The Annals of Mathematical Statistics
1970, Vol. 41, No. 1, 164–171

A MAXIMIZATION TECHNIQUE OCCURRING IN THE STATISTICAL ANALYSIS OF PROBABILISTIC FUNCTIONS OF MARKOV CHAINS

BY LEONARD E. BAUM, TED PETRIE, GEORGE SOULES, AND NORMAN WEISS

Institute for Defense Analyses, California Institute of Technology and Columbia University

1. Introduction. This paper lays bare a principle which underlies the effectiveness of an iterative technique which occurs in employing the maximum likelihood method in statistical estimation for probabilistic functions of Markov chains. We exhibit a general technique for maximizing a function $P(\lambda)$ when P belongs to a large class of probabilistically defined functions.

In the earlier note [2], an application of an inequality was made to ecology. The more general approach of this paper has allowed the use of a certain transformation and inequality in maximizing likelihood function in models for stock market behavior [3] and sunspot behavior. We also expect to apply these techniques to problems in weather prediction.

Let $A = (a_{ij})$ be an $s \times s$ stochastic matrix. Let $a = (a_i), i = 1, \cdots, s$ be a probability distribution. For each $i = 1, \cdots, s$ let $f_i(y)$ be a probability density: $\int f_i(y)\,dy = 1$. For the triple $A, a, f = \{f_i\}$ we define a stochastic process $\{Y_t\}$ with density

$$(1)\quad P(A,a,f)\{Y_1 = y_1, Y_2 = y_2 \cdots Y_T = y_T\} = \sum_{i_0,i_1,\cdots,i_T=1}^{s} a_{i_0} a_{i_0 i_1} f_{i_1}(y_1) a_{i_1 i_2} f_{i_2}(y_2) \cdots a_{i_{T-1} i_T} f_{i_T}(y_T).$$

For convenience we denote this expression by $P_{y_1 \cdots y_T}(A, a, f)$.

We call the process $Y = \{Y_t\}$ a probabilistic function of the Markov process $\{X_t\}$ determined by A. If a is chosen as a stationary distribution for the matrix A then Y will be a stationary stochastic process.

Let Λ be an open subset of Euclidean n space. Suppose that to each $\lambda \in \Lambda$, we have a smooth assignment $\lambda \to (A(\lambda), a(\lambda), f(\lambda))$. Specifically each $f_i(\lambda, \cdot)$ is a density in y and for each fixed y is a smooth function in λ. Under these assumptions, for each fixed $y_1, y_2, \cdots, y_T$, $P_{y_1 \cdots y_T}(\lambda) = P_{y_1 \cdots y_T}(A(\lambda), a(\lambda), f(\lambda))$ is a smooth function of λ. Given a fixed Y-sample $y = y_1, \cdots, y_T$ we seek a parameter value λ^0 which maximizes the likelihood $P_y(\lambda) = P_{y_1 \cdots y_T}(\lambda)$ determined from $A(\lambda)$, $a(\lambda)$, $f(\lambda)$ by (1).

One might suspect from the complicated nature of the expression (1) for $P_{y_1 \cdots y_T}(A, a, f)$ and the difficult analysis of maximizing this function of λ for very special choices of f presented in [2], [8] that a simple explicit procedure for maximization for a general f would be quite difficult; however, this is not the case. There is an extremely simple feature of this function which under mild hypothesis on f enable us to define a continuous transformation $\mathcal{T}$ mapping Λ into itself with

Received March 24, 1969.