

Mathematical Proceedings of the Cambridge Philosophical Society

<http://journals.cambridge.org/PSP>

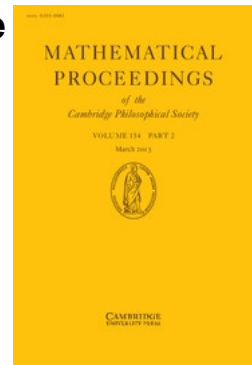
Additional services for *Mathematical Proceedings of the Cambridge Philosophical Society*:

Email alerts: [Click here](#)

Subscriptions: [Click here](#)

Commercial reprints: [Click here](#)

Terms of use : [Click here](#)



Maximum-likelihood estimation in non-standard conditions

P. A. P. Moran

Mathematical Proceedings of the Cambridge Philosophical Society / Volume 70 / Issue 03 / November 1971, pp 441 - 450

DOI: 10.1017/S0305004100050088, Published online: 24 October 2008

Link to this article: http://journals.cambridge.org/abstract_S0305004100050088

How to cite this article:

P. A. P. Moran (1971). Maximum-likelihood estimation in non-standard conditions. *Mathematical Proceedings of the Cambridge Philosophical Society*, 70, pp 441-450 doi:10.1017/S0305004100050088

Request Permissions : [Click here](#)

Maximum-likelihood estimation in non-standard conditions

By P. A. P. MORAN

The Australian National University, Canberra

(Received 13 November 1970)

The origin of the present paper is the desire to study the asymptotic behaviour of certain tests of significance which can be based on maximum-likelihood estimators. The standard theory of such problems (e.g. Wald (4)) assumes, sometimes tacitly, that the parameter point corresponding to the null hypothesis lies inside an open set in the parameter space. Here we wish to study what happens when the true parameter point, in estimation problems, lies on the boundary of the parameter space.

We suppose we have a sample X_1, \dots, X_n (scalars or vectors) from a distribution $F(x, \theta)$, where $\theta = (\theta^1, \dots, \theta^k)$ is a vector parameter lying in Euclidean space. The set of all such points such that $F(x, \theta)$ is a proper distribution is the 'natural parameter space Ω ' of the distribution. This may be open or closed or neither. In what follows we shall confine ourselves to estimates which are obtained by maximising the likelihood for variations of θ in a closed bounded space Ω_1 contained in Ω , and which may or may not be a proper subset of Ω . In particular we shall be concerned with cases where θ lies on the boundary of Ω_1 and usually on the boundary of Ω as well.

As an example of the type of problem to which we want to apply the theory consider the usual test of homogeneity for a sample from a Poisson distribution. Let the sample be X_1, \dots, X_n and suppose that the null hypothesis asserts that it comes from a Poisson distribution of the form

$$e^{-\lambda_0} \lambda_0^n (n!)^{-1}. \quad (1)$$

The alternative hypothesis is that the sample comes from a compound distribution of the form

$$(n!)^{-1} \int_0^\infty e^{-\lambda} \lambda^n f(\lambda) d\lambda, \quad (2)$$

where $f(\lambda)$ is the density of some distribution with mean λ_0 and small variation about λ_0 . For example, we might assume that $f(\lambda)$ is the log-normal distribution with mean λ_0 and variance σ^2 . The null hypothesis is then $\sigma^2 = 0$, which is certainly on the natural boundary of the parameter space, and the alternative hypothesis is that $\sigma^2 > 0$. We may now ask the following questions. (1) Is the usual test involving the ratio of the sample variance to the sample mean asymptotically equivalent to a test based on maximum-likelihood estimators? (2) Is the test based on a maximum-likelihood estimator of σ^2 asymptotically optimal in some sense? (3) If it is optimal, is this optimality robust against various forms of the distribution $f(\lambda)$?

It is proposed to investigate these problems in later papers but in order to do so we have first to investigate the asymptotic behaviour of maximum-likelihood estimators when the true value lies on the boundary of the parameter space.

We do this by proceeding along the lines of Wald(4). We suppose that $F(x, \theta)$ is either a purely discrete distribution with an enumerable number of points of increase, or an absolutely continuous distribution on the space of the x 's. In the first case we write $f(x, \theta)$ for the probability of a point, and in the second for the probability density at x .

We assume that θ takes values in a closed bounded set Ω_1 which is not necessarily the whole of the natural parameter space Ω of the distribution. In general Ω_1 may have some boundary points which are also boundary points of Ω . For reasons of convenience and simplicity we take Ω_1 to be the closed subset of Euclidean space defined by

$$0 \leq \theta^i \leq b_i \quad (b_i > 0) \quad (i = 1, \dots, k). \quad (3)$$

We shall be concerned with the situation when one or more of the θ^i are zero, and in particular $\theta^1 = 0$.

We have as likelihood,

$$\exp L = f(X_1, \theta) \dots f(X_n, \theta). \quad (4)$$

If θ_1 and θ_2 are parameter points (or other vectors) we write

$$\theta_1 - \theta_2 < A,$$

where $A = (A^1, \dots, A^k)$, to mean that the components satisfy

$$\theta_1^i - \theta_2^i < A^i \quad \text{for } i = 1, \dots, k.$$

We also define a distance function by

$$|\theta_1 - \theta_2| = \max_i |\theta_1^i - \theta_2^i|. \quad (5)$$

Assumption 1. $f(X, \theta)$ is a continuous function of θ in Ω_1 .

The maximum-likelihood estimator is defined to be any function $\hat{\theta}(X_1, \dots, X_n)$ which lies in Ω_1 and results in an absolute maximum of L in Ω_1 . By Assumption 1 at least one such function exists. If there exists more than one such function we select one by any convenient rule.

Assumption 2. Let D_n be the set of points in the sample space for which the derivatives

$$\frac{\partial^2 f(X_s, \theta)}{\partial \theta^i \partial \theta^j} \quad (s = 1, \dots, n; i, j = 1, \dots, k) \quad (6)$$

are continuous functions of θ . Then

$$\text{Prob}(D_n | \theta) = 1. \quad (7)$$

Assumption 3. $\hat{\theta}$ is a consistent estimator of θ , uniformly for θ in Ω_1 .

This will be true if the assumptions set out in Moran(3) are verified but will also hold in more general cases.

Following Wald we now define

$$\psi_{ij}(x, \theta_1, \delta) = \text{glb} \frac{\partial^2 \log f(x, \theta)}{\partial \theta^i \partial \theta^j}, \quad (8)$$

where the greatest lower bound is taken for values of θ satisfying $|\theta - \theta_1| \leq \delta$. Similarly we define

$$\phi_{ij}(x, \theta_1, \delta) = \text{lub} \frac{\partial^2 \log f(x, \theta)}{\partial \theta^i \partial \theta^j}, \tag{9}$$

where the least upper bound is taken for values of θ satisfying $|\theta - \theta_1| \leq \delta$.

Assumption 4. For any sequences $\{\theta_1(n)\}, \{\theta_2(n)\}$ belonging to Ω_1 , and any sequence $\{\delta_n\}$, for which

$$\lim_n \theta_1(n) = \lim_n \theta_2(n) = \theta \quad \text{in} \quad \Omega_1, \tag{10}$$

$$\lim_n \delta_n = 0, \tag{11}$$

we have

$$\begin{aligned} \lim_n E_{\theta_1(n)} \psi_{ij}(X, \theta_2(n), \delta_n) &= \lim_n E_{\theta_1(n)} \phi_{ij}(X, \theta_2(n), \delta_n) \\ &= E \left\{ \frac{\partial^2 \log f(X, \theta)}{\partial \theta^i \partial \theta^j} \right\} \end{aligned} \tag{12}$$

uniformly in θ , where E denotes the expectation.

Assumption 5. There exists $\epsilon > 0$ so that

$$E_{\theta_1} \{ \psi_{ij}(X, \theta_2, \delta)^2 \} \quad \text{and} \quad E_{\theta_1} \{ \phi_{ij}(X, \theta_2, \delta)^2 \}$$

are bounded functions of θ_1, θ_2 , and δ in the subset of Ω_1 defined by $|\theta_1 - \theta_2| \leq \epsilon$, and the interval $0 \leq \delta \leq \epsilon$.

Assumption 6. The greatest lower bound, with respect to θ in Ω_1 , of the determinant of the matrix

$$\left\{ -E_\theta \left(\frac{\partial^2 \log f(X, \theta)}{\partial \theta^i \partial \theta^j} \right) \right\} \tag{13}$$

is greater than zero.

Assumption 7.

$$\int \frac{\partial f(x, \theta)}{\partial \theta^i} dx = \int \frac{\partial^2 f(x, \theta)}{\partial \theta^i \partial \theta^j} dx = 0, \tag{14}$$

for all θ in Ω_1 , and $i, j = 1, \dots, k$. Here the integrals are interpreted as sums if $F(x, \theta)$ is discrete.

Assumption 8. There exists $\eta > 0$ so that

$$E_\theta \left| \frac{\partial}{\partial \theta^i} \log f(X, \theta) \right|^{2+\eta} \quad (i = 1, \dots, k), \tag{15}$$

are bounded functions of θ in Ω_1 .

Write

$$c_{ij} = -E_\theta \left\{ \frac{\partial^2 \log f(X, \theta)}{\partial \theta^i \partial \theta^j} \right\}, \tag{16}$$

and (σ_{ij}) for the inverse of the matrix (c_{ij}) . Write $z_n(\theta)$ for the vector whose components are

$$z_n^i = n^{\frac{1}{2}}(\hat{\theta}^i - \theta^i) \quad (i = 1, \dots, k) \tag{17}$$

where θ is the true value of the parameter and $\hat{\theta}$ is the maximum-likelihood estimator

THEOREM I. Suppose that $\theta^1 = 0$, and $0 < \theta^i < b_i$ for $i = 2, \dots, k$. Then the distribution function

$$\Phi_n(t, \theta) = \text{Prob}(z_n < t, \theta) \tag{18}$$

converges, uniformly in t and θ , towards the mixture of distributions

$$\text{Prob}(z < t, \theta) = \frac{1}{2}F_1(t, \theta) + \frac{1}{2}F_2(t, \theta), \tag{19}$$

where $F_1(t, \theta)$ is a k -dimensional multivariate distribution defined on the region

$$t^1 > 0, \quad -\infty < t^i < \infty \quad (i = 2, \dots, k), \tag{20}$$

and having in this region a probability density equal to twice the density of a multivariate normal distribution with means zero, and covariance matrix equal to (σ_{ij}) . $F_2(t, \theta)$ is a $(k-1)$ -dimensional distribution concentrated on the subspace $t^1 = 0$, $-\infty < t^i < \infty$ for $i = 2, \dots, k$, and such that the joint distribution of z^2, \dots, z^k is that of the quantities

$$z^i = \sum_{j=2}^k \sigma_{ij}^{(1)} y^j \quad (i = 2, \dots, k), \tag{21}$$

where y^1, \dots, y^k are jointly normally distributed with zero means and covariance matrix (c_{ij}) , the distribution of y^2, \dots, y^k being taken conditional on the inequality

$$y^1 - \sum_{j=2}^k c_{1j} \sum_{s=2}^k \sigma_{js}^{(1)} y^s \leq 0, \tag{22}$$

where the $\sigma_{ij}^{(1)}$ are the elements of the matrix

$$(\sigma_{ij}^{(1)}) = \begin{pmatrix} c_{22} & \cdot & \cdot & c_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{k2} & \cdot & \cdot & c_{kk} \end{pmatrix}^{-1}. \tag{23}$$

Furthermore, the convergence to the limiting distribution is uniform in t and in the subset of Ω_1 given by $\theta_1 = 0$.

Proof. This follows Wald(4) very closely. From Assumption 7 we have

$$E_\theta \left\{ \frac{\partial \log f(X, \theta)}{\partial \theta^i} \right\} = 0, \tag{24}$$

and
$$E_\theta \left\{ \frac{\partial \log f(X, \theta)}{\partial \theta^i} \cdot \frac{\partial \log f(X, \theta)}{\partial \theta^j} \right\} = -E_\theta \left\{ \frac{\partial^2 \log f(X, \theta)}{\partial \theta^i \partial \theta^j} \right\} = c_{ij}. \tag{25}$$

Hence the matrix (c_{ij}) is positive definite by Assumption 4, and so also are all submatrices obtained by striking out rows and columns with the same indices.

θ is the true value of the parameter, and $\hat{\theta}$ the maximum-likelihood estimator. By Assumption 3, as n tends to infinity, the probability tends to unity that $\hat{\theta}$ is in any neighbourhood of $\theta = (0, \theta^2, \dots, \theta^k)$ which is open relative to Ω_1 . Since $0 < \theta^i < b_i$, for $i = 2, \dots, k$, and the derivatives exist we have

$$\sum_s \frac{\partial \log f(X_s, \hat{\theta})}{\partial \theta^1} \leq 0, \tag{26}$$

and
$$\sum_s \frac{\partial \log f(X_s, \hat{\theta})}{\partial \theta^i} = 0, \quad \text{for } i = 2, \dots, k. \tag{27}$$

If $\theta^1 = 0$, the derivative in (26) is taken to the right whilst if $\theta^1 > 0$, (26) becomes an equality. Consider first the distribution of $\hat{\theta}$ conditional on $\theta^1 > 0$. Then for all sample points in the set D of assumption 2 we have

$$\sum_s \frac{\partial \log f(X_s, \hat{\theta})}{\partial \hat{\theta}^i} = \sum_s \frac{\partial \log f(X_s, \theta)}{\partial \theta^i} + \sum_{j=1}^k (\hat{\theta}^j - \theta^j) \sum_s \frac{\partial^2 \log f(X_s, \bar{\theta}_i)}{\partial \hat{\theta}^i \partial \theta^j} = 0, \tag{28}$$

where $\theta^1 = 0$, and $\bar{\theta}_i$, depending on i , is a vector lying on the segment joining θ to $\hat{\theta}$. Writing y_n for the vector with components

$$y_n^i = n^{-\frac{1}{2}} \sum_s \frac{\partial \log f(X_s, \theta)}{\partial \theta^i}, \tag{29}$$

and using (17), we get

$$y_n^i + \sum_{j=1}^k z_n^j n^{-1} \sum_s \frac{\partial^2 \log f(X_s, \bar{\theta}_i)}{\partial \hat{\theta}^i \partial \theta^j} = 0. \tag{30}$$

Let ν be arbitrarily small and define $Q_n(\theta)$ to be the subset of D_n for which

$$\left| n^{-1} \sum_s \frac{\partial^2 \log f(X_s, \bar{\theta})}{\partial \hat{\theta}^i \partial \theta^j} + c_{ij} \right| < \nu. \tag{31}$$

Then $\text{Prob}\{Q_n(\theta), \theta\}$ (32)

converges to unity as n tends to infinity, uniformly in θ . This is proved by Wald(4) p. 431, and it is unnecessary to repeat the proof here. Now using the fact that the determinant of (c_{ij}) has a positive lower bound, we conclude from (30) and (31) that conditional on $z_n^1 > 0$,

$$z_n^i = \sum_j y_n^i \{\sigma_{ij} + \nu \epsilon_{ij}\}, \tag{33}$$

where ϵ_{ij} is a bounded function of n, θ , and ν , and the sample is restricted to points of $Q_n(\theta)$. From the assumptions made and the standard central limit theorem it follows that, independently of any conditions on the signs of the z_n^i , the distribution of (y_n^1, \dots, y_n^k) converges, uniformly in θ , towards the multivariate normal distribution with means zero, and covariance matrix (c_{ij}) .

Neyman's $C(\alpha)$ tests are based on the asymptotic distribution of these derivatives and although in his papers he does impose the condition that the true parameter point is internal to an open set in the parameter space, it is clear from the above that this is an unnecessary condition. It is only for tests based on maximum-likelihood estimators that a special investigation is necessary when the null hypothesis asserts that the parameter lies on an essential boundary of the natural parameter space.

From the fact that ν in (33) can be made arbitrarily small by choosing n large, and from (32), it follows that the joint distribution of the z_n^i , conditional on $z_n^1 > 0$, converges, uniformly in θ , to a k -variate distribution whose density is zero for $z_n^1 \leq 0$, and for $z_n^1 > 0$ is twice that of a multivariate normal distribution with means zero and covariance matrix $(\sigma_{ij}(\theta))$.

Now consider what happens when $\theta_1 = 0$. Since $\hat{\theta}$ gives an absolute maximum to

the likelihood in Ω_1 , we have (26) and (27). Then arguing as above, with an error which converges to zero uniformly in θ ,

$$y_n^1 - \sum_{j=2}^k z_n^j c_{1j} \leq 0, \tag{34}$$

$$y_n^i = \sum_{j=2}^k z_n^j c_{ij}, \quad \text{for } i = 2, \dots, k. \tag{35}$$

Thus the limiting distribution of (z_n^2, \dots, z_n^k) is that of the quantities z_i in (21) subject to the inequality (22). This limiting distribution, which is not multivariate normal, is approached uniformly for θ in the subset of Ω_1 determined by $\theta^1 = 0$.

The condition (22) can be written in a simpler form which we need later. Write $(c_{ij}^{(1)})$ for the matrix obtained by deleting the first row and column of (c_{ij}) . Write b_{ij} for the minor of (c_{ij}) complementary to c_{ij} , and similarly $b_{ij}^{(1)}$ for the minor of $(c_{ij}^{(1)})$ complementary to c_{ij} . Then the coefficient of y_s ($s \geq 2$) in (22) is

$$\begin{aligned} \sum_{t=2}^k c_{1t} \sigma_{st}^{(1)} &= \sum_{t=2}^k c_{1t} b_{ts}^{(1)} (-1)^{t+s} |c_{ij}^{(1)}|^{-1} \\ &= |c_{ij}^{(1)}|^{-1} b_{1s} (-1)^s \\ &= -\sigma_{11}^{-1} \sigma_{1s}, \end{aligned} \tag{36}$$

so that (22) can be written

$$\sigma_{11}^{-1} \sum_{s=1}^k \sigma_{1s} y^s \leq 0. \tag{37}$$

We now consider the case where more than one of the θ^i lie on the boundary. For simplicity we confine ourselves to the situation where $\theta^1 = \theta^2 = 0$, $0 < \theta^i < b_i$ for $i = 3, \dots, k$.

THEOREM II. *In the above circumstances the quantities z_n^i have a distribution which converges, uniformly in θ , towards the mixture of distributions*

$$a_1 F_1(t, \theta) + a_2 F_2(t, \theta) + a_3 F_3(t, \theta) + a_4 F_4(t, \theta) \tag{38}$$

where, t being the vector (t^1, \dots, t^k) ,

(1) $F_1(t, \theta)$ is a multivariate distribution on the space $t^1 > 0, t^2 > 0, -\infty < t^i < \infty$ for $i = 3, \dots, k$ which has the probability density, in this region, of the k -variate normal distribution with means zero and covariance matrix (σ_{ij}) , conditional on the inequalities

$$z^1 = A = \sum_j \sigma_{1j} y^j > 0, \tag{39}$$

$$z^2 = B = \sum_j \sigma_{2j} y^j > 0, \tag{40}$$

and a_1 is the probability that these inequalities hold in the joint distribution of the y^i .

(2) $F_2(t, \theta)$ is the $(k-1)$ -variate distribution on the space $t^1 = 0, t^2 > 0, -\infty < t^i < \infty$ for $i = 3, \dots, k$ for which the density is that of the quantities

$$z^i = \sum_{j=2}^k \sigma_{ij}^{(1)} y^j, \tag{41}$$

conditional on

$$C = y^1 - \sum_{j=2}^k c_{1j} \sum_{s=2}^k \sigma_{js}^{(1)} y^s \leq 0, \tag{42}$$

and

$$D = z^2 = \sum_{j=2}^k \sigma_{2j}^{(1)} y^j > 0, \tag{43}$$

where the y^i are distributed normally with zero means and covariance matrix (c_{ij}) , and $(\sigma_{ij}^{(1)})$ is defined by (23). a_2 is the probability that $C \leq 0$, and $D > 0$.

(3) $F_3(t, \theta)$ is similar to $F_2(t, \theta)$ with z^1 and z^2 interchanged, and $(\sigma_{ij}^{(1)})$ replaced by $(\sigma_{ij}^{(2)})$, the inverse of the matrix obtained by striking out the second row and column of (c_{ij}) . The conditions are

$$E = y^2 - \sum_j c_{2j} \sum_s \sigma_{js}^{(2)} y^s \leq 0, \tag{44}$$

and

$$F = z^1 = \sum \sigma_{ij}^{(2)} y^j > 0, \tag{45}$$

where the sums are taken over $j, s, = 1, 3, \dots, k$. a_3 is the probability that $E \leq 0, F > 0$.

(4) $F_4(t, \theta)$ is a $(k-2)$ -variate distribution on the space $t^1 = t^2 = 0, -\infty < t^i < \infty$ for $i = 3, \dots, k$, for which the density is that of the quantities

$$z^i = \sum_{j=3}^k \sigma_{ij}^{(3)} y^j, \tag{46}$$

where $(\sigma_{ij}^{(3)})$ is the inverse of the matrix obtained by striking out the first and second rows and columns of (c_{ij}) , and the distribution is taken conditional on

$$G = y^1 - \sum_{j=3}^k c_{1j} \sum_{s=3}^k \sigma_{js}^{(3)} y^s \leq 0, \tag{47}$$

and

$$H = y^2 - \sum_{j=3}^k c_{2j} \sum_{s=3}^k \sigma_{js}^{(3)} y^s \leq 0. \tag{48}$$

a_4 is the probability that $G \leq 0, H \leq 0$.

Proof. The proof follows exactly the same modification of Wald's proof required to prove Theorem 1. Each of the above cases occurs conditionally on two inequalities involving the quantities y^i .

It is not obvious that given any set of values of (y^1, \dots, y^k) that the above set of conditions are both exclusive and exhaustive. A discussion of this point throws some light on what happens. Just as in the discussion in Theorem I we can prove at once that $C \leq 0$ is equivalent to

$$\sigma_{11}^{-1} A \leq 0,$$

or

$$A \leq 0, \tag{49}$$

since $\sigma_{11} > 0$. Similarly $E \leq 0$ is equivalent to

$$B \leq 0.$$

Next we can prove in exactly the same way that $G \leq 0$ and $H \leq 0$ are respectively equivalent to $F \leq 0, D \leq 0$. We also prove that D is a multiple of $\sigma_{11} B - \sigma_{12} A$. To do this we use a theorem of Jacobi (Mirsky (2), p. 25).

THEOREM III. *Let L be the matrix (l_{ij}) . Let M be a two-rowed minor of L , and L_{ij} the co-factor of l_{ij} in L . Write $L^* = (L_{ij})$ and define M' as the cofactor of M in L , and M^* as the minor of L^* corresponding to M . Then*

$$M^* = |L| M'. \tag{50}$$

We take the matrix L as (c_{ij}) . Then L_{ij} is $\sigma_{ij} |c|$, so that L^* is $(\sigma_{ij} |c|)$. Take M as the minor

$$\begin{vmatrix} c_{11} & c_{1j} \\ c_{21} & c_{2j} \end{vmatrix}. \tag{51}$$

Then M^* is
$$\begin{vmatrix} \sigma_{11} |c| & \sigma_{1j} |c| \\ \sigma_{21} |c| & \sigma_{2j} |c| \end{vmatrix}. \tag{52}$$

M' is then the cofactor of c_{2j} in the matrix whose inverse occurs on the right-hand side of (23) and is therefore equal to $\sigma_{2j}^{(1)} |c^{(1)}|$. From this it follows that

$$\begin{aligned} \sigma_{11} \sigma_{2j} - \sigma_{21} \sigma_{1j} &= |c|^{-1} |c^{(1)}| \sigma_{2j}^{(1)} \\ &= \sigma_{11} \sigma_{2j}^{(1)}. \end{aligned} \tag{53}$$

Hence
$$D = \sigma_{11}^{-1} (\sigma_{11} B - \sigma_{12} A). \tag{54}$$

Similarly, we can show

$$F = \sigma_{22}^{-1} (\sigma_{22} A - \sigma_{12} B). \tag{55}$$

We can now write the conditions involved in the four cases as follows.

$$F_1(t, \theta). \quad A > 0, \quad B > 0. \tag{56}$$

$$F_2(t, \theta). \quad A \leq 0, \quad \sigma_{11} B - \sigma_{12} A > 0. \tag{57}$$

$$F_3(t, \theta). \quad B \leq 0, \quad \sigma_{22} A - \sigma_{12} B > 0. \tag{58}$$

$$F_4(t, \theta). \quad \sigma_{11} B - \sigma_{12} A \leq 0, \quad \sigma_{22} A - \sigma_{12} B \leq 0. \tag{59}$$

There is no restriction in putting $\sigma_{11} = \sigma_{22} = 1$ and $\sigma_{12} = \rho$. Then for any set of values of the y^i , (59) is clearly incompatible with (57) and (58). Adding the two equations in (56), and the two in (59) we get $A + B > 0$, and $(1 - \rho)A + (1 - \rho)B \leq 0$ which are incompatible. (56) is obviously incompatible with (57) and (58), whilst from (57) and (58) we get $A + B \leq 0$, $(1 - \rho)A + (1 - \rho)B > 0$. Thus the four sets of conditions are incompatible.

To prove they exhaust all possibilities suppose that (59) is false. Then we must have one of the following possibilities:

- (1) $B - \rho A > 0, \quad A - \rho B > 0;$ (2) $B - \rho A > 0, \quad A - \rho B \leq 0;$
- (3) $B - \rho A \leq 0, \quad A - \rho B > 0.$

Consider case (1). If $B \leq 0$ we get (58). If $B > 0$ we get (56) if $A > 0$, and (57) if $A \leq 0$. In case (2) if $A \leq 0$ we get (57). Suppose $A > 0$. We have

$$(A - \rho B) - (B - \rho A) = (1 + \rho)(A - B).$$

The left-hand side is negative and therefore $B > 0$, so that we get (56). Case (3) is symmetric to case (2) and hence the above set of conditions exhaust all possibilities.

It is also instructive to look at the situation in another way. The quantities y^i are asymptotically distributed in a multivariate normal distribution with means zero and covariance matrix (c_{ij}) . The set of equations

$$y^i = \sum_{j=1}^k c_{ij} v^j \quad (i = 1, \dots, k), \tag{60}$$

therefore has a unique solution $v = (v^1, \dots, v^k)$, which has a k -variate normal distribution with means zero and covariance matrix (σ_{ij}) . We may call v the ‘pseudo-maximum likelihood point’, for if the v_i are such that the set of quantities $\theta^i + n^{-\frac{1}{2}}v^i$ lie in Ω_1 , they coincide, asymptotically, with the rescaled maximum likelihood point. Define the function

$$L_1 = -\frac{1}{2} \sum_{ij} c_{ij} (u^i - v^i) (u^j - v^j) \tag{61}$$

as a function of the point (u^1, \dots, u^k) in the whole Euclidean space. This is proportional to the logarithmic likelihood which can therefore be studied by studying L_1 . Suppose that either or both of v^1 and v^2 are negative and consider the hyperellipsoid

$$\sum_{ij} c_{ij} (u^i - v^i) (u^j - v^j) = K, \tag{62}$$

for varying values of K increasing from zero. Let K increase to the three values where (62) is tangential to the spaces $u^1 = 0$ ($k - 1$ dimensional), $u^2 = 0$ ($k - 1$ dimensional), and $u^1 = u^2 = 0$ ($k - 2$ dimensional). The first of these to occur which satisfies $u^1 \geq 0$, $u^2 \geq 0$ will have a tangent point which is, asymptotically, the maximum-likelihood point with respect to Ω_1 .

Assume first that $v^1 < 0$ and $v^2 > 0$. Then the first point of tangency lying in Ω_1 will be of the form $(0, u^2, \dots, u^k)$. If $u^2 > 0$ the conditions (42) and (43) will be satisfied. The only other alternative is that $u^2 = 0$ when the conditions (47) and (48) will be satisfied. The case $v^1 > 0$, $v^2 < 0$ is symmetric. If $v^1 < 0$, $v^2 < 0$ the first point of tangency with the space Ω_1 can have $u^1 > 0$, $u^2 = 0$, or $u^1 = u^2 = 0$, or $u^1 = 0$, $u^2 > 0$. Notice that this shows that, contrary to the situation in Theorem 1, we can have $v^1 < 0$ and $\theta^1 > 0$.

Theorems I and II are not adequate for a discussion of the asymptotic power of tests of hypotheses involving points on the boundary because to obtain the asymptotic power we also need to consider the asymptotic behaviour of maximum likelihood when the true parameter is inside Ω_1 , but approaches the boundary at the rate $n^{-\frac{1}{2}}$ as the sample size increases. It is here that uniformity of convergence becomes essential.

THEOREM IV. *Suppose that Assumptions 1-8 are satisfied and that $\theta^2, \dots, \theta^k$ have fixed values in open intervals $0 < \theta^i < b_i$ ($i = 2, \dots, k$), whilst $\theta^1 = a n^{-\frac{1}{2}}$, where $0 \leq a < a_0$. Then uniformly for a in this interval the joint distribution of z^1, \dots, z^k will tend to a mixture of distributions*

$$\alpha F_1(t) + (1 - \alpha) F_2(t), \tag{63}$$

where

(1) $F_1(t)$ is a k -variate distribution on the space $t^1 > -a$, $-\infty < t^i < \infty$ ($i = 2, \dots, k$), whose probability density is that of a multivariate normal distribution with means zero and covariance matrix (σ_{ij}) , conditional on $z^1 > -a$.

(2) $F_2(t)$ is a $(k-1)$ variate distribution on the space $t^1 = -a$, $-\infty < t_1 < \infty$ ($i = 2, \dots, k$), which is such that the joint distribution of z^2, \dots, z^k is obtained in the following way. The z^i are the solutions of the set of equations

$$\begin{aligned} z^1 &= -a, \\ y^i &= ac_{1i} + \sum_{j=2}^k c_{ij}z^j \quad (i = 2, \dots, k), \end{aligned} \quad (64)$$

where y^1, \dots, y^k are jointly normally distributed with zero means and covariance matrix (c_{ij}) conditional on

$$y^1 - ac_{11} - \sum_{j=2}^k c_{1j}z^j \leq 0. \quad (65)$$

Proof. This runs exactly as in the previous theorems. α is obtained as follows. Consider the distribution of z^1 . When $z^1 > 0$ this is a distribution with a density equal to that of a normal distribution with mean zero and variance σ_{11} . Thus we must have

$$\begin{aligned} \alpha &= \text{Prob}(z^1 > -a) \\ &= \frac{1}{(2\pi)^{\frac{1}{2}} \sigma_{11}^{\frac{1}{2}}} \int_{-a}^{\infty} \exp\left(-\frac{u^2}{2\sigma_{11}}\right) du. \end{aligned}$$

Theorem IV is easily generalized to cases where more than one of the θ^i are zero.

I am indebted to Dr W. Du Mouchel for some helpful criticism. Part of the above, work was carried out in the Department of Statistics, University of California Berkeley under a grant (GM 10525(07)) from the U.S. National Institutes of Health, Bethesda, Maryland.

REFERENCES

- (1) BRUNK, H. D. On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.* **29** (1958), 437-454.
- (2) MIRSKY, L. *An Introduction to Linear Algebra*. (Oxford University Press, 1955.)
- (3) MORAN, P. A. P. The uniform consistency of maximum-likelihood estimators. *Proc. Cambridge Philos. Soc.* **70** (1971), 435-439.
- (4) WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** (1943), 426-482.