PROBABILISTIC FUNCTIONS OF FINITE-STATE MARKOV CHAINS*

BY T. PETRIE

INSTITUTE FOR DEFENSE ANALYSES, VON NEUMANN HALL, PRINCETON, NEW JERSEY

Communicated by David Blackwell, January 9, 1967

These papers* are statistically motivated; the content is mathematical. The motivation is this: Given is an $s \times s$ stochastic matrix $A = ((a_{ij}))$ and an $s \times r$ stochastic matrix $B = ((b_{jk}))$ where A generates a stationary Markov process $\{X_t\}$ according to $a_{ij} = P[X_{t+1} = j | X_t = i]$ and B generates a process $\{Y_t\}$ described by $P[Y_t = k | X_t = j] = b_{jk}$, so if R is the set of integers $1, 2 \ldots r$ and $R^\infty = \Pi_{t-1}^{\infty} R_t$, $R_t = R$ (a point $Y \in R^\infty$ has coordinates Y_t), then the matrices A and B define a measure $P_{(A,B)}$ on R^∞, for $k_i \in R$

$$P_{(A,B)}\{Y_1 = k_1, Y_2 = k_2 \ldots Y_n = k_n\} = \Sigma_{i_0 = 1, i_1, i_2 \ldots i_n = 1}^{s \ldots s} a_{i_0} a_{i_0 i_1} b_{i_1 k_1} a_{i_1 i_2} b_{i_2 k_2} \ldots a_{i_{n-1} i_n} b_{i_n k_n}$$

where $\{a_{i_0}\}$ is the stationary absolute distribution for A. The resulting process $\{Y_t\}$ is called a probabilistic function of the Markov process $\{X_t\}$. Let Λ_1 be the space of $s \times s$ ergodic stochastic matrices, Λ_2 the space of $s \times r$ stochastic matrices and $\Pi = \Lambda_1 \times \Lambda_2$. The above associates to $\pi = (A, B) \in \Pi$ and a stationary vector a for A a measure P_π on R^∞.

The Problem.—Fix $\pi_0 \in \Pi$ and let a sample $Y_1, Y_2 \ldots Y_n$ be generated according to the distribution P_{π_0}. From the sample $Y \ldots Y_n$ obtain an estimate $\Pi_n(Y)$ of π_0 so that $\Pi_n(Y) \rightarrow \pi_0$ a.e., P_{π_0}. Throughout this paper π_0 is fixed and π varies in Π.

The Mathematics.—Part I (classification of equivalent processes) demonstrates that the problem has a solution in the following sense: Let $M[\pi_0] = \{\pi \in \Pi | P_\pi = P_{\pi_0}$ as measures on $R^\infty\}$. Clearly the points of $M[\pi_0]$ cannot 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 symmetric group of degree s operating on 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 $\sigma \in \mathfrak{S}_s$. Observe that $P_{\sigma\pi} = P_\pi$ as measures on R^∞. The main result of part I is

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

Part II (limit theorems and statistical analysis) extends and generalizes the results of reference 1. For each n and each $Y \in R^\infty$ there is a function $H_n[\pi, Y]$ on Π defined by $H_n[\pi, Y] = \dfrac{1}{n} \log P_\pi\{Y_1, Y_2 \ldots Y_n\}$; thus, each $H_n[\pi, \quad]$ is a random variable on the probability space (R^∞, P_{π_0}). 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. $\lim\limits_{n \to \infty} H_n[\pi, Y] = H_{\pi_0}(\pi)$ *exists a.e.,* P_{π_0}.

THEOREM 3. $H_{\pi_0}(\pi) \leq H_{\pi_0}(\pi_0)$ *and* $H_{\pi_0}(\pi) = H_{\pi_0}(\pi_0)$ *iff* $\pi \in M[\pi_0]$.

Define $\Pi_n(Y) = \{\pi' \in \Pi | \pi'$ maximizes $H_n[\pi, Y]\}$.

THEOREM 4. $\Pi_n(Y) \rightarrow M[\pi_0]$ *a.e.,* P_{π_0}.

Theorems 1 through 4 theoretically solve our problem. Note in particular the importance of the function $H_{\pi_0}(\pi)$ in view of Theorems 1 and 3.

Part III (Morse theory) makes a further study of the function $H_{\pi_0}(\pi)$ for $\pi \in \Pi_\delta =$

$\{(A,B) \in \pi | a_{ij} \geq \delta, b_{jk} \geq \delta\}$, $\delta > 0$ and ties the theory together with the following theorem of reference 2: $\exists$ a class of functions $\mathfrak{I}$ on Π such that if $f \in \mathfrak{I}$, there is a transformation $\tau_f: \Pi \rightarrow \Pi$ with the property that $f\tau_f(\pi) \geq f(\pi)$ and $f\tau_f(\pi) = f(\pi)$ iff π is a critical point of f. The class $\mathfrak{I}$ contains each $H_n[\pi,Y]$; thus, a procedure which is naturally suggested for dealing with the problem is: Given $Y_1 \ldots Y_n$, let $f = H_n[\pi,Y]$ and take $\Pi^n(\pi,Y) = \lim_{k \to \infty} \tau_f^k(\pi)$ for any $\pi \in \Pi_\delta$. (In great generality this limit exists; for complete validity let $\Pi^n(\pi,Y)$ be the accumulation points of $\{\tau_f^k(\pi)\}$.) How good is this estimate of π_0? The main theorem of part III answers this with

THEOREM 5. *Let $\pi_0 \in \Pi_0 \cap \Pi_\delta$. $\exists$ an open set U_{π_0} containing $M[\pi_0]$ such that given $\epsilon > 0$, $\exists$ $N(\epsilon) \ni P\{Y | \Pi^n(\pi',Y) \in M[\pi,\epsilon]\} > 1 - \epsilon$ for $n > N(\epsilon)$, $\forall$ $\pi' \in U_{\pi_0}$.* ($M[\pi_0,\epsilon]$ is the set of points of Π_δ whose Euclidean distance from some point of $M[\pi_0]$ is less than ϵ.)

The proof of Theorem 5 rests on a study of the critical points of $H_{\pi_0}(\pi)$, in particular,

THEOREM 6. *$H_{\pi_0}(\pi)$ is an analytic function of the coordinates of π for $\pi \in \Pi_\delta$.*

THEOREM 7. *The critical point set $M'[\pi_0]$ of $H_{\pi_0}(\pi)$ (as a function of π) is an analytic variety, and the points of $M'[\pi_0]$ which are absolute maxima, i.e., the points of $M[\pi_0]$, are isolated critical points if $\pi_0 \in \Pi_0 \cap \Pi_\delta$.*

Conjecture.—All critical points of $H_{\pi_0}(\pi)$ are nondegenerate for $\pi_0 \in \Pi_0 \cap \Pi_\delta$, and the critical points which are local maxima are precisely the elements of $M[\pi_0]$.

Corollary of Conjecture.—Theorem 5 holds for *any* $\pi' \in \Pi_\delta$.

* Petrie, T., "Probabilistic functions of finite-state Markov chains: I. Classification of equivalent processes; II. Limit theorems and statistical analysis; III. Morse theory," to appear.

[1] Baum, L. E., and T. Petrie, "Statistical inference for probabilistic functions of finite-state Markov chains," *Ann. Math. Stat.*, to appear.

[2] Baum, L. E., and J. A. Eagon, "An inequality with applications to statistical estimations for probabilistic functions of Markov processes and to a model for ecology," *Bull. Am. Math. Soc.*, to appear.

[3] Billingsley, Patrick, *Statistical Inference for Markov Processes* (Chicago, Illinois: University of Chicago Press, 1961).