STOCHASTIC PROCESSES IN EPIDEMIOLOGY

This page is intentionally left blank

STOCHASTIC PROCESSES IN EPIDEMIOLOGY

**HIV/AIDS, Other Infectious
Diseases and Computers**

**Charles J Mode
Candace K Sleeman**

Drexel University, USA

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

STOCHASTIC PROCESSES IN EPIDEMIOLOGY

Copyright © 2000 by World Scientific Publishing Co. Pte. Ltd.

All rights reserved. This book, or parts thereof, may not be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or any information storage and retrieval system now known or to be invented, without written permission from the Publisher.

For photocopying of material in this volume, please pay a copying fee through the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923, USA. In this case permission to photocopy is not required from the publisher.

ISBN 981-02-4097-X

This book is printed on acid and chlorine free paper.

Printed in Singapore by World Scientific Printers

Dedication

To the idea that stochastic models, coupled with computer intensive methods, will in future play a significant role in man's quest to understand and control epidemics of infectious diseases.

This page is intentionally left blank

Preface

AIDS, acquired immune deficiency syndromes, is a devastating disease caused by HIV, human immunodeficiency virus, which may be transmitted by either sexual or other contacts in which body fluids are exchanged. Following a few recognized cases among homosexual men in the United States in the early 1980's, cases of AIDS were subsequently reported in a majority of countries throughout the world among heterosexuals, intravenous drugs users and others, indicating that HIV/AIDS was a global pandemic. It is indeed a pandemic of such proportions that it ranks as one of the most destructive microbial scourges in human history and has posed formidable challenges to the biomedical research and public health communities of the world. In response to these challenges, a voluminous biomedical literature on HIV/AIDS has been generated during the last two decades, including mathematical papers in numerous journals of applied mathematics, applied probability, biomathematics and biostatistics. When this book was first conceived, the original intention was to confine attention to those mathematical and statistical techniques underlying the models used to understand the dynamics of epidemics of HIV/AIDS as they developed in populations, with an emphasis of computer intensive methods. But as the development of ideas progressed, it became apparent that many of the techniques would be applicable to the populations dynamics of other infectious diseases. Consequently, the scope of the book was broadened to include other infectious diseases, although the main focus of the book has remained that of HIV/AIDS and other sexually transmitted diseases.

Mathematical models of epidemics of infectious diseases may be classified into two broad classes, deterministic and stochastic. Ordinary non-linear differential equations are among the principal tools used in the formulation of most deterministic models of epidemics, but when ages of individuals are accommodated in models, some authors have based their formulations on first order non-linear partial differential equations, belonging to the McKendrick-von Foerster class. Even though there is an extensive literature on deterministic models of epidemics, all deterministic models are incomplete in the sense that the variation and uncertainty that characterizes the development of most

epidemics in populations cannot be accommodated in purely deterministic formulations. A widely recognized need to accommodate this variation and uncertainty has given rise to a rather large literature on stochastic models of epidemics, which takes into account variability in the development of epidemics and quantifies the uncertainties as to what course an epidemic may take in terms of probabilities. Because, for the most part, the mathematics underlying stochastic formulations is more difficult to penetrate than that used in deterministic formulations, this difficulty has in the past proven to be a barrier to applying stochastic models in practical situations. But, with the help of computer intensive methods designed to compute sample realizations of an epidemic, practical illustrations of the variability inherent in the evolution of a stochastic process are provided, and the barriers to practical application may, in part, be removed.

Computer intensive methods, whose aim is to compute random samples from probability distributions or stochastic processes, are often referred to collectively as Monte Carlo simulation. Contained within the literature of epidemiology published during the last three decades, as well as other fields of the biological sciences, are numerous papers devoted to reporting the results of Monte Carlo simulation experiments designed to gain some understanding of the intrinsic variability and uncertainty inherent in the evolution of most biological phenomena. Although the intent of these papers is impeccable, most of them are unsatisfactory to the mathematical scientist because they usually lack a formal account of the mathematical structures underlying their computer experiments. This lack of any formal account of mathematical structures on which their simulations are based is a serious impediment to understanding and interpreting the results of their computer experiments, for it can be demonstrated by examples that the results one obtains in Monte Carlo simulation experiments can depend significantly on the assumptions going into the design of the stochastic process from which the sample of realizations was supposed to have come. In short, this lack of formal documentation of the stochastic model underlying some Monte Carlo simulation models is analogous to asking a chemist or experimental biologist to evaluate the credibility of a laboratory experiment in which only the final results are reported,

but the techniques used to obtain the results are withheld from his or her scrutiny so that any attempt to duplicate the results would be very difficult and time consuming.

One of the distinguishing features of this book is that a concerted attempt has been made to make clear the assumptions going into the design of the stochastic processes underlying all reported computer simulation experiments. No attempt has been made, however, to include listings of the computer code used in the implementations of the stochastic processes developed in this book. The reasons for this omission are two fold. Firstly, because it is a succinct and powerful medium in which to develop experimental code for not only computing Monte Carlo samples of realizations of stochastic processes, but also for computing informative summary statistics of these samples, such as the trajectories of order statistics, the programming language APL has been used to write the code for all the implementations of stochastic processes presented in this book. But, unfortunately, even though the authors have had over a decade of experience programming in this language, and there is a sizeable international community devoted to it, most readers would be unfamiliar with the succinct, but powerful notation that characterizes this language. Consequently, a second reason for not listing the computer code is that relatively few people would be able read it without a specialized knowledge of the symbols peculiar to APL. In principle, however, because an attempt has been made to carefully outline all computational procedures, the experimental results could be duplicated, using such widely used programming languages as C++, FORTRAN, or even MATLAB, a computer software package with numerous capabilities that is being used by an increasing number of engineers and scientists.

Another distinguishing feature of this book is that deterministic and stochastic models are not discussed in isolation, but are presented in a framework that synthesizes the two approaches in formulating models of epidemics. This synthesis is accomplished by systematically using a scheme of embedding deterministic models in a stochastic process by taking conditional expectations of present values of the sample functions, given the past, and using other approximation schemes borrowed from theories of statistical estimation. An advantage of this embedding

scheme is that it allows for comparison of the trajectory of some feature of an epidemic based on the embedded deterministic model with the corresponding trajectories of statistical summaries of a sample of Monte Carlo realizations of the process. Included in these summaries are the trajectories of chosen quantiles and the mean trajectory of a Monte Carlo sample. By inspecting graphs of these summary trajectories, it becomes possible to assess to what degree a projection of an epidemic made solely on the basis of a deterministic model would be misleading or under what circumstances a deterministic projection may be adequate. Background and motivational material for developing deterministic and stochastic models within a unified formulation amenable to using computer intensive methods is considered in Chapters 1 through to 9, but, for the most part, computer intensive methods were used extensively only in Chapters 10, 11 and 12. Consequently, an overview of the themes used in the development of these chapters will be provided.

Four themes underlie the development of the new non-linear stochastic models of epidemics of HIV/AIDS and other sexually transmitted diseases presented in Chapters 10, 11, and 12, which accommodate one or more risk groups or behavioral classes and states of disease in definitions of the types of individuals. The same themes will underlie the age dependent models outlined in Chapter 13, which are extensions of the model developed in Chapter 12. As yet, however, the rather difficult task of developing software for these more complex models has not been undertaken. The first of these themes is to define state spaces for semi-Markovian life cycle models, which include types of individuals, along with matrices of competing latent risks, governing transitions among states. For the case of the one-sex model presented in Chapter 10, the life cycle model pertains to the evolution of individuals in the population, but, for models accommodating marital partnership or couples in Chapters 11 and 12, life cycle models for singles and couples are included. Among the latent risks in all life cycle models are those governing the infection of susceptibles due to sexual contacts with infected individuals.

Whether a susceptible individual becomes infected during any time interval depends upon his or her choices of sexual partners among

the types of individuals present in the population at any time. Consequently, a second theme that underlies the models presented in Chapters 10, 11 and 12 is that of the use of an acceptance probability, which is a parametric function expressing the probability that a person of one type finds another acceptable as a sexual partner. In order for a sexual contact to occur, both partners must find each other acceptable, and by using set of probability arguments, including the law of total probability and Bayes' theorem, it is possible to express the latent risk that a susceptible person becomes infected during any time interval as non-linear functions of the sample functions of the process at any time, as well as functions of the parameters of the acceptance probabilities. By varying these parameters, it is possible to consider random and highly assortative mixing patterns in choosing sexual partners. For the partnership models of Chapters 11 and 12, the idea of acceptance probabilities is also used to model couple formation.

A third theme common to Chapters 10, 11, and 12 is the use of matrices of competing latent risks in the life cycle models to systematically compute vectors of conditional probabilities for chains of multinomial distributions that are used in the recursive computation of Monte Carlo realizations of the sample functions of the process on a time lattice of equally spaced points. Such computational schemes are sometimes referred to as chain multinomial models. According to these models, given values of the sample functions at some time point on the lattice, the conditional distribution of the sample functions at the next point in time is a multinomial distribution whose probabilities depend on the values of the sample functions of the preceding point in time. By taking conditional expectations of these vectors of multinomial random variables, given the past, along with other operations, it becomes possible to systematize procedures for embedding non-linear difference equations in the stochastic epidemic models on the discrete time lattice. Given a numerical specification of any point in the parameter space of the model and a set of initial conditions, it is then possible to compute trajectories describing various aspects of the evolution of an epidemic as it would develop in time according to the embedded deterministic model. Furthermore, by using computer generated graphs, the trajectories determined by the embedded difference equations may be

compared visually with various statistical trajectories, such as selected quantiles, summarizing a Monte Carlo sample computed according to a chain multinomial model.

Unlike many stochastic models of epidemics, branching process approximations cannot be used to derive threshold conditions of the non-linear models of Chapter 10, 11, and 12, particularly in those cases where the parameters in the acceptance probabilities are chosen such that the selection of sexual partners is highly assortative. Consequently, it became necessary to find other approaches to finding threshold conditions for the stochastic models. Let $h > 0$ denote the distance between any two points in a discrete time lattice. Then, by letting $h \downarrow 0$, it can be shown, under rather general conditions, that the embedded non-linear difference equations give rise to a system of ordinary non-linear differential equations. Thus, a fourth theme common to Chapters 10, 11, and 12 is that of finding threshold conditions by deriving formulas for the partial derivatives that arise as elements in the Jacobian matrix for the embedded differential equations, and testing whether this matrix is stable or unstable when evaluated at a stationary population state vector for the case where the population is free of infected individuals. The stability or non-stability of this Jacobian matrix provides a useful indicator as to whether an epidemic will or will not develop according to a stochastic model, following the introduction of a few infectives into a population of susceptibles, and this is demonstrated empirically in the computer experiments reported in these chapters. Because many types of individuals in a population give rise to large Jacobian matrices, and the parameter spaces of the models are of high dimensions, it was not practically feasible to derive stability conditions symbolically. Consequently, software was written to compute Jacobian matrices and their eigenvalues numerically at any parameter point so that it could be numerically determined whether all real parts of the eigenvalues were negative. Furthermore, search engines were written to explore what regions in a parameter space would yield stable or unstable Jacobian matrices.

Among the countless numbers of numerical examples that could be have been chosen to illustrate biologically interesting realizations of the stochastic models, only a select sample of experiments was chosen

for presentation, due to space limitations. It seems appropriate, therefore, to offer an explanation as to why these computer experiments were chosen for presentation. A theme common to most of the reported computer experiments was that the evolution of an epidemic was started by either the invasion of a few initial infectives into a population of susceptibles in demographic equilibrium, or by a recurrent stream of infective recruits that could enter the population during any time interval with low probability. In such experiments, it was observed that trajectories of the epidemic determined by the embedded deterministic model were often representative of only the worst cases of the epidemic that would occur in a Monte Carlo sample, especially in those cases where there was a positive probability of extinction or infective recruits entered the population with low probability during any time interval. Thus, if investigators confined their attention to using only deterministic models to forecast an epidemic, they could be seriously misled as to the severity of an epidemic. Epidemics of HIV/AIDS develop slowly in populations, and it seems reasonable to suppose that in many parts of the world, the pandemic is still in its early stages. This supposition motivated the selection of most of the computer experiments that have been presented in an attempt to highlight the importance of taking stochastic fluctuations into account in forecasting the future course of an epidemic still in its early stages.

Mention should also be made of the results of some computer experiments that were not reported due to space limitations. Among these experiments were those in which it was assumed that an epidemic had reached maturity in a population and that strategies of prevention to rid the population of infectives over time had been implemented. The effectiveness of a set of such strategies could be expressed by assigning values to parameters and testing whether the Jacobian matrix of the embedded differential equations was stable at a parameter point in question, when evaluated at a stationary vector for a population containing only susceptibles. By adjusting such parameters as the probability that a susceptible was infected per sexual contact with an infective, as well as those in acceptance probabilities so that a susceptible finds an infective acceptable as a sexual partner with low probability, one could attain stability of the Jacobian matrix, which would suggest

that eventually the epidemic would become extinct. However, it was found in a number of experiments that if an investigator relied solely on the embedded deterministic model to forecast the evolution of an epidemic or its path to extinction, he or she could be seriously misled if the stochastic fluctuations exhibited in a sample of Monte Carlo projections were ignored. In conclusion, based on the evidence collected from the large number of computer experiments conducted so far, significant stochastic fluctuations will occur when a population undergoes a transition from one point of equilibrium to another. Indeed, it is this transient behavior with its stochastic fluctuations, rather than the existence of points equilibria, that produces the phenomena of most interest in computer experiments designed to study the development and control of epidemics of infectious diseases.

Acknowledgments

A number of people have provided help, encouragement and inspiration while the authors were working on this book. Among them is Guenther Hasibeder, who in 1994 and 1996 invited the senior author, C.J.M., to give a series lectures on Stochastic Processes in Epidemiology in the Institute of Algebra and Computational Mathematics of the Technical University of Vienna. Words of thanks are also due Dietmar Dornenger, Head of the Department of Algebra, and his colleagues, who with customary Viennese hospitality and camaraderie, made the author's visits to Austria most rewarding and pleasant. It was during these visits that most of the material presented in Chapters 6 and 7 was developed.

Three professional colleagues have read initial drafts of Chapters 6 and 7 and have offered constructive criticisms. John Jacquez, a late Professor Emeritus of the University of Michigan Medical School, read a draft of Chapter 6 and made a number of valuable suggestions. Frank Ball, Department of Mathematics, University of Nottingham, United Kingdom, read drafts of Chapters 6 and 7 and offered a number of valuable technical and historical comments that improved the presentation. Finally, Ingemar Nasell, Department of Mathematics, The Royal Institute of Technology, Stockholm, Sweden, read Chapter 7 and made a number of insightful comments that were included in the final version of the chapter.

Words of thanks are due also to Ms. Sook Chen Lim of World Scientific Publishing Company, who read preliminary drafts of the book and made many useful annotations, pointing out grammatical and typographical errors that were subsequently corrected.

The senior author, C.J.M., also wishes to extend a special word of thanks to his wife Eleanore, who with patience and forbearance agreed to postpone recreational travel until work on the book was completed.

The junior author, C.K.S., wishes to thank her father Richard Sleeman, Professor Emeritus of the Massachusetts College of Liberal Arts, for all his support and counsel, as well as her husband Ralph Fife for all his love and encouragement.

This page is intentionally left blank

Contents

1	Biology and Epidemiology of HIV/AIDS	1
1.1	Introduction	1
1.2	Emergence of a New Disease	1
1.3	A New Virus as a Causal Agent	2
1.4	On the Evolutionary Origins of HIV	4
1.5	AIDS Therapies and Vaccines	7
1.6	Clinical Effects of HIV Infection	10
1.7	An International Perspective of the AIDS Epidemic	12
1.8	Evolution of Antibiotic Resistance	16
1.9	Mathematical Models of the HIV/AIDS Epidemic	18
1.10	References	20
2	Models of Incubation and Infectious Periods	23
2.1	Introduction	23
2.2	Distribution Function of the Incubation Period	23
2.3	The Weibull and Gamma Distributions	25
2.4	The Log-Normal, Log-Logistic and Log-Cauchy Distributions	29
2.5	Quantiles of a Distribution	32
2.6	Some Principles and Results of Monte Carlo Simulation	37
2.7	Compound Distributions	42
2.8	Models Based on Symptomatic Stages of HIV Disease	47

2.9	CD4⁺ T Lymphocyte Decline	53
2.10	Concluding Remarks	57
2.11	References	58
3	Continuous Time Markov and Semi-Markov Jump Processes	60
3.1	Introduction	60
3.2	Stationary Markov Jump Processes	61
3.3	The Kolmogorov Differential Equations	64
3.4	The Sample Path Perspective of Markov Processes	70
3.5	Non-Stationary Markov Processes	74
3.6	Models for the Evolution of HIV Disease	80
3.7	Time Homogeneous Semi-Markov Processes	86
3.8	Absorption and Other Transition Probabilities	95
3.9	References	100
4	Semi-Markov Jump Processes in Discrete Time	102
4.1	Introduction	102
4.2	Computational Methods	103
4.3	Age Dependency with Stationary Laws of Evolution	110
4.4	Discrete Time Non-Stationary Jump Processes	118
4.5	Age Dependency with Time Inhomogeneity	123
4.6	On Estimating Parameters From Data	127
4.7	References	129
5	Models of HIV Latency Based on a Log-Gaussian Process	131
5.1	Introduction	131
5.2	Stationary Gaussian Processes in Continuous Time	131
5.3	Stationary Gaussian Processes in Discrete Time	140
5.4	Stationary Log-Gaussian Processes	147
5.5	HIV Latency Based on a Stationary Log-Gaussian Process	150

5.6	HIV Latency Based on the Exponential Distribution	157
5.7	Applying the Model to Data in a Monte Carlo Experiment	159
5.8	References	166
6	The Threshold Parameter of One-Type Branching Processes	168
6.1	Introduction	168
6.2	Overview of a One-Type CMJ-Process	170
6.3	Life Cycle Models and Mean Functions	175
6.4	On Modeling Point Processes	180
6.5	Examples with a Constant Rate of Infection	185
6.6	On the Distribution of the Total Size of an Epidemic	192
6.7	Estimating HIV Infectivity in the Primary Stage of Infection	199
6.8	Threshold Parameters for Staged Infectious Diseases	201
6.9	Branching Processes Approximations	208
6.10	References	215
7	A Structural Approach to SIS and SIR Models	218
7.1	Introduction	218
7.2	Structure of SIS Stochastic Models	219
7.3	Waiting Time Distributions for the Extinction of an Epidemic	225
7.4	Numerical Study of Extinction Time of Logistic SIS	232
7.5	An Overview of the Structure of Stochastic SIR Models	237
7.6	Algorithms for SIR-Processes with Large State Spaces	244
7.7	A Numerical Study of SIR-Processes	251
7.8	Embedding Deterministic Models in SIS-Processes	255

7.9	Embedding Deterministic Models in SIR-Processes	262
7.10	Convergence of Discrete Time Models	268
7.11	References	271
8	Threshold Parameters For Multi-Type Branching Processes	274
8.1	Introduction	274
8.2	Overview of the Structure of Multi-Type CMJ-Processes	275
8.3	A Class of Multi-Type Life Cycle Models	279
8.4	Threshold Parameters for Two-Type Systems	286
8.5	On the Parameterization of Contact Probabilities	292
8.6	Threshold Parameters for Malaria	295
8.7	Epidemics in a Community of Households	302
8.8	Highly Infectious Diseases in a Community of Households	309
8.9	References	314
9	Computer Intensive Methods for the Multi-Type Case	316
9.1	Introduction	316
9.2	A Simple Semi-Markovian Partnership Model	317
9.3	Linking the Simple Life Cycle Model to a Branching Process	320
9.4	Extinction Probabilities for the Simple Life Cycle Model	326
9.5	Computation of Threshold Parameters for the Simple Model	329
9.6	Extinction Probabilities and Intrinsic Growth Rates	332
9.7	A Partnership Model for the Sexual Transmission of HIV	333
9.8	Latent Risks for the Partnership Model of HIV/ AIDS	336

9.9	Linking the Partnership Model to a Branching Process	340
9.10	Some Numerical Experiments with the HIV Model	342
9.11	Stochasticity and the Development of Major Epidemics	347
9.12	On Controlling an Epidemic	354
9.13	References	356
10	Non-linear Stochastic Models in Homosexual Populations	357
10.1	Introduction	357
10.2	Types of Individuals and Contact Structures . . .	358
10.3	Probabilities of Susceptibles Being Infected . . .	362
10.4	Semi-Markovian Processes as Models for Life Cycles	365
10.5	Stochastic Evolutionary Equations for the Population	369
10.6	Embedded Non-linear Difference Equations . . .	372
10.7	Embedded Non-linear Differential Equations . . .	376
10.8	Examples of Coefficient Matrices	379
10.9	On the Stability of Stationary Points	385
10.10	Jacobian Matrices in a Simple Case	392
10.11	Jacobian Matrices in a More Complex Case . . .	395
10.12	On the Probability an Epidemic Becomes Extinct	400
10.13	Software for Testing Stability of the Jacobian . .	405
10.14	Invasion Thresholds: One-Stage Model, Random Assortment	410
10.15	Invasion Thresholds: One-Stage Model, Positive Assortment	421
10.16	Recurrent Invasions by Infectious Recruits . . .	432
10.17	References	443
11	Stochastic Partnership Models in Homosexual Populations	445

11.1	Introduction	445
11.2	Types of Individuals and Partnerships	447
11.3	Life Cycle Model for Couples with One Behavioral Class	450
11.4	Couple Types for Two or More Behavioral Classes	455
11.5	Couple Formation	459
11.6	Probabilities of Being Infected by Extra-Marital Contacts	462
11.7	Stochastic Evolutionary Equations for the Population	466
11.8	Embedded Non-linear Difference Equations	471
11.9	Embedded Non-linear Differential Equations	473
11.10	Examples of Coefficient Matrices for One Behavioral Class	478
11.11	Stationary Vectors and Structure of the Jacobian Matrix	481
11.12	Overview of the Jacobian for Extra-Marital Contacts	489
11.13	General Form of the Jacobian for Extra-Marital Contacts	496
11.14	Jacobian Matrix for Couple Formation	506
11.15	Couple Formation for Cases $m \geq 2$ and $n \geq 2$	516
11.16	Invasion Thresholds for $m = 2$ and $n = 1$	522
11.17	Invasion Thresholds of Highly Sexually Active Infectives	527
11.18	Mutations and the Evolution of Epidemics	536
11.19	References	544
12	Heterosexual Populations with Partnerships	545
12.1	Introduction	545
12.2	Types of Individuals and Partnerships	547
12.3	Matrices of Latent Risks for Life Cycle Models	549
12.4	Marital Couple Formation	557
12.5	Probabilities of Being Infected by Extra-Marital Contacts	562

12.6 Stochastic Evolutionary Equations	567
12.7 Embedded Non-Linear Difference Equations . . .	572
12.8 Embedded Non-Linear Differential Equations . .	575
12.9 Coefficient Matrices for the Two-Sex Model . . .	583
12.10 The Jacobian Matrix and Stationary Points . . .	588
12.11 Overview of the Jacobian for Extra-Marital Contacts	593
12.12 General Form of the Jacobian for Extra-Marital Contacts	602
12.13 Jacobian Matrix for Couple Formation	614
12.14 Couple Formation for $m \geq 2$ and $n \geq 2$	623
12.15 Invasion Thresholds for $m = n = 1$	631
12.16 Four-Stage Model Applied to Epidemics of HIV/AIDS	640
12.17 Highly Active Anti-Retroviral Therapy of HIV/AIDS	649
12.18 Epidemics of HIV/AIDS Among Senior Citizens	656
12.19 Invasions of Infectives for Elderly Heterosexuals	662
12.20 Recurrent Invasions of Infectious Recruits	670
12.21 References	679
13 Age-Dependent Stochastic Models with Partnerships	681
13.1 Introduction	681
13.2 Parametric Models of Human Mortality	685
13.3 Latent Risks for Susceptible Infants and Adolescents	694
13.4 Couple Formation in a Population of Susceptibles	700
13.5 Births in a Population of Susceptibles	704
13.6 Latent Risks with Infectives	709
13.7 References	716
14 Epilogue - Future Research Directions	718
14.1 Modeling Mutations in Disease Causing Agents .	718
14.2 References	722