expressions is bounded because of Assumption V. From (138) it follows that

\[
\left| \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leq 2\epsilon' \sum_i \left| \left| \phi_{ij}(x, \theta_2, \delta') \right| + \left| \psi_{ij}(x, \theta_2, \delta') \right| \right|
\]

and

\[
\left| \frac{\partial \log f(x, \theta_1)}{\partial \theta^i} \right| \leq 2\epsilon' \sum_i \left| \left| \phi_{ij}(x, \theta_2, \delta') \right| + \left| \psi_{ij}(x, \theta_2, \delta') \right| \right|
\]

where \( \delta' = \delta + \sum_i |\theta^1_i - \theta^2_i| \). From the above inequalities and the fact that \( E_\delta [\phi_{ij}(x, \theta_2, \delta)]^2 \) and \( E_\delta [\psi_{ij}(x, \theta_2, \delta)]^2 \) are bounded in \( D_\epsilon \), it follows that for sufficiently small \( \epsilon \) the expressions \( E_\delta [\rho(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial \theta^1]^2 \) and \( E_\delta [\mu_i(x, \theta_2, \delta) - \partial \log f(x, \theta_1)/\partial \theta^i]^2 \) are bounded in \( D_\epsilon \). Hence our statement is proved.

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

Assumption IV* is an immediate consequence of Assumption IV.

We have

\[
(154) \quad \frac{\partial \log f^*(x, \xi)}{\partial \xi^i} = \sum_i \frac{\partial \log f(x, \theta)}{\partial \theta^i} \frac{\partial \theta^i}{\partial \xi^i}.
\]

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

\[
(155) \quad E_\delta [\rho(x, \theta)]^{k \epsilon^*}
\]

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

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

\[
(156) \quad c^*_{ij}(\theta) = \sum_{l} \sum_m c_{lm}(\theta) \frac{\partial \theta^l}{\partial \xi^i} \frac{\partial \theta^m}{\partial \xi^j}.
\]

Denote by \( A \) the matrix \( \| \partial \theta^j / \partial \xi^i \| (i, j = 1, \ldots , k) \) and let \( \bar{A} \) be the trans-
posed of $A$. Then equation (156) can be written as

$$\|c_i^*(\theta)\| = \overline{A}\|c_i(\theta)\|.$$  

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

$$\|\sigma^*_{ij}(\theta)\| = B\|\sigma^*_{ij}(\theta)\|B^t,$$

where $\|\sigma^*_{ij}(\theta)\| = ||c^*_{ij}(\theta)||^{-1}$. Equation (158) can be written as

$$\sigma^*_{ij}(\theta) = \sum_i \sum_m \frac{\partial^t_\theta \xi^i}{\partial \theta^m} \sigma_{im}(\theta).$$

Let

$$\|c^*_{pq}(\theta)\| = ||\sigma^*_{pq}(\theta)||^{-1} \quad (p, q = 1, \ldots, r).$$

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

$$n \sum_{q=1}^r \sum_{j=1}^r \xi^q(\theta_n) \xi^q(\theta_n) c^*_{pq}(\theta_n) \geq d_n,$$

where the constant $d_n$ is chosen so that

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

The point $\theta$ is restricted to points of the set $\omega$ defined by equations (134).

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

$$n \sum_{q=1}^r \sum_{p=1}^r \xi^q(\theta)^t c^*_{pq}(\theta) = c,$$

$$\sum_{j=1}^k \gamma_{ij}(\theta) \xi^i(\theta) = \sum_{j=1}^k \gamma_{ij}(\theta) \xi^j(\theta) \quad (t = r + 1, \ldots, k),$$

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

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

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

Consider the transformation of the parameter space given by

$$\theta^t = \beta_{p1}(\theta) \xi^1(\theta) + \cdots + \beta_{pr}(\theta) \xi^r(\theta) \quad (p = 1, \ldots, r),$$

$$\theta^t = \gamma_{11}(\theta) \xi^1(\theta) + \cdots + \gamma_{rk}(\theta) \xi^k(\theta) \quad (l = r + 1, \ldots, k),$$
where $\theta$ denotes the point of $\omega$ for which $\theta$ lies on $S_c(\theta)$ for some value of $c$. The matrix $\|b_P(\theta)\|$ is chosen so that (163) is fulfilled. The transformation (164) transforms $S_c(\theta)$ into the sphere $S'_c(\theta)$ given by

$$\sum_{p=1}^{r} (\theta^p)^2 = c, \quad \theta^t = \sum_{i=1}^{h} \gamma_{si}(\theta) \xi^i(\theta).$$

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

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

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 $W_\theta^*$ 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 $\xi(\theta) = \cdots = \xi^t(\theta) = 0$ the sequence $\{W_\theta^*\}$

(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 $H_\theta$, Neyman and Pearson introduced a statistic(4), called likelihood ratio, defined as follows: The density function in the sample space is given by $\prod_{a=1}^{n} f(x_a, \theta)$. Denote by $P(x_1, \cdots, x_n)$ the maximum of this function with respect to $\theta^1, \cdots, \theta^k$, and let $P_\theta(x_1, \cdots, x_n)$ be the conditional maximum with respect to $\theta^1, \cdots, \theta^k$, subject to the condition that $\theta$ must be a point of $\omega$. Then the likelihood ratio for testing the hypothesis $H_\theta$ is given by $\lambda_n(\omega, E_n) = P_\theta(x_1, \cdots, x_n)/P(x_1, \cdots, x_n)$. 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 $H_\theta$ is rejected if the value of $\lambda_n(\omega, E_n)$ is less than a certain constant $\lambda_\omega(\omega)$. Denote the region $\lambda_n(\omega, E_n) < \lambda_\omega(\omega)$ by $L_n(\omega)$. In all that follows we choose the constant $\lambda_n(\omega)$ so that

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

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

(4) 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).
\[ \sum_a \log f(x_a, \theta) = \sum_a \log f(x_a, \theta_n) + \frac{1}{2} \sum_n \sum a (\delta_n^i - \theta^i)(\delta_n^i - \theta^i) \sum_a \frac{\partial^2 \log f(x_a, \theta_n)}{\partial \theta_i \partial \theta^i} \frac{1}{n} \]

where \( \theta_n \) lies in the interval \([\theta_n, \theta]\). Since

\[ \sum_a \log f(x_a, \theta_n) - \sum_a \log f(x_a, \theta) \geq - \log \lambda_n(\omega, E_n), \]

we have

\[ -2 \log \lambda_n(\omega, E_n) \leq - \sum a (\delta_n^i - \theta^i)(\delta_n^i - \theta^i) \frac{1}{n} \sum a \frac{\partial^2 \log f(x_a, \theta_n)}{\partial \theta_i \partial \theta^i} \]

Since \( n^{-1} \sum a \partial^2 \log f(x_a, \theta_n)/\partial \theta_i \partial \theta^i \) converges stochastically to \(-c_i(\theta)\) under the assumption that \( \theta \) is the true parameter point, it follows easily from Proposition 1 that for any \( \epsilon > 0 \) there exists a positive value \( A(\epsilon) \) such that for any \( \omega \)

\[ \limsup_{n \to \infty} \left\{ \text{l.u.b.} P \left[-2 \log X_n(\omega, E_n) \geq A(\epsilon) | \theta \right] \right\} \leq \epsilon \]

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

\[ \limsup_{n \to \infty} \left\{ \text{l.u.b.} P \left[-2 \log \lambda_n(\omega, E_n) \geq A(\epsilon) | \theta \right] \right\} \leq \epsilon. \]

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

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

Let \( \{ \theta_n \} \) be a sequence of parameter points such that \( \lim \delta(\theta_n, \omega) = 0 \) and \( \lim_{n \to \infty} n^{1/2} \delta(\theta_n, \omega) = +\infty \). We shall prove that there exists a positive \( \nu_0 \) such that for any constant \( A \)

\[ \lim_{n \to \infty} P \left[ \sum_a \log f(x_a, \theta_n) - \sum_a \log f(x_a, \theta^*) > A | \theta_n \right] = 1, \]

where \( \theta^* \) denotes a point of \( \omega \) for which \( \left| \theta^* - \theta_n \right| \leq \nu_0 \) and \( \sum_a \log f(x_a, \theta^*) \geq \sum_a \log f(x_a, \theta) \) for all \( \theta \) in the domain \( \left| \theta - \theta_n \right| \leq \nu_0 \). Consider the Taylor expansion

\[ \sum_a \log f(x_a, \theta^*) - \sum_a \log f(x_a, \theta_n) = \frac{n}{2} \sum a (\delta_n^i - \theta^i)(\delta_n^i - \theta^i) \frac{1}{n} \sum a \frac{\partial^2 \log f(x_a, \theta_n)}{\partial \theta_i \partial \theta^i}, \]
where \( \bar{\theta}_n \) lies in the interval \([\theta_0, \theta^*]\). Because of conditions (a) and (b) of Assumption III for any \( \epsilon > 0 \) the positive number \( \nu_0 \) can be chosen so that

\[
(169) \quad \lim_{n \to \infty} P \left[ \left| c_{ij}(\theta^*) + \frac{1}{n} \sum_a \frac{\partial^2}{\partial \theta^i \partial \theta^j} f(x_n, \bar{\theta}_n) \right| < \epsilon \mid \theta_n \right] = 1.
\]

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, \( \nu_0, (167) \) holds. Hence our statement is proved.

We say that the likelihood ratio test is uniformly consistent if for any positive \( \nu \)

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

uniformly in \( \omega \) and \( \theta \) over the domain \( \delta(\theta, \omega) \geq \nu \). 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 \( \omega_n(\theta) \) be the intersection of \( \omega \) with the set of all points \( \theta' \) for which \( \sum_{i=1}^k |\theta' - \theta'| \geq \nu \). From Assumption VII it follows that for any positive \( \nu \)

\[
(170) \quad \lim_{n \to \infty} P \left\{ -2 \log \lambda_n[\omega_n(\theta), E_n] \geq -2 \log \Lambda_n[\omega_n(\theta)] \mid \theta \right\} = 1
\]

uniformly in \( \theta \).

Let \( \{\theta_n\} \) be a sequence of parameter points such that \( \lim_{n \to \infty} \delta(\theta_n, \omega) = 0 \) and \( \lim_{n \to \infty} n^{1/2} \delta(\theta_n, \omega) = +\infty \). Denote by \( \omega_n \) the set of all points \( \theta \) for which \( |\theta - \theta_n| \leq \nu_n \). Since \( -2 \log \lambda_n[\omega_n(\theta)] \) has a finite upper bound it follows from (167) that for a sufficiently small \( \nu_0 \)

\[
(171) \quad \lim_{n \to \infty} P \left\{ -2 \log \lambda_n[\omega_n, E_n] \geq -2 \log \Lambda_n[\omega_n(\theta_n)] \mid \theta_n \right\} = 1.
\]

Obviously

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

From (170) and (171) we obtain

\[
(172) \quad \lim_{n \to \infty} P \left\{ -2 \log \lambda_n(\omega, E_n) \geq -2 \log \Lambda_n[\omega_n(\theta_n)] \mid \theta_n \right\} = 1.
\]

Since \( -2 \log \lambda_n[\omega_n(\theta_n), E_n] \geq -2 \log \lambda_n(\omega, E_n) \), we have \( -2 \log \Lambda_n[\omega_n(\theta_n)] \geq -2 \log \Lambda_n(\omega) \). Hence from (172) we obtain

\[
(173) \quad \lim_{n \to \infty} P \left[ L_n(\omega) \mid \theta_n \right] = 1.
\]
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 \( \omega \) is the set given by \( \omega = \omega_0 \). Let \( \{\theta_n\} \) be a sequence of parameter points for which the sequence \( n^{1/2} \delta(\theta_n, \omega) \) is bounded. Denote by \( \theta_n(\omega) \) the parameter point in \( \omega \) for which

\[
\theta_n(\omega) = \inf_{\theta \in \omega} \left\{ \theta : \sum_{i=1}^n \langle \theta_i - \bar{\theta}_n, \theta - \bar{\theta}_n \rangle \leq 2 \right\}.
\]

Let \( \theta_n \) be the point for which \( \delta(\theta_n, \omega) = \delta(\theta_n, \omega_0) \). Denote by \( T_n(\nu) \) the region given by the inequalities

\[
|\theta_n - \bar{\theta}_n| < \nu,
\]

\[
|\theta_n(\omega) - \bar{\theta}_n| < \nu.
\]

We shall prove that

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

Consider the Taylor expansion

\[
\sum_{a} \log f(x_a, \theta_n) = \sum_{a} \log f(x_a, \theta_n) + \frac{n}{2} \sum_{a} \sum_{i,j} (\theta_n^i - \bar{\theta}_n^i)(\theta_n^j - \bar{\theta}_n^j) \frac{1}{n} \sum_{a} \frac{\partial^2 \log f(x_a, \bar{\theta}_n)}{\partial \theta_i \partial \theta_j}.
\]

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

\[
\lim sup_n P\left\{ \sum_{a} \log f(x_a, \theta_n) - \sum_{a} \log f(x_a, \theta_n) \geq B_\epsilon \mid \theta_n \right\} = \epsilon.
\]

Since \( -2 \log \lambda_n(\omega, E_n) \leq 2 \sum_{a} \log f(x_a, \theta_n) - 2 \sum_{a} \log f(x_a, \theta_n) \) we have

\[
\lim sup_n P\left\{ -2 \log \lambda_n(\omega, E_n) \geq 2 B_\epsilon \mid \theta_n \right\} \leq \epsilon.
\]

Denote by \( \omega_n \) the subset of \( \omega \) in which \( |\theta - \theta_n| \geq \nu \). From Assumption VII it follows that

\[
\lim_{n \to \infty} P\left\{ -2 \log \lambda_n(\omega_n, E_n) \geq -2 \log \lambda_n(\omega_n) \mid \theta_n \right\} = 1.
\]

Since \( -2 \log \lambda_n(\omega) \leq -2 \log \lambda_n(\omega_n) \) we have

\[
\lim_{n \to \infty} P\left\{ -2 \log \lambda_n(\omega_n, E_n) \geq -2 \log \lambda_n(\omega) \mid \theta_n \right\} = 1.
\]

For any given constant \( B \) there exists a positive \( \alpha < 1 \) such that \( -2 \log \lambda_n(\omega) \).
$B$ if l.u.b.a $P[L_n(\omega) \mid \theta] = \alpha$. Hence from (179) we obtain

$$\lim_{n \to \infty} P[-2 \log \lambda_n(\omega, E_n) \geq B \mid \theta_n] = 1$$

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

$$\lim_{n \to \infty} \{P[-2 \log \lambda_n(\omega, E_n) \geq B \mid \theta_n] - P[\mid \theta_n(\omega) - \theta_n \mid \geq \nu \mid \theta_n]\} \geq 0$$

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

$$\lim_{n \to \infty} P[\mid \theta_n(\omega) - \theta_n \mid \geq \nu \mid \theta_n] = 0.$$ 

Hence (175) is proved.

Consider the Taylor expansion

$$\sum \log f(x, \theta_n) = \sum \log f[x, \theta_n(\omega)]$$

$$+ \frac{n}{2} \sum_{r=1}^{k} \sum_{s=1}^{k} [\theta_n(\omega) - \theta_n]^i [\theta_n(\omega) - \theta_n]^j \frac{1}{n} \sum_{\alpha} \frac{\partial^2 \log f(x, \theta_n)}{\partial \theta^i \partial \theta^j}$$

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

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

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

$$\lim_{n \to \infty} P[(1 + \varepsilon)q_n - (1 - \varepsilon)\bar{q}_n \geq -2 \log \lambda_n(\omega, E_n)]$$

$$\geq (1 - \varepsilon)q_n - (1 + \varepsilon)\bar{q}_n \mid \theta_n] = 1,$$

where

$$q_n = \sum_{i=1}^{k} \sum_{r=1}^{k} n(\theta_n^i - \theta_n)(\theta_n^i - \theta_n)c_{ij}(\theta_n)$$

and

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

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

$$\lim_{n \to \infty} P[(1 + \varepsilon)q_n - (1 - \varepsilon)\bar{q}_n \geq 0 \mid \theta_n] = 1.$$
Since the sequence \( \{ n^{1/2} \delta(\theta, \theta_n) \} \) is bounded, the expression \( q_n \) is bounded in the probability sense, that is for each positive \( \rho \) there exists a positive value \( A_\rho \) such that

\[
\lim_{n \to \infty} P(q_n > A_\rho \mid \theta_n) \leq \rho.
\]

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

\[
\lim P[q_n - \tilde{q}_n + \epsilon \geq - 2 \log \lambda_n(\omega, E_n) \geq q_n - \tilde{q}_n - \epsilon \mid \theta_n] = 1
\]

for any \( \epsilon > 0 \). From the Taylor expansions

\[
y_n^i(\theta_n) = \frac{1}{n^{1/2}} \sum \frac{\partial \log f(x_n, \theta_n)}{\partial \theta^i}
\]

(188)

\[
= - \sum_{j=1}^k n^{1/2} (\theta^*_n - \theta_n^i) \frac{1}{n} \sum \frac{\partial^2 \log f(x_n, \theta_n^* \mid \theta_n^i)}{\partial \theta^j \partial \theta^i} \quad (i = 1, \ldots, k)
\]

and

\[
y_n^i(\theta_n) = - \sum_{j=r+1}^k n^{1/2} [\theta^*_n - \theta_n^i] \frac{1}{n} \sum \frac{\partial^2 \log f(x_n, \theta_n^* \mid \theta_n^i)}{\partial \theta^j \partial \theta^i} \quad (i = r + 1, \ldots, k)
\]

(189)

we obtain

\[
y_n^i(\theta_n) = + \sum_{j=1}^k n^{1/2} [\theta^*_n - \theta_n^i] [c_{ij}(\theta_n) + \epsilon_{ij}(E_n)],
\]

(190)

\[
n^{1/2} (\theta^*_n - \theta_n^i) = \sum_{j=1}^k [\sigma_{ij}(\theta_n) + \eta_{ij}(E_n)] y_n^i(\theta_n) \quad (i = 1, \ldots, k),
\]

and

\[
n^{1/2} [\theta^*_n - \theta_n^i] = \sum_{j=r+1}^k [\bar{\sigma}_{ij}(\theta_n) + \bar{\eta}_{ij}(E_n)] y_n^i(\theta_n) \quad (i = r + 1, \ldots, k),
\]

(191)

where

\[
\| \sigma_{ij}(\theta_n) \| = \| c_{ij}(\theta_n) \|^{-1} \quad (i, j = 1, \ldots, k),
\]

\[
\| \bar{\sigma}_{ij}(\theta) \| = \| \bar{c}_{ij}(\theta) \|^{-1} \quad (i, j = r + 1, \ldots, k),
\]

and for any positive \( \nu \) we have

\[
\lim_{n \to \infty} P[| \epsilon_{ij}(E_n) | < \nu \mid \theta_n] = \lim_{n \to \infty} P[| \eta_{ij}(E_n) | < \nu \mid \theta_n] = \lim_{n \to \infty} P[| \bar{\epsilon}_{ij}(E_n) | < \nu \mid \theta_n] = 1.
\]

Hence
\[
\sum_{r+1}^{k} \sum_{r+1}^{k} n \left[ \theta^i_n (\omega) - \theta^i_n \right] \left[ \theta^j_n (\omega) - \theta^j_n \right] c_{ij}(\theta_n)
\]

\[
= \sum_{r+1}^{k} \sum_{r+1}^{k} \sum_{r+1}^{k} \sum_{r+1}^{k} \left[ c_{ij}(\theta_n) \bar{c}_{ij}(\theta_n) \bar{c}_{jm}(\theta_n) + \rho_{ijmn}(E_n) \right] y^i_n(\theta_n) y^m_n(\theta_n)
\]

\[
= \sum_{r+1}^{k} \sum_{r+1}^{k} \left[ \bar{c}_{ij}(\theta_n) + \rho_{ijmn}'(E_n) \right] y^i_n(\theta_n) y^j_n(\theta_n)
\]

\[
= \sum_{r+1}^{k} \sum_{r+1}^{k} \sum_{r+1}^{k} \sum_{r+1}^{k} \left[ \bar{c}_{ij}(\theta_n) c_{ij}(\theta_n) c_{jm}(\theta_n) + \rho_{ijmn}'(E_n) \right] n \left( \theta^i_n - \theta^i_n \right) (\theta^m_n - \theta^m_n),
\]

where

\[
\lim_{n \to \infty} P \left[ \left| \rho_{ijmn}'(E_n) \right| < \nu \right| \theta_n \right] = 1
\]

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

\[
\sum_{j=r+1}^{k} \sum_{l=r+1}^{k} \bar{c}_{ij}(\theta_n) c_{ij}(\theta_n) c_{jm}(\theta_n) = c_{mj}(\theta_n).
\]

Hence

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

The coefficient

\[
A_{im}(\theta_n) = c_{im}(\theta_n) - \sum_{r+1}^{k} \sum_{l=r+1}^{k} \bar{c}_{ij}(\theta_n) c_{ij}(\theta_n) c_{jl}(\theta_n)
\]

can be written as the following ratio:

\[
A_{im}(\theta_n) = -\frac{\left| \begin{array}{cccc}
  c_{im}(\theta_n) & c_{i r+1}(\theta_n) & \cdots & c_{i k}(\theta_n) \\
  c_{m r+1}(\theta_n) & c_{r+1 r+1}(\theta_n) & \cdots & c_{r+1 k}(\theta_n) \\
  \vdots & \vdots & \ddots & \vdots \\
  c_{mk}(\theta_n) & c_{k r+1}(\theta_n) & \cdots & c_{kk}(\theta_n)
\end{array} \right|}{\left| \begin{array}{cccc}
  c_{r+1 r+1}(\theta_n) & \cdots & c_{r+1 k}(\theta_n) \\
  \vdots & \ddots & \vdots \\
  c_{k r+1}(\theta_n) & \cdots & c_{kk}(\theta_n)
\end{array} \right|}.
\]  

(193)

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

\[
(194) \quad M = \begin{vmatrix}
\sigma_{1m}(\theta_n) & \sigma_{1r+1}(\theta_n) & \cdots & \sigma_{1k}(\theta_n) \\
\sigma_{r+1 m}(\theta_n) & \sigma_{r+1 r+1}(\theta_n) & \cdots & \sigma_{r+1 k}(\theta_n) \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{km}(\theta_n) & \sigma_{kr+1}(\theta_n) & \cdots & \sigma_{kk}(\theta_n)
\end{vmatrix}.
\]

Then we have

\[
(195) \quad M' = A^{k-r+1} M',
\]

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

\[
(196) \quad M_1' = A^{k-r} M_1,
\]

where \( M_1 \) denotes the algebraic complement of \( M_1 \).

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

\[
A_{im}(\theta_n) = \frac{M}{M_1} = c_{im}(\theta_n) \quad (l, m = 1, \ldots, r),
\]

where \( ||c_{im}(\theta_n)|| = ||\sigma_{im}(\theta_n)||^{-1} \) \( (l, m = 1, \ldots, r) \). Hence

\[
g_n - \tilde{g}_n = \frac{n}{r} \sum_{q=1}^r \sum_{p=1}^r (\theta_n^q - \theta_n^p)(\theta_n^q - \theta_n^p) \left[ \varepsilon_{pq}(\theta_n) + \eta_{pq}(E_n) \right] + \rho_n(E_n),
\]

where for any positive \( \nu \)

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

From (187) it follows that for any positive \( \epsilon \)

\[
(197) \quad \lim_{n \to \infty} \left\{ -2 \log \lambda_n(\omega, E_n) - \epsilon \leq \sum_{i=1}^r \sum_{p=1}^r n(\theta_n^p - \theta_n^i)(\theta_n^p - \theta_n^i) \left[ \varepsilon_{pq}(\theta_n) + \eta_{pq}(E_n) \right] \leq -2 \log \lambda_n(\omega, E_n) + \epsilon | \theta_n \right\} = 1.
\]
Since \( \theta_n^p = \theta_0^p (p = 1, \ldots, r) \) from (197) we obtain because of the boundedness of the sequence \( \{ n^{1/2} (\theta_n - \theta_0) \} \)

\[
\lim_{n \to \infty} P \left[ -2 \log \lambda_n(\omega, E_n) - \epsilon \leq \sum_{p=1}^{r} \sum_{q=1}^{s} n (\theta_n^p - \theta_0^p) (\theta_n^q - \theta_0^q) \xi_{pq}(\theta_n) \right] \leq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n \right] = 1.
\]

The above equation remains valid if we substitute \( \theta_n \) for \( \theta_0 \), that is

\[
\lim_{n \to \infty} P \left[ -2 \log \lambda_n(\omega, E_n) - \epsilon \leq n \sum_{p=1}^{r} \sum_{q=1}^{s} (\theta_n^p - \theta_0^p) (\theta_n^q - \theta_0^q) \xi_{pq}(\theta_n) \right] \leq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n \right] = 1.
\] (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)

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

uniformly in \( \theta \).

Now we consider the general case where \( \omega \) is given by \( r \) equations

\[
\hat{\xi}^i(\theta) = \cdots = \hat{\xi}^r(\theta) = 0
\]

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

\[
\xi^1 = \hat{\xi}^1(\theta), \ldots, \xi^r = \hat{\xi}^r(\theta),
\]

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

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

uniformly in \( \theta \), 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(*) has derived the large sample distribution of the likelihood ratio \( \lambda_n(\omega, E_n) \) if \( \omega \) is

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 $\omega$ satisfying Assumption VI in both cases, when the hypothesis to be tested is true, and when it is not true.

Let $u_1, \ldots, u_r$ be $r$ independently and normally distributed variates with unit variances. Denote the expected value of $u_p$ by $\mu_p$. 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,

$$(201) \quad P(U^2 < t) = F_r(\lambda^2, t) \quad (\lambda^2 = \mu_1^2 + \cdots + \mu_r^2).$$

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

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

$$(202) \quad V^2 = \sum_{s=1}^r \sum_{p=1}^r \lambda_{pq} v_p v_q,$$

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

$$(203) \quad P(V^2 < t) = F_r(\lambda^2, t),$$

where

$$(204) \quad \lambda^2 = \sum_{s=1}^r \sum_{p=1}^r \lambda_{pq} \mu_p \mu_q.$$

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

$$(205) \quad Q_n = n \sum_{s=1}^r \sum_{p=1}^r \xi^p(\theta_n) \xi^q(\theta_n) c_{pq}^*(\theta_n).$$

The joint distribution of the variates $n^{1/2} \left[ \xi^p(\theta_n) - \xi^p(\theta) \right], \ldots, n^{1/2} \left[ \xi^q(\theta_n) - \xi^q(\theta) \right]$ converges with $n \to \infty$ uniformly towards the cumulative normal distribution with zero means and covariance matrix $\|\sigma_{pq}^*(\theta)\| = \|\sigma_{pq}(\theta)\|^{-1}$. Since $\theta_n$ is a uniformly consistent estimate of $\theta$ and since $c_{ij}(\theta)$ is a uniformly continuous function of $\theta$, the statistic

$$(206) \quad \overline{Q}_n = n \sum \sum \xi^p(\theta_n) \xi^q(\theta_n) \overline{c}_{pq}(\theta)$$

has the same limit distribution as $Q_n$, that is,

\begin{equation}
\lim_{n \to \infty} \{ P[Q_n < t \mid \theta] - P[Q_n < t \mid \theta]\} = 0
\end{equation}

uniformly in $\theta$ and $t$.

It is easy to see that

\begin{equation}
\lim_{n \to \infty} \{ P[Q_n < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0
\end{equation}

uniformly in $\theta$ and $t$, where

\begin{equation}
\lambda_n^2(\theta) = n \sum \xi(\theta) = \xi(\theta) \xi^* \xi(\theta) = \xi^* \xi(\theta).
\end{equation}

Hence, because of (207) we have

\begin{equation}
\lim_{n \to \infty} \{ P[Q_n < t \mid \theta] - F_r[\lambda_n^2(\theta), t]\} = 0
\end{equation}

uniformly in $\theta$ and $t$.

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

\begin{equation}
\lim_{n \to \infty} P[-2 \log \lambda_n(\omega, E_n) - \epsilon \leq Q_n \leq -2 \log \lambda_n(\omega, E_n) + \epsilon \mid \theta_n] = 1
\end{equation}

for any positive $\epsilon$. From (210) and (211) it follows that

\begin{equation}
\lim_{n \to \infty} \{ P[-2 \log \lambda_n(\omega, E_n) < t \mid \theta_n] - F_r[\lambda_n^2(\theta_n), t]\} = 0
\end{equation}

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)

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

uniformly in $\theta$ and $t$. Hence we have proved the following theorem.

\textbf{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 $\lambda^2(\omega) = \cdots = \lambda^2(\theta) = 0$. Let furthermore $\lambda_n^2(\theta)$ be the expression defined in (209). Then, if Assumption VI holds, we have

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

uniformly in $t$ and $\theta$. If the hypothesis to be tested is true, that is if $\theta$ is a point of $\omega$, $\lambda_n^2(\theta) = 0$ and therefore the limit distribution of $-2 \log \lambda_n(\omega, E_n)$ is the $\chi^2$-distribution with $r$ degrees of freedom.
15. Summary. Let \( f(x_1, \ldots, x_m, \theta_1, \ldots, \theta_k) \) be the joint probability density function of the variates \( x_1, \ldots, x_m \) involving \( k \) unknown parameters \( \theta_1, \ldots, \theta_k \). Any set of values \( \theta_1, \ldots, \theta_k \) can be represented by a point \( \theta \) of the \( k \)-dimensional Cartesian space with the coordinates \( \theta_1, \ldots, \theta_k \). Let \( \omega \) be the subset of the parameter space defined by the equations

\[
\xi^r(\theta) = 0, \quad (r \leq k),
\]

that is, \( \omega \) is the set of all points \( \theta \) for which the above equations are fulfilled. Denote by \( H_\omega \) the hypothesis that the true parameter point \( \theta \) is an element of \( \omega \). In this paper the question of an appropriate test of the hypothesis \( H_\omega \) is discussed when the number of observations is large.

The following notations have been introduced. The point \( \hat{\theta}_n \) denotes the point with the coordinates \( \hat{\theta}_{1n}, \ldots, \hat{\theta}_{kn} \) where \( \hat{\theta}_{in} \) is the maximum likelihood estimate of \( \theta_i \) based on \( n \) independent observations on \( x_1, \ldots, x_m \). The expected value of \( -\partial^2 \log f(x_1, \ldots, x_m, \theta_1, \ldots, \theta_k) \partial \theta_i \partial \theta_j \) is denoted by \( c_{ij}(\theta) \) and \( ||\sigma_{ij}(\theta)|| = ||c_{ij}(\theta)||^{-1} \). Furthermore \( \sum_{q=1}^r \sum_{p=1}^r (\partial^2 \xi^p / \partial \theta_i \partial \theta_j) (\partial^2 \xi^q / \partial \theta_i \partial \theta_j) \sigma_{pq}(\theta) \) \((p, q = 1, \ldots, r)\) is denoted by \( \sigma_{pq}^*(\theta) \) and

\[
||\xi_{pq}^*(\theta)|| = ||\sigma_{pq}^*(\theta)||^{-1}, \quad (p, q = 1, \ldots, r).
\]

The region \( W_n^* \) denotes the critical region defined by the inequality

\[
n \sum_{q=1}^r \sum_{p=1}^r \xi_{pq}(\hat{\theta}_n) \xi_{pq}^*(\hat{\theta}_n) \xi_{pq}^*(\hat{\theta}_n) \geq d_n,
\]

where \( n \) is the number of independent observations on \( x_1, \ldots, x_m \), 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 \( \omega \), is equal to a given positive \( \alpha < 1 \).

Let \( \lambda_n \) be the likelihood ratio statistic for testing \( H_\omega \) and let \( L_n \) be the critical region defined by the inequality

\[
\lambda_n \leq \bar{\lambda}_n,
\]

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 \( \omega \), is equal to \( \alpha \).

Under certain assumptions on \( f(x_1, \ldots, x_m, \theta_1, \ldots, \theta_k) \) and the functions \( \xi^1(\theta), \ldots, \xi^r(\theta) \) the following results have been obtained:

I. For testing the hypothesis \( H_\omega \) 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 \( n \sum_{q=1}^r \sum_{p=1}^r \xi_{pq}(\hat{\theta}_n) \xi_{pq}^*(\hat{\theta}_n) \xi_{pq}^*(\hat{\theta}_n) \) have the
same limit distribution. The limit distribution of $-2 \log \lambda_n$ is the $\chi^2$-distribution with $r$ degrees of freedom if the hypothesis to be tested is true. If the true parameter point $\theta_*$ is not an element of $\omega$, 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 $u_1, \ldots, u_r$ are independently and normally distributed with unit variances and

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

Columbia University,
New York, N. Y.