The Annals of Mathematical Statistics
1969, Vol. 40, No. 1, 97–115

PROBABILISTIC FUNCTIONS OF FINITE STATE MARKOV CHAINS

BY T. PETRIE

The Institute for Defense Analysis

0. Introduction. This paper is statistically motivated; the content mathematical. The motivation is this: Given is an $s \times s$ ergodic stochastic matrix $A = ((a_{ij}))$ and an $s \times r$ stochastic matrix $B = ((b_{jk}))$. A generates a stationary Markov process $\{W_t\}$ according to $a_{ij} = P[W_{t+1} = j \mid W_t = i]$ and B generates a process $\{Y_t\}$ described by $P[Y_t = k \mid Wt = j] = b_{jk}$. If R is the set of integers $1, 2 \cdots r$, $R_t = R$, and $R^\infty = \prod_{t=1}^{\infty} R_t$ (a point $Y \varepsilon R^\infty$ has coordinates Y_t), then the matrices A and B define a measure $P_{(A,B)}$ on R^∞ by

$$(0.1) \quad P_{(A,B)}\{Y_1 = k_1, Y_2 = k_2 \cdots Y_n = k_n\} = \sum_{i_0, \cdots, i_n \varepsilon S} a_{i_0} a_{i_0 i_1} b_{i_1 k_1} a_{i_1 i_2} b_{i_2 k_2} \cdots a_{i_{n-1} i_n} b_{i_n k_n}$$

where $S = \{1, \cdots, S\}$ and $\{a_{i_0}\}$ is the stationary absolute distribution for A, and $k_i \varepsilon R$. The resulting process $\{Y_t\}$ is called a probabilistic function of the Markov process $\{W_t\}$. Let Λ_1 be the space of $s \times s$ ergodic stochastic matrices, $\tilde{\Lambda}_1$ be the space of $s \times s$ stochastic matrices, Λ_2 the space of $s \times r$ stochastic matrices, $\Pi = \Lambda_1 \times \Lambda_2$ and $\tilde{\Pi} = \tilde{\Lambda}_1 \times \Lambda_2$. The above associates to $\pi = (A, B) \varepsilon \Pi$ and the stationary vector $\mathbf{a}$ for A a stationary measure P_π on R^∞. We write $P_\pi(Y_1, Y_2 \cdots Y_n)$ for that function on R^∞ whose value at $Y_1 = k_1, \cdots Y_n = k_n$ is given by (0.1) if $\pi = (A, B)$.

We also find it necessary to introduce $R_{-\infty} = \Pi_{t=0}^{-\infty} R_t$ and define the measure $P_{(A,B)}$ on $R_{-\infty}$ by $P_{(A,B)}$ $(Y_{-n} = k_{-n}, Y_{-n+1} = k_{-n+1}, \cdots Y_0 = k_0) = P_{(A,B)}$ $(Y_1 = k_{-n}, Y_2 = k_{-n+1}, \cdots Y_{n+1} = k_0)$.

The *problem*: Fix $\pi_0 \varepsilon \Pi$ and let a sample $Y_1, Y_2 \cdots Y_n$ be generated according to the distribution P_{π_0}. From the sample $Y_1, Y_2 \cdots Y_n$ obtain estimators $\theta_n(Y)$ of π_0 so that $\theta_n(Y) \rightarrow \pi_0$ a.e. P_{π_0}. Throughout this paper π_0 is fixed and π varies in Π.

The *mathematics*: Chapter I (Classification of Equivalent Processes) demonstrates that the problem has a solution in the following sense: Let $M[\pi_0] = \{\pi \varepsilon \Pi \mid P_\pi = P_{\pi_0}$ as measures on $R^\infty\}$. Clearly the points of $M[\pi_0]$ can't be distinguished by any finite or infinite sample. The description of $M[\pi_0]$ is crucial in our study. Let $\mathfrak{S}_S$ be the group of permutations of the integers 1 through s. $\mathfrak{S}_S$ acts on Π by $\sigma(A, B) = (\sigma A, \sigma B)$, $(\sigma A)_{ij} = a_{\sigma(i),\sigma(j)}$, $(\sigma B)_{jk} = b_{\sigma(j)k}$ for $\sigma \varepsilon \mathfrak{S}_S$. Observe that $P_{\sigma\pi} = P_\pi$ as measures on R^∞.

The subset $\sum_{j=1}^{S} a_{ij} = 1$, $a_{ij} \geqq 0$, is part of an $s - 1$ dimensional hyperplane in Euclidean s space and has finite non zero $(s-1)$-dimensional Lebesgue measure $\delta_{(s-1)}$; similarly the set $\sum_{k=1}^{r} b_{jk} = 1$, $b_{jk} \geqq 0$, has $(r-1)$-dimensional Lebesgue measure $\lambda_{(r-1)} \neq 0$. It follows that Π with the product measure has

Received 16 February 1967; revised 23 August 1968.

measure $\delta^s_{(s-1)}\lambda^r_{(r-1)}$. We normalize by this constant and, henceforth, assume Π has measure 1. The main result of Chapter I is

THEOREM 1.3. *There is an open subset* Π_0 *of* Π *of measure* 1 *such that for* $\pi_0 \varepsilon \Pi_0$, $M[\pi_0] = \mathfrak{S}_S\pi_0$, i.e., π_0 *is distinguishable up to permutation by the measure* $P_{\pi_0}(\mathfrak{S}_S\pi_0 = \{\sigma\pi_0 \mid \sigma \varepsilon \mathfrak{S}_S\})$.

Chapter II (Limit Theorems and Statistical Analysis) extends and generalizes the results of [1]. For each n and each $Y \varepsilon R^\infty$ define the function $H_n[\pi, Y]$ on Π by $H_n[\pi, Y] = n^{-1} \log P_\pi[Y_1, Y_2 \cdots Y_n]$. Each $H_n[\pi, \cdot]$ is a random variable on the probability space (R^∞, P_{π_0}). For fixed Y the value $H_n[\pi, Y]$ is a function on Π. These random variables hold the solution to our problem as the following shows.

THEOREM 2.1. *For each* π *in* Π, $\lim_{n\to\infty} H_n[\pi, Y] = H_{\pi_0}(\pi)$ *exists* a.e. P_{π_0}.

Proposition 2.2 *and Theorem* 2.5 $H_{\pi_0}(\pi) \leqq H_{\pi_0}(\pi_0)$ *and*

$$H_{\pi_0}(\pi) = H_{\pi_0}(\pi_0) \text{ iff } \pi \varepsilon M[\pi_0].$$

Define $\Pi_n(Y) = \{\pi' \varepsilon \Pi \mid H_n[\pi, Y] \text{ is maximized at } \pi'\}$.

THEOREM 2.8. $\Pi_n(Y) \to M[\pi_0]$ a.e. P_{π_0}.

These theorems theoretically solve our problem. Note in particular, the importance of the function $H_{\pi_0}(\pi)$ in view of Theorems 2.1 and 2.8.

Chapter III (Morse Theory) makes a further study of the function $H_{\pi_0}(\pi)$ for $\pi \varepsilon \Pi_{\delta_1} = \{(A, B) \varepsilon \Pi \mid a_{ij} \geqq \delta_1, b_{jk} \geqq \delta_1\}$, $\delta_1 > 0$, and ties the theory together with the following theorem of [2] and [6]: There exists a class of functions $\mathfrak{J}$ on $\tilde{\Pi}$ such that if $f \varepsilon \mathfrak{J}$ there is a transformation τ_f: $\tilde{\Pi} \to \tilde{\Pi}$ with the property that $f[\tau_f(\pi)] \geqq f(\pi)$ and $f[\tau_f(\pi)] = f(\pi)$ iff π is a critical point of f. The transformations τ_f are given explicitly in [2] and [4]. The class $\mathfrak{J}$ contains each $H_n[\pi, Y]$. Thus, a procedure which is naturally suggested for dealing with the *problem* is: Given $Y_1 \cdots Y_n$ let $f = H_n[\cdot, Y]$ and take $\theta_n(\pi, Y) = \lim_{k\to\infty} \tau_f^k(\pi)$ for any $\pi \varepsilon \Pi_\delta$.[1] How good an estimate of π_0 is $\theta_n(\pi, Y)$? The main theorem of Chapter III answers this with

THEOREM 3.18. *Let* $\pi_0 \varepsilon \Pi_0 \cap \Pi_{\delta_1}$. *There exists an open set* U_{π_0} *containing* $M[\pi_0]$ *such that given* $\epsilon > 0$, *there exists* $N = N(\epsilon)$ *such that* $P\{Y \mid \theta_n(\pi', Y) \varepsilon M[\pi, \epsilon]\} > 1 - \epsilon$ *for* $n > N(\epsilon)$, *for all* $\pi' \varepsilon U_{\pi_0}$. ($M[\pi_0, \epsilon]$ *is the set of points of* Π_δ *whose Euclidean distance from some point of* $M[\pi_0]$ *is less than* ϵ.)

I. CLASSIFICATION OF EQUIVALENT PROCESSES

1. Exhibiting Π_0. The major aim of this chapter is to prove Theorem 1.3. We first exhibit the Π_0 of that theorem.

Let V be an $s \times x$ matrix and $P(\lambda, V) = \det(\lambda I - V)$ where I is the $s \times s$ identity matrix. $P'(\lambda, V) = (d/d\lambda)P(\lambda, V)$. A stochastic matrix A is ergodic iff it has a unique stationary vector $\mathbf{a}$ i.e., $\mathbf{a}^T A = \mathbf{a}^T$ and $\sum a_i = 1$. Observe that the set of ergodic stochastic matrices contains an open subset of the set of all stochastic matrices, namely $\{A \mid P'(1, A) \neq 0, A \text{ stochastic}\}$. Since $\mathbf{a}$ is unique for A, the distribution $P_{(A,B)}$ is uniquely defined by A and B; moreover, $P_{(A,B)}(Y_1 = k_1, \cdots Y_n = k_n)$ is a rational function of the coordinates of A

[1] In great generality this limit exists; for complete validity let $\theta_n(\pi, Y)$ be the set of accumulation points of $\{\tau_f^k(\pi)\}$.

and B because the coordinates a_i of $\mathbf{a}$ are rational functions of the coordinates of A.

For each $i \varepsilon R$, we will exhibit in Section 3 a rational function $P_i(A, B)$ in the coordinates of A and B. Let $f(A, B) = P'(1, A)(\det A)(\Pi_{jk} b_{jk})(\Pi_l P_l(A, B)) \cdot (\sum_{\alpha=1}^{r} \Pi_{\beta \neq \beta'} (b_{\beta\alpha} - b_{\beta'\alpha})^2)$. Modulo the facts that the $P_l(A, B)$ and $\det A$ are not identically zero on Π, it follows that $f(A, B) \not\equiv 0$ on Π; thus $\Pi_1 = \{(A, B) \varepsilon \Pi \mid f(A, B) = 0\}$ is a closed subset of Π of measure zero. Let $\Pi_0 = \Pi - \Pi_1$. It will be seen that Π_0 is the desired set.

2. The probabilistic to deterministic mappings. Via (0.1) we have seen how to obtain a process $\{Y_t\}$ for each point $\pi = (A, B) \varepsilon \Pi$. The process $\{Y_t\}$ is called a *probabilistic function* of the s state Markov process whose transition matrix is A. The $\{Y_t\}$ process can also be considered as a *deterministic function* of an $s \cdot r$ state Markov process as follows: let S' be a new state space whose states are pairs $\langle i, j \rangle$ $i \varepsilon S$, $j \varepsilon R$. Let A' be the transition matrix $A'_{\langle i,j \rangle, \langle i',j' \rangle} = a_{ii'} b_{i'j'}$ for a new Markov process with state space S'. Let $h: S' \to R$ be defined by $h\langle i, j \rangle = j$. The Markov matrix A' together with its unique stationary vector $\mathbf{a}'$ induce a distribution $P'_{A'}$ on S'^{∞}. The map h defines a map $h_\infty: S'^\infty \to R^\infty$. The induced measure $h_\infty P'_{A'} = P''_{A'}$ is precisely P_π. For later convenience we formalize this. Let Γ be the space of $sr \times sr$ stochastic matrices. There is a mapping $g: \Pi \to \Gamma$ such that $P''_{g(\pi)} = P_\pi$. In fact if $\pi = (A, B)$ and if $A * B(k)$ denotes the $s \times s$ matrix whose ijth entry is $a_{ij} b_{jk}$, then the matrix $g(\pi)$ is

$$\begin{pmatrix} A * B(1) & A * B(2) & \cdots & A * B(r) \\ A * B(1) & A * B(2) & \cdots & A * B(r) \\ \vdots & \vdots & & \vdots \\ A * B(1) & A * B(2) & \cdots & A * B(r) \end{pmatrix}$$

Here we assume the states of S' ordered as $\langle 1, 1 \rangle, \langle 2, 1 \rangle \cdots \langle s, 1 \rangle, \langle 1, 2 \rangle, \langle 2, 2 \rangle, \cdots$. Since the mapping g is an imbedding we can consider Π as a subspace of Γ if we so desire.

3. Regular functions. Let $\Sigma(R)$ denote the free semigroup generated by the states of R; so that if $\alpha \varepsilon \Sigma(R)$, $\alpha = k_1 k_2 \cdots k_n$ for some n. $\Sigma(R)$ has a unit $\varnothing$, the empty sequence. For $\gamma \varepsilon \Gamma$ let $P_\gamma''(\alpha) = P_\gamma''\{Y_1 = k_1, Y_2 = k_2 \cdots Y_n = k_n\}$. By previous remarks $P_\pi(\alpha) = P_\pi''(\alpha)$ the latter being defined when π is considered as an element of Γ via g.

DEFINITION 3.1. $\gamma \varepsilon \Gamma$ is regular if for each $k \varepsilon R$ there exist $s_i(k), t_j(k) \varepsilon \Sigma(R)$, $i, j = 0, 1, \cdots s - 1$ such that $\det P_\gamma''(s_i(k) k t_j(k)) \neq 0$.

We can now state the main result of Gilbert [7] tempered to our needs.

THEOREM [7]. *Let $\gamma, \gamma' \varepsilon \Gamma$, γ regular. A necessary and sufficient condition that γ and γ' be equivalent* i.e., $P_\gamma'' = P''_{\gamma'}$ *as measures on R^∞, is that there exists a non singular matrix Z of the form*

$$Z = \begin{pmatrix} X(1) & & & \\ & X(2) & & 0 \\ & & \ddots & \\ 0 & & & X(r) \end{pmatrix}$$

where $X(i)$ *is a nonsingular* $s \times s$ *matrix with* $\sum_{j=1}^{s} X(i)_{lj} = 1, l = 1, \cdots s$ *and* $\gamma' = Z^{-1}\gamma Z$.

The main task of this section is to show that the set of regular points of Π is an open set of measure 1.

For any $k \varepsilon R$, $k^j \varepsilon \Sigma(R)$. The functions $P_k(A, B)$ mentioned in the introduction are now defined by $P_k(A, B) = \det P_\pi(k^i k k^j)$, $i, j = 0, 1, \cdots s - 1$ and $\pi = (A, B)$. Let $G(\pi) = \prod_{k \varepsilon R} P_k(A, B)$. Clearly if $G(\pi) \neq 0$, π is a regular point of Π. Since $G(\pi)$ is a rational function of the coordinates of π the object of this section will be completed if we can exhibit one $\pi \varepsilon \Pi$ such that $G(\pi) \neq 0$.

Let $x_1, x_2, \cdots x_s$ be indeterminants and $R[x_1, x_2, \cdots x_s]$ the ring of polynomials with real coefficients. Let $\theta_r(x_1, x_2 \cdots x_s) = \sum_{i=1}^{s} x_{i+1}x_{i+2} \cdots x_{i+r}$ where the subscripts are taken mod s. Let

$$\alpha(x_1, x_2, \cdots, x_s) = \det \theta_{i+j+1}(x_1 \cdots x_s), \qquad i, j = 0, 1, \cdots s - 1.$$

Thus $\alpha(x_1, x_2, \cdots, x_s) \varepsilon R[x_1, x_2 \cdots x_s]$.

LEMMA 1.1. $\alpha(x_1, x_2, \cdots, x_s)$ *is not the zero polynomial.*

PROOF. Let ζ be a primitive sth root of unity. Choose $x_i = \zeta^{i-1}$, $i = 1, 2, \cdots, s$. Then $\theta_j = \theta_j(1, \zeta, \zeta^2 \cdots \zeta^{s-1}) = 0$ for $j < s$ and $\theta_s = \theta_s(1, \zeta, \zeta^2 \cdots \zeta^{s-1}) \neq 0$. Also observe that $\theta_{s+i} = s^{-1}\theta_s\theta_i$ for $i > 0$; thus, $\alpha(1, \zeta, \zeta^2 \cdots \zeta^{s-1}) = [-\theta_s(1, \zeta, \zeta^2 \cdots \zeta^{s-1})]^s \neq 0$.

Let A^0 be the $s \times s$ matrix with $A^0_{ij} = 0$ if $j \not\equiv i + 1 \bmod s$, $A^0_{i,i+1} = 1$. Observe that $\det A^0 \neq 0$ and 1 is a simple eigenvalue of A^0. Moreover, $A^0 * B(k)$ is the matrix $(A^0 * B(k))_{ij} = 0$ unless $j = i + 1 \bmod s$ and $(A^0 * B(k))_{i,i+1} = b_{i+1,k}$. Note that for $\pi = (A^0, B)$

$$(3.2) \quad G(\pi) = \prod_{k=1}^{r-1} \alpha(b_{1k}, b_{2k}, \cdots, b_{sk}) \cdot\alpha\left(1 - \sum_{k=1}^{r-1} b_{1k}, 1 - \sum_{k=1}^{r-1} b_{2k}, \cdots, 1 - \sum_{k=1}^{r-1} b_{sk}\right).$$

In view of (3.2) and Lemma 1.1 it is not hard to see that there is a $B \varepsilon \Lambda_2$ such that for $\pi = (A^0, B)$, $G(\pi) \neq 0$. Formally,

PROPOSITION 1.2. *There is a* $\pi \varepsilon \Pi_0$ *such that* $G(\pi) \neq 0$.

4. Proof of Theorem 1.3. Let π and $\pi' \varepsilon \Pi_0 \subset \Gamma$. Suppose that $\pi' \varepsilon M[\pi]$. We show that $\pi' = \sigma\pi$ for some $\sigma \varepsilon \mathfrak{S}_s$. Let $\pi = (A, B)$ and $\pi' = (A', B')$. Considered as an element of Γ

$$\pi = \begin{pmatrix} A * B(1) & A * B(2) & \cdots & A * B(r) \\ A * B(1) & A * B(2) & \cdots & A * B(r) \\ \vdots & \vdots & & \vdots \\ A * B(1) & A * B(2) & \cdots & A * B(r) \end{pmatrix}.$$

Similarly for π'. Let Z be the matrix given by Gilbert's theorem so that

$$(4.1) \qquad Z^{-1}\pi Z = \pi'.$$

This implies in particular that

$$(4.2) \quad X(j)^{-1}A * B(1)X(1) = X(1)^{-1}A * B(1)X(1) \quad \text{for} \quad j = 1 \cdots r.$$

Since $\det A * B(1) = (\prod_{j=1}^{s} b_{j1}) \det A \neq 0$ it follows that $X(j) = X(1)$. Let X denote the common value of the $X(j)$'s. It follows from (4.1) that

$$X^{-1}A * B(k)X = A' * B'(k), \qquad k = 1 \cdots r. \tag{4.3}$$

Using the fact that $\sum_{k=1}^{r} A * B(k) = A$, we deduce from (4.3) that $X^{-1}AX = A'$ hence (4.3) becomes

$$X^{-1}\{A * B(k)\}X = \{X^{-1}AX\} * B'(k). \tag{4.4}$$

Multiply (4.4) by $A^{-1}X$ on the left obtaining

$$(b_{jk} - b'_{uk})X_{ju} = 0, \qquad j = 1 \cdots s, \quad u = 1 \cdots s. \tag{4.5}$$

Since $\pi \varepsilon \Pi_0$ there is a k' such that the $b_{jk'}$, $j = 1, 2, \cdots, s$, are pairwise distinct. Examining (4.5) with $k = k'$ we see that for each u there is at most one j such that $b_{jk'} = b'_{uk'}$; hence, the matrix X has at most one non zero entry in each column; however, X is non singular; hence, X has at least one non zero entry in each row and column. These two facts imply that X has exactly one non zero entry in each row and column. Since $\sum_j X_{ij} = 1$ for each i this non zero entry is one and so X is a permutation matrix. Let $X = \sigma$. Then $X^{-1}AX = A'$ shows that $A' = \sigma A$. $(X^{-1}A * B(k)X)_{i,j} = a_{\sigma(i)\sigma(j)}b_{\sigma(j)k} = a_{\sigma(i)\sigma(j)}b'_{jk}$ implies $b'_{jk} = b_{\sigma(j)k}$; hence $B' = \sigma B$ and the theorem is established.

II. LIMIT THEOREMS AND STATISTICAL ANALYSIS

1. Preliminaries. It is a well known fact, see [5], p. 175, that for any stochastic matrix A, $\lim n^{-1} \sum_{k=1}^{n} A^k = A^\infty$ exists. Λ_0 is the subspace of $s \times s$ stochastic matrices consisting of those A's for which the rows of A^∞ are identical and all a_{ij}^∞ are positive; thus $a_{ij}^\infty = \mathbf{a}_j > 0$. In this case A has a uinque stationary vector $\mathbf{a} = \{a_j\}$ and A defines an ergodic Markov process. Let $\Pi_2 \subset \Pi$ be $\Lambda_0 \times \Lambda_2$. For each $\pi \varepsilon \Pi_2$, P_π is an ergodic measure on R^∞.

Convention. In order to dispense with carrying the subscript throughout we use Π for Π_2 throughout this chapter.

Let M denote the semi-ring of real $s \times s$ matrices with non-negative entries. For any subset X of Π, $C_M[X]$ denotes the semi-ring of continuous functions on X with values in M. The product, of course, is defined by $(f \cdot g)\pi = f(\pi) \cdot g(\pi)$. In like manner $C_{\mathfrak{R}}[X]$ denotes the semi-ring of continuous functions on X with values in the real numbers $\mathfrak{R}$.

Define two linear functions L and L_1 from $C_M[X]$ to $C_{\mathfrak{R}}[X]$ by

$$L_1(x)[\pi] = \sum_{i,j} x_{ij}(\pi), \qquad x \varepsilon C_M[X], \qquad \pi \varepsilon X,$$
$$L(x)[\pi] = \sum_{i,j} \mathbf{a}_i x_{ij}(\pi), \qquad x \varepsilon C_M[X],$$

$\pi = (A, B) \varepsilon X$. Again $\mathbf{a}$ is the unique stationary vector for A. Observe that

$$L_1(x \cdot y)[\pi] \leqq L_1(x)[\pi] \cdot L_1(y)[\pi]. \tag{1.1}$$

Also note that for $\pi \varepsilon \Pi$

$$L(x \cdot y)[\pi] \leqq L(x)[\pi] L_1(y)[\pi], \tag{1.2}$$

(1.3) $$L(x\cdot y)[\pi] \leqq L(x)[\pi]L(y)[\pi]\alpha_\pi ,$$

where

$$\alpha_\pi = \max_j a_j^{-1} < \infty.$$

(1.4) $$0 < \min_i \mathbf{a}_i \leqq L(x)[\pi]/L_1(x)[\pi] \leqq \max_i \mathbf{a}_i .$$

2. The basic construction and asymptotic behavior of $n^{-1} \log P_\pi[Y_1, \cdots, Y_n]$. We single out a particular point $\pi_0 \varepsilon \Pi$. This point determines a distribution P_{π_0} on R^∞ as before and also a probability space $\Gamma = (R^\infty, P_{\pi_0})$. Define random variables ψ_n on Γ with values in $C_M[\Pi]$ by $\psi_n(Y)[\pi]_{ij} = P_\pi[W_t = j, Y_n \mid W_{t-1} = i] = a_{ij}b_{jY_n}$. If $\Omega_X(\Pi)$ denotes $\prod_{t=1}^\infty C_M[\Pi]_t$, $C_M[\Pi]_t = C_M[\Pi]$ then we have a mapping $\psi = \prod_{t=1}^\infty \psi_t$ from R^∞ to $\Omega_X(\Pi)$. Let $\nu = \psi P_{\pi_0}$ denote the induced probability measure on $\Omega_X(\Pi)$. Let x_n denote the coordinate functions on $\Omega_X(\Pi)$. $\{x_n\}$ is a stationary ergodic process on $(\Omega_X(\Pi), \nu)$. Observe that

(2.1) $$\sum_{Y_i,\cdots,Y_{i+t-1}\varepsilon R} \psi_i(Y)\psi_{i+1}(Y) \cdots \psi_{i+t-1}(Y)[\pi] = A^t$$

and

(2.2) $$[\psi_1(Y)\psi_2(Y) \cdots \psi_n(Y)]_{ij}[\pi] = P_\pi[Y_1, Y_2, \cdots, Y_n, W_n = j \mid W_0 = i]$$

(2.2′) $$L(\psi_1(Y) \cdots \psi_n(Y))[\pi] = P_\pi[Y_1, Y_2, \cdots, Y_n].$$

In view of (2.2′) we wish to determine the asymptotic behavior of $n^{-1} \log L(\psi_1(Y) \cdots \psi_n(Y))$ as an element of $C_{\mathfrak{R}}[\Pi]$; however, by (1.4) this is the same as determining the asymptotic behavior of $n^{-1} \log L_1(\psi_1(Y) \cdots \psi_n(Y))$.

Let $\pi_1 \varepsilon \Pi$. The restriction mapping $C_M[\Pi] \to C_M[\pi_0 \cup \pi_1]$ induces a map $r:\Omega_X[\Pi] \to \Omega_X[\pi_0 \cup \pi_1]$. Let $\nu' = r\nu$ be the induced measure. We now follow closely [5], Section 2. Let $\Omega_c[\pi_0 \cup \pi_1]$ be the subset of $\prod_{t=1}^\infty [C_M[\pi_0 \cup \pi_1] \times C_M[\pi_0 \cup \pi_1]]$ defined as $\{(x_n, z_n) \mid n \geqq 1, L(z_n x_{n+1})z_{n+1} = z_n x_{n+1}, L(z_n) = 0 \text{ or } 1\}$. $\Omega_c^0[\pi_0 \cup \pi_1]$ is the subset of $\Omega_c[\pi_0 \cup \pi_1]$ defined by

$$z_1(\pi) = x_1(\pi)/L(x_1)[\pi] \quad \text{if} \quad x_1(\pi) \neq 0,$$

$z_1(\pi) = 0$ otherwise, and $z_{n+1}[\pi] = 0$ if $z_n x_{n+1}[\pi] = 0$. The projection mapping $\rho:\Omega_c^0[\pi_0 \cup \pi_1] \to \Omega_x[\pi_0 \cup \pi_1]\rho\{(x_n, z_n)\} = \{x_n\}$ is a $1-1$ correspondence; hence, we can carry the measure ν' over to $\Omega_c^0[\pi_0 \cup \pi_1]$ and trivially extend it to a measure μ_1 on $\Omega_c[\pi_0 \cup \pi_1]$. X^n, Z^n are defined to be the coordinate functions.

Let $T:\Omega_c[\pi_0 \cup \pi_1] \to \Omega_c[\pi_0 \cup \pi_1]$ be the shift operator $T\{(x_n, z_n)\} = \{(x_{n+1}, z_{n+1})\}$ and T^{-1} its set valued inverse. The measure μ_1 is generally not an invariant measure so we construct one as follows: Define measures μ_k on $\Omega_c[\pi_0 \cup \pi_1]$ by $\mu_k(\Omega') = \mu_1(T^{-k+1}\Omega')$ for $\Omega' \subset \Omega_c[\pi_0 \cup \pi_1]$ and $\nu_n(\Omega') = n^{-1} \sum_{k=1}^n \mu_k(\Omega')$.

According to [6], Lemma 1, we have two measures μ_1 and μ on $\Omega_c[\pi_0 \cup \pi_1]$ such that μ is stationary, and on subsets of $\Omega_c[\pi_0 \cup \pi_1]$ defined by the X^n alone $\mu_1 = \mu$.

Define evaluation maps $\lambda_i: (\Omega_c[\pi_0 \cup \pi_1], \Omega_c^0[\pi_0 \cup \pi_1]) \to (\Omega_M, \Omega_M^0)$ by $\lambda_i\{(x_n, z_n)\} = \{x_n(\pi_i), z_n(\pi_i)\} i = 0$ or 1. Ω_M and Ω_M^0 are defined in the same manner as $\Omega_c[\pi_0 \cup \pi_1]$ and $\Omega_c^0[\pi_0 \cup \pi_1]$ except that the coordinates are in M. Via the λ_i we obtain measures $\lambda_i\mu_1$ and $\lambda_i\mu$ on Ω_M. Let $\mathcal{E}_i: C_{\mathfrak{R}}[\pi_0 \cup \pi_1] \to \mathfrak{R}$ be the evaluation functions.

There is a 1—1 correspondence between random variables α on $(\Omega_c[\pi_0 \cup \pi_1], \mu)$ with values in $C_{\mathcal{R}}[\pi_0 \cup \pi_1]$ and pairs of random variables $\lambda_i\alpha$ on $(\Omega_M, \lambda_i\mu)$ with the property that $\mathcal{E}_i\alpha = (\lambda_i\alpha)\lambda_i$. Moreover, $\mathcal{E}_iE_\mu(\alpha) = E_{\lambda_i\mu}(\lambda_i\alpha)$, $i = 0, 1$.

Define convergence in $C_{\mathcal{R}}[\pi_0 \cup \pi_1]$ to be pointwise convergence of functions. Let α_n be a sequence of random variables on $\Omega_c[\pi_0 \cup \pi_1]$ with values in $C_{\mathcal{R}}[\pi_0 \cup \pi_1]$. Then α_n converges a.e. μ iff the $\lambda_i\alpha_n$ converge a.e. $\lambda_i\mu$, $i = 0$ and 1.

Consider the sequence of random variables α_n on $\Omega_c[\pi_0 \cup \pi_1]$ with values in $C_{\mathcal{R}}[\pi_0 \cup \pi_1]$ defined by $\lambda_i\alpha_n = \alpha_n(\pi_i) = n^{-1} \log L_1(X^1X^2 \cdots X^n)[\pi_i]$. In [6] it was shown that $\lim_{n\to\infty} \lambda_i\alpha_n$ exists a.e., $\lambda_i\mu$. By (1.4) it follows that $\lim_{n\to\infty} n^{-1} \log L(X^1X^2 \cdots X^n)[\pi_i]$ exists a.e. $\lambda_i\mu$. A diligent look at the argument of [5], Section 2, together with (1.4) and (1.2) shows that $\lim_{n\to\infty} \lambda_i n^{-1} \log L(X^1 \cdots X^n) = E_{\lambda_i\mu}(\lambda_i \log L(Z^1X^2))$ and this is equal to $\mathcal{E}_iE_\mu(\log L(Z^1X^2)) = E_\mu \log L(Z^1X^2) \cdot [\pi_i]$ by the above discussion.

Since $\lim n^{-1} \log L(X^1X^2 \cdots X^n)[\pi_i]$ exists a.e. $\lambda_i\mu$ it follows that $\lim n^{-1} \log P_{\pi_i}(Y_1 \cdots Y_n) = \lim n^{-1} \log L(\psi_1(Y) \cdots \psi_n(Y))[\pi_i]$ exists a.e. P_{π_0} and equals $E_\mu(\log L(Z^1X^2)[\pi_i]$; thus,

THEOREM 2.1. *Let* $\pi_1 \in \Pi$. *Then* $\lim_{n\to\infty} n^{-1} \log P_{\pi_1}[Y_1 \cdots Y_n] = H_{\pi_0}[\pi_1]$ *exists* a.e. P_{π_0} *and* $H_{\pi_0}[\pi_1] = E_\mu[\log L(Z^1X^2)[\pi_1]]$.

PROPOSITION 2.2.

$$H_{\pi_0}[\pi_1] \leqq H_{\pi_0}[\pi_0].$$

PROOF.

$$\begin{aligned} H_{\pi_0}[\pi_1] - H_{\pi_0}[\pi_0] &= \lim n^{-1}E_{P_{\pi_0}}[\log P_{\pi_1}[Y_1 \cdots Y_n]] - E_{P_{\pi_0}}[\log P_{\pi_0}[Y_1 \cdots Y_n]] \\ &= \lim n^{-1}E_{P_{\pi_0}} \log P_{\pi_1}[Y_1 \cdots Y_n]/P_{\pi_0}[Y_1 \cdots Y_n] \leqq 0 \end{aligned}$$

by Jensen's inequality.

LEMMA 2.3. $$E_\mu \log L(Z^1X^2)[\pi_1]/L(Z^1X^2)[\pi_0] \leqq 0$$

and equality holds iff $L(Z^1X^2)[\pi_1] = L(Z^1X^2)[\pi_0]$ a.e. μ. *If equality holds* $L(Z^1X^2 \cdots X^k)[\pi_1] = L(Z^1X^2 \cdots X^k)[\pi_0]$ a.e. μ.

PROOF. Let the reader note that for any function f on $\Omega_c[\pi_0 \cup \pi_1]$ we have $\int_{\Omega_c[\pi_0\cup\pi_1]} f\, d\mu_1 = \int_{\Omega_c^0[\pi_0\cup\pi_1]} f\, d\mu_1 = \int_{R^\infty} f\rho^{-1}r\psi\, dP_{\pi_0}$ by construction of μ_1. The inequality in the statement of the lemma was established in Proposition 2.2. Since $-H_{\pi_0}[\pi_0]$ is the entropy of the P_{π_0} process, $H_{\pi_0}[\pi_0] > -\infty$; so if

$$E_\mu \log L(Z^1X^2)[\pi_1]/L(Z^1X^2)[\pi_0] = 0$$

then $H_{\pi_0}[\pi_1] > -\infty$ and consequently

$$\inf_n n^{-1}E_{P_{\pi_0}} \log L(\psi_1(Y) \cdots \psi_n(Y))[\pi_1] > -\infty.$$

This implies that $P_{\pi_0}\{L(\psi_1(Y) \cdots \psi_n(Y))[\pi_i] = 0\} = 0$. These statements imply that $\mu_1\{L(x_1 \cdots x_n)[\pi_i] = 0\} = 0$ and $\mu_1\{L(x_1)[\pi_i] = 0\} = 0$. On the support of μ_1 we have $z_1 = x_1/L(x_1)$ and by induction using $L(z_nx_{n+1})z_{n+1} =$

$z_n x_{n+1}$ and the fact that $L(x_1 \cdots x_n)[\pi_i]$ vanishes with probability zero that

$$z_n = x_1 \cdots x_n / L(x_1 \cdots x_n);$$

hence, on the support of μ_1

$$L(z_k x_{k+1}) = L(x_1 \cdots x_{k+1})/L(x_1 \cdots x_k).$$

Let f_{n_i} be the real valued function on $\Omega_c[\pi_0 \cup \pi_1]$ defined by

$$f_{n_i}\{(x_n , z_n)\} = n_i^{-1} \sum_{k=1}^{n_i} L(z_k x_{k+1})[\pi_1]/L(z_k x_{k+1})[\pi_0].$$

Recall that

$$\nu_{n_i} = n_i^{-1} \sum_{k=1}^{n_i} \mu_1(T^{-k+1}).$$

Let

$$f\{(x_n , z_n)\} = L(z_1 x_2)[\pi_1]/L(z_1 x_2)[\pi_0].$$

Then $E_{\mu_1}(f_{n_i}) = E_{\nu_{n_i}}(f)$; furthermore,

$$\begin{aligned} E_{\mu_1}(f_{n_i}) &= n_i^{-1} \sum_{k=1}^{n_i} \int_{\Omega_c[\pi_0 \cup \pi_1]} L(z_k x_{k+1})[\pi_1]/L(z_k x_{k+1})[\pi_0] \, d\mu_1 \\ &= n_i^{-1} \sum_{k=1}^{n_i} \int_{R^\infty} P_{\pi_1}[Y_{k+1} \mid Y_k \cdots Y_1]/P_{\pi_0}[Y_{k+1} \mid Y_k \cdots Y_1] \, dP_{\pi_0} = 1 \end{aligned}$$

because

$$L(z_k x_{k+1})[\pi_i] = L(x_1 x_2 \cdots x_{k+1})[\pi_i]/L(x_1 x_2 \cdots x_k)[\pi_i]$$

and

$$P_{\pi_i}[Y_{k+1} \mid Y_k \cdots Y_1] = L(\psi_1(Y) \cdots \psi_{k+1}(Y))[\pi_i]/L(\psi_1(Y) \cdots \psi_k(Y))[\pi_i].$$

Thus we've shown that $E_{\nu_{n_i}}(f) = 1$. Since $f \geqq 0$, $E_\mu(f) \leqq \lim_{n_i \to \infty} E_{\nu_{n_i}}(f) = 1$ (because $\nu_{n_i} \to \mu$ by construction) so that

$$E_\mu \log L(Z^1X^2)[\pi_1]/L(Z^1X^2)[\pi_0] \leqq \log E_\mu L(Z^1X^2)[\pi_1]/L(Z^1X^2)[\pi_0] \leqq 0$$

and equality holds iff $L(Z^1X^2)[\pi_1] = L(Z^1X^2)[\pi_0]$, a.e. μ. Since μ is a stationary measure, the same reasoning shows that $L(Z^kX^{k+1})[\pi_1] = L(Z^kX^{k+1})[\pi_0]$ for all k, a.e. μ. Since $E_\mu(\log L(Z^kX^{k+1})[\pi_1]) = H_{\pi_0}[\pi_1] > -\infty$, $L(Z^kX^{k+1})[\pi_1] \neq 0$ a.e. μ. If we have shown that $L(Z^1X^2 \cdots X^k)[\pi_i] \neq 0$ a.e. μ then

$$(\dagger) \qquad Z^kX^{k+1} = Z^1X^2 \cdots X^{k+1}/L(Z^1X^2 \cdots X^k)$$

and $L(Z^kX^{k+1})[\pi_i] \neq 0$ a.e. μ implies the same for $L(Z^1X^2 \cdots X^{k+1})[\pi_i]$. Then $(\dagger)$ holds a.e. μ for all k,

$$L(Z^kX^{k+1}) = L(Z^1 \cdots X^{k+1})L(Z^1X^2 \cdots X^k)^{-1}$$

and if $L(Z^1X^2 \cdots X^k)[\pi_1] = L(Z^1X^2 \cdots X^k)[\pi_0]$ and $L(Z^kX^{k+1})[\pi_1] = L(Z^kX^{k+1})[\pi_0]$ then $L(Z^1X^2 \cdots X^{k+1})[\pi_1] = L(Z^1X^2 \cdots X^{k+1})[\pi_0]$. Thus this equality holds for all k.

PROPOSITION 2.4. *Let* $\gamma_\pi = \max_{i,j} a_{ij}/a_j$. *Then*

1. $P_\pi[Y_1 , Y_2 \cdots Y_{n+m}] \leqq \gamma_\pi P_\pi[Y_1 \cdots Y_n] P_\pi[Y_{n+1} \cdots Y_{n+m}]$.

Equivalently

$$L(\psi_1(Y) \cdots \psi_{n+m}(Y))[\pi] \leqq \gamma_\pi L(\psi_1(Y) \cdots \psi_n(Y))[\pi] L(\psi_{n+1}(Y) \cdots \psi_{n+m}(Y))[\pi].$$

PROOF.

$$P_\pi[Y_1 \cdots Y_{n+m}]/P_\pi[Y_1 \cdots Y_n]P_\pi[Y_{n+1} \cdots Y_{n+m}]$$
$$= \sum_{i,j,l} P_\pi[W_n = jY_n \cdots Y_1 \mid W_0 = i]a_i a_{jl} P_\pi[Y_{n+m} \cdots Y_{n+1} \mid W_{n+1} = l]$$
$$\cdot\{\sum_{i,j,l} P_\pi[W_n = jY_n \cdots Y_1 \mid W_0 = i]a_i P_\pi[Y_{n+m} \cdots Y_{n+1} \mid W_{n+1} = l]a_l\}^{-1}$$
$$\leqq \max_{i,j,l} a_{jl}/a_l = \gamma_\pi .$$

Define $\bar{M}[\pi_0] = \{\pi \,\varepsilon\, \Pi \mid H_{\pi_0}[\pi] = H_{\pi_0}[\pi_0]\}$.

THEOREM 2.5. $M[\pi_0] = \bar{M}[\pi_0]$.

PROOF. We have just seen that for every t, $L(Z^1X^2 \cdots X^{t+1})[\pi_1] = (L(Z^1X^2 \cdots X^{t+1})[\pi_0]$ a.e. μ if $\pi_1 \,\varepsilon\, M[\pi_0]$. Let $k_1, k_2, \cdots, k_{t+1}$ be any sequence with $k_i \,\varepsilon\, R$. It is convenient to regard each k_i as that element of $C_M[\pi]$ such that $k_i(\pi) = (a_{lm}b_{mk_i}))$, $\pi = (a, b)$. Then if $\Lambda = \{(X_n, Z_n) \mid X_i = k_i, i = 2, 3, \cdots t+1\}$

$$\int_\Lambda \{L(Z^1X^2 \cdots X^{t+1})[\pi_1]^{-1} d\mu = \int_\Lambda \{L(Z^1X^2 \cdots X^{t+1})[\pi_0]\}^{-1} d\mu.$$

For convenience we introduce the following notation:

$\mathcal{K} = k_2 \cdot k_3 \cdots k_{t+1}$, ${}_r\chi^j = X^rX^{r+1} \cdots X^{r+j}$ and $\psi^k(Y) = \psi_1(Y) \cdot \psi_2(Y) \cdots \psi_k(Y)$.

Then

$$\int_\Lambda L(Z_2^1 \chi^{t-1})^{-1}[\pi_0]\, d\mu$$
$$= \lim n_i^{-1} \sum_{k=1}^{n_i} \int_{T^{-k+1}\Lambda} L(Z^k{}_{k+1}\chi^{t-1})^{-1}[\pi_0]\, d\mu_1$$
$$= \lim n_i^{-1} \sum_{k=1}^{n_i} \int_{T^{-k+1}\Lambda} L({}_1\chi^{k-1})[\pi_0]/L({}_1\chi^{k-1+t})[\pi_0]\, d\mu_1$$
$$= \lim n_i^{-1} \sum_{k=1}^{n_i} \int_{\Gamma_{-k+1}} L(\psi^k(Y))[\pi_0]/L(\psi^k(Y)\mathcal{K})[\pi_0]\, dP_{\pi_0} = 1$$

where $\Gamma_{-k+1} = \{Y \mid Y_{k+1} = k_2, Y_{k+2} = k_3 \cdots Y_{k+t} = k_{t+1}\}$. The last equality follows from the fact that $L(\psi^k(Y))[\pi_0] = P_{\pi_0}(Y_1, Y_2, \cdots, Y_k)$ and $L(\psi^k(Y)\mathcal{K})[\pi_0] = P_{\pi_0}(Y_1, Y_2, \cdots, Y_k, Y_{k+1} = k_2, \cdots, Y_{k+t} = k_{t+1})$. On the other hand

$$(*) \quad \int_\Lambda L(Z_2^1 \chi^{t-1})^{-1}[\pi_1]\, d\mu = \lim n_i^{-1} \sum_{k=1}^{n_i} \int_{\Gamma_{-k+1}} L(\psi^k(Y))[\pi_1]/L(\psi^k(Y)\mathcal{K})[\pi_1]\, dP_{\pi_0} .$$

By Proposition 2.4 $L(\psi^k(Y)\mathcal{K})[\pi_1] \leqq \gamma_{\pi_1} L(\psi^k(Y))[\pi_1]L(\mathcal{K})[\pi_1]$. Plugging this in $(*)$, we obtain

$$\int_\Lambda L(Z_2^1 \chi^{t-1})^{-1}[\pi_1]\, d\mu \geqq \gamma_{\pi_1}^{-1} P_{\pi_0}(Y_{k+1} = k_2 \cdots Y_{k+t} = k_{t+1})/ P_{\pi_1}(Y_{k+1} = k_2 \cdots Y_{k+t} = k_{t+1}).$$

Thus we have

$$1 = \int_\Lambda L(Z_2^1 \chi^{t-1})^{-1}[\pi_1]\, d\mu \ \geqq\ \gamma_{\pi_1}^{-1} P_\pi(Y_{k+1} = k_2 \cdots Y_{k+1} = k_{t+1})/$$
$$P_{\pi_0}(Y_{k+1} = k_2, \cdots Y_{k+t} = k_{t+1}).$$

Since both P_{π_1} and P_{π_0} are ergodic measures $P_{\pi_1} = P_{\pi_0}$.[2]

3. Convergence of the maximum likelihood estimator. Let $S(\pi', \epsilon)$ be the open sphere about $\pi' \varepsilon \Pi$ of radius ϵ. For $x \varepsilon C_M[\Pi]$ define

$$L_\epsilon(x)[\pi'] = \sup_{\pi \varepsilon S(\pi', \epsilon)} \sum x(\pi)_{i,j}.$$

Then

$$(3.1) \qquad L_\epsilon(x \cdot y)[\pi'] \leqq L_\epsilon[\pi'] L_\epsilon(yZ[\pi'].$$

For $Y \varepsilon R^\infty$ let $H_n[\pi, Y, \epsilon] = n^{-1} \log L_\epsilon(\psi_1(Y) \cdots \psi_n(Y))[\pi]$ and $H_n[\pi, Y] = n^{-1} \log L(\psi_1(y) \cdots \psi_n(Y))[\pi]$.

PROPOSITION 2.6. $\lim_{n\to\infty} E_{P_{\pi_0}}(H_n[\pi, Y, \epsilon])$ *exists and*

$$\limsup H_n[\pi, Y, \epsilon] \leqq \lim_{n\to\infty} E_{P_{\pi_0}}(H_n[\pi, Y, \epsilon]) = \mathcal{H}[\pi, \epsilon].$$

PROOF. Stationarity and (3.1) imply the first statement. Ergodicity, stationarity and (3.1) imply the second. See [6], Theorem 1.

PROPOSITION 2.7. *If* $H_{\pi_0}[\pi] < H_{\pi_0}[\pi_0]$, *then there exists* $\epsilon > 0$ *such that* $\mathcal{H}[\pi, \epsilon] < H_{\pi_0}[\pi_0]$.

PROOF. Define $\mathcal{H}_n[\pi, \epsilon] = E_{P_{\pi_0}}(H_n[\pi, Y, \epsilon])$. Then

(1) $\mathcal{H}_{n+m}[\pi, \epsilon] \leqq n(n+m)^{-1}\mathcal{H}_n[\pi, \epsilon] + m(n+m)^{-1}\mathcal{H}_m[\pi, \epsilon]$ so that

(2) $\mathcal{H}_{r\cdot n}[\pi, \epsilon] \leqq \mathcal{H}_n[\pi, \epsilon]$ for any r and n.

Suppose that $H_{\pi_0}[\pi] < H_{\pi_0}[\pi_0] - \alpha$, $\alpha > 0$. Since $\mathcal{H}_n[\pi] = E_{P_{\pi_0}}[H_n[\pi, Y]]$ converges to $H_{\pi_0}[\pi]$, there exists N such that for $n > N$, $|\mathcal{H}_n[\pi] - H_{\pi_0}[\pi]| < \alpha/3$. Thus, for $n > N$ $\mathcal{H}_n[\pi] < H_{\pi_0}[\pi] + \alpha/3$. Let $n_0 > N$ be fixed. Since $\mathcal{H}_{n_0}[\pi, \epsilon] \to \mathcal{H}_{n_0}[\pi]$ at $\epsilon \to 0$, there exists $\epsilon > 0$ such that $\mathcal{H}_{n_0}[\pi, \epsilon] < \mathcal{H}_{n_0}[\pi] + \alpha/3 < H_{\pi_0}[\pi_0]$. By Property 2 above we see that $\mathcal{H}[\pi, \epsilon] \leqq \mathcal{H}_{n_0}[\pi, \epsilon]$ and therefore $\mathcal{H}[\pi, \epsilon] < H_{\pi_0}[\pi_0]$.

Let K be any compact subset of Π and for any $Y \varepsilon R^\infty$ define $\prod_K^n [Y] = \{\pi' \varepsilon K \mid \max H_n[\pi, Y] = H_n[\pi', Y]\}$.

THEOREM 2.8. $\prod_K^n [Y] \to M[\pi_0] \cap K$ a.e. P_{π_0}.

PROOF. Let $M_K[\pi_0] = M[\pi_0] \cap K$. Let $\epsilon > 0$ be given. By the preceding theorem, for any $\pi' \varepsilon K - M_K[\pi_0]$ there exists an $\epsilon_{\pi'}$ such that $\mathcal{H}[\pi', \epsilon_{\pi'}] < H_{\pi_0}[\pi_0]$. Let $M_K[\pi_0, \epsilon] = \{\pi \varepsilon K \mid d(\pi, M_K[\pi_0]) < \epsilon\}$ where d denotes the Euclidean distance. Cover $K - M_K[\pi_0, \epsilon]$, which is compact, by spheres $S(\pi', \epsilon_{\pi'})$ such that $\mathcal{H}[\pi', \epsilon_{\pi'}] < H_{\pi_0}[\pi_0]$. There is a finite subset $\pi_1 \cdots \pi_r$ and $\epsilon_{\pi_1}, \cdots, \epsilon_{\pi_r}$ such that the spheres $S(\pi_i, \epsilon_{\pi_i})$ cover $K - M_K[\pi_0, \epsilon]$.

By Proposition 3.2 we have a.e. P_{π_0}, $\limsup_{\pi \varepsilon S(\pi_i, \epsilon_{\pi_i})} H_n[\pi, Y] \leqq \mathcal{H}[\pi_i, \epsilon_{\pi_i}] <$

[2] I wish to thank Leonard E. Baum for useful comments on this proof.

$H_{\pi_0}[\pi_0]$; thus, if N is chosen so that for the finite number of points $\pi_1, \cdots, \pi_r$, $\sup_{\pi \varepsilon S(\epsilon_i, \epsilon_{\pi_i})} H_n[\pi, Y] < H_{\pi_0}[\pi_0]$ for $n > N$ a.e. P_{π_0}, we have for $n > N$, $\sup_{\pi \varepsilon K - M_K[\pi_0, \epsilon]} H_n[\pi, Y] < H_{\pi_0}[\pi_0]$ a.e. P_{π_0} i.e., $\prod_K^n (Y) \varepsilon M[\pi_0, \epsilon]$ for $n > N$ a.e. P_{π_0}.

III. MORSE THEORY

1. Preliminaries. For a fixed $Y \varepsilon R^\infty$, $H_n[\pi, Y]$ is an analytic function on Π_{δ_1}, $\lim_{n\to\infty} H_n[\pi, Y] = H_{\pi_0}(\pi)$ exists a.e. P_{π_0} and is a C^3 function of $\pi \varepsilon \Pi_{\delta_1}$ ([1], Corollary 4.3). Here we show that $H_{\pi_0}(\pi)$ is an analytic function on Π_{δ_1} and give a structure theorem for the set of critical points $M'[\pi_0]$ of $H_{\pi_0}(\pi)$.

We see from the theorems emphasized in the introduction that a theoretical solution to the problem is effected by choosing any π' which maximizes $H_n[\pi, Y]$ (n large) as an estimate of π_0. A *practical* method for obtaining such a π' is to take $\pi' = \theta_n(\pi, Y)$ for almost any $\pi \varepsilon \Pi$. (See the introduction for the definition of $\theta_n(\pi, Y)$.) This chapter is devoted to justifying this *practical* method which has been found to work in all the examples studied to date. The reader is referred to a forthcoming paper concerning an application of these ideas to stock market prediction [3].

$\theta_n(\pi, Y)$ does not necessarily yield an absolute maximum of $H_n[\pi, Y]$. It does provide a local maximum; in particular it is a critical point of this function. Thus an understanding of the relation between the set of absolute maxima of $H_{\pi_0}(\pi)$ (i.e., $M[\pi_0]$) and its full set of critical points $M'[\pi_0]$ and its connection with the corresponding sets for the functions $H_n[\pi, Y]$ is of utmost importance. These theorems are set forth in Theorem 3.18 and culminate in (3.19) which we offer as a first step towards proving this

CONJECTURE. *There is a subset* $\Gamma \subset \Pi$ *of measure zero such that given* $\epsilon > 0$ *there is an* $N(\epsilon)$ *such that* $P\{Y \mid \theta_n(\pi', Y) \varepsilon M[\pi_0, \epsilon]\} > 1 - \epsilon$ *for* $N(\epsilon)$ *and any* $\pi' \varepsilon \Pi - \Gamma$. (Recall that $\theta_n(\pi', Y) = \lim_{k\to\infty} t^k_{H_n[\pi, Y]}(\pi')$.)

For any sequence of real numbers $\{X_i\}$, let $\bar{X}_-$ denote its maximum and $\underline{X}_-$ denote its minimum. We shall also deal with doubly indexed sequences $\{X_{ij}\}$ and the corresponding notation will be $\bar{X}_{-j} = \max_i X_{ij}$, $\underline{X}_{-j} = \min_i X_{ij}$. If c is any complex number Re c and $I(c)$ are respectively the real and imaginary parts of c.

A *complex s vector* A is an s triple $(A_1, A_2 \cdots A_s)$ where A_i is a complex number. A *complex* $[s, \gamma, \theta)]$ stochastic vector N is a complex s vector $N = (N_1, N_2, \cdots, N_s)$ such that: $\operatorname{Re} N_j \geqq \gamma > 0$, $\sum_c I(N_j) \leqq \theta$ where c is the set of j such that $I(N_j) > 0$, $\sum_j N_j = 1$.

LEMMA 3.1. *Let A be a complex s vector and let $N_i = (N_{i1} \cdots N_{is})$ be a sequence of complex $[s, (\gamma, \theta)]$ stochastic vectors. Define $A_i' = \sum_j N_{ij} A_j$ then*

(i) $\overline{\operatorname{Re}} A_-' \leqq (1 - \gamma)\, \overline{\operatorname{Re}} A_- + \gamma\, \underline{\operatorname{Re}} A_- + \theta(\bar{I}(A_-) - \underline{I}(A_-))$,

(ii) $\underline{\operatorname{Re}} A_-' \geqq \gamma\, \overline{\operatorname{Re}} A_- + (1 - \gamma)\, \underline{\operatorname{Re}} A_- - \theta(\bar{I}(A_-) - \underline{I}(A_-))$,

(iii) $\bar{I}(A_-') \leqq (1 - \gamma)\bar{I}(A_-) + \gamma \underline{I}(A_-) + \theta(\overline{\operatorname{Re}} A_- - \underline{\operatorname{Re}} A_-)$,

(iv) $\underline{I}(A_-') \geqq \gamma \bar{I}(A_-) + (1 - \gamma)\underline{I}(A_-) - \theta(\overline{\operatorname{Re}} A_- - \underline{\operatorname{Re}} A_-)$.

PROOF. Omitted

COROLLARY 3.2. *Under the hypothesis of Lemma* 3.1 *we have*

(i) $\max(\overline{\operatorname{Re}} A_-' - \underline{\operatorname{Re}} A_-', \bar{I}(A_-') - \underline{I}(A_-))$

$$\leqq (1 - 2\gamma + 2\theta) \max(\overline{\operatorname{Re}} A_- - \underline{\operatorname{Re}} A_-, \bar{I}(A_-) - \underline{I}(A_-)),$$

(ii) $\max(|\overline{\operatorname{Re}}\ A_-' - \underline{\operatorname{Re}} A_-|, |\bar{I}(A_-') - \underline{I}(A_-')|, |\overline{\operatorname{Re}}\ A_- - \underline{\operatorname{Re}} A_-'|, |\bar{I}(A_-) - \underline{I}(A_-')|) \leqq (1 - \gamma + \theta) \max(\overline{\operatorname{Re}} A_- - \underline{\operatorname{Re}} A_-, \bar{I}(A_-) - \underline{I}(A_-))$.

The situation in which this corollary is applied is this: Let A and B be complex stochastic matrices i.e., $\sum_j A_{ij} = \sum_j B_{ij} = 1$ for which $\operatorname{Re} A_{ij} \geqq \gamma$ and $|I(A_{ij})|$, $|I(B_{ij})|$ are sufficiently small. Corollary 3.2 implies in particular that the elements of any column of the product BA are closer together than the elements of any column of A. Precisely:

COROLLARY 3.3. *Let A and B be complex $s \times s$ stochastic (γ, θ) matrices. Then if $A' = BA$ we have*

(i) $\max(\overline{\operatorname{Re}} A_{-j}' - \underline{\operatorname{Re}} A_{-j}', \bar{I}A_{-j}' - \underline{I}A_{-j}')$

$$\leqq (1 - 2\gamma + 2\theta) \max(\overline{\operatorname{Re}} A_{-j} - \underline{\operatorname{Re}} A_{-j}, \bar{I}A_{-j} - \underline{I}A_{-j})$$

(ii) $\max(|\operatorname{Re} A_{ij}' - \operatorname{Re} A_{ij}|, |IA_{ij}' - IA_{ij}|)$

$$\leqq (1 - \gamma + \theta) \max(\overline{\operatorname{Re}} A_{-j} - \underline{\operatorname{Re}} A_{-j}, \bar{I}A_{-j} - \underline{I}A_{-j}).$$

For any $s \times s$ matrix M, $M_{i*} = \sum_{j=1}^{s} M_{ij}$ and $M_{*j} = \sum_{i=1}^{s} M_{ij}$. Let K and δ be positive constants and let $\mathcal{K}$ be the set of real $s \times s$ matrices M satisfying $\delta \leqq M_{ij} \leqq K$. We shall be concerned with sequences $M^0, M^{-1}, M^{-2}, \cdots$ of matrices $M^{-1} \varepsilon \mathcal{K}$ and the behavior of the sequences $(M^{-k}M^{-k+1} \cdots M^0)_{rs}/(M^{-k}M^{-k+1} \cdots M^0)_{r*}$ as k approaches ∞.

Let M_{-k}^0 denote the matrix $(M^{-k}M^{-k+1} \cdots M^0)$. The following theorem is probably folklore. Because we need the ideas involved in the proof, we include it.

THEOREM 3.4. *Let $M^0, M^{-1}, \cdots, M^{-k}, \cdots$ denote any sequence with $M^{-i} \varepsilon \mathcal{K}$. Then $\lim_{k\to\infty} (M_{-k}^0)_{rs}/(M_{-k}^0)_{r*}$ exists, is independent of r and the rate of convergence is exponential.*

PROOF.

$$(M_{-n-1}^0)_{rs}/(M_{-n-1}^0)_{r*} = \sum_j [M_{rj}^{-n-1}(M_{-n}^0)_{j*}/(M^{-n-1}M_{-n}^0)_{r*}](M_{-n}^0)_{js}/(M_{-n}^0)_{j*}.$$

Let $a_{rj} = M_{rj}^{-n-1}(M_{-n}^0)_{j*}/(M^{-n-1}M_{-n}^0)_{r*}$. Then

$$a_{rj}/a_{rj'} \leqq \max_{r,j,j',k} M_{rj}^{-n-1}M_{jk}^{-n}/M_{rj'}^{-n-1}M_{j'k}^{-n} \leqq K^2/\delta^2$$

since $\sum_j a_{rj} = 1$, it follows that $1 + \sum_{j\neq k} a_{rj}/a_{rk} = 1/a_{rk} \leqq 1 + (s-1)K^2/\delta^2$ so $a_{rj} \geqq (1 + (s-1)K^2/\delta^2)^{-1} = \lambda_0$. (Similarly it can be established that for every k, $(M_{-k}^0)_{rs}/(M_{-k}^0)_{r*} \geqq \lambda_0$.)

Thus

$$\begin{aligned}(M_{-n-1}^0)_{rs}/(M_{-n-1}^0)_{r*} &= \sum_j a_{rj}(M_{-n}^0)_{js}/(M_{-n}^0)_{j*} \leqq (1 - \lambda_0) \max_j (M_{-n}^0)_{js}/(M_{-n}^0)_{j*} \\ &\quad + \lambda_0 \min_j (M_{-n}^0)_{js}/(M_{-n}^0)_{j*}\end{aligned}$$

and

$$\begin{aligned}(M_{-n-1}^0)_{rs}/(M_{-n-1}^0)_{r*} \\ \geqq \lambda_0 \max_j (M_{-n}^0)_{rs}/(M_{-n}^0)_{j*} + (1 - \lambda_0) \min_j (M_{-n}^0)_{js}/(M_{-n}^0)_{j*}.\end{aligned}$$

Thus $\max_r (M^0_{-n-1})_{rs}/(M^0_{-n-1})_{r*} - \min_r (M^0_{-n-1})_{rs}/(M^0_{-n-1})_{r*} \leqq (1 - 2\lambda_0)\cdot(\max_j (M^0_{-n})_{rs}/(M^0_{-n})_{r*} - \min (M^0_{-n})_{rs}/(M^0_{-n-1})_{r*}) \leqq (1 - 2\lambda_0)^{n+2}$ by induction.

The two main ingredients to this proof are the positivity of the M_{rj} *and* the fact that we could easily establish a lower bound for the a_{rj} independent of n. We seek a complex analog of Theorem 3.4. Its proof is more complicated precisely because we are not able to obtain a simple lower bound for the real part of the a_{rj}.

Let $\mathcal{H}$ be the set of $s \times s$ matrices M with complex entries satisfying the following

$$|\geqq | M_{rj}|, \qquad 1 - \delta \geqq \operatorname{Re} M_{rj} \geqq \delta.$$

Let $\mathcal{H}^\infty$ be the product of $\mathcal{H}$ with itself a countable number of times. An element of $\mathcal{H}^\infty$ will be denoted by $m = (M^0, M^{-1} \cdots)$ with $M^{-i} \varepsilon \mathcal{H}$. Let $\mathcal{H}_\alpha = \{M \varepsilon \mathcal{H} \mid |I(M_{ij})| \leqq \alpha\}$ and let $\mathcal{H}_\alpha^\infty \subset \mathcal{H}^\infty$ be the subspace with all $M^{-i} \varepsilon \mathcal{H}_\alpha$.

Define these functions on $\mathcal{H}^\infty$:

(i) $R^n(r, m)_{ij} = (M^{-n})_{rj}(M^{-1}_{-n+1})_{ji}/(M^{-1}_{-n})_{ri}$;

(ii) $H^n(m)_{ri} = (M^{-1}_{-n})_{ri}/(M^{-1}_{-n})_{r*}$;

(iii) $\Gamma^n(m)_{ri} = (M^{-1}_{-n})_{ri}M^0_{i*}/(M^0_{-n})_{r*}$;

(iv) $h^n(m)_{rj} = (M^{-n})_{rj}(M^{-n+1} \cdots M^{-1})_{j*}/(M^{-n} \cdots M^{-1})_{r*}$;

(v) $p^n(m)_{rj} = (M^{-n})_{rj}(M^{-n+1} \cdots M^0)_{j*}/(M^{-n} \cdots M^0)_{r*}$;

(vi) $P^n(m)_{ri} = (M^0_{-n})_{ri}/(M^0_{-n})_{r*}$.

Let $T:\mathcal{H}^\infty \to \mathcal{H}^\infty$ be the shift operator $T(M^0, M^{-1} \cdots) = (M^{-1}, M^{-2}, \cdots)$. (Note that T maps $\mathcal{H}_\alpha$ onto itself.) Then $h^n(m) = p^{n-1}(Tm)$ and $H^n(m) = P^{n-1}(Tm)$. Thus if we have established a property of $p^{n-1}(m)$ or for $P^{n-1}(m)$ for all $m \varepsilon \mathcal{H}^\infty$ we have established that property for $h^n(m)$ and $H^n(m)$ for all m.

We record the basic relations among these functions

(1) $\sum_i H^n(m)_{ri}R^n(r, m)_{ij} = h^n(m)_{rj}$;

(2) $\sum_i \Gamma^n(m)_{ri}R^n(r, m)_{ij} = p^n(m)_{rj}$;

(3) $p^n(m)\cdot P^{n-1}(m) = P^n(m)$ as matrices;

(4) $\Gamma^n(m)_{ri} = H^n(m)_{ri}M^0_{i*}/\sum_j H^n(m)_{rj}M^0_{j*}$;

(5) $R^n(r, m)_{ij} = H^{n-1}(\theta^{n-1}m)_{ij}\tilde{M}^{-n}_{jr}/\sum_j H^{n-1}(\theta^{n-1}m)_{ij}\tilde{M}^{-n}_{jr}$.

Here $\tilde{M}$ is the transpose of M and $\theta^{n-1}:\mathcal{H}^\infty \to \mathcal{H}^\infty$ is the mapping which takes

$$(M^0, M^{-1} \cdots M^{-n+1}, M^{-n} \cdots) \quad \text{to} \quad (M^0, \tilde{M}^{-n+1}, \tilde{M}^{-n+2}, \cdots \tilde{M}^{-1}, M^{-n}, M^{-n-1} \cdots).$$

Observe that θ^{n-1} maps $\mathcal{H}_\alpha$ onto itself.

LEMMA 3.5. *There is an* $\epsilon_{2k} > 0$ *such that if* $M \varepsilon \mathcal{H}$, $m \varepsilon \mathcal{H}^\infty |I(M^{-i}_{rj})| \leqq \epsilon_{2k}$ *for all* i, $|I(M_{rj})| \leqq \epsilon_{2k}$ *and* $P^k(m)_{Al} = s^{-1} \sum_{i=1}^s P^k(m)_{jl}$, *then*

$$[2(1 - \delta)]^{-1} \leqq |[\textstyle\sum_l P^k(m)_{Al}M_{lr}]^{-1}| \leqq 2\delta^{-1}.$$

PROOF. If $I(M^{-i}_{rj}) = 0$ and $I(M_{rj}) = 0$, then $(1 - \delta)^{-1} \leqq [\sum_l P^k(m)_{Al}M_{lr}]^{-1} \leqq 2\delta^{-1}$. The result now follows from the fact that P^k is a continuous function on $\mathcal{H}^\infty$ depending on only $k + 1$ coordinates of m.

Let $\epsilon_3 > 0$ be such that if x is any complex number satisfying $[2(1 - \delta)]^{-1} \leqq |x^{-1}| \leqq 2\delta^{-1}$ and $|\beta| < \epsilon_3$ then $[4(1 - \delta)]^{-1} \leqq |(x + \beta)^{-1}| \leqq 4\delta^{-1}$. Suppose that

$0 < \rho < 1$ and k has been chosen so large that $8s\rho^k(1 - \delta) < \frac{1}{2}$ and $2s\rho^k/(1 - \rho) < \epsilon_3$.

LEMMA 3.6. *Suppose that for all m in $\mathcal{K}_\alpha^\infty$ and M in $\mathcal{K}_\alpha$ we have*

(i) $[2(1 - \delta)]^{-1} \leqq |[\sum_l P^k(m)_{Al}M_{lr}]^{-1}| \leqq 2\delta^{-1}$,

(ii) $|P_{Al}^{n+k-2}(m) - P_{il}^{n+k-2}(m)| \leqq 2\rho^{n+k}$ *for all i and l,*

(iii) $|P_{Al}^{n+k-2}(m) - P_{Al}^k(m)| \leqq 2\rho^k/1 - \rho$ *for all l.*

Then for all m in $\mathcal{K}_\alpha^\infty$ and M in $\mathcal{K}_\alpha$, $[\sum_l P^{n+k-2}(m)_{il}M_{lr}]^{-1} = 1/K_r + x_{ir}/(1 - x_{ir}/K_r)$ and $[4(1 - \delta)]^{-1} \leqq |1/K_r| \leqq 4\delta^{-1}$ and

$$|x_{ir}/(1 - x_{ir}/K_r)| \leqq 4s\rho^{n+k}.$$

PROOF. We suppress the reference to m in expressions like $P^{n+k-2}(m)_{il}$ and write P_{il}^{n+k-2}. Then

$$[\textstyle\sum_l P_{il}^{n+k-2}M_{lr}]^{-1} = [\sum_i P_{Al}^{n+k-2}M_{lr} - \sum_l (P_{Al}^{n+k-2} - P_{il}^{n+k-2})M_{lr}]^{-1}$$
$$= [K_r - x_{ir}]^{-1}$$

where

$$1/K_r = [\textstyle\sum_l P_{Al}^k M_{lr} + \sum_l (P_{Al}^{n+k-2} - P_{Al}^k)M_{lr}]^{-1}.$$

Now $|[\sum_l (P_{Al}^{n+k-2} - P_{Al}^k)M_{lr}]| \leqq 2s\rho^k/(1 - \rho) < \epsilon_3$ so that $[4(1 - \delta)]^{-1} \leqq |K_r^{-1}| \leqq 4\delta^{-1}$. Also

$$|x_{ir}/(1 - x_{ir}/K_r)| \leqq |x_{ir}|/(1 - |x_{ir}/K_r|) \leqq 2s\rho^{n+k}/(1 - 8s\rho^{n+k}(1 - \delta)) \leqq 4s\rho^{n+k}.$$

COROLLARY 3.7. *Under the conditions of Lemma 3.6 we have*

$$R^{n+k}(r, m)_{ij} = H^{n+k-1}(\theta^{n+k-1}m)_{ij}(K_{rj}^n + K_{rij}^n)$$

where $K_{rj}^n = M_{rj}^{-n-k}/K_r$ and $K_{rij}^n = M_{rj}^{-n-k}(x_{ir}/K_r^2)/(1 - x_{ir}/K_r)$ satisfy $|K_{rj}^n| \leqq 4/\delta$, $|K_{rij}^n| \leqq 64\delta^{-2}s\rho^{n+k}$.

PROOF.

$$R^{n+k}(r, m)_{ij}/H^{n+k-1}(\theta^{n+k-1}m)_{ij} = M_{rj}^{-n-k}[\textstyle\sum_j H^{n+k-1}(\theta^{n+k-1}m)_{ij}M_{rj}^{-n}]^{-1}$$
$$= M_{rj}^{-n-k}[\textstyle\sum_j P^{n+k-2}(T\theta^{n+k-1}m)_{ij}M_{rj}^{-n}]^{-1}.$$

LEMMA 3.8. *Let $1 > \alpha > 0$, $1 > \beta > 0$ and $\bar{\epsilon} > 0$ be real numbers. Then there is an $\epsilon_1 = \epsilon_1(\alpha, \beta) > 0$ such that if $H = (H_{rj})$ and $M = (M_{jk})$ are matrices satisfying: $\sum_j H_{rj} = 1$, $1 > \operatorname{Re} H_{rj} > \alpha$, $1 > \operatorname{Re} M_{rj} \geqq \beta$, $|I(M_{rj})| < \epsilon_1$, $|I(H_{rj})| < \epsilon_1$ then $\operatorname{Re}(H_{rj}M_{j*}[\sum_j H_{rj}M_{j*}]^{-1}) \geqq \alpha\beta$ and*

$$|I(H_{rj}M_{j*}[\textstyle\sum_j H_{rj}M_{j*}]^{-1})| \leqq \bar{\epsilon}/s.$$

PROOF. Omitted.

LEMMA 3.9. *Fix the constants $K > 0$ and $1 - 2\lambda_0 < \rho < 1$. There is an $\epsilon_{4k} >$ 0 such that if $|I(M_{rj}^{-i})| \leqq \epsilon_{4k}$, $i = 0, 1, \cdots, k$, then*

(i) $\operatorname{Re} P^k(m)_{ij} \geqq \lambda_0 - \rho^{k+1}$,

(ii) $|IP^k(m)_{ij}| \leqq \rho^{k+1}$,

(iii) $\max_j (\overline{\operatorname{Re}}\, P^k(m)_{-j} - \underline{\operatorname{Re}}\, P^k(m)_{-j}) < \rho^{k+1}$,

(iv) $\max_j (\bar{I}(P^k(m)_{-j}) - \underline{I}(P^k(m)_{-j})) < \rho^{k+1}$,

(v) $\operatorname{Re} p^k(m)_{rj} \geqq \lambda_0 - K\rho^k$,
(vi) $|Ip^k(m)_{rj}| \leqq K\rho^k$.

PROOF. The statements follow from the fact that P^k and p^k are continuous functions depending on only $k + 1$ coordinates of m and if all M_{rj}^{-i} are real numbers, $P^k(m)_{rj} \geqq \lambda_0$, $p^k(m)_{rj} \geqq \lambda_0$ and $\max_j (\bar{P}^k(m)_{-j} - \underline{P}^k(m)_{-j}) \leqq (1 - 2\lambda_0)^{k+1} < \rho^{k+1}$. (Refer to the proof of Theorem 3.4.)

Let $K = 250s^2\delta^{-2}$. Choose $\underline{\lambda} < \delta\lambda_0/2$ and $\bar{\epsilon} < \underline{\lambda}$. Let $\rho = (1 - \underline{\lambda} + \bar{\epsilon}) < 1$ and choose k so that

(i) $\lambda_0 - K\rho^k/(1 - \rho) > \frac{1}{2}\lambda_0$, $\min(\epsilon_1(\lambda_0/2, \delta), \bar{\epsilon}) > sK\rho^k/(1 - \rho)$,
(ii) $2s\rho^k 4(1 - \delta) < \frac{1}{2}$, $2s\rho^k/(1 - \rho) < \epsilon_3$.

For this fixed k choose $\epsilon < \min(\epsilon_1(\lambda_0/2), \epsilon_{2k}, \epsilon_{4k})$ with the additional condition that if c is a complex number with $|\operatorname{Re} c| < 1 - \underline{\lambda}$ and $|I(c)| < \epsilon$ we have $|c| < 1$.

THEOREM 3.10. *For all $m \in \mathcal{H}_\epsilon^\infty$ and all $n \geqq 1$ we have*

(i) $\max_j (\overline{\operatorname{Re}}\, P^{n+k-1}(m)_{-j} - \underline{\operatorname{Re}}\, P^{n+k-1}(m)_{-j},$
$$\bar{I}P^{n+k-1}(m)_{-j} - \underline{I}P^{n+k-1}(m)_{-j}) \leqq \rho^{n+k},$$
(ii) $\operatorname{Re} P^{n+k-1}(m)_{ij} \geqq \lambda_0 - \rho^{k+1}(1 + \rho + \cdots + \rho^{n-1})$,
(iii) $|IP^{n+k-1}(m)_{ij}| \leqq \rho^{k+1}(1 + \rho + \cdots \rho^{n-1})$,
(iv) $\operatorname{Re} p^{n+k-1}(m)_{ij} \geqq \lambda_0 - K\rho^k(1 + \rho + \cdots + \rho^{n-1})$,
(v) $|Ip^{n+k-1}(m)_{ij}| \leqq K\rho^k(1 + \rho + \cdots + \rho^{n-1})$.

PROOF. The theorem is true for $n = 1$ by Lemma 3.9. Assume the truth of these statements for n. We establish them for $n + 1$. The induction hypothesis (ii) and (iii) and the fact that $\rho^{k+1}(1/(1 - \rho)) < \bar{\epsilon}/s$ imply that $\operatorname{Re} H^{n+k}(m)_{ij} \geqq \underline{\lambda}$ and $\sum_{c_i} IH^{n+k}(m)_{ij} \leqq \bar{\epsilon}$. We also show that these inequalities hold for the matrix $\Gamma^{n+k}(m)$. Recall that $\Gamma^{n+k}(m)_{rj} = H^{n+k}(m)_{rj}M_{j*}^0[\sum_j H^{n+k}(m)_{rj}M_{i*}^0]^{-1}$. The inductive assumptions (ii) and (iii) imply that $\operatorname{Re} H^{n+k}(m)_{rj} \geqq \lambda_0 - \rho^{k+1}(1 + \cdots + \rho^{n-1}) > \lambda_0/2$ and $|IH^{n+k}(m)_{rj}| \leqq \rho^{k+1}(1 + \cdots + \rho^{n-1}) \leqq \epsilon_1(\lambda_0/2, \delta)$. Since $|I(M_{sj}^0)| < \epsilon(\lambda_0/2, \delta)$, we have by Lemma 3.8 the desired inequalities for $\Gamma^{n+k}(m)$.

Now we establish the conditions of Lemma 3.6. The condition (i) of that lemma is satisfied because $\epsilon < \epsilon_{2k}$. See Lemma 3.5. Condition (ii) follows from these facts:

$$P^{n+k-2}(m)_{Al} = \sum_{i=1}^{s} s^{-1}P^{n+k-2}(m)_{il},$$

induction hypothesis i applied to $P^{n+k-2}(m)$ and Corollary 3.2 statement (ii). Condition (iii) holds for these reasons: $p^{r+k}(m)P^{r+k-1}(m) = P^{r+k}(m)$, the inductive assumptions, and Corollary 3.3 shows that $|P^{r+k}(m)_{ij} - P^{r+k-1}(m)_{ij}| \leqq 2\rho^{r+k-1}$, $r \leqq n - 2$.

Having satisfied the conditions of Lemma 3.6 we have the conclusion of Corollary 3.7, i.e., $R^{n+k}(r, m)_{ij} = H^{n+k-1}(\theta^{n+k-1}m)_{ij}(K_{rj}^n + K_{rij}^n)$ where $|K_{rj}^n| \leqq 4\delta^{-1}$ and $|K_{rij}^n| \leqq \rho^{n+k}64s\delta^{-2}$. The relations (1) and (2) of this section with n replaced by $n + k$ thus gives

(1′) $\sum_i H^{n+k}(m)_{ri}H^{n+k-1}(\theta^{n+k-1}m)_{ij}(K_{rj}^n + K_{rij}^n) = h^{n+k}(m)_{rj}$,
(2′) $\sum_i \Gamma^{n+k}(m)_{ri}H^{n+k-1}(\theta^{n+k-1}m)_{ij}(K_{rj}^n + K_{rij}^n) = p^{n+k}(m)_{rj}$.

The inductive assumption (ii) applied to $H^{n+k-1}(\theta^{n+k-1}m) = P^{n+k-2}(T\theta^{n+k-1}m)$

together with Corollary 3.2 and the bounds on K_{rj}^n and K_{rsj}^n give $|h^{n+k}(m)_{rj} - p^{n+k}(m)_{rj}| \leqq (2|K_{rj}| + 128s^2\delta^{-2})\rho^{n+k} \leqq (8\delta^{-1} + 128s^2\delta^{-2})\rho^{n+k}$. The equality $h^{n+k}(m) = p^{n+k-1}(Tm)$ together with the inductive assumptions (iv) and (v) show that

(iv′) $\operatorname{Re} p^{n+k}(m)_{rj} \geqq \lambda_0 - K\rho^k(1 + \rho + \cdots + \rho^n)$,

(v′) $|Ip^{n+k}(m)_{rj}| \leqq K\rho^k(1 + \rho + \cdots + \rho^n)$.

The lower bound in (iv′) exceeds λ while the upper bound in (v′) is less than $\bar{\epsilon}/s$; hence, the relation $p^{n+k}(m)P^{n+k-1}(m) = P^{n+k}(m)$ together with Corollary 3.3 and the inductive assumption(iv) imply the inductive assumption (i) with n replaced by $n + 1$.

Corollary 3.3 also implies that

$$|\operatorname{Re} P^{n+k-1}(m)_{rj} - \operatorname{Re} P^{n+k}(m)_{rj}| \leqq \rho^{n+k-1},$$
$$|IP^{n+k-1}(m)_{rj} - IP^{n+k}(m)_{rj}| \leqq \rho^{n+k+1}.$$

These two inequalities together with inductive assumptions (ii) and (iii) for n imply these assumptions for $n + 1$.

Corollary 3.11. *For all $m \in \mathcal{H}_\epsilon^\infty$ we have*

$$|P^n(m)_{rs} - P^{n-1}(m)_{rs}| \leqq 2\rho^{n+1}.$$

Corollary 3.12. *$\lim_{n\to\infty} P^n(m)$ exists uniformly for all $m \in \mathcal{H}_\epsilon^\infty$ and for each m the rows of the limiting matrix are equal.*

2. Applications. Let $\Lambda_{1\delta_1}$ be the set of $s \times s$ matrices $A = (a_{ij})$ such that $\sum_j a_{ij} = 1$, $1 - \delta_1 \geqq \operatorname{Re} a_{ij} \geqq \delta_1$, $|I(a_{ij})| \leqq \alpha$; let $\Lambda_{2\delta_1}$ be the set of $s \times r$ matrices $B = (b_{jk})$ such that $\sum_k b_{jk} = 1$, $1 - \delta_1 \geqq \operatorname{Re} b_{jk} \geqq \delta_1$, $|I(b_{jk})| \leqq \alpha$; and let $\Pi_{\delta_1} = \Lambda_{1\delta_1} \times \Lambda_{2\delta_1}$. If $\delta < \delta_1^2$ and α is sufficiently small, then for every $\pi \in \Pi_{\delta_1}$, $\pi = (A, B)$ the r matrices $(a_{ij}b_{jk})k = 1, 2, \cdots r$ are in $\mathcal{H}_\epsilon$. With this assumption we define for each $Y \in R_{-\infty}$ a mapping $f(\cdot, Y):\Pi_{\delta_1} \to \mathcal{H}_\epsilon^\infty$ by $f(\pi, Y)^{-k} = a_{ij}b_{jY_{-k}}$ if $\pi = (A, B)$. From Corollaries 3.11 and 3.12 we conclude that

Corollary 3.13. *For every $Y \in R_{-\infty}$ and each $\pi \in \Pi_{\delta_1}$ $\lim_{n\to\infty} P^n(f(\pi, Y))_{ij} = P^\infty[\pi, Y]_j$ exists uniformly in the coordinates of π in Π_{δ_1}.*

Corollary 3.14. *For every $Y \in R_{-\infty}$ the function $P^\infty[\cdot, Y]_j$ is analytic in the coordinates of π.*

Proof. This follows from the fact that for every n, $P^n(f(\cdot, Y))_{ij}$ is analytic in the coordinates of π and the fact that $P^n(f(\cdot, Y))_{ij} \to P^\infty(f(\cdot, Y))_j$ uniformly.

Observe that if $\pi = (A, B)$ and the coordinates of π are real, then $P^\infty[\pi, Y]_j = P_\pi[W_0 = j \mid Y_0Y_{-1}Y_{-2}\cdots]$ so

Corollary 3.15. *For every $Y \in R_{-\infty}$, the function $P_\pi[W_0 = j \mid Y_0Y_{-1}Y_{-2}\cdots]$ is analytic for all $\pi \in \Pi_{\delta_1}$.*

Corollary 3.16. *For every $Y_1, Y_0, Y_{-1}, \cdots$ the function $P_\pi[Y_1 \mid Y_0, Y_{-1}, \cdots]$ is analytic for all $\pi \in \Pi_{\delta_1}$.*

Proof. $P_\pi[Y_1 \mid Y_0, Y_{-1}, \cdots] = \sum_{i,j} a_{ij}b_{jY_1}P_\pi[W_0 = i \mid Y_0, Y_{-1}\cdots]$.

By stationarity, for every $Y \varepsilon R_{-\infty}$, $P_\pi[Y_0 \mid Y_{-1}, Y_{-2} \cdots]$ is an analytic function of $\pi \varepsilon \Pi_{\delta_1}$.

Define $H_{\pi_0}(\pi) = E_{P_{\pi_0}}(\log P_\pi[Y_0 \mid Y_{-1} \cdots])$. The reader ought to check that this definition coincides with that of Theorem 2.1 (compare [1]). In [1], Corollary 4.3, it was shown that $H_n[\pi, Y] \to H_{\pi_0}(\pi)$ a.e. P_{π_0} and $H_{\pi_0}(\pi)$ is a C^3 function of the coordinates of π.

COROLLARY 3.17. *$H_{\pi_0}(\pi)$ is an analytic function of the coordinates of $\pi \varepsilon \Pi_{\delta_1}$.*

3. The critical point set $M'[\pi_0]$ of the function $H_{\pi_0}(\pi)$. Let $M'[\pi_0]$ be the set of critical points of $H_{\pi_0}(\pi)$ in the interior of $\Pi_{\delta_1} \cap \Pi_0$. If π_0 is in this set, then $M[\pi_0] = \mathfrak{S}_s\pi_0 \subset$ interior $\Pi_{\delta_1} \cap \Pi_0$ and $M[\pi_0] \subset M'[\pi_0]$ because $H_{\pi_0}(\pi)$ assumes its maximum on the set $M[\pi_0]$.

Optimally we seek to determine the set $M[\pi_0] = \{\pi \varepsilon \Pi_{\delta_1} \mid P_\pi = P_{\pi_0}\} = \mathfrak{S}_s\pi_0 = \{\pi' \mid H_{\pi_0}(\pi)$ is maximized at $\pi = \pi'\}$. One method of accomplishing this is to determine $M'[\pi_0]$ and then characterize the subset $M[\pi_0] \subset M'[\pi_0]$.

The reason for working so hard to establish the analyticity of $H_{\pi_0}(\pi)$ is this. The critical point set of an analytic function is in particular an analytic variety i.e., the simultaneously zero set of a finite number of analytic functions. In this case that set is the set of $\pi \varepsilon \Pi_{\delta_1}$ such that $(\partial/\partial\pi_i)H_{\pi_0}(\pi) = 0$. Now an analytic variety in Euclidean n space is by no means an arbitrary closed subset. We use the known structure theorems about analytic varieties to study the set $M'[\pi_0]$. We also warn the reader that it is important that the function $H_{\pi_0}(\pi)$ is analytic and not just C^∞. For in dim 1 an arbitrary closed subset of R' is the critical point set of a C^∞ function. This is far from true for an analytic function. I don't know the corresponding statements for the higher dimensional case of the critical point set of a C^∞ function but I suppose it's not any better.

With this motivation we set down a structure theorem for the set $M'[\pi_0]$. Let $M'[\pi_0, a] = M'[\pi_0] \cap \{\pi \mid H_{\pi_0}(\pi) = a\}$. Let $\bar{a} = H_{\pi_0}(\pi_0)$; so that $M[\pi_0] = M'[\pi_0, \bar{a}]$ and $M'[\pi_0, a]$ is empty if $a > \bar{a}$.

THEOREM 3.18. *For $\pi_0 \epsilon$ interior $\Pi_\delta \cap \Pi_0$, $M'[\pi_0] = \sum_{a<\bar{a}} M'[\pi_0, a] + M[\pi_0]$. The summation is finite. Each $M'[\pi_0, a]$ is an analytic variety. $\mathfrak{S}_s M'[\pi_0, a] = M'[\pi_0, a]$ and $M[\pi_0] = \mathfrak{S}_s\pi_0$. In particular the elements of $M[\pi_0]$ are isolated critical points of $H_{\pi_0}(\pi)$.*

PROOF. The reader is referred to the fine book by Gunning and Rossi [8] for the details of analytic varieties. The fact that the summation is finite is due to the fact that any compact set intersects only a finite number of connected components of an analytic variety and to the fact that an analytic function is constant on the connected components of its critical point set. (This statement is false for a C^2 function.) The fact that the points of $M[\pi_0]$ are isolated critical points follows from the fact that $M[\pi_0] = \mathfrak{S}_s\pi_0$ for $\pi_0 \varepsilon \Pi_0$ and thus is finite (Theorem 1.1) and from the fact that $H_{\pi_0}(\pi) > H_{\pi_0}(\pi')$ if $\pi \varepsilon M[\pi_0]$ and $\pi' \not\varepsilon M[\pi_0]$ (Theorem 2.2); thus $\pi' \varepsilon M'[\pi_0, a]$ for $a < \bar{a}$ and $M'[\pi_0, a]$ is separated from $M'[\pi_0, \bar{a}]$. The invariance of $M'[\pi_0, a]$ under the action of $\mathfrak{S}_s$ comes from a straight forward check.

4. The connection with the physical problem. Recall the fundamental property of the Baum-Eagan theorem discussed in the introduction. It is: $f[t_f(\pi)] \geqq f(\pi)$ and $f[t_f(\pi)] = f(\pi)$ iff π is a critical point of f. Also recall that $\mathfrak{J}$ contains the functions $H_n[\pi, Y]$.

An extremely useful application of the above theorem would be the conjecture mentioned in the preliminary section of this chapter. The obstacle to proving the conjecture is a lack of knowledge concerning the set of critical points of $H_{\pi_0}(\pi)$ in Π. Another complication is that this set is not finite. In fact if $\pi_0 = (A^0, B^0)$, $\mathbf{a}^0$ is a stationary vector for A^0 and B' is the $s \times r$ matrix with $B'_{jk} = b_k' = \sum_{i\varepsilon s} a_i b_{ik}^0$ then any π of the form $(A. B')$ with A an arbitrary stochastic matrix is a critical point of $H_{\pi_0}(\pi)$. Of course, this set of points has measured zero in Π would be included in Γ.

Here, however, is a positive step towards the theorem. There is an open set U_{π_0} containing $M[\pi_0]$ and contained in Π_δ such that if n is large and $\pi' \varepsilon U_{\pi_0}$ then the probability that $t^k_{H_n[\pi,Y]}[\pi']$ is near $M[\pi_0]$ for k large is large. Precisely: Let $d(\pi_1, \pi_2)$ be the Euclidean distance between π_1 and π_2 and set $M[\pi_0, \epsilon] = \{\pi' \varepsilon \Pi_\delta \mid d(\pi, \pi') < \epsilon \text{ for some } \pi \varepsilon M[\pi_0]\}$.

THEOREM 3.19. *Let $\pi_0 \varepsilon \Pi_0 \cap \Pi_\delta$. There exists an open set U_{π_0} containing $M[\pi_0]$ such that given $\epsilon > 0$ there is an $N(\epsilon) = N$ such that $P_{\pi_0}\{Y \mid \theta_n(\pi, Y) \varepsilon M[\pi_0, \epsilon]\} > 1 - \epsilon$ for $n > N(\epsilon)$ for all π contained in U_{π_0}.*

PROOF. Set $M'[\pi_0, \epsilon] = \{\pi' \varepsilon \Pi_{\delta_1} \mid d(\pi, \pi') < \epsilon \text{ for some } \pi' \varepsilon M'[\pi_0]\}$ and $M_n[\pi_0, Y] = \{\pi' \varepsilon \Pi_\delta \mid \pi \text{ maximize } H_n[\pi, Y]$ and $M_n'[\pi, Y] = \{\pi' \varepsilon \Pi_{\delta_1} \mid \pi' \text{ is a critical point of } H_n[\pi, Y]\}$. Let $a_1 = \sup$ over $\pi \varepsilon M'[\pi_0] - M[\pi_0]$ of $H_{\pi_0}(\pi)$ so that $a_1 < \bar{a}$. For positive $r > 1$, let $a_r = a_1 + (\bar{a} - a_1)/r$. Let $U_{\pi_0} = \{\pi \varepsilon \Pi_{\delta_1} \mid H_{\pi_0}(\pi) > a_3\}$, $U'_{\pi_0} = \{\pi \varepsilon \Pi_{\delta_1} \mid H_{\pi_0}(\pi) > a_5\}$; so $U_{\pi_0} \subset U'_{\pi_0}$. From [1], Corollary 4.3, it follows that given $\epsilon > 0$ there is an $N(\epsilon)$ such that if $n > N(\epsilon)$ and if the set C is defined by $C = \{Y \mid M_n'[\pi_0, Y] \subset M'[\pi_0, \epsilon]$ and $|H_n[\pi, Y] - H_{\pi_0}(\pi)| < \epsilon$ for all $\pi\}$, then $P(C) > 1 - \epsilon$. For the remainder of the proof we restrict Y to the set C and set $t = t_{H_n[\pi,Y]}$. Let $U[n, Y] = \{\pi \varepsilon \Pi_{\delta_1} \mid H_n[\pi, Y] > a_4\}$; thus, $U_{\pi_0} \subset U[n, Y] \subset U'_{\pi_0}$ for small ϵ. Also observe that if ϵ is sufficiently small all the critical points of $H_n[\pi, Y]$ in $U[n, Y]$ lie in $M[\pi_0, \epsilon]$ because $M_n'[\pi_0, Y] \subset M'[\pi_0, \epsilon]$, $U[n, Y] \subset U'_{\pi_0}$, and $U'_{\pi_0} \cap (\sum_{a<a_1} M'[\pi_0, \epsilon, a]) = \varphi$. Here $M'[\pi_0, \epsilon, a] = \{\pi \varepsilon \Pi_{\delta_1} \mid d(\pi, \pi') < \epsilon$ for some $\pi' \varepsilon M'[\pi_0, a]\}$. By the fundamental property of t, $tU[n, Y] \subset U[n, Y]$; thus $tU_{\pi_0} \subset U[n, Y] \subset U'_{\pi_0}$. Since $\theta_n[U_{\pi_0}, Y] \subset U[n, Y] \subset U'_{\pi_0}$ and since $\theta_n[U_{\pi_0}, Y]$ is contained in the critical point set of $H_n[\pi, Y]$ i.e., in $M_n'[\pi_0, Y]$ we have $\theta_n[U_{\pi_0}, Y] \subset M[\pi_0, \epsilon]$.

REMARKS. At present an explicit form for $N(\epsilon)$ in terms of $\pi_0 = (A, B)$ appears quite difficult. It is now determined by Monte Carlo methods for each application.

REFERENCES

[1] BAUM, L. E. and PETRIE, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.* **30** 1554–1563.

[2] BAUM, L. E. and EAGON, J. A. (1967). An inequality with applications to statistical

estimations for probabilistic functions of Markov processes and a model for ecology. *Bull. Amer. Math. Soc.* **73** 360–363.

[3] BAUM, L., PETRIE, T., GAINS, S. and SIMONS, J. (1968). *Stock Market Prediction.* To appear.

[4] BAUM, L., PETRIE, T., SOULES, GEORGE and WEISS, N. (1968). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. To appear.

[5] DOOB, J. L. (1953). *Stochastic Processes.* Wiley, New York.

[6] FURSTENBERG, H. and KESTEN, H. (1960). Products of random matrices. *Ann. Math. Statist.* **31** 457–469.

[7] GILBERT, E. J. (1959). On the identifiability problem for functions of finite Markov chains. *Ann. Math. Statist.* **30** 688–697.

[8] GUNNING, R. and ROSSI, H. (1965). *Analytic Functions of Several Complex Variables.* Prentice Hall, Englewood Cliffs.