The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1542–1556

ESTIMATION OF PARAMETERS IN A TRANSIENT MARKOV CHAIN ARISING IN A RELIABILITY GROWTH MODEL[1]

BY M. DUBMAN AND B. SHERMAN

Rocketdyne, North American Rockwell Corporation

1. Introduction. We consider the following reliability growth model. Initially a device has probability p of failure. We subject it to a sequence of trials, making no changes if there is no failure. If there is a failure on any trial, then changes are made in the device which cause the probability of failure on the next trial to be reduced by the factor β, where $0 < \beta < 1$. Thus if there have been k failures the probability of failure on the next trial is $p\beta^k$. Let y_i be the number of failures up to and including trial i, with $y_0 = 0$. Then y_i, $i \geqq 0$, is a Markov chain, which may be regarded as a random walk on the nonnegative integers with the transition probabilities

$$(1.1)\quad P(y_{i+1} = k + 1 \mid y_i = k) = p\beta^k, \qquad P(y_{i+1} = k \mid y_i = k) = 1 - p\beta^k.$$

In this paper we prove that the likelihood equations for the Markov chain y_i, $i \geqq 0$ have solutions, $\hat{p}$ and $\hat{\beta}$, which converge in probability to the true parameter values, p_0 and β_0, and which are asymptotically jointly normally distributed.

The Markov chain (1.1) is clearly transient, and we may regard this work as an example of the theory of estimation in chains of that type. Billingsley [2] has developed a general theory of estimation in Markov processes but his results do not apply here since his basic assumption is that the Markov process possesses a unique stationary distribution.

An estimation problem for a sequence of independent but not identically distributed random variables ξ_k which is closely related to the estimation problem for the Markov chain (1.1) is arrived at by defining ξ_k, $k \geqq 0$, to be the number of times state k is occupied in the infinite sequence $y_0, y_1, y_2, \cdots$. Clearly, the random variables ξ_k so defined are independent and have the geometric distributions

$$(1.2)\qquad P(\xi_k = x) = p\beta^k(1 - p\beta^k)^{x-1}, \qquad x = 1, 2, \cdots.$$

The sequence ξ_k, $k \geqq 0$ provides an alternative description of the reliability growth model considered here, in that for any $k \geqq 1$ the partial sum $\xi_0 + \xi_1 + \cdots + \xi_{k-1}$ represents the first trial (i) for which the accumulated number of failures (y_i) equals k. In this paper we first consider the estimation problem for a sample ξ_k, $k = 0, 1, \cdots, n - 1$ and prove that the likelihood

Received 11 February 1966; revised 7 March 1969.

[1] This research was supported by the United States Air Force through the Aerospace Research Laboratories, Office of Aerospace Research, under Contract No. AF33(615)-2818, Project 7071.

equations have roots, $\hat{p}$ and $\hat{\beta}$, which are jointly consistent and asymptotically normally distributed. We then prove the corresponding results for a Markov chain sample y_i, $i = 1, 2, \cdots, N$ by showing that if we take n to be the integral part of $-\log N/\log \beta_0$, then the likelihood equations for the Markov chain sample, when suitably normalized, differ from the likelihood equations for the sample ξ_k, $k = 0, 1, \cdots, n - 1$ by terms which are asymptotically negligible (in probability).

The method of proof used for the independent sequence ξ_k, $k \geqq 0$ is an extension of the proof of Aitchison and Silvey [1] who treat the situation of independent and identically distributed random variables having a density $f(x, \theta)$ where θ is a vector parameter whose components are subject to restraints. For the sequence ξ_k, $k \geqq 0$ we are concerned with a parameter $\theta = (p, \beta)$ which has two components; there are no restraints (which simplifies matters) but the observables ξ_k are not identically distributed. Hoadley [8] has developed a general theory of maximum likelihood estimation of vector parameters in situations where the observations are independent but not identically distributed. His approach is similar to that taken by Wald [14] and Wolfowitz [15] who treat the independent and identically distributed cases. In [8], [14] and [15] the authors prove that the maximum likelihood estimator of a vector parameter converges in probability (or in Wald's treatment, converges with probability one) to the true parameter value. For the sequence ξ_k, $k \geqq 0$, however, the assumptions of Hoadley do not appear to be satisfied.

For a discussion of other reliability growth models we refer to [3], [9] and [13].

We wish to thank the associate editor for his very detailed and extensive critiques of earlier versions of this paper. His efforts amounted very nearly to a collaboration.

2. Summary. The following notation is used in the paper. Column vectors and matrices are designated by boldface letters. If $\mathbf{x}$ is a vector with transpose $\mathbf{x}' = (x_1, x_2, \cdots, x_d)$ and $\mathbf{A} = (a_{ij})$ is a $d \times d$ matrix, then $\|\mathbf{x}\| \equiv [\sum_{i=1}^d x_i^2]^{\frac{1}{2}}$ while $\|\mathbf{A}\|$ stands for the Euclidean matrix norm $\|\mathbf{A}\| \equiv [\sum_{i,j=1}^d a_{ij}^2]^{\frac{1}{2}}$. We use the notation o_p and O_p, discussed by Mann and Wald [11], Chernoff [4] and Pratt [12]. For example, if $\mathbf{x}_n$, $n \geqq 1$ are random vectors, $\mathbf{x}_n = o_p(1)$ means that $\|\mathbf{x}_n\|$ converges in probability to 0 as $n \to \infty$, while $\mathbf{x}_n = O_p(1)$ means that the sequence $\|\mathbf{x}_n\|$, $n \geqq 1$ is stochastically bounded (see Feller [7]). That is, $\mathbf{x}_n = O_p(1)$ if for any given $\epsilon > 0$ there exist a constant $K_\epsilon > 0$ and a positive integer n_ϵ such that $P(\|\mathbf{x}_n\| \geqq K_\epsilon) \leqq 1 - \epsilon$ for all $n \geqq n_\epsilon$.

We consider now the maximum likelihood estimation of the parameters p and β for the sequence of independent random variables ξ_k, $k \geqq 0$ where ξ_k has the geometric distribution (1.2). Let θ denote the two-dimensional parameter $\theta = (p, \beta)$ and let ξ (without subscripts) stand for the n-dimensional random point $\xi = (\xi_0, \xi_1, \cdots, \xi_{n-1})$. The likelihood function of ξ is denoted by $L(\xi, \theta)$ and its logarithm is given by

$$\log L(\xi, \theta) = \sum_{k=0}^{n-1} [\log (p\beta^k) + (\xi_k - 1) \log (1 - p\beta^k)]. \tag{2.1}$$

Let p_0, β_0, and θ_0 denote the true values of p, β, and θ. In order to avoid consideration of trivial special cases, we assume

$$0 < p_0 < 1, \qquad 0 < \beta_0 < 1. \tag{2.2}$$

In Section 3 we prove the following theorem concerning the existence and asymptotic properties of solutions of the likelihood equations

$$\partial \log L(\xi, \theta)/\partial p = 0, \qquad \partial \log L(\xi, \theta)/\partial \beta = 0. \tag{2.3}$$

THEOREM 1. *Subject to* (2.2), *there exist random variables* $\hat{\theta}_n = (\hat{p}_n, \hat{\beta}_n)$ *such that*

(a) *With probability tending to* 1 *as* $n \to \infty$, $\hat{\theta}_n$ *satisfies the likelihood equations* (2.3).

(b) $\hat{\theta}_n - \theta_0 = o_p(1)$.

(c) *The random variables* $n^{\frac{1}{2}}(\hat{p}_n - p_0)$ *and* $n^{\frac{3}{2}}(\hat{\beta}_n - \beta_0)$ *are asymptotically jointly normally distributed with zero means and variance-covariance matrix*

$$\Sigma \equiv \begin{bmatrix} 4p_0^2 & -6p_0\beta_0 \\ -6p_0\beta_0 & 12\beta_0^2 \end{bmatrix} \tag{2.4}$$

It is of interest to note that the $\hat{\theta}_n$ can be chosen so as to satisfy the additional property.

(d) *With probability tending to* 1 *as* $n \to \infty$, $\hat{\theta}_n$ *maximizes* $L(\xi, \theta)$ *for* θ *ranging over a certain neighborhood* U_n *of* θ_0 (*where* $U_n \to \{\theta_0\}$ *as* $n \to \infty$). A proof of (d), which is omitted from the present paper, can easily be constructed by modifying appropriately arguments of [1, Section 4].

The method of proof used in Section 3 for Theorem 1 is a generalization of the Taylor expansion procedure of Aitchison and Silvey [1]. This procedure, as applied to the problem of proving the existence of consistent solutions of the likelihood equations for estimating a vector parameter $\boldsymbol{\theta}' = (\theta_1, \theta_2 \cdots, \theta_d)$, is in turn a generalization of Cramér's proof [5, Section 33.3] for a single real-valued parameter and may be described briefly as follows. Starting from the likelihood function $L(x, \theta)$ based on a sample $x = (x_1, x_2, \cdots x_n)$ of independent and identically distributed observations, one expands $\partial \log L(x, \theta)/\partial\theta_m$, $m = 1, 2, \cdots, d$ about the true parameter point θ_0 up to quadratic terms. After normalizing the expanded equations (multiplying by $1/n$), one shows that as $n \to \infty$ the constant terms converge in probability to 0, the coefficients of the linear terms (when arranged in a matrix) converge in probability to a symmetric negative definite matrix and the coefficients of the quadratic terms are stochastically bounded. It follows that for θ belonging to a sufficiently small neighborhood $U_\delta = \{\theta: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leqq \delta\}$ of the true parameter point θ_0 and for n sufficiently large one may apply a result (Lemma 2 of [1]) equivalent to the Brouwer fixed point theorem to conclude that the likelihood equations will (with probability arbitrarily close to 1) have a solution $\hat{\theta}$ belonging to the neighborhood U_δ.

In Section 3 of this paper we adapt the Aitchison and Silvey procedure to account for the lack of identical distribution of the random variables ξ_k, $k \geqq 0$ in the following way. The likelihood equations (2.3) are expanded about the true parameter point θ_0 up to quadratic terms but now we normalize the expanded equations by n^{-1} and n^{-2}, respectively. In addition, the fixed neighborhood U_δ is replaced by neighborhoods $U(\delta, n)$ which depend on n and which we define as follows. Given n and $\tau = (\tau_1, \tau_2)$ where τ_1 and τ_2 are real, let

$$(2.5a) \qquad p_\tau = p_0(1 + \tau_1), \qquad \beta_\tau = \beta_0(1 + \tau_2 n^{-1}), \qquad \theta_\tau = (p_\tau, \beta_\tau).$$

Then we take $U(\delta, n)$ to be the elliptical neighborhood

$$(2.5b) \qquad U(\delta, n) = \{\theta_\tau : \|\boldsymbol{\tau}\| \leqq \delta\}.$$

It is shown that when the expanded, normalized likelihood equations are considered for points belonging to $U(\delta, n)$ and as functions of τ_1 and τ_2, the coefficients with respect to τ_1 and τ_2 behave asymptotically as do the corresponding coefficients in the case of independent and identically distributed observations. We then apply Lemma 2 of [1] to establish the existence of solutions $\hat{\theta}$ of the likelihood equations which belong to the neighborhoods $U(\delta, n)$ for sufficiently small δ and sufficiently large n.

The asymptotic normality (part (c), Theorem 1) of the solutions of the likelihood equations (2.3) is proved in a similar manner by making the appropriate modifications in the arguments of Section 5 of [1].

We consider now the problem of jointly estimating the parameters p and β for the Markov chain y_i ($i \geqq 0, y_0 \equiv 0$) with the transition probabilities (1.1). The likelihood function of $y \equiv (y_1, y_2, \cdots, y_N)$ is denoted by $L^*(y, \theta)$ and its logarithm is given by

$$(2.6) \quad \log L^*(y, \theta) = \textstyle\sum_{i=0}^{N-1} [(y_{i+1} - y_i) \log (p\beta^{y_i}) + (1 - y_{i+1} + y_i) \log (1 - p\beta^{y_i})].$$

In Section 4 we prove the following theorem concerning the existence and asymptotic properties of solutions of the likelihood equations[2]

$$(2.7) \qquad \partial \log L^*(y, \theta)/\partial p = 0, \qquad \partial \log L^*(y, \theta)/\partial \beta = 0.$$

THEOREM 2. *For given N let $n = n(N)$ denote the integral part of $(-\log \beta_0)^{-1}(\log N)$. Then, subject to* (2.2), *there exist random variables $\hat{\theta}_N = (\hat{p}_N, \hat{\beta}_N)$ such that*

(a) *With probability tending to* 1 *as $N \to \infty$, $\hat{\theta}_N$ satisfies the likelihood equations* (2.7).

(b) $\hat{\theta}_N - \theta_0 = o_p(1)$.

(c) *The random variables $n^{\frac{1}{2}}(\hat{p}_N - p_0)$ and $n^{\frac{3}{2}}(\hat{\beta} - \beta_0)$ are asymptotically jointly normally distributed with zero means and variance-covariance matrix Σ given by* (2.4).

The method of proof used for Theorem 2 is as follows. Starting from the

[2] A result similar to (d) of Theorem 1 above also holds for the Markov chain y_i, $i \geq 0$.

Markov chain y_i with transition probabilities (1.1) we *define* ξ_k for $k = 0, 1, 2, \cdots$ to be the number of times state k is occupied in the infinite sequence $y_0, y_1, y_2, \cdots$. It is clear that the random variables ξ_k so defined are independent and have the geometric distributions (1.2). The independence of the ξ_k, $k \geqq 0$ is a consequence of the basic properties of the Markov chain that $y_0 \equiv 0$ and $y_{i+1} - y_i = 0$ or 1. It is also easy to see that (cf. (2.1))

$$(2.8)\quad \log L^*(y, \theta) = \sum_{j=0}^{y_N-1} [\log (p\beta^j) + (\xi_j - 1) \log (1 - p\beta^j) + (N - \sum_{j=0}^{y_N-1} \xi_j) \log (1 - p\beta^{y_N}).$$

In Section 4 we prove that for $N \to \infty$

$$(2.9)\quad y_N = -\log N/\log \beta_0 + O_p(1).$$

Let n denote the integral part of $-\log N/\log \beta_0$. Using (2.9), we subsequently show that

$$(2.10)\quad \sup_{\theta \varepsilon U(\delta,n)} |n^{-1}\partial/\partial p[\log L^*(y, \theta) - \log L(\xi, \theta)]| = o_p(1),$$
$$\sup_{\theta \varepsilon U(\delta,n)} |n^{-2}\partial/\partial \beta[\log L^*(y, \theta) - \log L(\xi, \theta)]| = o_p(1).$$

The result (2.10) permits us to establish Theorem 2 by making simple modifications in the arguments used in Section 3 to prove Theorem 1.

It is worthwhile noting that the existence of consistent solutions of the likelihood equations (2.7) may be proved without explicit reference to the ξ_k by means of a Taylor expansion procedure applied directly to the left sides of (2.7). This alternative proof is carried out in [6] and uses the following results. For $N \to \infty$,

$$(2.11)\quad E(y_N^m) = O(\log^m N), \qquad m = 1, 2, \cdots;$$

$$(2.12)\quad y_N(-\log \beta_0)/\log N = 1 + o_p(1).$$

The proof of (2.11) is straightforward. The result (2.12) follows from (2.9) or, alternatively, from the asymptotic results

$$(2.13)\quad E(y_N) = -\log N/\log \beta_0 + O(1),$$
$$E(y_N^2) = (\log N/\log \beta_0)^2 + O(\log N).$$

Proofs of (2.13) are given in [13] with the aid of the following explicit expressions for the first two moments of y_N. Let $\pi_0 = 1$ and for $k \geqq 1$ let

$$\pi_k = \prod_{j=1}^{k} (1 - \beta_0^j), \qquad \rho_k = \sum_{j=1}^{k} \beta_0^j(1 - \beta_0^j)^{-1},$$

where we assume $0 < \beta_0 < 1$. Then

$$E(y_N) = \sum_{k=1}^{N} (-1)^{k-1} \binom{N}{k} \pi_{k-1} p_0^k,$$
$$E(y_N^2) = E(y_N) + 2 \sum_{k=2}^{N} (-1)^k \binom{N}{k} \pi_{k-1} \rho_{k-1} p_0^k.$$

3. Proof of Theorem 1. Throughout this section we use $\theta_\tau = (p_\tau, \beta_\tau)$ to denote an arbitrary point of the neighborhood $U(\delta, n)$ defined by (2.5). Also we assume that $0 < \delta \leqq \delta_0$, where δ_0 is fixed and is sufficiently small ($\delta_0 \leqq \frac{1}{2} \min \{1, p_0^{-1} - 1, \beta_0^{-1} - 1\}$ will do) so that p_τ and β_τ are bounded away from 0 and from 1 for $\|\boldsymbol{\tau}\| \leqq \delta_0$ and $n \geqq 1$. This assumption is possible by virtue of the hypothesis (2.2). We now begin the proof of Theorem 1 with the following lemma which is comparable to Theorem 1 of [1] and which establishes the existence of solutions of the likelihood equations (2.3).

LEMMA 1. *Let ϵ be a given positive number less than* 1. *If $\delta > 0$ is sufficiently small and n sufficiently large, say $n \geqq n(\delta, \epsilon)$, then the likelihood equations* (2.3) *will, with probability exceeding $1 - \epsilon$, have a solution $\hat{\theta} = (\hat{p}, \hat{\beta})$ which belongs to the interior of the neighborhood $U(\delta, n)$ of θ_0.*

PROOF. It is convenient (in order to obtain symmetry in explicit expressions for Taylor expansion coefficients) to consider instead of (2.3) the equations

$$l_1(\xi, \theta) \equiv p[\partial \log L(\xi, \theta)/\partial p] = 0, \tag{3.1}$$

$$l_2(\xi, \theta) \equiv \beta[\partial \log L(\xi, \theta)/\partial \beta] = 0.$$

These have the same solutions as (2.3) in $U(\delta, n)$ since, by assumption, $p > 0$ and $\beta > 0$ if $\theta = (p, \beta) \varepsilon U(\delta, n)$. Following the Taylor expansion procedure discussed in the previous section, we expand the left sides of (3.1) about $\theta = \theta_0$ up to quadratic terms in p and β, and then we multiply the resulting equations by n^{-1} and n^{-2}, respectively. When this is done we are able to write (3.1), considered for points $\theta_\tau \varepsilon U(\delta, n)$, in the form

$$n^{-m} l_m(\xi, \theta_\tau) = a_m(\theta_0) + \tau_1 b_m(\theta_0) + \tau_2 b_{m+1}(\theta_0) + \tfrac{1}{2}\boldsymbol{\tau}' \mathbf{C}_m(\theta_\tau{}^m)\boldsymbol{\tau} = 0, \quad m = 1, 2. \tag{3.2}$$

The various coefficients that occur in (3.2) are identified as follows. For $m = 1, 2$

$$a_m(\theta) \equiv n^{-m} l_m(\xi, \theta), \tag{3.3}$$

where from the definitions of $l_1(\xi, \theta)$ and $l_2(\xi, \theta)$ and from (2.1) we have

$$l_m(\xi, \theta) = \textstyle\sum_{k=0}^{n-1} k^{m-1}(1 - \xi_k p\beta^k)(1 - p\beta^k)^{-1}. \tag{3.4}$$

Also

$$b_1(\theta) \equiv n^{-1} p[\partial l_1(\xi, \theta)/\partial p],$$

$$b_2(\theta) \equiv n^{-2}\beta[\partial l_1(\xi, \theta)/\partial \beta] = n^{-2} p[\partial l_2(\xi, \theta)/\partial p], \tag{3.5}$$

$$b_3(\theta) \equiv n^{-3}\beta[\partial l_2(\xi, \theta)/\partial \beta]$$

and from (3.4) we have, after simple computations, that for $m = 1, 2, 3$

$$b_m(\theta) = n^{-m} \textstyle\sum_{k=0}^{n-1} k^{m-1}(1 - \xi_k)p\beta^k(1 - p\beta^k)^{-2}. \tag{3.6}$$

The matrix $\mathbf{C}_m(\theta)$ is defined for arbitrary θ to be the 2×2 matrix

$$\mathbf{C}_m(\theta) \equiv (c_{m;ij}(\theta)) \tag{3.7}$$

whose components are given by

$$(3.8)\qquad \begin{aligned} c_{m;11}(\theta) &\equiv n^{-m}p_0^2[\partial^2 l_m(\xi,\theta)/\partial p^2],\\ c_{m;12}(\theta) &\equiv c_{m;21}(\theta) \equiv n^{-m-1}p_0\beta_0[\partial^2 l_m(\xi,\beta)/\partial p\partial\beta],\\ c_{m;22}(\theta) &\equiv n^{-m-2}\beta_0^2[\partial^2 l_m(\xi,\theta)/\partial\beta^2]. \end{aligned}$$

From (3.4) it is easily seen that

$$(3.9)\qquad \begin{aligned} c_{m;11}(\theta) &= 2n^{-m}\textstyle\sum_{k=0}^{n-1} k^{m-1}(1-\xi_k)(p_0\beta^k)^2(1-p\beta^k)^{-3},\\ c_{m;12}(\theta) &= (\beta_0/\beta)n^{-m-1}\textstyle\sum_{k=0}^{n-1} k^m(1-\xi_k)p_0\beta^k(1+p\beta^k)(1-p\beta^k)^{-3},\\ c_{m;22}(\theta) &= (p/p_0)(\beta_0/\beta)^2 n^{-m-2}\textstyle\sum_{k=0}^{n-1} k^m(1-\xi_k)p_0\beta^k\\ &\quad\cdot[(k+1)p\beta^k+k-1](1-p\beta^k)^{-3}. \end{aligned}$$

In (3.2) the matrices $\mathbf{C}_m(\theta)$, $m = 1, 2$ are evaluated at random points θ_τ^m lying on the line segment joining θ_0 with the point θ_τ. Clearly, θ_τ^1 and θ_τ^2 belong to the neighborhood $U(\delta, n)$ for any choice of θ_τ.

We prove below that for given ϵ, with $0 < \epsilon < 1$, there exist a $\delta > 0$ and a positive integer $n(\delta, \epsilon)$ such that if $n \geqq n(\delta, \epsilon)$ then equations (3.2) will, with probability exceeding $1 - \epsilon$, have a solution $\hat{\tau} = (\hat{\tau}_1, \hat{\tau}_2)$ satisfying $\|\hat{\boldsymbol{\tau}}\| < \delta$, where $\hat{\boldsymbol{\tau}}' = (\hat{\tau}_1, \hat{\tau}_2)$. This is the same as proving Lemma 1, since for points $\theta_\tau \varepsilon U(\delta, n)$ equations (3.2) are equivalent to the likelihood equations (2.3) under the transformation $\tau \to \theta_\tau$ defined by (2.5a). Thus, given such solutions $\hat{\tau}$ of (3.2) we obtain corresponding solutions $\hat{\theta} = (\hat{p}, \hat{\beta}) \varepsilon U(\delta, n)$ of (2.3) which satisfy the conclusions of Lemma 1 by defining $\hat{\theta} = \theta_{\hat{\tau}}$; that is,

$$(3.10)\qquad \hat{p} = p_0(1+\hat{\tau}_1),\qquad \hat{\beta} = \beta_0(1+\hat{\tau}_2 n^{-1}),\qquad \hat{\theta} = (\hat{p}, \hat{\beta}).$$

In order to carry out the program just described we require the following results concerning the asymptotic behavior of the coefficients in (3.2). As $N \to \infty$,

$$(3.11)\qquad a_m(\theta_0) = o_p(1),\ m = 1, 2,$$

$$(3.12)\qquad b_m(\theta_0) = b_m + o_p(1),\ m = 1, 2, 3, \text{ where } B \equiv \begin{pmatrix} b_1 & b_2 \\ b_2 & b_3 \end{pmatrix}$$

$\cdot$ is a negative definite matrix,

$$(3.13)\qquad \sup_{\|\boldsymbol{\tau}\|\leqq\delta_0} \|\mathbf{C}_m(\theta_\tau)\| = O_p(1).$$

Proofs of (3.11)–(3.13) are given below. At this point we shall use these results to prove the existence of solutions $\hat{\tau}$ of (3.2) along the lines of [1]. Let

$$\mathbf{a}(\theta) \equiv \begin{pmatrix} a_1(\theta) \\ a_2(\theta) \end{pmatrix},\quad \mathbf{B}(\theta) \equiv \begin{pmatrix} b_1(\theta) & b_2(\theta) \\ b_2(\theta) & b_3(\theta) \end{pmatrix}.$$

Then we may write (3.2) in the vector form

$$(3.14)\qquad \mathbf{f}(\boldsymbol{\tau}) \equiv \mathbf{a}(\theta_0) + \mathbf{B}(\theta_0)\,\boldsymbol{\tau} + \frac{1}{2}\begin{pmatrix} \boldsymbol{\tau}'\mathbf{C}_1(\theta_\tau^1)\boldsymbol{\tau} \\ \boldsymbol{\tau}'\mathbf{C}_2(\theta_\tau^2)\boldsymbol{\tau} \end{pmatrix} = 0.$$

We now refer to Lemma 2 of [1], a result which is equivalent to the Brouwer fixed point theorem and which is stated here for convenience as

PROPOSITION 1. *Let* $\mathbf{f}(\mathbf{x})$, *with* $\mathbf{x}' = (x_1, x_2, \cdots, x_d)$, *be a continuous mapping of* $\{\mathbf{x}: \|\mathbf{x}\| \leqq a\}$, *where* $a > 0$, *into* R^d *which satisfies* $\mathbf{x}'\mathbf{f}(\mathbf{x}) < 0$ *for every* $\mathbf{x}$ *such that* $\|\mathbf{x}\| = a$. *Then there exists a point* $\hat{\mathbf{x}}$ *such that* $\|\hat{\mathbf{x}}\| < a$ *and* $\mathbf{f}(\hat{\mathbf{x}}) = 0$.

We wish to apply this proposition to show that (3.14), hence (3.2), has a root $\hat{\tau}$ satisfying $\|\hat{\boldsymbol{\tau}}\| < \delta$, provided δ is sufficiently small and n is sufficiently large. To this end we first note that (3.11)–(3.13) imply

$$\mathbf{f}(\boldsymbol{\tau}) = \mathbf{B}\boldsymbol{\tau} + o_p(1) + \|\boldsymbol{\tau}\|^2 O_p(1) \tag{3.15}$$

and, therefore,

$$\boldsymbol{\tau}'\mathbf{f}(\boldsymbol{\tau}) = \boldsymbol{\tau}'\mathbf{B}\boldsymbol{\tau} + o_p(1) + \|\boldsymbol{\tau}\|^3 O_p(1).$$

Let λ_0 denote the maximum eigenvalue of the matrix $\mathbf{B}$. Since $\mathbf{B}$ is negative definite we have $\lambda_0 < 0$ and

$$\boldsymbol{\tau}'\mathbf{f}(\boldsymbol{\tau}) \leqq \lambda_0 \|\boldsymbol{\tau}\|^2 + o_p(1) + \|\boldsymbol{\tau}\|^3 O_p(1). \tag{3.16}$$

From (3.16) it is clear that for given ϵ, with $0 < \epsilon < 1$, there exist a $\delta > 0$ and a positive integer $n(\delta, \epsilon)$ such that

$$\sup_{\|\boldsymbol{\tau}\|=\delta} \boldsymbol{\tau}'\mathbf{f}(\boldsymbol{\tau}) \leqq \lambda_0\delta^2 + o_p(1) + \delta^3 O_p(1) < 0$$

with probability exceeding $1 - \epsilon$ provided $n \geqq n(\delta, \epsilon)$. Then the conditions of Proposition 1 are satisfied and (3.14) has, with probability greater than $1 - \epsilon$, a root $\hat{\tau}$ such that $\|\hat{\boldsymbol{\tau}}\| < \delta$. This completes the proof of Lemma 1, except for the verification of (3.11)–(3.13).

To prove (3.11) we first note that the independent random variables ξ_k have means $E(\xi_k) = (p_0\beta_0^k)^{-1}$ and variances $\operatorname{Var}(\xi_k) = (1 - p_0\beta_0^k)(p_0\beta_0^k)^{-2}$. Thus, using (3.3) and (3.4), we have $E[a_m(\theta_0)] = 0$ and

$$\operatorname{Var}[a_m(\theta_0)] = n^{-2m} \textstyle\sum_{k=0}^{n-1} k^{2m-2}(1 - p_0\beta_0^k)^{-1} \leqq n^{-1}(1 - p_0)^{-1} \to 0$$

as $n \to \infty$. These results clearly imply (3.11).

In a similar fashion we find from (3.6) that $E[b_m(\theta_0)] \to -1/m$ and $\operatorname{Var}[b_m(\theta_0)] \to 0$ for $m = 1, 2, 3$. Hence, $b_m(\theta_0) = -1/m + o_p(1)$ and (3.12) holds with

$$\mathbf{B} \equiv \begin{pmatrix} b_1 & b_2 \\ b_2 & b_3 \end{pmatrix} = -\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}. \tag{3.17}$$

Recalling the assumption concerning δ_0 made at the beginning of this section, we conclude from an inspection of (3.9) that there exists a constant $K > 0$ not depending on τ, δ_0 or n such that

$$|c_{m;ij}(\theta_\tau)| \leqq Kn^{-1} \textstyle\sum_{k=0}^{n-1} \xi_k p_0 \beta_\tau^k$$

for $m = 1, 2$ and $i, j = 1, 2$. Now $\beta_\tau = \beta_0(1 + \tau_2 n^{-1})$ and, therefore, $(\beta_\tau/\beta_0)^k \leqq$

$(1 + n^{-1})^n \leqq e$ for $k = 0, 1, \cdots, n-1$. Thus, if $K' \equiv Ke$, we have (for $\|\boldsymbol{\tau}\| \leqq \delta_0$)

$$|c_{m;\,ij}(\theta_\tau)| \leqq K'n^{-1} \textstyle\sum_{k=0}^{n-1} \xi_k p_0 \beta_0^{\,k} \equiv G(\xi). \tag{3.18}$$

The right side of the inequality (3.18) has mean K' and variance tending to 0 as $n \to \infty$. Hence $G(\xi) = K' + o_p(1)$ and (3.13) follows. This concludes the proof of Lemma 1. We now proceed to the proof of Theorem 1.

From this point on we suppose that δ is small enough and $n(\delta, \epsilon)$ is sufficiently large for Lemma 1 to apply. Let $\hat{\theta}$ and $\hat{\tau}$ be as in the proof of Lemma 1 and set

$$\mathbf{D}(\theta_\tau) = -\mathbf{B}(\theta_0) - \frac{1}{2}\begin{pmatrix} \boldsymbol{\tau}'\mathbf{C}_1(\theta_\tau^{\,1}) \\ \boldsymbol{\tau}'\mathbf{C}_2(\theta_\tau^{\,2}) \end{pmatrix}.$$

Since $\hat{\tau}$ is (with probability greater than $1 - \epsilon$) a solution of (3.14) we have, using (3.10),

$$\mathbf{D}(\hat{\theta}) \begin{pmatrix} p_0^{-1}(\hat{p} - p_0) \\ \beta_0^{-1} n(\hat{\beta} - \beta_0) \end{pmatrix} = \mathbf{a}(\theta_0) \equiv \begin{pmatrix} n^{-1} l_1(\xi, \theta_0) \\ n^{-2} l_2(\xi, \theta_0) \end{pmatrix}. \tag{3.19}$$

From (3.12) and (3.13) it follows that $\mathbf{D}(\theta_\tau) = -\mathbf{B} + o_p(1) + \|\boldsymbol{\tau}\|\, O_p(1)$. Since the matrix $\mathbf{B}$ is nonsingular, we may assume that δ is sufficiently small and $n(\delta, \epsilon)$ sufficiently large so that if $\|\boldsymbol{\tau}\| \leqq \delta$ and $n \geqq n(\delta, \epsilon)$ then the conclusion that $\mathbf{D}(\hat{\theta})$ is also nonsingular may be added to Lemma 1. This being so, we have, upon multiplying (3.19) by $n^{\frac{1}{2}}\mathbf{D}^{-1}(\hat{\theta})$,

$$\begin{pmatrix} p_0^{-1} n^{1/2}(\hat{p} - p_0) \\ \beta_0^{-1} n^{3/2}(\hat{\beta} - \beta_0) \end{pmatrix} = \mathbf{D}^{-1}(\hat{\theta}) \begin{pmatrix} n^{-1/2} l_1(\xi, \theta_0) \\ n^{-3/2} l_2(\xi, \theta_0) \end{pmatrix}. \tag{3.20}$$

Using the results obtained so far and arguments similar to those in Section 5 of [1], it is not difficult to construct a sequence of random variables $\hat{\theta}_n$ which satisfies parts (a) and (b) of Theorem 1 and which also satisfies the following. Let $\hat{\mathbf{P}}_n = \mathbf{D}^{-1}(\hat{\theta}_n)$ if the inverse exists and let $\hat{\mathbf{P}}_n = -\mathbf{B}^{-1}$ otherwise. Then as $n \to \infty$, we have $\hat{\mathbf{P}}_n = -\mathbf{B}^{-1} + o_p(1)$ and, with probability approaching 1,

$$\begin{pmatrix} p_0^{-1} n^{1/2}(\hat{p}_n - p_0) \\ \beta_0^{-1} n^{3/2}(\hat{\beta}_n - \beta_0) \end{pmatrix} = \hat{\mathbf{P}}_n \begin{pmatrix} n^{-1/2} l_1(\xi, \theta_0) \\ n^{-3/2} l_2(\xi, \theta_0) \end{pmatrix}. \tag{3.21}$$

From (3.17) we have

$$\mathbf{B}^{-1} = \begin{pmatrix} -4 & 6 \\ 6 & -12 \end{pmatrix}.$$

Using (3.21) and appealing to a theorem of Cramér [5, Section 20.6] extended to a multivariate situation we may conclude that part (c) of Theorem 1 is also satisfied by the sequence $\hat{\theta}_n$ once we prove

LEMMA 2. *The random variables* $n^{-\frac{1}{2}}l_1(\xi, \theta_0)$ *and* $n^{-\frac{3}{2}}l_2(\xi, \theta_0)$ *are asymptotically jointly normally distributed with zero means and variance-covariance matrix* $-\mathbf{B}$.

PROOF. Given real constants c_1 and c_2, set $\mathbf{c}' = (c_1, c_2)$ and define

$$Z_n = c_1 n^{-\frac{1}{2}} l_1(\xi, \theta_0) + c_2 n^{-\frac{3}{2}} l_2(\xi, \theta_0).$$

From the continuity theorem for multivariate characteristic functions it follows that Lemma 2 holds if for every choice of c_1 and c_2 (not both zero) the random variable Z_n is asymptotically normally distributed with mean 0 and variance $\sigma^2 \equiv -\mathbf{c}'\mathbf{B}\mathbf{c} = c_1^2 + c_1c_2 + c_2^2/3$. Note that $\sigma > 0$ whenever $\mathbf{c} \neq 0$ since the matrix $\mathbf{B}$ is negative definite.

Fixing $\mathbf{c} \neq 0$, we have from (3.4)

$$Z_n = \sum_{k=0}^{n-1} \zeta_{nk},$$

where

$$\zeta_{nk} \equiv n^{-\frac{1}{2}}(c_1 + c_2 k n^{-1})(1 - \xi_k p_0 \beta_0^k)(1 - p_0\beta_0^k)^{-1}.$$

The random variables ζ_{nk}, $k = 0, 1, \cdots, n-1$ are independent with means 0, and is is not difficult to verify that as $n \to \infty$

$$\text{(3.22)} \qquad \operatorname{Var}(Z_n) \equiv \sigma_n^2 = \sigma^2 + o(1), \qquad \sum_{k=0}^{n-1} E|\zeta_{nk}|^3 = o(1).$$

Thus, if F_{nk} denotes the distribution function of ζ_{nk}/σ_n, then for every $\epsilon > 0$ we have

$$\sum_{k=0}^{n-1} \int_{|z| \geq \epsilon} z^2\, dF_{nk}(z) \leqq (1/\epsilon\sigma_n^3) \sum_{k=0}^{n-1} E|\zeta_{nk}|^3 \to 0$$

as $n \to \infty$. From the normal convergence criterion given in [10, Section 21.2] it follows that Z_n/σ_n converges in law to the standard normal distribution $N(0, 1)$. Consequently, since $\sigma_n \to \sigma$, Z_n converges in law to $N(0, \sigma^2)$. This completes the proof of Lemma 2.

4. Proof of Theorem 2. We begin with a proof of the asymptotic result (2.9) which is restated here as

LEMMA 3. *Suppose* (2.2) *holds and for any positive integer* N *let* $n = n(N)$ *denote the integral part of* $(-\log\beta_0)^{-1}(\log N)$. *Then if* $y_i (i \geqq 0,\ y_0 \equiv 0)$ *is the Markov chain with transition probabilities* (1.1), *with* $(p, \beta) = (p_0, \beta_0)$, *we have for* $N \to \infty$

$$\text{(4.1)} \qquad y_N = n + O_p(1).$$

PROOF. For $k \geqq 0$ let ξ_k be the number of times state k is occupied by the Markov chain. Define $S_0 \equiv 0$ and $S_k \equiv \xi_0 + \xi_1 + \cdots + \xi_{k-1}$, $k \geqq 1$. Let $\mu_k = E(S_k)$. We shall first show that for any given ϵ satisfying $0 < \epsilon < 1$ there exists a constant $K_\epsilon > 0$ such that for $k = 1, 2, \cdots$

$$\text{(4.2)} \qquad P(S_k \geqq \mu_k K_\epsilon) \leqq \epsilon/2,$$

$$\text{(4.3)} \qquad P(S_k \leqq \mu_k \leqq \epsilon/2.K_\epsilon^{-1})$$

In words these results assert that the sequence S_k/μ_k, $k \geqq 1$ is stochastically bounded (above) and also stochastically bounded away from 0. The first result, i.e., the existence of a $K_\epsilon > 0$ for which (4.2) is satisfied, is a simple consequence of the Markov inequality [10] according to which one has $P(X \geqq a) \leqq a^{-r}E|X|^r$ for any random variable X and for any choice of constants $a > 0$, $r > 0$. Upon taking $X = S_k/\mu_k$, $a = K_\epsilon$ and $r = 1$ in the Markov inequality, we obtain (4.2) provided only that $K_\epsilon^{-1} \leqq \epsilon/2$.

To prove (4.3) we note first that for arbitrary $a > 0$ one has

$$(4.4) \qquad P(S_k \leqq a) \leqq P(\xi_{k-1} \leqq a) \leqq 1 - (1 - p_0\beta_0^{k-1})^a.$$

Since the random variables ξ_j are geometrically distributed with means $(p_0\beta_0^j)^{-1}$, we have

$$(4.5) \qquad \mu_k \equiv E(S_k) = p_0^{-1}(\beta_0^{-1} - 1)^{-1}(\beta_0^{-k} - 1).$$

Thus if $K > 0$,

$$(1 - p_0\beta_0^{k-1})^{\mu_k K^{-1}} \to \exp[-K^{-1}/(1 - \beta_0)] \quad \text{as} \quad k \to \infty.$$

Upon taking $a = \mu_k K_\epsilon^{-1}$ in (4.4) we obtain

$$(4.6) \qquad \limsup_{k\to\infty} P(S_k \leqq \mu_k K_\epsilon^{-1}) \leqq 1 - \exp[-K_\epsilon^{-1}/(1 - \beta_0)].$$

We now choose $K_\epsilon > 0$ so that K_ϵ satisfies the previous requirement $K_\epsilon^{-1} \leqq \epsilon/2$ and so that, in addition, the right side of (4.6) is no larger than $\epsilon/4$. Then we may conclude that (4.3), as well as (4.2), holds for all sufficiently large k. By increasing K_ϵ if necessary we may further conclude that (4.2) and (4.3) hold for all $k \geqq 1$.

We shall now use the results (4.2) and (4.3) just established to prove that for given ϵ satisfying $0 < \epsilon < 1$ there exist a positive integer N_ϵ and a constant $J_\epsilon > 0$ such that if $N \geqq N_\epsilon$ then

$$(4.7) \qquad P(y_N < n - J_\epsilon) \leqq \epsilon/2, \qquad P(y_N > n + J_\epsilon) \leqq \epsilon/2.$$

The inequalities (4.7) imply $P(|y_N - n| > J_\epsilon) \leqq \epsilon$ for $N \geqq N_\epsilon$ and, since ϵ is arbitrary, (4.1) follows.

Beginning the proof of (4.7) we note first that for $k \geqq 0$ one has $y_N > k$ if and only if $S_{k+1} \leqq N$. Hence, if J is a given positive integer and if N is sufficiently large so that $n > J$, then

$$(4.8) \quad P(y_N < n - J) = P(S_{n-J} > N) \leqq P(S_{n-J} \geqq \mu_{n-J}\{N/\mu_{n-J}\})$$

and also

$$(4.9) \quad P(y_N > n + J) = P(S_{n+J+1} \leqq N) = P(S_{n+J+1} \leqq \mu_{n+J+1}\{N/\mu_{n+J+1}\}).$$

Now from (4.5) and from the definition of n as the integral part of $(-\log \beta_0)^{-1}(\log N)$ it is easily seen that

$$(4.10) \qquad N/\mu_{n-J} \geqq p_0(\beta_0^{-1} - 1)\beta_0^{-J}, \qquad N/\mu_{n+J+1} \leqq \beta_0^{J-1}.$$

Given ϵ satisfying $0 < \epsilon < 1$, let $K_\epsilon > 0$ be such that (4.2) and (4.3) are satisfied for all $k \geqq 1$. Next, choose a constant $J_\epsilon > 0$, say a positive integer, so that for $J = J_\epsilon$ the right sides of the inequalities (4.10) are, respectively, larger than K_ϵ and less than K_ϵ^{-1}. Finally, select N_ϵ sufficiently large so that $n - J_\epsilon > 0$ when $N \geqq N_\epsilon$. Then if $N \geqq N_\epsilon$ we have from (4.8) and (4.2)

$$P(y_N < n - J_\epsilon) \leqq P(S_{n-J_\epsilon} \geqq \mu_{n-J_\epsilon} K_\epsilon) \leqq \epsilon/2$$

and from (4.9) and (4.3)

$$P(y_N > n + J_\epsilon) \leqq P(S_{n+J_\epsilon+1} \leqq \mu_{n+J_\epsilon+1} K_\epsilon^{-1}) \leqq \epsilon/2.$$

This completes the proof of (4.7) and, therefore, of Lemma 3.

Henceforth in this section n will be used to denote the integral part of $(-\log \beta_0)^{-1}(\log N)$, where N is the number of observations $y_1, y_2, \cdots, y_N$ on which the likelihood function $L^*(y, \theta)$ depends. Also, ξ will stand for the n-dimensional random point $\xi = (\xi_0, \xi_1, \cdots, \xi_{n-1})$. From (2.8) we have

$$\log L^*(y, \theta) = \log L(\xi, \theta) + u(y, \xi, \theta), \tag{4.11}$$

where

$$u(y, \xi, \theta) = \operatorname{sgn}(y_N - n) \sum_{j=\min(n,y_N)}^{\max(n,y_N)-1} [\log (p\beta^j) + (\xi_j - 1)\log(1 - p\beta^j)] + (N - \textstyle\sum_{j=0}^{y_N-1} \xi_j) \log (1 - p\beta^{y_N}) \tag{4.12}$$

and where sgn (x) denotes the signum function. Thus if

$$w_1(y, \xi, \theta) = p[\partial u(y, \xi, \theta)/\partial p], \qquad w_2(y, \xi, \theta) \equiv \beta[\partial u(y, \xi, \theta)/\partial \beta] \tag{4.13}$$

and if $l_1(\xi, \theta)$, $l_2(\xi, \theta)$ are as in (3.1), we may express the likelihood equations (2.7) in the form

$$\begin{aligned} n^{-1}p[\partial \log L^*(y, \theta)/\partial p] &= n^{-1}[l_1(\xi, \theta) + w_1(y, \xi, \theta)] = 0, \\ n^{-2}\beta[\partial \log L^*(y, \theta)/\partial \beta] &= n^{-2}[l_2(\xi, \theta) + w_2(y, \xi, \theta)] = 0. \end{aligned} \tag{4.14}$$

For given δ and N, let $U(\delta, n)$ denote the neighborhood of the true parameter point $\theta_0 = (p_0, \beta_0)$ defined by (2.5). Let δ_0 be as in the preceding section. We now prove

LEMMA 4. *For* $N \to \infty$

$$\sup_{\theta \varepsilon U(\delta_0, n)} |n^{-m} w_m(y, \xi, \theta)| = o_p(1), \qquad m = 1, 2. \tag{4.15}$$

PROOF. From (4.12) and (4.13) we may write

$$n^{-m} w_m(y, \xi, \theta) = v_m^1(y, \xi, \theta) + v_m^2(y, \xi, \theta),$$

where

$$v_m^1(y, \xi, \theta) = -n^{-m}[N - \textstyle\sum_{j=0}^{y_N-1} \xi_j] y_N^{m-1} p\beta^{y_N}(1 - p\beta^{y_N})^{-1} \tag{4.16}$$

and

$$\begin{aligned} &v_m^2(y, \xi, \theta) \\ &\quad = \operatorname{sgn}(y_N - n) n^{-m} \textstyle\sum_{j=\min(n,y_N)}^{\max(n,y_N)-1} [j^{m-1}(1 - \xi_j p\beta^j)(1 - p\beta^j)^{-1}] \end{aligned} \tag{4.17}$$

and where the sum over j in (4.16) is taken to be 0 if $y_N = 0$. It is easily seen that there exists a constant $K_1 > 0$, not depending on δ_0, such that for $\theta \varepsilon U(\delta_0, n)$

$$|v_m^1(y, \xi, \theta)| \leqq n^{-1}[K_1(y_N/n)^{m-1}(1 + n^{-1})^{y_N} N\beta_0^{y_N}]. \tag{4.18}$$

In arriving at (4.18) we used the inequality

$$S_{y_N} \equiv \sum_{j=0}^{y_N-1} \xi_j \leqq N$$

which follows from the definition of the ξ_j in terms of the Markov chain y_i. Now Lemma 3 implies that as $N \rightarrow \infty$

$$y_N/n = 1 + o_p(1), \qquad (1 + n^{-1})^{y_N} = e + o_p(1), \qquad N\beta_0^{y_N} = O_p(1). \tag{4.19}$$

Thus, the expression enclosed within brackets in (4.18) is stochastically bounded, and we conclude that for $N \rightarrow \infty$

$$\sup_{\theta \varepsilon U(\delta_0, n)} |v_m^1(y, \xi, \theta)| = o_p(1).$$

In order to complete the proof of Lemma 4 we need only show now that a similar result holds for $v_m^2(y, \xi, \theta)$. From (4.17) and (2.5a) it follows that for points $\theta \varepsilon U(\delta_0, n)$

$$\begin{aligned} &|v_m^2(y, \xi, \theta)| \\ &\quad \leqq n^{-1}[(1 - p)^{-1}(1 + y_N/n)^{m-1}\{|y_N - n| + 2(1 + n^{-1})^{n+y_N} \Sigma_j \xi_j p_0 \beta_0^j\}] \end{aligned} \tag{4.20}$$

where the sum over j ranges from $j = \min(n, y_N)$ to $j = \max(n, y_N) - 1$. The number of terms involved in $\Sigma_j \xi_j p_0 \beta_0^j$ is equal to $|y_N - n|$ and, by Lemma 3, $|y_N - n|$ is stochastically bounded. Also, since $E(\xi_j) = (p_0\beta_0^j)^{-1}$, it follows from the Markov inequality that $P(\xi_j p_0 \beta_0^j \geqq a) \leqq a^{-1}$ for any $a > 0$. Thus, the sum over j in (4.20) is stochastically bounded. From the first two results in (4.19) it then follows that the entire expression enclosed within brackets in (4.20) remains stochastically bounded as $N \rightarrow \infty$. Hence,

$$\sup_{\theta \varepsilon U(\delta_0, n)} |v_m^2(y, \xi, \theta)| = o_p(1).$$

We shall now sketch a proof of the Markov chain version of Lemma 1.

LEMMA 5. *Let ϵ be a given positive number less than* 1, *and for given N let n denote the integral part of* $(-\log \beta_0)^{-1}(\log N)$. *If $\delta > 0$ is sufficiently small and N sufficiently large, say $N \geqq N(\delta, \epsilon)$, then the likelihood equations* (2.7) *will, with probability exceeding* $1 - \epsilon$, *have a solution $\hat{\theta} = (\hat{p}, \hat{\beta})$ which belongs to the interior of the neighborhood $U(\delta, n)$* of θ_0.

PROOF. We consider (4.14) for points $\theta = \theta_\tau$ belonging to $U(\delta, n)$, where $\delta \leqq \delta_0$, and we expand $n^{-m} l_m(\xi, \theta_\tau)$, $m = 1, 2$ about $\tau = (\tau_1, \tau_2) = (0, 0)$ as in (3.2). Then, using the notation introduced in Section 3, we may write (4.14) in the form

$$\mathbf{f}^*(\boldsymbol{\tau}) \equiv \mathbf{f}(\boldsymbol{\tau}) + \begin{pmatrix} n^{-1} w_1(y, \xi, \theta_\tau) \\ n^{-2} w_2(y, \xi, \theta_\tau) \end{pmatrix} = 0. \tag{4.21}$$

From (3.15) and (4.15) it follows that for $N \to \infty$

$$\mathbf{f}^*(\boldsymbol{\tau}) = \mathbf{B}\boldsymbol{\tau} + o_p(1) + \|\boldsymbol{\tau}\|^2 O_p(1). \tag{4.22}$$

The rest of the proof of Lemma 5 is essentially the same as the proof of Lemma 1.

Upon modifying the arguments of Section 3 in the appropriate manner we are able to use Lemma 5 to construct a sequence of random variables $\hat{\theta}_N = (\hat{p}_N, \hat{\beta}_N)$ which satisfy parts (a) and (b) of Theorem 2 and which also satisfy (with probability approaching 1 as $N \to \infty$) an equation of the form (cf. (3.21))

$$\begin{pmatrix} p_0^{-1} n^{1/2}(\hat{p}_N - p_0) \\ \beta_0^{-1} n^{3/2}(\hat{\beta}_N - \beta_0) \end{pmatrix} = \hat{\mathbf{P}}_N \begin{pmatrix} n^{-1/2} l_1(\xi, \theta_0) + n^{-1/2} w_1(y, \xi, \hat{\theta}_N) \\ n^{-3/2} l_2(\xi, \theta_0) + n^{-3/2} w_2(y, \xi, \hat{\theta}_N) \end{pmatrix}, \tag{4.23}$$

where $\hat{\mathbf{P}}_N = -\mathbf{B}^{-1} + o_p(1)$. Now from the proof of Lemma 4 above it is clear that (4.15) may be strengthened to

$$\sup_{\theta \varepsilon U(\delta_0, n)} |n^{-m+\frac{1}{2}} w_m(y, \xi, \theta)| = o_p(1), \qquad m = 1, 2. \tag{4.24}$$

Consequently, the random variables

$$n^{-\frac{1}{2}} l_1(\xi, \theta_0) + n^{-\frac{1}{2}} w_1(y, \xi, \hat{\theta}_N), \qquad n^{-\frac{3}{2}} l_2(\xi, \theta_0) + n^{-\frac{3}{2}} w_2(y, \xi, \hat{\theta}_N)$$

have the same asymptotic joint distribution as do $n^{-\frac{1}{2}} l_1(\xi, \theta_0)$ and $n^{-\frac{3}{2}} l_2(\xi, \theta_0)$. An application of Lemma 2 of Section 3 and the multivariate analogue of the theorem of Cramér [5, Section 20.6] mentioned earlier, now permits us to conclude from (4.23) that the random variables $\hat{\theta}_N = (\hat{p}_N, \hat{\beta}_N)$ satisfy part (c) of Theorem 2.

REFERENCES

[1] AITCHISON, J. and SILVEY, S. D. (1958). Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.* **29** 813–828.

[2] BILLINGSLEY, P. (1961). *Statistical Inference for Markov Processes*. Univ. of Chicago Press.

[3] BRESENHAM, J. E. (1964). Reliability growth models. Technical Report No. 74, Department of Statistics, Stanford Univ.

[4] CHERNOFF, H. (1956). Large-sample theory: parametric case. *Ann. Math. Statist.* **27** 1–27.

[5] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*. Princeton Univ. Press.

[6] DUBMAN, M. R. and SHERMAN, B. (1968). Final report for Contract No. AF(33)615-2818, Project 7071, Aerospace Research Laboratories, United States Air Force, period 1 Nov. 1967–31 Oct. 1968.

[7] FELLER, W. (1966). *An Introduction to Probability Theory and its Applications* **2.** Wiley, New York.

[8] HOADLEY, A. B. (1968). Asymptotic properties of maximum likelihood estimates when the observations are independent but not identically distributed. Bell Telephone Laboratories Memorandum.

[9] LLOYD, D. K. and LIPOW, M. (1962). *Reliability: Management, Methods and Mathematics.* Prentice-Hall, New Jersey.

[10] LOÈVE, M. (1963). *Probability Theory,* (3rd ed.). Van Nostrand, Princeton.

[11] MANN, H. B. and WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[12] PRATT, J. W. (1959). On a general concept of "in probability." *Ann. Math. Statist.* **30** 549–558.

[13] SHERMAN, B. (1966). Consistency of maximum likelihood estimators in some reliability growth models. Aerospace Research Laboratories Report No. 66-0084, United States Air Force.

[14] WALD, A. (1949). Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 595–601.

[15] WOLFOWITZ, J. (1949). On Wald's proof of the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20** 601–602.