MAXIMUM-LIKELIHOOD ESTIMATION OF PARAMETERS SUBJECT TO RESTRAINTS

By J. Aitchison and S. D. Silvey

University of Glasgow

Summary. The estimation of a parameter lying in a subset of a set of possible parameters is considered. This subset is the null space of a well-behaved function and the estimator considered lies in the subset and is a solution of likelihood equations containing a Lagrangian multiplier. It is proved that, under certain conditions analogous to those of Cramér, these equations have a solution which gives a local maximum of the likelihood function. The asymptotic distribution of this 'restricted maximum likelihood estimator' and an iterative method of solving the equations are discussed. Finally a test is introduced of the hypothesis that the true parameter does lie in the subset; this test, which is of wide applicability, makes use of the distribution of the random Lagrangian multiplier appearing in the likelihood equations.

1. Introduction. Quite frequently in statistical theory the natural way of building up a mathematical model of an experiment leads to the description of the experiment by a random variable X whose distribution function F depends on s parameters $\theta_1, \theta_2, \cdots, \theta_s$, which are not mathematically independent but satisfy r functional relationships $h_i(\theta_1, \theta_2, \cdots, \theta_s) = 0$, $i = 1, 2, \cdots, r$, $r < s$. In many cases where such a natural description arises it is possible to solve the r equations $h_i(\theta_1, \theta_2, \cdots, \theta_s) = 0$ for r of the parameters in terms of the remaining $s - r$, to express the distribution function F in terms of these remaining parameters only and, given observations on X, to estimate these $s - r$ unrestricted parameters by the method of maximum likelihood. This procedure has two disadvantages. First, it may be impossible to express r of the parameters explicitly in terms of the remaining $s - r$ and second, interest may lie in estimating all of the parameters simultaneously, in which case a symmetrical procedure for so doing is certainly desirable. The natural symmetric method for maximum-likelihood estimation in this case is achieved by the introduction of Lagrangian multipliers and it is this method that we will consider in this paper.

2. Formulation of the problem. In this section we will formulate more precisely the problem to be considered.

We will denote m-dimensional Euclidian space by $\mathfrak{R}^m$, $m = 1, 2, 3, \cdots$. A point in $\mathfrak{R}^s$, denoted by $\theta = (\theta_1, \theta_2, \cdots, \theta_s)$ will represent a value of a parameter. There is a particular point $\theta_0 = (\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_s^{(0)})$ in $\mathfrak{R}^s$ which is the true, though unknown, parameter value. Corresponding to each θ in some neighbour-

Received February 18, 1957.

hood of θ_0, say in $U_\alpha = \{\theta: \|\theta - \theta_0\| \leqq \alpha\}$, is a probability density function f_θ defined on $\mathcal{R}^1$ and we will denote the value of f_θ at the point $t \varepsilon \mathcal{R}^1$ by $f(t, \theta)$. The probability density function f_{θ_0} defines a probability measure on $\mathcal{R}^1$ and we will assume that, with respect to this measure, for almost all t, the partial derivatives $\partial \log f(t, \theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, exist for every θ in U_α.

There is given a continuous function h from $\mathcal{R}^s$ into $\mathcal{R}^r$, $r < s$, defined by $h(\theta) = (h_1(\theta), h_2(\theta), \cdots, h_r(\theta))$, which is such that, for every θ in U_α, the partial derivatives $\partial h_j(\theta)/\partial\theta_i$, $i = 1, 2, \cdots, s, j = 1, 2, \cdots, r$, exist. The function h has the further property that $h(\theta_0) = 0$.

A point in $\mathcal{R}^n$ denoted by $x = (x_1, x_2, \cdots, x_n)$ will be regarded as representing a set of n independent observations on a random variable whose probability density function is f_{θ_0} and we use the fact that points in $\mathcal{R}^n$ are being so regarded to define, in the usual way, a probability measure on $\mathcal{R}^n$, for each n. Subsequent statements regarding the probabilities of sets in $\mathcal{R}^n$ will refer to this particular probability measure.

It will be convenient to use also matrix representation for points in $\mathcal{R}^m$ and for linear operators from one Euclidian space to another and we will use the convention that, for example, $\boldsymbol{\theta}$ is the $s \times 1$ column vector representing the point θ in $\mathcal{R}^s$, and $\mathbf{H}$, an $s \times r$ matrix, represents a linear operator H from $\mathcal{R}^r$ into $\mathcal{R}^s$.

The log-likelihood function L is defined on a subset of $\mathcal{R}^n \times \mathcal{R}^s$ by

$$L(x, \theta) = \sum_{i=1}^{n} \log f(x_i, \theta).$$

If $\mathbf{H}_\theta$ denotes the $s \times r$ matrix $(\partial h_j(\theta)/\partial\theta_i)$, and if λ is a Lagrangian multiplier in $\mathcal{R}^r$, then we propose to estimate the unknown parameter θ_0 by a solution, if such exists, of the equations

$$\ell(x, \theta) + H_\theta\lambda = 0 \tag{2.1}$$

$$h(\theta) = 0, \tag{2.2}$$

where $\ell(x, \theta)$ is the point in $\mathcal{R}^s$ whose ith component is $\partial L(x, \theta)/\partial\theta_i$.

We will show that, under certain fairly general conditions, if x belongs to a set whose probability measure tends to 1 as $n \to \infty$, these equations have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$, where $\hat{\theta}(x)$ is near θ_0 and $\hat{\theta}(x)$ maximises $L(x, \theta)$ subject to the condition $h(\theta) = 0$. The definition of $\hat{\theta}$ and $\hat{\lambda}$ will then be extended in a natural way to the whole of $\mathcal{R}^n$ and we will show that the random variables thus defined are asymptotically jointly normally distributed. We will then consider an iterative procedure for solving the equations (2.1) and (2.2). Finally tests of the adequacy of the model will be introduced.

3. Existence of a solution. The proof that we will give of the existence of a solution of the equations (2.1) and (2.2) is based on the same principle as a proof given by Cramér [2] of the existence of a maximum likelihood estimate of a parameter in $\mathcal{R}^1$. However the presence of the restraining condition $h(\theta) = 0$ in the situation we are discussing makes our proof more intricate in detail than a straightforward generalisation of Cramér's proof to a parameter in $\mathcal{R}^s$ would be. And we start by indicating the main lines of the proof.

We set out to show that, under certain conditions, if δ is a sufficiently small given number and if n is sufficiently large, then, for a set of x whose probability measure is near 1, the equations (2.1) and (2.2) have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$, where $\hat{\theta}(x) \varepsilon U_\delta$. We will demand that in U_α the function $\log f(x, \cdot)$ should possess partial derivatives of the third order and the components of the function h should possess partial derivatives of the second order. Then it will be possible, by expanding the components of $\ell(x, \theta)$ and $h(\theta)$ about θ_0 to express the equations (in matrix notation) in the form

$$\mathbf{l}(x, \theta_0) + \mathbf{M}_{x, \theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{v}^{(1)}(x, \theta) + \mathbf{H}_\theta\boldsymbol{\lambda} = \mathbf{0}, \tag{3.1}$$

$$\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \mathbf{v}^{(2)}(\theta) = \mathbf{0}, \tag{3.2}$$

where

(i) $\mathbf{M}_{x, \theta_0}$ is the matrix $(\partial^2 L(x, \theta_0)/\partial\theta_i\partial\theta_j)$,

(ii) $\mathbf{v}^{(1)}(x, \theta)$ is a vector whose mth component may be expressed in the form $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{L}_m(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $\mathbf{L}_m$ being the matrix $(\partial^3 L(x, \theta^{(m,1)})/\partial\theta_m\partial\theta_i\partial\theta_j)$, $i, j = 1, 2, \cdots, s$, and $\theta^{(m, 1)}$ a point such that $\| \theta^{(m, 1)} - \theta_0 \| < \| \theta - \theta_0 \|$.

(iii) $\mathbf{v}^{(2)}(\theta)$ is a vector whose mth component is $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{H}_m(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$, $\mathbf{H}_m$ being the matrix $(\partial^2 h_m(\theta^{(m,2)})/\partial\theta_i\partial\theta_j)$, $i, j = 1, 2, \cdots, s$, and $\theta^{(m, 2)}$ a point such that $\| \theta^{(m, 2)} - \theta_0 \| < \| \theta - \theta_0 \|$.

Further conditions imposed on f, which are almost a straightforward generalisation of Cramér's conditions [2], will ensure that, for large enough n, there is a set of x whose probability measure is near 1 such that, if x belongs to this set,

(i) $\| (1/n)\ell(x, \theta_0) \|$ is small,

(ii) $-(1/n)\mathbf{M}_{x, \theta_0}$ is near a certain positive definite matrix $\mathbf{B}_{\theta_0}$ and

(iii) the elements of $(1/n)\mathbf{L}_m$ are bounded for $\theta \varepsilon U_\delta$. By dividing (3.1) by n we will then be able to express this equation in the form

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{n}\mathbf{H}_\theta\boldsymbol{\lambda} + \delta^2\mathbf{v}^{(3)}(x, \theta) = \mathbf{0} \tag{3.3}$$

where $\| v^{(3)}(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$. In addition we will demand that, for $\theta \varepsilon U_\alpha$, the second order derivatives of the components of h should be bounded. Then we will be able to express (3.2) in the form

$$\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta^2\mathbf{v}^{(4)}(\theta) = \mathbf{0} \tag{3.4}$$

where $\| v^{(4)}(\theta) \|$ is bounded for $\theta \varepsilon U_\delta$.

If the equations (3.3) and (3.4) have a solution, then by pre-multiplying (3.3) by $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}$ and substituting for $\mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)$ from (3.4) we find that the values of θ and λ satisfying these equations also satisfy an equation of the form

$$\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta\left(\frac{1}{n}\boldsymbol{\lambda}\right) + \delta^2\mathbf{v}^{(5)}(x, \theta) = \mathbf{0}. \tag{3.5}$$

We will impose conditions on h which ensure that the matrix $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta$ is non-singular and the elements of its inverse are bounded functions of θ for $\theta \varepsilon U_\delta$. Then it will be possible to solve equation (3.5) for λ in terms of θ and on substitution in (3.3) we will obtain the result that any value of θ in U_δ for which equations (3.3) and (3.4) are satisfied is also a solution of an equation of the form

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta^2 \mathbf{v}(x, \theta) = \mathbf{0} \tag{3.6}$$

where $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$.

Conversely it will be shown that if the equation (3.6) has a solution $\hat{\theta}(x) \varepsilon U_\delta$ then $\hat{\theta}(x)$ leads to a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ of equations (2.1) and (2.2). We will then use the fact that $\mathbf{B}_{\theta_0}$ is a positive definite matrix to prove that, if δ is sufficiently small, (3.6) has a solution in U_δ.

This outline of the method of proof to be adopted provides the motivation for the introduction of conditions on f and h which we now discuss.

Conditions on f. The following conditions on the function f appear complicated and restrictive from the mathematical point of view. In fact they will be satisfied in most practical estimation problems.

$\mathfrak{F}1$. *For every $\theta \varepsilon U_\alpha$ and for almost all $t \varepsilon \mathcal{R}^1$ (almost all with respect to the probability measure on $\mathcal{R}^1$ defined by f_{θ_0}), the derivatives*

$$\frac{\partial \log f(t, \theta)}{\partial \theta_i}, \quad \frac{\partial^2 \log f(t, \theta)}{\partial \theta_i \partial \theta_j} \quad \textit{and} \quad \frac{\partial^3 \log f(t, \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \qquad i, j, k = 1, 2, \cdots, s,$$

exist, and the first and second order derivatives are continuous functions of θ.

$\mathfrak{F}2$. *For every $\theta \varepsilon U_\alpha$ and for $i, j = 1, 2, \cdots, s$, $| \partial f(t, \theta)/\partial \theta_i | < F_1(t)$ and $| \partial^2 f(t, \theta)/\partial \theta_i \partial \theta_j | < F_2(t)$, where F_1 and F_2 are finitely integrable over $(-\infty, \infty)$.*

$\mathfrak{F}3$. *For every $\theta \varepsilon U_\alpha$ and $i, j, k = 1, 2, \cdots, s$, $| \partial^3 \log f(t, \theta)/\partial \theta_i \partial \theta_j \partial \theta_k | < F_3(t)$, where $\int_{-\infty}^{\infty} F_3(t) f(t, \theta_0)\, dt$ is finite and equal to κ_1, say.*

$\mathfrak{F}4$. $$b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta_0)}{\partial \theta_i} \frac{\partial \log f(t, \theta_0)}{\partial \theta_j} f(t, \theta_0)\, dt$$

is finite for $i, j = 1, 2, \cdots, s$, and the matrix $\mathbf{B}_{\theta_0} = (b_{ij})$ is positive definite with minimum latent root μ_0.

The conditions $\mathfrak{F}3$ and $\mathfrak{F}4$ are apparently less stringent than a straightforward generalisation of Cramér's corresponding conditions would be. In §6 we return to this point.

If f satisfies these conditions then for any given positive numbers $\delta < \alpha$ and $\epsilon < 1$ and for sufficiently large n, say $n \geqq n(\delta, \epsilon)$, there exists a set $X_n \subset \mathcal{R}^n$ with the properties

$\mathfrak{X}1$. $\Pr \{X_n\} > 1 - \epsilon$.

$\mathfrak{X}2$. $\left\| \frac{1}{n} \ell(x, \theta_0) \right\| < \delta^2$, if $x \varepsilon X_n$.

$\mathfrak{X}3$. $\frac{1}{n} \mathbf{M}_{x, \theta_0}$ can be expressed in the form $-\mathbf{B}_{\theta_0} + \delta \mathbf{m}_{x,\theta_0}$, where $\mathbf{m}_{x,\theta_0}$ is an $s \times s$ matrix the moduli of whose elements are bounded by 1, if $x \varepsilon X_n$.

$\mathfrak{X}4$. For *every* $\theta \varepsilon U_\alpha$ and $i, j, k = 1, 2, \cdots, s,$

$$\left| \frac{1}{n} \frac{\partial^3 L(x, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < 2\kappa_1$$

if $x \varepsilon X_n$.

The proof of these results is similar to the proof of the corresponding results given by Cramér [2] in the case of a parameter in $\mathfrak{R}^1$ and we merely remark that the conditions $\mathfrak{F}1$–4 imply (as they are designed to imply) that

(i) $(1/n)\ell(\cdot, \theta_0)$ converges in probability to $0 \varepsilon \mathfrak{R}^s$,

(ii) $(1/n)\mathbf{M}_{\cdot,\theta_0}$ converges in probability to $-\mathbf{B}_{\theta_0}$, and

(iii) if $G(x) = 1/n\sum_{i=1}^n F_3(x_i)$, then the random variable G converges in probability to κ_1 and

$$\frac{1}{n}\left| \frac{\partial^3 L(x, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| = \frac{1}{n}\left| \sum_{l=1}^{n} \frac{\partial^3 \log f(x_l, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < G(x),$$

by $\mathfrak{F}3$.

In future when we refer to a set X_n we imply that n is sufficiently large for the existence of a set in $\mathfrak{R}^n$ with the properties $\mathfrak{X}1$–4 and that the set X_n referred to has these properties.

As has already been indicated, one of the main purposes of the introduction of the conditions $\mathfrak{F}$ was to ensure that (3.1) could be expressed in the form (3.3). Now if the conditions $\mathfrak{F}$ are satisfied, if $x \varepsilon X_n$ and $\theta \varepsilon U_\delta$, it is easily verified that

(i) by $\mathfrak{X}2$, $(1/n\delta^2)\| \ell(x, \theta_0) \| < 1$,

(ii) by $\mathfrak{X}3$, $(1/\delta)\| m_{x,\theta_0}(\theta - \theta_0) \| \leqq s^2$,

(iii) by $\mathfrak{X}4$, $(1/n\delta^2)\| v^{(1)}(x, \theta) \| < (1/\delta^2)s^3\kappa_1\| \theta - \theta_0 \|^2 \leqq s^3\kappa_1$.

It follows that (3.1) can then be expressed in the form (3.3) and

$$\| v^{(3)}(x, \theta) \| < 1 + s^2 + s^3\kappa_1, \quad \text{when} \quad x \varepsilon X_n \quad \text{and} \quad \theta \varepsilon U_\delta.$$

Conditions on h. We impose the following conditions on the function h.

$\mathfrak{H}1$. *For every* $\theta \varepsilon U_\alpha$ *the partial derivatives* $\partial h_k(\theta)/\partial\theta_i$, $i = 1, 2, \cdots, s$, $k = 1, 2, \cdots, r$, *exist and these are continuous functions of* θ.

$\mathfrak{H}2$. *For every* $\theta \varepsilon U_\alpha$ *the partial derivatives* $\partial^2 h_k(\theta)/\partial\theta_i\partial\theta_j$, $i, j = 1, 2, \cdots,$ s, $k = 1, 2, \cdots, r$, *exist and* $| \partial^2 h_k(\theta)/\partial\theta_i\partial\theta_j | < 2\kappa_2$, *a given constant, for all* i, j *and* k.

$\mathfrak{H}3$. *The* $s \times r$ *matrix* $\mathbf{H}_{\theta_0}$ *is of rank* r.

The condition $\mathfrak{H}2$ is introduced to ensure that when (3.2) is expressed in the form (3.4), $\| v^{(4)}(\theta) \|$ is bounded for $\theta \varepsilon U_\delta$. It is clear that it does ensure this since, as is easily verified, by $\mathfrak{H}2$, $\| v^{(2)}(\theta) \| < s^3\kappa_2 \| \theta - \theta_0 \|^2$ and so $\| v^{(4)}(\theta) \| = (1/\delta^2)\| v^{(2)}(\theta) \| < s^3\kappa_2$ if $\theta \varepsilon U_\delta$.

Also the condition $\mathcal{H}3$ implies that the matrix $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_{\theta_0}$ is positive definite, since the matrix $\mathbf{B}_{\theta_0}^{-1}$ is positive definite. Since the elements of $\mathbf{H}_\theta$ are, by $\mathcal{H}1$, continuous functions of θ it follows that there exists a neighbourhood of θ_0 in which det $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)$ is bounded away from zero, and we may assume that this neighbourhood is U_α. (This assumption merely involves choosing α small enough initially). This means that when $\theta \varepsilon U_\alpha$ we can solve the equation (3.5) for λ in terms of θ. Furthermore the elements of the matrix $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ are then continuous functions on U_α since the elements of $\mathbf{H}_\theta$ are continuous and det $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)$ is bounded away from zero. Since U_α is a closed set it follows that the elements of $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ are uniformly bounded on U_α. This result, together with the results that $\| v^{(3)}(x, \theta) \|$ and $\| v^{(4)}(\theta) \|$ are bounded on U_δ, enable us to prove that when λ is eliminated from (3.3) and (3.4), and (3.6) is obtained, then in (3.6) $\| v(x, \theta) \|$ is bounded on U_δ, if $x \varepsilon X_n$.

We have now gone a considerable way towards proving the main part of the following lemma.

LEMMA 1. *Subject to the conditions $\mathcal{F}$ and $\mathcal{H}$, if $\delta < \alpha$ and $\epsilon < 1$ are given positive numbers and if $x \varepsilon X_n$, then the equations* (2.1) *and* (2.2) *have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$, if and only if $\hat{\theta}(x)$ satisfies a certain equation of the form $-B_{\theta_0}(\theta - \theta_0) + \delta^2 v(x, \theta) = 0$. In this equation $v(x, \cdot)$ is a continuous function on U_δ and $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$ by a positive number κ_3, say.*

PROOF. The fact that the condition is necessary has virtually been established already. On eliminating λ from (2.1) and (2.2) by the method outlined at the beginning of §3 we obtain, in matrix notation, the following explicit expression for (3.6)

$$-\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \mathbf{H}_\theta(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}\{\mathbf{v}^{(2)}(\theta) + \mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{v}^{(6)}(x, \theta)\} + \mathbf{v}^{(6)}(x, \theta) = \mathbf{0}, \tag{3.7}$$

where

$$\mathbf{v}^{(6)}(x, \theta) = \delta^2\mathbf{v}^{(3)}(x, \theta) = \frac{1}{n}\mathbf{l}(x, \theta) + \mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0), \tag{3.8}$$

and

$$\mathbf{v}^{(2)}(\theta) = \delta^2\mathbf{v}^{(4)}(\theta) = \mathbf{h}(\theta) - \mathbf{H}'_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{3.9}$$

Hence in (3.6),

$$\mathbf{v}(x, \theta) = -\mathbf{H}_\theta(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}\{\mathbf{v}^{(4)}(\theta) + \mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{v}^{(3)}(x, \theta)\} + \mathbf{v}^{(3)}(x, \theta). \tag{3.10}$$

The fact that $v(x, \cdot)$ is a continuous function on U_δ and that $\| v(x, \theta) \|$ is bounded for $\theta \varepsilon U_\delta$ follows from (3.8), (3.9) and (3.10), in virtue of the discussion of $v^{(3)}(x, \theta)$, $v^{(4)}(\theta)$ and $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_\theta)^{-1}$ above.

Turning to the sufficiency of the condition we now suppose that the equation

(3.7) has a root $\hat{\theta}(x) \varepsilon U_\delta$. Then, writing $\hat{\theta}$ instead of $\hat{\theta}(x)$ for brevity, we obtain on premultiplication of (3.7) by $\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}$,

$$-\mathbf{H}'_{\theta_0}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \mathbf{v}^{(2)}(\hat{\theta}) = \mathbf{0}, \tag{3.11}$$

i.e., by (3.9),

$$\mathbf{h}(\hat{\theta}) = \mathbf{0}.$$

Substitution for $\mathbf{v}^{(2)}(\hat{\theta})$ from (3.11) and for $\mathbf{v}^{(6)}(x, \hat{\theta})$ from (3.8), in (3.7) gives

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{H}_{\hat{\theta}}(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_{\hat{\theta}})^{-1}\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{l}(x, \hat{\theta}),$$

or, if we write $\mathbf{Q}_{\hat{\theta}}$ for $(\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}\mathbf{H}_{\hat{\theta}})^{-1}\mathbf{H}'_{\theta_0}\mathbf{B}_{\theta_0}^{-1}$,

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{H}_{\hat{\theta}}\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta}). \tag{3.12}$$

If we now define $\hat{\boldsymbol{\lambda}}(x)$ by

$$\hat{\boldsymbol{\lambda}}(x) = -\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta}),$$

then

$$\mathbf{l}(x, \hat{\theta}) + \mathbf{H}_{\hat{\theta}}\hat{\boldsymbol{\lambda}}(x) = \mathbf{0},$$

and $\hat{\theta}(x)$, $\hat{\lambda}(x)$ satisfy the equations (2.1) and (2.2).

In order to prove that the equation (3.6) has a root in U_δ, if δ is sufficiently small, we will require the following lemma.

Lemma 2. *If g is a continuous function mapping $\mathcal{R}^s$ into itself with the property that, for every θ such that $\| \theta \| = 1$, $\boldsymbol{\theta}'\mathbf{g}(\theta) < 0$, then there exists a point $\hat{\theta}$ such that* $\| \hat{\theta} \| < 1$ and $g(\hat{\theta}) = 0$.

Proof. For the proof of this result we are indebted to Mr. J. M. Michael who has proved that this result is equivalent to Brouwer's fixed point theorem [4]. A direct proof from the latter theorem is as follows.

We suppose that $g(\theta) \neq 0$ for any θ such that $\| \theta \| \leqq 1$. Then the function g_1, defined on the unit sphere in $\mathcal{R}^s$ by

$$g_1(\theta) = \frac{g(\theta)}{\| g(\theta) \|},$$

is a continuous function mapping this unit sphere into itself. Hence by the fixed point theorem there is a point θ^* in the unit sphere such that $\theta^* = g_1(\theta^*)$. Also since $\| g_1(\theta) \| = 1$ for every θ in the unit sphere, it follows that $\| \theta^* \| = 1$, and $\boldsymbol{\theta}^{*\prime}\mathbf{g}_1(\theta^*) = \boldsymbol{\theta}^{*\prime}\boldsymbol{\theta}^* = 1 > 0$. But this contradicts the fact that $\boldsymbol{\theta}'\mathbf{g}(\theta) < 0$ (and consequently that $\boldsymbol{\theta}'\mathbf{g}_1(\theta) < 0$) for every θ such that $\| \theta \| = 1$.

Hence there is a point $\hat{\theta}$ in the unit sphere such that $g(\hat{\theta}) = 0$. It is obvious that $\| \hat{\theta} \| \neq 1$. Hence $\| \hat{\theta} \| < 1$.

We are now in a position to prove the following existence theorem.

Theorem 1. *Subject to the conditions $\mathfrak{F}$ and $\mathfrak{H}$, if δ is a sufficiently small given* *positive number, ϵ is a given positive number less than* 1 *and if $x \varepsilon X_n$, then the equations* (2.1) *and* (2.2) *have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$.*

PROOF. We suppose $\delta < \alpha$ and $x \varepsilon X_n$. We consider (3.6) and define a function g on the unit sphere in $\mathcal{R}^s$ by

$$g\left(\frac{\theta - \theta_0}{\delta}\right) = -B_{\theta_0}(\theta - \theta_0) + \delta^2 v(x, \theta).$$

By Lemma 1, $v(x, \cdot)$ is a continuous function on U_δ. Hence g is a continuous function on the unit sphere in $\mathcal{R}^s$. Also

$$\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}\left(\frac{\theta - \theta_0}{\delta}\right) = -\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{v}(x, \theta)$$

$$\leqq -\frac{1}{\delta}\mu_0 \| \theta - \theta_0 \|^2 + \delta\kappa_3 \| \theta - \theta_0 \|,$$

if $\theta \varepsilon U_\delta$, since $\mathbf{B}_{\theta_0}$ is positive definite with minimum latent root μ_0 and, by Lemma 1, $\| v(x, \theta) \| < \kappa_3$ when $\theta \varepsilon U_\delta$. Hence for every θ such that $\| \theta - \theta_0 \| = \delta$, we have

$$\frac{1}{\delta}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)'\mathbf{g}\left(\frac{\theta - \theta_0}{\delta}\right) \leqq \delta(\delta\kappa_3 - \mu_0)$$

$$< 0, \quad \text{if} \quad \delta < \frac{\mu_0}{\kappa_3}.$$

Hence if $\delta < \mu_0/\kappa_3$, it follows by Lemma 2 that there exists a point $\hat{\theta}(x)$ such that $\hat{\theta}(x) \varepsilon U_\delta$ and $g((\hat{\theta}(x) - \theta_0) / \delta) = 0$, i.e., $\hat{\theta}(x)$ is a solution of (3.6). The result follows by application of Lemma 1.

4. Existence of a maximum of $L(x, \theta)$. In this paragraph we will show that for sufficiently small δ, if $x \varepsilon X_n$, any solution of (3.6) in U_δ maximises $L(x, \theta)$ subject to the condition $h(\theta) = 0$.

We suppose that $x \varepsilon X_n$, that δ is small enough for Theorem 1 to apply and that $\hat{\theta}(x)$, written $\hat{\theta}$ for typographical brevity, is a solution in U_δ of (3.6). We let θ be a point in a neighbourhood of $\hat{\theta}$ contained in U_δ, such that $h(\theta) = 0$. (Such a neighbourhood exists since $\hat{\theta}$ is an interior point of U_δ.) Then by expanding $L(x, \theta)$ about $\hat{\theta}$ we have

$$L(x, \theta) - L(x, \hat{\theta}) = \mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \tfrac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{M}_{x,\theta^*}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{4.1}$$

where $\mathbf{M}_{x,\theta^*} = (\partial^2 L(x, \theta^*) / \partial\theta_i\partial\theta_j)$ and $\theta^* \varepsilon U_\delta$.

We now consider separately the two terms in the right hand side of (4.1). By (3.12)

$$\mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \mathbf{l}'(x, \hat{\theta})\mathbf{Q}_{\hat{\theta}}'\mathbf{H}_{\hat{\theta}}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).$$

Now

$$\mathbf{0} = \mathbf{h}(\theta) - \mathbf{h}(\hat{\theta}) = \mathbf{H}_{\hat{\theta}}'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \mathbf{r}(\theta),$$

where, because of $\mathcal{H}2$, by the same argument as was applied to $v^{(2)}(\theta)$ in (3.2), (4.2) $$\| r(\theta)\| < s^3\kappa_2 \| \theta - \hat{\theta} \|^2.$$

Hence

(4.3) $$\mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = -[\mathbf{Q}_{\hat{\theta}}\mathbf{l}(x, \hat{\theta})]'\mathbf{r}(\theta).$$

By (3.8)

$$\frac{1}{n}\mathbf{l}(x, \hat{\theta}) = -\mathbf{B}_{\theta_0}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathbf{v}^{(6)}(x, \hat{\theta}),$$

and so

$$\frac{1}{n}\| \ell(x, \hat{\theta}) \| < \kappa_4\delta + \kappa_5\delta^2, \quad \text{since} \quad \hat{\theta} \,\varepsilon\, U_\delta,$$

where κ_4 is a positive number depending only on the elements of $\mathbf{B}_{\theta_0}$, and, as above, $\kappa_5 = 1 + s^2 + s^3\kappa_1$. Also the elements of $\mathbf{Q}_{\hat{\theta}}$ are bounded by a number independent of δ, since $\hat{\theta} \,\varepsilon\, U_\alpha$. Hence

(4.4) $$\frac{1}{n}\| Q_\theta l(x, \hat{\theta}) \| < \kappa_6\delta + \kappa_7\delta^2,$$

where κ_6, κ_7 are positive numbers independent of δ. From (4.2), (4.3) and (4.4) it follows that

(4.5) $$\frac{1}{n}\,| \mathbf{l}'(x, \hat{\theta})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \,| < (\kappa_6\delta + \kappa_7\delta^2)s^3\kappa_2 \| \theta - \hat{\theta} \|^2.$$

We now consider the second term of (4.1). By expanding the elements of $\mathbf{M}_{x,\theta^*}$ about θ_0 we find that

$$\frac{1}{n}\mathbf{M}_{x,\theta^*} = \frac{1}{n}\mathbf{M}_{x,\theta_0} + \mathbf{m}^*_{x,\theta^*},$$

where, as is easily shown using $\mathfrak{X}4$, the moduli of the elements of the matrix $\mathbf{m}^*_{x,\theta^*}$ are less than $2s\kappa_1\delta$. Also by $\mathfrak{X}3$,

$$\frac{1}{n}\mathbf{M}_{x,\theta_0} = -\mathbf{B}_{\theta_0} + \delta\mathbf{m}_{x,\theta_0},$$

and so

$$\frac{1}{n}\mathbf{M}_{x,\theta^*} = -\mathbf{B}_{\theta_0} + \delta\mathbf{m},$$

say, where $\mathbf{m}$ is a matrix whose elements are bounded by a number independent of δ. Hence

(4.6) $$\frac{1}{2n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{M}_{x,\theta^*}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = -\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{B}_{\theta_0}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})$$

$$+ \frac{1}{2}\delta(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{m}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) < -\frac{1}{2}\mu_0 \| \theta - \hat{\theta} \|^2 + \kappa_8\delta \| \theta - \hat{\theta} \|^2,$$

since $\mathbf{B}_{\theta_0}$ is positive definite with minimum latent root μ_0, and the elements of $\mathbf{m}$ are bounded. Here κ_8 is a positive number depending only on the elements of $\mathbf{m}$. Using (4.5) and (4.6) in (4.1) we find that there exist positive numbers κ_9, κ_{10}, independent of δ, such that

$$\frac{1}{n}[L(x, \theta) - L(x, \hat{\theta})] < \left(-\frac{1}{2}\mu_0 + \kappa_9\delta + \kappa_{10}\delta^2\right) \| \theta - \hat{\theta} \|^2.$$

It follows that if δ is sufficiently small then $L(x, \theta) < L(x, \hat{\theta})$, i.e., $L(x, \hat{\theta})$ is a maximum value of $L(x, \theta)$ subject to $h(\theta) = 0$.

We have thus established the fact that, if the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied, there exists a consistent maximum likelihood estimator $\hat{\theta}$ of θ_0 satisfying the condition $h(\hat{\theta}) = 0$.

5. Asymptotic distributions. We return now to consideration of (3.1) and (3.2). We suppose that $x \in X_n$ and that $\hat{\theta}(x)$, $\hat{\lambda}(x)$ is a solution of these equations with $\hat{\theta}(x) \in U_\delta$, δ being small enough for such a solution to exist. Then, considering the equations from a slightly different viewpoint we have,

$$\text{(5.1)} \qquad \frac{1}{n}\mathbf{l}(x, \theta_0) - [\mathbf{B}_{\theta_0} + \hat{\mathbf{b}}(x)][\hat{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}_0] + [\mathbf{H}_{\theta_0} + \hat{\mathbf{h}}(x)]\frac{1}{n}\hat{\boldsymbol{\lambda}}(x) = \mathbf{0},$$

$$\text{(5.2)} \qquad [\mathbf{H}'_{\theta_0} + \hat{\mathbf{h}}^*(x)][\hat{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}_0] = \mathbf{0},$$

where $\hat{\mathbf{b}}(x)$, $\hat{\mathbf{h}}(x)$ and $\hat{\mathbf{h}}^*(x)$ are matrices whose elements tend to 0 as δ (and hence $\| \hat{\theta}(x) - \theta_0 \|$) $\to 0$. We now prove the following lemma.

LEMMA 3. *The partitioned matrix*

$$\begin{bmatrix} \mathbf{B}_{\theta_0} & -\mathbf{H}_{\theta_0} \\ -\mathbf{H}'_{\theta_0} & \mathbf{0} \end{bmatrix}$$

is non-singular.

PROOF. For brevity we omit the suffix θ_0. Then we wish to find a matrix

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix}$$

such that, in the usual notation,

$$\begin{bmatrix} \mathbf{B} & -\mathbf{H} \\ -\mathbf{H}' & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_r \end{bmatrix}$$

and this requires

$$\text{(5.3)} \qquad \mathbf{BP} - \mathbf{HQ}' = \mathbf{I}_s,$$

$$\text{(5.4)} \qquad \mathbf{BQ} - \mathbf{HR} = \mathbf{0},$$

$$\text{(5.5)} \qquad \mathbf{H}'\mathbf{P} = \mathbf{0},$$

$$\text{(5.6)} \qquad -\mathbf{H}'\mathbf{Q} = \mathbf{I}_r.$$

These equations are easily solved since $\mathbf{B}$ is positive definite and $\mathbf{H}$ is of rank

r so that $\mathbf{H}'\mathbf{B}^{-1}\mathbf{H}$ is non-singular. We obtain

$$\mathbf{R} = -(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1},$$
$$\mathbf{Q} = -\mathbf{B}\mathbf{H}(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1},$$
$$\mathbf{P} = \mathbf{B}^{-1}[\mathbf{I}_s - \mathbf{H}(\mathbf{H}'\mathbf{B}^{-1}\mathbf{H})^{-1}\mathbf{H}'\mathbf{B}^{-1}].$$

We note at this stage, though we do not require this result immediately, that the matrix $\mathbf{P}$ has rank $s - r$. For, from (5.5) since rank $(\mathbf{H}') = r$, rank $(\mathbf{P}) \leqq s - r$. While from (5.3) we have $s =$ rank $(\mathbf{P} - \mathbf{H}\mathbf{Q}') \leqq$ rank $(\mathbf{P})$ + rank $(\mathbf{H}\mathbf{Q}') \leqq$ rank $(\mathbf{P}) + r$, and so rank $(\mathbf{P}) \geqq s - r$.

We return now to equations (5.1) and (5.2). If δ is sufficiently small then the matrix

$$\begin{bmatrix} \mathbf{B}_{\theta_0} + \hat{\mathbf{b}}(x) & -[\mathbf{H}_{\theta_0} + \hat{\mathbf{h}}(x)] \\ -[\mathbf{H}'_{\theta_0} + \hat{\mathbf{h}}^*(x)] & \mathbf{0} \end{bmatrix}.$$

also will be non-singular and we will write

$$\begin{bmatrix} \mathbf{B}_{\theta_0} + \hat{\mathbf{b}}(x) & -[\mathbf{H}_{\theta_0} + \hat{\mathbf{h}}(x)] \\ -[\mathbf{H}'_{\theta_0} + \hat{\mathbf{h}}^*(x)] & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \hat{\mathbf{P}}(x) & \hat{\mathbf{Q}}_1(x) \\ \hat{\mathbf{Q}}_2(x) & \hat{\mathbf{R}}(x) \end{bmatrix}.$$

Hence, from (5.1) and (5.2), for sufficiently small δ, if $x \varepsilon X_n$, we have

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}(x) - \boldsymbol{\theta}_0 \\ \frac{1}{n}\hat{\boldsymbol{\lambda}}(x) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{P}}(x) & \hat{\mathbf{Q}}_1(x) \\ \hat{\mathbf{Q}}_2(x) & \hat{\mathbf{R}}(x) \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta_0) \\ \mathbf{0} \end{bmatrix}. \tag{5.7}$$

If the functions $\hat{\theta}$ and $\hat{\lambda}$ were defined for the whole of $\mathcal{R}^n$ we could now discuss immediately the asymptotic distribution of these functions. However this is not necessarily so, and we go through the formality of extending the definition of these functions to the whole of $\mathcal{R}^n$. We will then show that the random variables thus defined are asymptotically normally distributed and, in this sense, we may say that a consistent maximum likelihood estimator $\hat{\theta}$ of θ_0 is asymptotically normally distributed.

We let (δ_m), (ϵ_m) be decreasing sequences of positive numbers, such that $\epsilon_1 < 1$, $\delta_1 < \mu_0/\kappa_3$ (see Theorem 1), and $\delta_m \to 0$ and $\epsilon_m \to 0$ as $m \to \infty$. We then define an increasing sequence (n_m) of integers such that, if $n \geqq n_m$, there exists a set in $\mathcal{R}^n$ with the properties $\mathfrak{X}1$ to $\mathfrak{X}4$ for $\epsilon = \epsilon_m$ and $\delta = \delta_m$. For $m = 1, 2, \cdots$, if $n_m \leqq n < n_{m+1}$ we choose a set X_n with the properties $\mathfrak{X}1$ to $\mathfrak{X}4$ for $\epsilon = \epsilon_m$ and $\delta = \delta_m$. Hence Pr $\{X_n\} \to 1$ as $n \to \infty$ and if $n_m \leqq n < n_{m+1}$ and $x \varepsilon X_n$, the likelihood equations (2.1) and (2.2) have a solution $\hat{\theta}_n(x)$, $\hat{\lambda}_n(x)$ such that $\| \hat{\theta}_n(x) - \theta_0 \| < \delta_m$. Moreover for sufficiently large m, $\hat{\theta}_n(x)$ is a *maximum* likelihood estimate of θ_0, by §4. We now extend the definition of $\hat{\theta}_n$ and $\hat{\lambda}_n$ to $\mathcal{R}^n$ by letting

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_n(x) - \boldsymbol{\theta}_0 \\ \frac{1}{n}\hat{\boldsymbol{\lambda}}_n(x) \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta_0) \\ \mathbf{0} \end{bmatrix}, \quad \text{if} \quad x \not\varepsilon X_n.$$

We have thus defined sequences $(\hat{\theta}_n)$, $(\hat{\lambda}_n)$, $n = n_m, n_{m+1}, \cdots$ of random variables which have the property that θ_n converges in probability to θ_0 and with probability tending to 1 as $n \to \infty$, $\hat{\theta}_n$, $\hat{\lambda}_n$ satisfy the likelihood equations (2.1) and (2.2).

THEOREM 2. *The random variables* $n^{1/2}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ *are asymptotically jointly normally distributed with variance-covariance matrix*

$$\begin{bmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{0} & -\mathbf{R} \end{bmatrix}.$$

PROOF. If $x \in X_n$, we define $\hat{\mathbf{P}}(x) = \mathbf{P}$, $\hat{\mathbf{Q}}_1(x) = \mathbf{Q}$, $\hat{\mathbf{Q}}_2(x) = \mathbf{Q}'$ and $\hat{\mathbf{R}}(x) = \mathbf{R}$. Then for sufficiently large n, by (5.7) we may write

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \\ \dfrac{1}{\sqrt{n}}\hat{\boldsymbol{\lambda}}_n \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{P}} & \hat{\mathbf{Q}}_1 \\ \hat{\mathbf{Q}}_2 & \hat{\mathbf{R}} \end{bmatrix} \begin{bmatrix} \dfrac{1}{\sqrt{n}}\mathbf{l}(\cdot, \theta_0) \\ \mathbf{0} \end{bmatrix}.$$

The elements of the matrix

$$\begin{bmatrix} \hat{\mathbf{P}} & \hat{\mathbf{Q}}_1 \\ \hat{\mathbf{Q}}_2 & \hat{\mathbf{R}} \end{bmatrix}$$

are random variables which converge in probability to the corresponding elements of the matrix

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix},$$

since in (5.1) and (5.2) $\hat{\mathbf{b}}$, $\hat{\mathbf{h}}$ and $\hat{\mathbf{h}}^*$ tend to $\mathbf{0}$ as $\delta \to 0$. Also the s-dimensional random variable $n^{-1/2}\ell(\cdot, \theta_0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix $\mathbf{B}_{\theta_0}$ (Cramér [1]), and the $(s + r)$-dimensional random variable $(n^{-1/2}\ell(\cdot, \theta_0), 0)$ is asymptotically normally distributed with zero mean and variance-covariance matrix

$$\begin{bmatrix} \mathbf{B}_{\theta_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

It follows by an extension, to a multi-dimensional random variable, of a theorem of Cramér [2], that $\sqrt{n}(\hat{\theta}_n - \theta_0)$, $n^{-1/2}\hat{\lambda}_n$ are jointly asymptotically normally distributed with zero mean and variance-covariance matrix.

$$\begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{\theta_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{P}\mathbf{B}_{\theta_0}\mathbf{P} & \mathbf{P}\mathbf{B}_{\theta_0}\mathbf{Q} \\ \mathbf{Q}'\mathbf{B}_{\theta_0}\mathbf{P} & \mathbf{Q}'\mathbf{B}_{\theta_0}\mathbf{Q} \end{bmatrix}.$$

(We omit details of the proof of this extension though this result, in contrast to Cramér's result for real-valued random variables, is best obtained by considering characteristic functions). Now from (5.3), $\mathbf{P}\mathbf{B}_{\theta_0}\mathbf{P} - \mathbf{P}\mathbf{H}_{\theta_0}\mathbf{Q}' = \mathbf{P}$. Since $\mathbf{P}$ is symmetric, $\mathbf{P}\mathbf{H}_{\theta_0} = \mathbf{P}'\mathbf{H}_{\theta_0} = \mathbf{0}$ by (5.5). Hence $\mathbf{P}\mathbf{B}_{\theta_0}\mathbf{P} = \mathbf{P}$. Similarly $\mathbf{P}\mathbf{B}_{\theta_0}\mathbf{Q} = \mathbf{0}$ and $\mathbf{Q}'\mathbf{B}_{\theta_0}\mathbf{Q} = -\mathbf{R}$.

This completes the proof of the Theorem. We note, however, that, as might be expected, the asymptotic normal distribution of the s-dimensional random variable $n^{1/2}(\hat{\theta}_n - \theta_0)$ is improper, being by the note in Lemma 3 of rank $s - r$.

6. Numerical solution of likelihood equations. In this section we will discuss an iterative procedure for solving (2.1) and (2.2) numerically, which yields an estimate of the matrices **P** and **R**.

In any practical situation we do not know θ_0, and the only way in which we can verify that the conditions $\mathfrak{F}$ and $\mathfrak{H}$ are satisfied is to find that, for every θ belonging to some set U, *in which we know* θ_0 *lies*, the following conditions $\mathfrak{F}'$, $\mathfrak{H}'$ are satisfied.

$\mathfrak{F}'1$, $\mathfrak{F}'2$. For every $\theta \varepsilon U$, $\mathfrak{F}1$ and $\mathfrak{F}2$ are satisfied.

$\mathfrak{F}'3$ For every $\theta \varepsilon U$ and $i, j, k = 1, 2, \cdots, s$,

$$\left| \frac{\partial^3 \log f(t, \theta)}{\partial\theta_i \partial\theta_j \partial\theta_k} \right| < F_3(t)$$

and

$$\int_{-\infty}^{\infty} F_3(t) f(t, \theta)\, dt \leqq \kappa_1',$$

a finite number.

$\mathfrak{F}'4$. For every $\theta \varepsilon U$,

$$b_{ij} = \int_{-\infty}^{\infty} \frac{\partial \log f(t, \theta)}{\partial \theta_i} \frac{\partial \log f(t, \theta)}{\partial \theta_j} f(t, \theta)\, dt,$$

$i, j = 1, 2, \cdots, s$, are finite, the matrix $\mathbf{B}_\theta = (b_{ij}(\theta))$ is positive definite and, if μ_θ is the minimum latent root of $\mathbf{B}_\theta$, then $\mu_\theta \geqq \mu_0'$ where μ_0' is a given number greater than 0.

$\mathfrak{H}'1$, $\mathfrak{H}'2$. For every $\theta \varepsilon U$, $\mathfrak{H}1$ and $\mathfrak{H}2$ are satisfied.

$\mathfrak{H}'3$ For every $\theta \varepsilon U$, $\mathbf{H}_\theta$ is of rank r.

The conditions $\mathfrak{F}'$ are a straightforward generalization of Cramér's conditions [2].

We will now assume that the conditions $\mathfrak{F}'$ and $\mathfrak{H}'$ are satisfied, that x is such that the likelihood equations (2.1) and (2.2) have a solution $\hat{\theta}(x)$, $\hat{\lambda}(x)$ and that $\theta^{(1)}$ is an initial approximation to $\hat{\theta}(x)$ such that $\| \theta^{(1)} - \hat{\theta}(x) \|$ is small. Then to a first order of approximation

$$\mathbf{l}(x, \hat{\theta}) = \mathbf{l}(x, \theta^{(1)}) + \mathbf{M}_{x,\theta^{(1)}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^1),$$

$$\mathbf{h}(\hat{\theta}) = \mathbf{h}(\theta^{(1)}) + \mathbf{H}_{\theta^{(1)}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(1)}).$$

Also if n is large, $(1/n)\hat{\boldsymbol{\lambda}}(x)$ is near **0** for "most" x. We assume that x is a point for which $(1/n)\boldsymbol{\lambda}(x)$ is near **0**. Then we also have to a first order of approximation

$$\mathbf{H}_{\hat{\theta}} \frac{1}{n} \hat{\boldsymbol{\lambda}} = \mathbf{H}_{\theta^{(1)}} \frac{1}{n} \hat{\boldsymbol{\lambda}}.$$

Since $\hat{\theta}(x)$, $\hat{\lambda}(x)$ satisfy (2.1) and (2.2) then, approximately, we have

$$\begin{bmatrix} -\frac{1}{n}\mathbf{M}_{x,\theta^{(1)}} & -\mathbf{H}_{\theta^{(1)}} \\ -\mathbf{H}'_{\theta^{(1)}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(1)} \\ \frac{1}{n}\hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(1)}) \\ \mathbf{h}(\theta^{(1)}) \end{bmatrix}. \tag{6.1}$$

The normal situation, if n is large, is that $\hat{\theta}(x)$ is near θ_0. Consequently since $\theta^{(1)}$ is near $\hat{\theta}(x)$ the matrix $-(1/n)\mathbf{M}_{x,\theta^{(1)}}$ approximates $-(1/n)\mathbf{M}_{x,\theta_0}$ which in turn approximates $\mathbf{B}_{\theta_0}$. Then $\mathbf{B}_{\theta^{(1)}}$ approximates $\mathbf{B}_{\theta_0}$ and we propose to replace $-(1/n)\mathbf{M}_{x,\theta^{(1)}}$ in (6.1) by $\mathbf{B}_{\theta^{(1)}}$, and to obtain a correction to $\theta^{(1)}$, and an initial approximation to $(1/n)\hat{\lambda}$, by solving the equation

$$\begin{bmatrix} \mathbf{B}_{\theta^{(1)}} & -\mathbf{H}_{\theta^{(1)}} \\ -\mathbf{H}'_{\theta^{(1)}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{(1)} \\ \frac{1}{n}\hat{\boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(1)}) \\ \mathbf{h}(\theta^{(1)}) \end{bmatrix}. \tag{6.2}$$

The idea of replacing $-(1/n)\mathbf{M}_{x,\theta^{(1)}}$ by $\mathbf{B}_{\theta^{(1)}}$ is not original though the authors do not know where it originated.

Because of $\mathcal{F}'4$, $\mathcal{H}'3$, by Lemma 3, the matrix

$$\begin{bmatrix} \mathbf{B}_{\theta^{(1)}} & -\mathbf{H}_{\theta^{(1)}} \\ -\mathbf{H}'_{\theta^{(1)}} & \mathbf{0} \end{bmatrix}$$

is non-singular and we will denote its inverse by

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}'_1 & \mathbf{R}_1 \end{bmatrix}.$$

We define $\theta^{(2)}$, $\lambda^{(2)}$ by

$$\begin{bmatrix} \boldsymbol{\theta}^{(2)} \\ \frac{1}{n}\boldsymbol{\lambda}^{(2)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(1)} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}'_1 & \mathbf{R}_1 \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(1)}) \\ \mathbf{h}(\theta^{(1)}) \end{bmatrix}$$

and, more generally, $\theta^{(r)}$, $\lambda^{(r)}$ by (with the obvious definition of $\mathbf{P}_{r-1}$, $\mathbf{Q}_{r-1}$ and $\mathbf{R}_{r-1}$),

$$\begin{bmatrix} \boldsymbol{\theta}^{(r)} \\ \frac{1}{n}\boldsymbol{\lambda}^{(r)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\theta}^{(r-1)} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_{r-1} & \mathbf{Q}_{r-1} \\ \mathbf{Q}'_{r-1} & \mathbf{R}_{r-1} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \theta^{(r-1)}) \\ \mathbf{h}(\theta^{(r-1)}) \end{bmatrix}.$$

If the sequences $(\theta^{(r)})$, $(\lambda^{(r)})$ converge then they converge to a solution of the likelihood equations, as is easily verified. We do not attempt to give rigorous conditions under which these sequences do converge. However the fact that we may expect them to converge in most practical situations follows from the heuristic argument leading to (6.2).

We have thus established an iterative procedure for solving the likelihood equations. The heaviest part of the computation involved in this method is the inversion of a matrix and computation will normally be reduced by considering the sequences $(\hat{\theta}^{(r)})$, $(\hat{\lambda}^{(r)})$ defined by

$$\begin{bmatrix} \hat{\theta}^{(r)} \\ \frac{1}{n}\hat{\lambda}^{(r)} \end{bmatrix} = \begin{bmatrix} \hat{\theta}^{(r-1)} \\ \frac{1}{n}\hat{\lambda}^{(r-1)} \end{bmatrix} + \begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_1' & \mathbf{R}_1 \end{bmatrix} \begin{bmatrix} \frac{1}{n}\mathbf{l}(x, \hat{\theta}^{(r-1)}) + \mathbf{H}_{\hat{\theta}^{(r-1)}} \frac{1}{n}\hat{\lambda}^{(r-1)} \\ \mathbf{h}(\hat{\theta}^{(r-1)}) \end{bmatrix}$$

$r = 1, 2, \cdots$, where $\hat{\theta}^{(0)} = \theta^{(2)}$ and $\hat{\lambda}^{(0)} = \lambda^{(2)}$. Again if these sequences converge, they converge to a solution of the likelihood equations since

$$\begin{bmatrix} \mathbf{P}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_1' & \mathbf{R}_1 \end{bmatrix}$$

is non-singular. And again we do not attempt to give conditions under which they do converge. The main justifications we put forward for this computational procedure are

(i) the similarity between this method and Newton's method, and

(ii) the fact that similar modifications of Newton's method have been used successfully elsewhere, for example in probit analysis [3]. The main advantage of this method of solving the likelihood equations is that it involves inversion of only one matrix.

7. Tests of the model. In a situation such as is outlined in §1 two natural questions arise in practice regarding the adequacy of the model introduced to describe an experimental situation.

(i) Does the true parameter point θ_0 satisfy the condition $h(\theta_0) = 0$?

(ii) Is the true parameter point some hypothetical point θ^* such that

$$h(\theta^*) = 0?$$

And this is the natural order for these questions since the second would be asked only if the first were answered in the affirmative. We now propose a procedure for answering these questions in this order.

(i) The most natural approach to the first question would be as follows. We would calculate an unrestrained maximum likelihood estimate $\hat{\theta}_u(x)$ of θ_0, and for $\hat{\theta}_u(x)$ we would have $\ell(x, \hat{\theta}_u(x)) = 0$. If $h(\hat{\theta}_u(x))$ were in some sense "near enough" $0 \varepsilon \mathcal{R}^r$ then we would decide that in fact $h(\theta_0) = 0$. Dually, we might calculate a maximum likelihood estimate $\hat{\theta}(x)$ subject to the restraint

$$h(\hat{\theta}(x)) = 0$$

and then decide that $h(\theta_0) = 0$ if $\ell(x, \hat{\theta}(x))$ were "near enough" $0 \varepsilon \mathcal{R}^s$. And the test we propose is based on the second possibility. We note that, by (2.1),

$$\mathbf{H}_{\hat{\theta}}\hat{\lambda}(x) = -\mathbf{l}(x, \hat{\theta}(x))$$

and it seems reasonable to decide that $h(\theta_0) = 0$ if $\hat{\lambda}(x)$ is in some sense 'near enough' $0 \varepsilon \mathcal{R}^r$.

We have seen in Theorem 2 that when $h(\theta_0) = 0$, $n^{-1/2}\dot{\lambda}$ is normally distributed asymptotically with variance-covariance matrix $-\mathbf{R}$, which is of rank r. Consequently $-(1/n)\hat{\lambda}'\mathbf{R}^{-1}\hat{\lambda}$ is asymptotically distributed as χ^2 with r degrees of freedom, when $h(\theta_0) = 0$, and, in obvious notation, $-(1/n)\hat{\lambda}'\mathbf{R}_{\hat{\theta}}^{-1}\hat{\lambda}$ also is approximately, for large n, distributed as χ^2 with r degrees of freedom. We propose to choose as a region of acceptance of the hypothesis that $h(\theta_0) = 0$ the set of x for which

$$-\frac{1}{n}\hat{\lambda}'(x)\mathbf{R}_{\hat{\theta}(x)}^{-1}\hat{\lambda}(x) \leqq k,$$

where k is determined by

$$\Pr\{\chi^2_{[r]} \leqq k\} = 0.95.$$

This gives a test of size 95% of the hypothesis that $h(\theta_0) = 0$.

(ii) The natural corollary of using the asymptotic distribution of $\hat{\lambda}$ in this way is to use the asymptotic distribution of $\hat{\theta}$ as established in Theorem 2 to answer the second question. If $\theta^* = \theta_0$ then $n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)'\mathbf{B}_{\theta^*}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ is approximately distributed as χ^2 with $s - r$ degrees of freedom if n is large. This is easily established by noting that a consequence of equations (5.3)–(5.6) is that $\mathbf{B}^{-1} = \mathbf{PBP} - \mathbf{QR}^{-1}\mathbf{Q}'$, and hence that

$$\frac{1}{n}\mathbf{l}'\mathbf{B}^{-1}\mathbf{l} = n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\mathbf{B}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \frac{1}{n}\hat{\lambda}'\mathbf{R}^{-1}\hat{\lambda}.$$

We use this fact as in the previous paragraph to establish a region of acceptance of the hypothesis that the true parameter point is θ^*.

Here no attempt is made to justify this test on other than an intuitive basis. Since the Lagrangian multiplier test seems to be of wide applicability and of considerable importance in practical statistics, it will be fully discussed both from the theoretical and practical points of view in subsequent papers.

REFERENCES

[1] H. Cramér, "Random variables and probability distributions," *Cambridge University Press*, 1937.
[2] H. Cramér, "Mathematical methods of statistics," *Princeton University Press*, 1949.
[3] D. J. Finney, "Probit analysis," *Cambridge University Press*, 1947.
[4] S. Lefschetz, "Introduction to topology," *Princeton University Press*, 1949.