PARAMETER ESTIMATION OF A PROBABILISTIC AUTOMATA MODEL OF DNA MEIOSIS

BY

KUAN CHAN WU

B.S., Chung Cheng Institute of Technology, Taiwan, 1963.

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements of the degree

MASTER OF SCIENCE

Department of Computer Science

Kansas State University
Manhattan, Kansas

1976.

Approved:

Major Professor

TABLE OF CONTENTS

ILLEGIBLE

THE FOLLOWING DOCUMENT (S) IS ILLEGIBLE DUE TO THE PRINTING ON THE ORIGINAL BEING CUT OFF

ILLEGIBLE

Acknowledgement

I wish to express my sincere gratitude to my major professor, Dr. Fred J. Maryanski for his guidance, creative suggestions and encouragement during the writting of this report.

LIST OF TABLES

LIST OF FIGURES

Introduction

Automata theory is the study of the dynamic behavior of discrete parameter information systems. Within this area, we define not only problems that deal with digital computers but also problems associated with such topics as describing the behavior of nerve networks, the represention of the properties of languages etc. (1)

A Probabilistic Automaton is an automaton in which the characteristics of the mapping can be described only in a probabilistic rather than a deterministic manner. (2)

The purpose of this report is to estimate parameters of a probabilistic Finite-state Automaton model to fit as closely as possible. The model DNA (Deoxynucleic Acid) Meiosis, with the sample data taken from experiments conducted by Mortimer (3).

Background

For this study we will need the following basic properties.

I. Formal grammars and probabilistic grammars.

Definition I - 1

Grammar: A formal grammar is an ordered quadruple

$G = (V_n, V_t, X_o, F)$

where

V_n - is a finite collection of variables or nonterminal symbols.

V_t - is a finite alphabet of terminal symbols such that $V_n \cap V_t = \phi$

X_o - is the initial letter (start sample) $X_o \in V_n$.

F - is a finite set of ordered pairs (P, Q) elements (P, Q) of F are called rewriting rules or productions and are written $P \rightarrow Q$, $P \in (V_n \cup V_t)^+$, $Q \in (V_n \cup V_t)^*$

P is the premise of the rule.

Q is the consequence of the rule.

$(V_n \cup V_t)^+ = (V_n \cup V_t)^* - \lambda$

λ is the empty string.

Note:

$(V_n \cup V_t)^*$ means any string of zero or more nonterminal or terminal symbols.