Spring 5-31-2011

Markovian and stochastic differential equation based approaches to computer virus propagation dynamics and some models for survival distributions

Lianzhe Xu

New Jersey Institute of Technology

Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

MARKOVIAN AND STOCHASTIC DIFFERENTIAL EQUATION BASED APPROACHES TO COMPUTER VIRUS PROPAGATION DYNAMICS AND SOME MODELS FOR SURVIVAL DISTRIBUTIONS

by
Lianzhe Xu

This dissertation is divided in two Parts. The first Part explores probabilistic modeling of propagation of computer 'malware' (generally referred to as 'virus') across a network of computers, and investigates modeling improvements achieved by introducing a random latency period during which an infected computer in the network is unable to infect others. In the second Part, two approaches for modeling life distributions in univariate and bivariate setups are developed.

In Part I, homogeneous and non-homogeneous stochastic susceptible-exposed-infectious-recovered (SEIR) models are specifically explored for the propagation of computer virus over the Internet by borrowing ideas from mathematical epidemiology. Large computer networks such as the Internet have become essential in today's technological societies and even critical to the financial viability of the national and the global economy. However, the easy access and widespread use of the Internet makes it a prime target for malicious activities, such as introduction of computer viruses, which pose a major threat to large computer networks. Since an understanding of the underlying dynamics of their propagation is essential in efforts to control them, a fair amount of research attention has been devoted to model the propagation of computer viruses, starting from basic deterministic models with ordinary differential equations (ODEs) through stochastic models of increasing realism.

In the spirit of exploring more realistic probability models that seek to explain the time dependent transient behavior of computer virus propagation by exploiting the essential stochastic nature of contacts and communications among computers, the present study introduces a new refinement in such efforts to consider the suitability

and use of the stochastic SEIR model of mathematical epidemiology in the context of computer viruses propagation. We adapt the stochastic SEIR model to the study of computer viruses prevalence by incorporating the idea of a latent period during which computer is in an 'exposed state' in the sense that the computer is infected but cannot yet infect other computers until the latency is over. The transition parameters of the SEIR model are estimated using real computer viruses data. We develop the maximum likelihood (MLE) and Bayesian estimators for the SEIR model parameters, and apply them to the 'Code Red worm' data.

Since network structure can be a possibly important factor in virus propagation, multi-group stochastic SEIR models for the spreading of computer virus in heterogeneous networks are explored next. For the multi-group stochastic SEIR model using Markovian approach, the method of maximum likelihood estimation for model parameters of interest are derived. The method of least squares is used to estimate the model parameters of interest in the multi-group stochastic SEIR-SDE model, based on stochastic differential equations. The models and methodologies are applied to Code Red worm data.

Simulations based on different models proposed in this dissertation and deterministic/stochastic models available in the literature are conducted and compared. Based on such comparisons, we conclude that (i) stochastic models using SEIR framework appear to be relatively much superior than previous models of computer virus propagation – even up to its saturation level, and (ii) there is no appreciable difference between homogeneous and heterogeneous (multi-group) models. The 'no difference' finding of course may possibly be influenced by the criterion used to assign computers in the overall network to different groups. In our study, the grouping of computers in the total network into subgroups or, clusters were based on their geographical location only, since no other grouping criterion were available in the Code Red worm data.

Part II covers two approaches for modeling life distributions in univariate and bivariate setups. In the univariate case, a new partial order based on the idea of 'star-shaped functions' is introduced and explored. In the bivariate context; a class of models for joint lifetime distributions that extends the idea of univariate proportional hazards in a suitable way to the bivariate case is proposed. The expectation-maximization (EM) method is used to estimate the model parameters of interest. For the purpose of illustration, the bivariate proportional hazard model and the method of parameter estimation are applied to two real data sets.

MARKOVIAN AND STOCHASTIC DIFFERENTIAL EQUATION BASED APPROACHES TO COMPUTER VIRUS PROPAGATION DYNAMICS AND SOME MODELS FOR SURVIVAL DISTRIBUTIONS

by
Lianzhe Xu

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark

May 2011

APPROVAL PAGE

MARKOVIAN AND STOCHASTIC DIFFERENTIAL EQUATION BASED APPROACHES TO COMPUTER VIRUS PROPAGATION DYNAMICS AND SOME MODELS FOR SURVIVAL DISTRIBUTIONS

Lianzhe Xu

Manish C. Bhattacharjee, Ph.D., Dissertation Advisor — Date
Professor of Mathematical Sciences, NJIT

Reza Curtmola, Ph.D., Committee Member — Date
Assistant Professor of Computer Sciences, NJIT

Sunil K. Dhar, Ph.D., Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Aridaman K. Jain, Ph.D., Committee Member — Date
Senior University Lecturer of Mathematical Sciences, NJIT

Sundarraman Subramanian, Ph.D., Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

BIOGRAPHICAL SKETCH

Author: Lianzhe Xu

Degree: Doctor of Philosophy

Date: May 2011

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences,
  New Jersey Institute of Technology, Newark, NJ, 2011
- Master of Science in Applied Statistics,
  New Jersey Institute of Technology, Newark, NJ, 2006
- Bachelor of Science in Mathematics,
  Capital Normal University, Beijing, China, 2001

Major: Mathematical Sciences (Applied Probability and Statistics)

Presentations and Publications:

Lianzhe Xu, Manish C. Bhattacharjee, "Superiority of Stochastic SEIR Model for the Propagation of Computer Viruses," *The Dona Knox Student Research Showcase*, NJIT, April 6, 2011.

Lianzhe Xu, Manish C. Bhattacharjee, "Superiority of Stochastic SEIR Model for the Propagation of Computer Viruses," *The VI^{th} Annual Graduate Student Association Research Day*, NJIT, November 4, 2010.

Lianzhe Xu, "Worm Propagation Modeling and Analysis Using a Stochastic Epidemic Model," *Applied Mathematics Seminar*, Department of Mathematical Sciences, NJIT, July 28, 2009.

To my family

ACKNOWLEDGMENT

I feel extremely lucky that I met Professor Manish Bhattacharjee during my studies for a M.S. in Applied Statistics at New Jersey Institute of Technology. Apart from being a great teacher, Professor Bhattacharjee was the first person who advised me to consider undertaking doctoral studies, and in particular to come to the 'Applied Probability and Statistics' track of the Ph.D. (Mathematical Sciences) program at NJIT. I would like to express my deep and sincere gratitude to him for his insightful guidance and continuous encouragement throughout my acadimic sojourn at NJIT. His way of advising has been very effective for me, as he guided me to keep my efforts focused in the right direction without constraining my thoughts and ideas. He has been generous with his time when I had questions or concerns regarding my research, and also about my writing in English, for which I feel greatly thankful.

I would also like to extend my gratitude to the other members of my dissertation committee: Professor Reza Curtmola, Professor Sunil Dhar, Professor Aridaman Jain and Professor Sundarraman Subramanian for their encouragement and support throughout the process of research for this dissertation. In particular, I would like to thank Professor Reza Curtmola, of the faculty of Computer Science at NJIT, for his constructive criticisms. His comments, at different stages of this work, on the pragmatic aspects of computer malware propagation that are relevant for modeling purposes have been very useful in ensuring that our model explorations meet reasonable standards of realism.

I am also grateful to Dr. Ivan Zorych, a friend and an alumnus of the Department's doctoral program, for his encouragement, support and advice regarding Markov chain Monte Carlo simulation methods, which we have used.

I thank the chairperson, Dr. Daljit Ahluwalia, of Department of Mathematical Sciences, for financial support and encouragement. Many thanks are also due to the former and current office staff: Ms. Susan Sutton, Ms. Padma Gulati, Ms. Eileen

Michie, Ms. Fatima Ejallali and Ms. Judy Robinson. Their help and support made my life at NJIT easier, in many ways. I would like to express my sincere gratitude to my co-students and friends Jing Li, Rudrani Banerjee, and Hui Wu for their support and advice.

My acknowledgement would not be complete without recognizing the role of my family in the persuit of my academic ambitions. I am very much indebted to my husband, Junfeng Liu, for always being there for me and for giving me great support and encouragement. I don't even know how I would have completed the research without his constant love and belief in me. I would also like to thank my two beautiful boys: Bodong and Botian. My heartfelt thanks go to my parents and my parents-in-law for their kindness, love, patience, understanding, and support over the years.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF TABLES
(Continued)

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

OVERVIEW

1.1 Motivation

Since the advent of the world wide web, the Internet continues to play an increasingly important role in modern society. However, the easy access to the web has made all participating computers on the net a prime target for malicious attacks. The continuing upsurge in the incidents of such attacks in the form of computer viruses, worms and other 'malware' have become a major problem for large computer networks, which then requires considerable amounts of resources and time to be spent recovering from large-scale attacks. It is believed that a good and reliable model is necessary to analyze the characteristics of computer virus propagation over the Internet so that the efficient human countermeasures can be implemented in time. So far, available studies to understand the characteristics of computer virus prevalence have employed epidemiological models which are *either* deterministic *or* relatively simple stochastic epidemic models. Thus, motivated by the need for more realistic models to better understand the underlying dynamics of computer virus propagation, we explore adapting stochastic susceptible-exposed-infectious-recovered (SEIR) models by incorporating the idea of a latent period during which the virus in an infected computer remains dormant. The corresponding analysis of the propagation of a computer virus in both homogeneous and heterogeneous networks are considered.

The other topic of research presented here concern some aspects of survival distribution models in reliability. Comparisons between survival distribution are often facilitated via suitable partial orderings that have been responsible for the development of several nonparametric aging classes in reliability theory. Motivated by such considerations, we introduce and explore *a new partial ordering*. Its properties

among univariate survival distributions which are so ordered are introduced and investigated.

Proportional hazards among univariate survival distributions are well known and have played an important role in reliability and survival analysis. In the case of two component lifetimes which need not be independent, our work on *bivariate proportional hazard models* presented here is motivated by the observation that there is no unique way of extending the idea of proportional hazards to a bivariate setup; although there are different approaches to such generalization and corresponding results that are available in the literature. We introduce a framework for a new formulation of bivariate proportional hazard models that is different from the existing approaches and explore its properties, basic distribution theory and estimation of parameters along with its illustration to two real life data sets.

1.2 Outline of the Dissertation

The contents of this dissertation are organized as follows:

The introduction to Part I, stochastic epidemic models and inference for the propagation of computer virus, is presented in Chapter 2. Chapter 3 reviews computer virus propagation models in the literature and presents an investigation of epidemic models and associated statistical inference techniques that have been described in the literature. Moreover, the background of the '*Code Red worm*' and the previous works on its modeling and analysis are summarized.

In Chapter 4, stochastic *susceptible-exposed-infectious-recovered* (SEIR) model using a Markovian approach is adapted to the context of computer virus propagation in homogeneous networks. With a completely observed computer virus epidemic process, (i.e., both infection times and recovery times are observed during the epidemic process), the methods of maximum likelihood and Bayesian estimation are used to estimate the model parameters of interest. The proposed stochastic SEIR model and the methods of parameter estimations are applied to the *Code Red worm* data.

In reality, a computer virus propagation process is unlikely to be observed completely. For example, in many cases the actual times when computers are infected may not be available. With such missing infection times, methods of parameter estimations discussed in Chapter 4 cannot be implemented. In Chapter 5, we explore Markov chain Monte Carlo (MCMC) methods for Bayesian inference to deal with the parameter estimations in cases where the computer virus propagation is only partially observed. The proposed method is applied to Code Red worm data.

It has been shown that the network structure has impact on computer virus propagation [66, 67, 86]. Chapter 6 presents a multi-group stochastic SEIR model, using a Markovian approach for computer virus prevalence in heterogenous networks. Maximum likelihood estimators of model parameters of interest are developed. Furthermore, for multi-group stochastic SEIR setup, we develop a new model based on stochastic differential equations (SDE). The models and methods are comparatively applied to Code Red worm data.

Conclusion and a comparative discussion of the models and methodologies described are given in Chapter 7. Possible extensions of the research presented in Part I are indicated.

Chapter 8 introduces the broad theme in Part II of this dissertation, dealing with some aspects of survival distribution models, followed by a brief description of expectation-maximization (EM) algorithm used for estimations of parameters of the bivariate models introduced in Chapter 10.

Chapter 9 explores a new partial ordering called reverse star-ordering among life distributions on the half line $[0, \infty)$. We investigate its properties including the relationship of this ordering to several other partial orderings and non-parametric life distribution classes, and its preservation properties under some reliability operations.

Proportional hazard model plays an important role in reliability and survival analysis. Modeling the joint distribution of possibly mutually dependent lifetimes which may be considered as multivariate versions of the univariate proportional

hazard idea, for which there is no unique way to formulate such a notion, is our final research theme. In Chapter 10, a new bivariate proportional hazard model is introduced and its distribution theory is explored. In most of cases of such distributions, the maximum likelihood estimators cannot be expressed in closed explicit forms. We describe an expectation-maximization (EM) algorithm to compute the maximum likelihood estimators of the unknown parameters.

Chapter 11 concludes with a discussion and summary of our work and methodology, described in Part II. Possible extensions to the research presented in Part II are indicated.

1.3 Contributions of the Dissertation

1.3.1 Dynamics of Computer Virus Propagation

We adapt the stochastic *susceptible-exposed-infectious-recovered* (SEIR) models of mathematical epidemiology, which to the best of our knowledge has not been attempted before, to the study of computer virus propagation by incorporating the idea of a random *latent period* during which a computer is in an *'exposed state'* in the sense that the computer is infected but cannot yet infect other computers until the latency is over. Such use of SEIR models is explored (i) using the standard Markovian approach that use ordinary differential equations governing transition probabilities between states, and (ii) via a stochastic differential equations (SDE) based approach in a multi-group setup to model the propagation dynamic of a virus across a non-homogeneous network with possibly different rates of infection across component subnetworks that are homogeneous within themselves with corresponding latent periods and recovery rates. Estimators of the model parameters (maximum likelihood and Bayesian estimators in the Markovian model, and least squares estimators in the SDE-based model) are developed. Our methods and results are illustrated numerically using data for the well known Code Red worm of 2001. Simulation of the Code Red worm based on the estimates developed under various scenarios (single

and multi-group setups) for Markovian as well as SDE-based SEIR models have been carried out and compared to the results of other models in the literature. These simulation based comparisons show that outputs of our stochastic SEIR models for the Code Red worm outbreak matches the actual observed time trajectory of infections more closely than the corresponding outputs of all other models considered by other researchers.

Situations where a computer virus outbreak data are only partially observed are more challenging. In this dissertation, we specifically consider the case where the clock times when individual computers are infected by a virus are unknown/missing and only recovery times are observed. In the context of SEIR models for computer virus propagation, such a scenario is eminently realistic since the duration of the latency period (i.e., *sojourn time* in the 'Exposed' state until becoming 'infectious') is uncertain. For Markovian SEIR models, we use Markov chain Monte Carlo (MCMC) methods in a Bayesian framework to provide estimates of model parameters via their data driven posterior distributions. The method is again illustrated with the Code Red worm data, and the simulation results again show a very good fit to the actual propagation data for Code Red worm. By including a latency period parameter; the overall thirst of our findings point to superiority of the stochastic SEIR framework as models of computer virus propagation relative to other models considered in the literature.

1.3.2 Survival Distributions in Reliability

Our work here considers two new approaches for modeling univariate and bivariate survival (life) distributions, respectively. We introduce a new partial order among survival distributions which exploits the concept of star-ordered real valued functions [9]. The star-ordering developed here, called 'reverse star-order', is different from the classic star-order among life distribution (see Barlow and Proschan [9]). Various properties of life distributions which are star-ordered in our sense are developed

and their ramifications are explored. A surprising finding is the equivalence of our proposed star-ordering to the so called reversed hazard rate order considered by Block, Savits and Singh [10] and others. This equivalence also justifies an aging property interpretation of a life distribution F relative to another distribution G when they are mutually reverse star-ordered.

Univariate proportional hazard models have been found to be of considerable importance and applicability in the context of survival analysis. In the bivariate set up, we formulate a new notion of bivariate proportional hazard models that is distinct from the formulation of Clayton and Cuzick [17], Hougaard [46] and Oakes [61]. Since our focus is on developing the basic distribution theory for a new class of lifetimes models that can be interpreted to hare bivariate proportional hazards in a suitably well defined sense; we do not consider *frailty effects* via unobserved explanatory covariates, although such refinements may be important in applications and would be a legitimate topic for future research. Dependence between component lifetimes is achieved in our setup via *latent variables*, some or, all of which may be individually unobserved. We show that a distribution belonging to our proposed class of bivariate proportional hazard models (BPHM) can be decomposed into an absolutely continuous and a singular part. Maximum likelihood estimators of model parameters are developed, which can require using an expectation-maximization (EM) algorithm when a baseline distribution parameter is also unknown. We also give illustration of our BPHMs with two actual data sets.

Part I

STOCHASTIC EPIDEMIC MODELS AND INFERENCE FOR THE PROPAGATION OF COMPUTER VIRUS

CHAPTER 2

INTRODUCTION

The primary purpose of Part I is to explore a more realistic stochastic epidemic model and corresponding statistical inference for application to computer virus propagation data. It may be observed here that such a methodology provides a means of quantifying transmission and the estimated benefit of infection control interventions in terms of changed transmission rates.

Despite a large body of research regarding the propagation of computer virus, most of them use deterministic epidemic models which only indicate the average tendency of computer virus spread in the long run. However, the spread of a computer virus has an essentially stochastic nature, especially in the early phase of the propagation. Consequently, the use of deterministic models may not be sufficient to understand the dynamics of prevalence of a computer virus. Stochastic epidemic models and statistical inference methods are well-equipped to tackle such difficulties arising from the use of deterministic models. Furthermore, such methods are able to quantify the estimated effects of infection control interventions in terms of changed transmission rates.

In practice, if it is the case that the entire epidemic process has been completely observed, so that recorded computer virus spread data include all infection and recovery times; then maximum likelihood and Bayesian estimates of transition parameters of stochastic models can be obtained. These methods with complete computer virus propagation data are examined and illustrated in Chapter 4. However, the infection times when a computer is infected may not be observed in practice. In such cases, the epidemic process must be considered to have been only partially observed. It is still possible to obtain parameter estimates using Markov chain Monte Carlo (MCMC) methods in a Bayesian framework [37, 38, 65, 78]. MCMC methods are

highly versatile and popular (e.g. [37], [38], [65], [78]) and have been used to analyze data of partially observed biological epidemic processes. In Chapter 5, we investigate the use of MCMC methods in the context of computer virus propagation to conclude that such methods appear well suited to the computer virus prevalence data.

It has been shown that network structure has impact on the propagation of computer virus [62, 66, 67, 86]. We investigate the sensitivity of our models and methods when a network structure is superimposed by considering the corresponding multi-group stochastic models in Chapter 6. Simulation results for heterogeneous multi-group stochastic model are contrasted with those with homogeneous assumption. All illustrations of our methods and results are shown with reference to the data on Code Red worm outbreak of July 19, 2001.

This Part of our dissertation investigates the use of stochastic epidemic model and statistical inference techniques to describe the spreading of computer virus over homogeneous and heterogeneous networks in Chapters 4 – 5 and Chapter 6, respectively. The methods are applied to data describing the occurrence of Code Red worm on July 19, 2001.

CHAPTER 3

LITERATURE REVIEW

3.1 Computer Virus Propagation Modeling

The Internet has become critically important to most facets of human activities ranging from personal communication to industrial productivity and financial viability of national and global economy in modern technological society. However, the increasingly easy access and use of the Internet by most computers make them a prime target for malicious activities, such as computer viruses using the Internet as a communicating tool. There is a widespread agreement about a continuing upsarge in such incidents of introduction of new "viruses" which quickly propagate through computer networks targeting individual computers that are vulnerable to such attacks. There have been many previous studies to model and analyze the spread of computer virus over the Internet using deterministic epidemic models of mathematical biology as a basis. Since the prevalence of malicious codes has a stochastic nature, especially in the early phase of their propagation, the use of deterministic models is typically not adequate to fully understand the dynamics of the propagation of malicious codes over the Internet. In this Part, it is proposed to use stochastic epidemic models to quantitatively describe and analyze the propagation of malicious codes and compare the results of our analysis with those obtained using the deterministic and simple stochastic epidemic models in previous works.

3.1.1 What is a Computer Virus?

A computer virus is "*a malicious code which is a software program that is intentionally designed to move from computer to computer, or from network to network and modify the system without the consent of the user*" [40]. Major types of malicious code include virus, worm, Trojan, and rogue Internet content. The first malicious code is

a computer virus, developed by Fred Cohen [18, 19] for research purposes. Cohen's definition of computer virus is "a program that can 'infect' other programs by modifying them to include a version of itself". *This definition has been generally accepted as a standard definition of a computer virus.*

Even though *worms* are very similar to viruses in that they are computer programs that replicate themselves and often, but not always, contain some malicious functions that would disrupt the normal use of a computer system or a network [40], worms exist as separate entities: they do not attach themselves to any other files or programs. The first computer worm, *Morris*, was released on November 2, 1988 [76]. It utilized the TCP/IP protocols, common application layer protocols, operating system bugs, and a variety of system administration flaws to propagate itself. The "Morris worm" infected approximately three thousand computers during eight hours of activity [76]. Since worms can spread automatically over a computer network without the need of human intervention, they can potentially spread on the Internet with staggering speed and cause damage on the order of billions of dollars [59]. Famous worms include *Code Red* in 2001 and *SQL Slammer* in 2003.

Borrowing relevant ideas from quantitative models of biological epidemics, the stochastic models we propose in this dissertation to analyze the spread, over computer networks, of malicious codes/programs (generically referred to as 'malware'), which include computer viruses, worms and other variants of potentially harmful software programs. *Since our main goal is to model the spread of such malware over time* across computer networks *by stochastic model(s)* that adequately explain the observed time trajectory of such propagation – *for which the technical distinctions between various forms of malware* such as 'virus', 'worm' etc. *are not essential*; the term *virus* is used throughout this dissertation to denote a *generic malware that has an ability to infect any computer to which it can gain access.*

3.1.2 Modeling Computer Virus Dynamics

A good and reliable computer virus propagation model can help us to understand the life cycle of a self-replicating program. At a fundamental level, they explain how wide and how fast the propagation is. They can also be useful in suggesting countermeasure techniques [15] to mitigate the disruptive effects of virus attacks and to examine the effects of network traffic [75] factors and network topology [62, 66]. Kephart and White [49] built a Susceptible-Infected-Susceptible (SIS) model as an explanatory model of computer virus propagation, and used deterministic ordinary differential equations to approximate the SIS model. Hierarchical and spatial models were also presented in [49], after which they introduced the concept of a Kill signal (a warning signal as a countermeasure to reduce the spreading of computer virus). They built a model for virus propagation with the Kill signal and concluded that the Kill signal is effective in reducing the spread of the virus [50]. Staniford [77] constructed a deterministic Susceptible-Infected (SI) model based on the empirical data from the outbreak of the Code Red worm. Serazzi and Zanero [75] surveyed existing models of computer virus propagation for virus and worms and derived a compartment-based model that deals with the propagation inside and outside of an autonomous system which is a sub-network administered by a single authority. Zou [90] considered a model for Code Red worm propagation based on the classic Susceptible-Infected-Removed (SIR) model. Two factors that might affect the worm propagation were introduced into the model, which are a countermeasure effect and a decreased infection rate due to the Internet congestions caused by the worm. Pastor-Satorras and Vespigani [66] studied the effects of network topology on epidemic models.

All of the above models are deterministic, described by a system of ordinary differential equations [69], except that of Kephart and White [49], which uses a linear birth and death process to study the expected lifetime of the infection.

In most virus propagation models based on epidemiology, epidemic prevalence and propagation characteristics are described by a system of ordinary differential

equations (ODEs). Although ODEs can be safely used to approximate a stochastic process when the population size is large, no probabilistic event is considered. Moreover, the ODEs only describe the average tendency of virus propagation. Thus, deterministic models cannot represent rare events such as saturation and extinction of a virus. Such rare events are quite important to evaluate security strategies in a network, and the transmission of a virus from one computer to another is actually stochastic [73, 89]. It is thus natural to model the computer virus prevalence via stochastic models. The stochastic model can give us the probability that an event will happen instead of deterministic yes-or-no answer relying on the law of large numbers [4]. When the population size is large, it has been shown that a stochastic model converges to deterministic model [4]. It is believed that both deterministic and stochastic models are important to understand the propagation of a malicious program like a worm or a virus. Andersson and Britton [4] concluded that stochastic models are preferred when their analysis is possible. Wang [84] simulated computer virus propagation to evaluate security policies based on the simulation. Wierman and Marchette [88] propose an extended model from the stochastic Susceptible-Infectious-Susceptible (SIS) model, taking account of the probability of infection reintroduction at the virus-free state. Okamura [63] developed a new stochastic model that reformulates a deterministic model with kill signal [50] as a continuous-time Markov chain to evaluate the probabilistic behavior of the Internet worm propagation and its associated dependability measures. Rohloff and Basar [72] considered the stochastic properties of a special type of a computer worm called a Random Constant Scanning (RCS) worm, to present an idealized stochastic propagation model for RCS worms using ideas from the literature of epidemiology and public health (namely density-dependent Markov jump process model), and compared the results obtained using their model with those obtained by the standard deterministic simple epidemic model. The influence of network topology on computer virus propagation have also been considered by several authors [32, 62, 91].

3.2 Epidemic Models

The mathematical modeling of diseases and their propagation has a history of about three hundred years [22]. Epidemic models describe the dynamics of infectiousness as opposed to the dynamics of disease. The infectious process can be represented by a succession of states or 'compartments'. Such a representation is referred to as a '*compartmental model*' [47]. A compartmental model can be described by either deterministic or stochastic equations. In a deterministic model, the number of infections in a short time interval can be assumed proportional to the number of susceptible and infectious individuals and to the time interval. In a stochastic model, the probability of a new case in a small time interval is correspondingly proportional to the same quantity.

The term "epidemic" refers to outbreaks of undesirable events (i.e., new cases of infections by a biological agent or, computer malware) which can usually be attributed to a point source. Mathematical models are used to describe biological and transmission mechanisms, threshold densities and to predict the course of epidemics [5]. In particular, they are used to predict the initial conditions which lead to an epidemic, the shape of the epidemic curve, the number of cases at the peak of the epidemic, the duration of the total epidemic and the total number of cases [28].

Epidemic modeling has three main goals [22]: 1) to understand the disease spreading mechanism; 2) to predict the future course of the epidemic; 3) to understand how to control the spread of the epidemic. A good epidemic model should be able to capture the essential features of the epidemic, make reasonable predictions, and evaluate the effect of suggested control methods.

3.2.1 Deterministic Models

Deterministic models are generally based on the mass-action principle which states that the evolutionary course of an epidemic depends on the number of susceptible individuals and the contact rate between susceptible and infectious individuals. The

mass-action principle was formulated in discrete time by Hamer in 1906 and then in continuous time by Ross in 1908 [4]. In a deterministic model, the future state of the epidemic process can be determined from the initial numbers of susceptible and infectious individuals, together with the infection rate, recovery rate, birth rate and death rate [5].

The first mathematical model of epidemics is generally attributed to Kermack and McKendrick [52]. However, the set of governing equations were first published by Ross and Hudson in 1917 [28]. These equations describe the dynamics of a Susceptible-Infectious-Removed (SIR) disease in continuous time. It is assumed that the disease being modeled occurs in a large, closed, homogeneous and uniformly mixing population of equally susceptible individuals; contacts are made according to the law of mass action and infection triggers an autonomous process within the host [27]. These assumptions are summarized by the equation

$$S'(t) = \frac{dS}{dt} = S(t) \int_0^\infty \bar{A}(\tau) S'(t-\tau) d\tau, \tag{3.2.1}$$

where $S(t)$ is the spatial density of susceptible (i.e., the number of susceptible per unit area) at time t, $-S'(t)$ is the incidence (i.e., the number of infection events in a unit of time) at time t and $\bar{A}(\tau)$ is the expected infectivity of an individual that become infected τ time units ago [26]. Equation (3.2.1), referred to as the *Kermack and McKendrick model*, is usually expressed for the specific case in which infectivity has an exponential distribution. If β is the rate at which an infectious individual has contact (sufficient for transmission) with susceptible members and γ is the removal rate, then the Kermack and McKendrick model with exponential infectivity is

$$S'(t) = -\beta S(t) I(t), \tag{3.2.2a}$$

$$I'(t) = \beta S(t) I(t) - \gamma I(t), \tag{3.2.2b}$$

where $I(t)$ is the number of infectious individuals at time t. The Kermack and McKendrick model is often referred to as the general epidemic model. The term

'general' is used in the sense that the model is not confined to infection only, i.e. the possibility of removal is also considered [5]. Hethcote [44] provides a brief outline of possible generalizations.

Besides SIR model, another well-known epidemic model is Susceptible-Infectious-Susceptible (SIS) model [1]. In SIS epidemic model, individuals in the population are classified according to disease status, *either* as healthy and susceptible *or* as infected and therefore infectious (i.e., able to infect others). A susceptible individual, after a successful contact with an infectious individual, becomes infected and hence infectious, but does not develop the immunity to the disease. Therefore, after recovery, infected individual return to the susceptible class. The standard assumptions of SIR model continue to apply to SIS model; i.e., the disease being modeled occurs in a large, closed, homogeneous and uniformly mixing population of equally susceptible individuals; contacts are made according to the law of mass action and infection triggers an autonomous process within the host [27]. If β is the rate at which an infectious individual has contact (sufficient for transmission) with susceptible members and γ is the removal rate, then an SIS epidemic model has the following form:

$$S'(t) = -\beta S(t)I(t) + \gamma I(t), \tag{3.2.3a}$$

$$I'(t) = \beta S(t)I(t) - \gamma I(t). \tag{3.2.3b}$$

For SIS model, an explicit solution can be obtained, which is given by

$$I(t) = \frac{(1 - \frac{\gamma}{\beta})I(0)e^{(\beta-\gamma)t}}{1 - \frac{\gamma}{\beta} + I(0)(e^{(\beta-\gamma)t} - 1)}, \tag{3.2.4}$$

where $I(0)$ is the initial value of infectious individuals.

Standard epidemic models, such as the SIR model above, share the core concept of classifying the state of an individual unit into one of several categories referred to as 'compartments', and are thus known as compartment models. There are a number

of other compartment models for the spread of biological diseases [1, 2, 4]. Some of these models can be relevant for modeling computer virus outbreaks. The stochastic version of one such class of models that we have found to be profitably adaptable to the context of computer virus propagation is SEIR model, which is described in greater details in Chapter 4 in the context of our research.

Deterministic models approximate actual state changes by assuming that the number of susceptible, infected or recovered individuals vary continuously, while in reality they are integer valued. When numbers of susceptible and infective individuals are both large and mixing is reasonably homogeneous, the deterministic model is likely to be satisfactory as a first approximation [5]. The approximation will not be good when any of the integer-valued variables become sufficiently small for the population as a whole to be considered close to extinction [4, 22, 58]. Further details about deterministic epidemic models can be found in [2, 5, 22].

3.2.2 Stochastic Models

In a stochastic model, probability distributions of the numbers of susceptible and infectious individuals at any instant replace the corresponding values in a deterministic model. The majority of stochastic models are based on variants of the general epidemic model and the chain binomial model. Both of these classical models are special cases of a general SIR model of a closed and homogeneous mixing population in which contacts between pairs of individuals occur independently. The infectious periods are assumed to be independent and identically distributed. Assumptions concerning the distribution of the infectious periods differ between the two classical models. For example, within the general epidemic model the infectious period is assumed to be exponential and in the chain binomial model the infectious period is assumed to be of a pre-determined fixed length. The general epidemic model is usually described in continuous time dynamics and the chain binomial model in a discrete time framework.

An overview of stochastic epidemic models is available in [4, 5, 22], where asymptotic and exact distributions of the final size of the epidemic, the total area under the trajectory of infective individuals, and approximations are discussed. In addition, generalizations of the stochastic epidemic model to allow for several classes of susceptible and infected individuals are made and the phenomena of recurrence and competition and spatial aspects are considered.

The general stochastic epidemic model was first studied by McKendrick in 1926 and then ignored until 1949 when it was analyzed by Bartlett. Deterministic general epidemic models assume that the actual number of new infections in a time interval is proportional to the product of the susceptible and infective population sizes and the time interval. Stochastic epidemic models, on the other hand, assume that the probability of a new infection in a short interval is proportional to this same amount, i.e. the product of the susceptible and infective population sizes and the time interval.

The stochastic version of system of equations (3.2.2) thus expresses the probability of infection and removal as

$$P(S(t+\Delta t)=s-1, I(t+\Delta t)=i+1 \mid S(t)=s, I(t)=i)=\beta si\Delta t+o(\Delta t), \quad (3.2.5a)$$

$$P(S(t+\Delta t)=s, I(t+\Delta t)=i-1 \mid S(t)=s, I(t)=i)=\gamma i\Delta t+o(\Delta t). \quad (3.2.5b)$$

The *force-of-infection* is defined as the rate at which an individual is infected, i.e. $\beta I(t)$. This formulation for the force-of-infection is known as the '*pseudo* mass-action' assumption [23]. It is used when the number of effective transmission is expected to remain the same regardless of the population size. Another approach, '*true* mass-action' assumes that the probability of contact decreases as population size increases. In this case, β should be divided by the population size. The force-of-infection at a time is sometimes referred to as the *hazard rate* function $h(t)$. For the model described by the system of equations (3.2.5), the hazard rate function at time t is then given

by

$$h(t) = \beta I(t). \tag{3.2.6}$$

The system of equations (3.2.5) define an infection process in which contacts between uniformly mixing susceptible and infections individuals occur at times governed by a homogeneous Poisson process with constant intensity β. An implicit assumption is that the infectious periods (= time during which the infectious individual can infect the susceptible individual) are independently and exponentially distributed with mean γ^{-1}. A corresponding implication is that the conditional distribution of an infection time i_n given previous infection times $i_0, i_1, \ldots, i_{n-1}$, depends only on the most recent infection time i_{n-1}; the so called 'Markov property'. Time periods that are modeled as random variables form an exponential distribution are said to be Markovian.

Stochastic models that do not assume homogeneous mixing have also received considerable attention. Such models include multitype models (e.g. [6], [7], [4]), which divide the population into homogeneous subpopulations, and their extensions [8], social cluster models [74] and random network models [3].

There are significant similarities between the propagation of a computer virus and that of biological epidemic agent [51], although there can be significant differences as well. For example, the concepts of susceptible as well as infectious units are the same in both; but the idea of a recovered unit acquiring a permanent or limited temporary immunity may or may not be applicable for computer virus. In the context of computer networks, such immunity apart from being virus specific, may be permanent if the anti-virus software remains resident in the same computer after recovery, or be temporary in the event the anti-virus software is deleted, or not renewed after its license expires.

3.2.3 Comparison of Deterministic and Stochastic Models

The stochastic versus deterministic model debate is often centered around model simplicity and realism [4]. Deterministic models are generally simpler to analyse than their stochastic counterparts [5, 58]. For a stochastic model to be analytically trackable, simpler assumptions that may not be entirely realistic are often required. Although exact analytic solution can be difficult, suitable approximations can provide useful information about the system [58].

Deterministic models are unsuitable for small populations. For large population, the mean number of infectives in a stochastic model may not always be approximated satisfactorily by a corresponding deterministic model [4, 22, 58]. According to Renshaw [68], it should always be assumed that stochastic effects play an important role in any given process unless proven otherwise.

3.3 Statistical Inference Methods

3.3.1 Maximum Likelihood Estimation

Maximum likelihood estimation is a method used for fitting a parametric statistical model to data and providing estimates for the model parameters. In general, for given set of data and an underlying probability model, the method of maximum likelihood selects values of the model parameters such that the parameters maximize the likelihood function, or equivalently, loglikelihood function. Maximum likelihood estimation, which have good large sample properties, provides a unified approach for estimating parameters of interest, and are known to have good large sample properties.

3.3.2 Bayesian Estimation

A Bayesian estimator of a parameter maximizes the expected value of a parameter's posterior distribution. While constructing a Bayesian estimator, if there is no inherent reason to prefer one prior probability distribution over another, a *conjugate prior* is

sometimes chosen for analytical simplicity. A conjugate prior is defined as a prior distribution belonging to some parametric family, for which the resulting posterior distribution also belongs to the same family. This is an important property, since the Bayesian estimator, as well as its statistical properties (e.g. variance, confidence interval), can all be derived from the posterior distribution. Conjugate priors are especially useful for sequential estimation, where the posterior of the current measurement is used as the prior in the next measurement. In sequential estimation, unless a conjugate prior is used, the posterior distribution typically becomes more complex with each added measurement, and the Bayesian estimate cannot usually be computed without resorting to numerical methods.

3.3.3 Bayesian Inference using Markov Chain Monte Carlo (MCMC) Methods

Bayesian inference [34] is the process of fitting a model to data and summarizing the results via the posterior distribution of the parameters and unobserved quantities. In contrast to the frequentist framework; in a Bayesian model, parameters are treated as random variables and relevant data are considered as given, which contain information about the parameter(s). The posterior distributions obtained within a Bayesian framework provide information about parameter uncertainty and permit the formulation of direct probability statements about parameters which is appropriate for small size samples. A frequentist approach estimates only the standard errors of parameters rather than full probability distributions of parameters. Probability statements within a frequentist approach rely on indirect statements based on confidence intervals and p-values. Calculation of the frequentist confidence levels may require development of appropriate theoretical results and the usual conditions that require asymptotic normality of maximum likelihood estimators are often violated [64].

A Bayesian framework incorporates prior information on the model parameters $\boldsymbol{\theta}$ in the form of a prior distribution $P(\boldsymbol{\theta})$. This prior distribution along with the

likelihood of the data $\mathbf{D}$, $P(\mathbf{D}|\boldsymbol{\theta}) = L(\boldsymbol{\theta};\mathbf{D})$, defines the posterior distribution, denoted by $P(\boldsymbol{\theta}|\mathbf{D})$. The posterior distribution is given by

$$P(\boldsymbol{\theta}|\mathbf{D}) = \frac{P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta})\, d\boldsymbol{\theta}}. \tag{3.3.1}$$

The posterior distribution (3.3.1) is the distribution of the model parameters conditional on the data. Because the denominator of equation (3.3.1) is not a function of $\boldsymbol{\theta}$ and since integration is with respect to $\boldsymbol{\theta}$, the posterior distribution (3.3.1) is proportional to the product of the prior and likelihood distributions,

$$P(\boldsymbol{\theta}|\mathbf{D}) \propto P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta}). \tag{3.3.2}$$

When making inference about the parameters, one is usually concerned with point and interval summaries of the posterior distribution, such as mean, variance or quantiles. Point and interval summaries are expressed in terms of their expectation of a function of the unknown parameters,

$$E(f(\boldsymbol{\theta})|\mathbf{D}) = \frac{\int_{\boldsymbol{\theta}} f(\boldsymbol{\theta})P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}} P(\boldsymbol{\theta})P(\mathbf{D}|\boldsymbol{\theta})\, d\boldsymbol{\theta}}. \tag{3.3.3}$$

Except in the simplest cases, the integrals in (3.3.3) cannot be evaluated analytically. Given a realization of a Markov chain $\{\theta^{[g]}\}$, $g = 1, 2, \ldots$ whose stationary distribution is the posterior distribution, the integral $E(f(\boldsymbol{\theta})|\mathbf{D})$ can be estimated by Monte Carlo integration as:

$$E(f(\boldsymbol{\theta})|\mathbf{D}) \approx \frac{1}{G}\sum_{g=1}^{G} f(\boldsymbol{\theta}^{[g]}).$$

The essence and practical utility of Markov Chain Monte Carlo (MCMC) methods derives from the fact that it is possible to construct a Markov chain with a preassigned stationary distribution. This makes MCMC method an ideal vehicle to carry out large scale simulations of a Markov chain to numerically capture its long run

behavior, thus producing random samples from its stationary distribution which we originally assign as the target posterior distribution.

Markov chain realizations can be obtained using Markov chain Monte Carlo (MCMC) techniques [12, 36, 39] which iteratively generate samples from some target distribution $\pi(\boldsymbol{\theta})$, that is known only up to proportionality. In a Bayesian framework, the target distribution is the posterior distribution of the model parameters. Samples are drawn from the appropriate Markov chain and the process is continued until the chain converges to its stationary distribution $\pi(\boldsymbol{\theta})$. After discarding initial samples to remove dependence of the simulated chain on its starting location (*burn-in simulations*) and convergence is achieved in terms of satisfying convergence criteria; samples generated by simulating the Markov chain can be used to estimate functions of the target distribution [48].

The transition kernel $\mathcal{P}(\boldsymbol{\theta}^{[g+1]}|\boldsymbol{\theta}^{[g]})$ is the probability law of the next state of the chain lies within some set, given that the chain is currently in state $\boldsymbol{\theta}^{[g]}$ [12]. If the transition kernel satisfies the detailed balance condition,

$$\pi(\boldsymbol{\theta}^{[g]})\mathcal{P}(\boldsymbol{\theta}^{[g]}, \boldsymbol{\theta}^{[g+1]}) = \pi(\boldsymbol{\theta}^{[g+1]})\mathcal{P}(\boldsymbol{\theta}^{[g]}, \boldsymbol{\theta}^{[g+1]}),$$

then the Markov chain will have a stationary distribution π [12, 81, 82].

To ensure that Markov chain converges to the stationary state; regularity conditions of irreducibility, aperiodicity and positive recurrence are required. An irreducible positively occurrent Markov chain will reach any non-empty set of states with positive probability in a finite number of iterations. The aperiodicity condition prevents the Markov chain from oscillating between different states in a regular periodic fashion. If a Markov chain is positive recurrent and an initial value is sampled from a stationary distribution $\pi(\boldsymbol{\theta})$, then all subsequent iterations will also be distributed according to $\pi(\boldsymbol{\theta})$.

The Metropolis-Hastings algorithm [16, 42, 56] is a Markov chain simulation method used to draw samples from Bayesian posterior distribution [34]. The *Metropol-*

is sampler [56] and the *Gibbs sampler* [13, 33, 35, 80] are special cases of the *Metropolis-Hastings algorithm.*

The *Metropolis-Hastings algorithm* can be described as follows:

- start with an initial value $\boldsymbol{\theta}^{[0]}$
- obtain a realization $\boldsymbol{\theta}^{[1]}, \boldsymbol{\theta}^{[2]}, \ldots$ from a Markov chain by repeating the following steps for $g = 1, 2, \ldots$:

  1) sample a point $\boldsymbol{\theta}^{[*]}$ from some proposal distribution $q(\boldsymbol{\theta}^{[*]}|\boldsymbol{\theta}^{[g]})$,

  2) evaluate the following value

$$\alpha(\boldsymbol{\theta}^{[g]}, \boldsymbol{\theta}^{[*]}) = \min(1, \frac{\pi(\boldsymbol{\theta}^{[*]})q(\boldsymbol{\theta}^{[g]}|\boldsymbol{\theta}^{[*]})}{\pi(\boldsymbol{\theta}^{[g]})q(\boldsymbol{\theta}^{[*]}|\boldsymbol{\theta}^{[g]})}),$$

  3) set $\boldsymbol{\theta}^{[g+1]}$ equal to $\boldsymbol{\theta}^{[*]}$ with probability $\alpha(\boldsymbol{\theta}^{[g]}, \boldsymbol{\theta}^{[*]})$, otherwise set $\boldsymbol{\theta}^{[g+1]}$ equal to $\boldsymbol{\theta}^{[g]}$.

The transition kernel for the Metropolis-Hastings algorithm is:

$$\mathcal{P}(\boldsymbol{\theta}^{[g+1]}|\boldsymbol{\theta}^{[g]}) = \begin{cases} q(\boldsymbol{\theta}^{[g+1]}|\boldsymbol{\theta}^{[g]})\alpha(\boldsymbol{\theta}^{[g]}, \boldsymbol{\theta}^{[g+1]}) & \text{if } \boldsymbol{\theta}^{*} \text{ is accepted such that } \boldsymbol{\theta}^{[g+1]} = \boldsymbol{\theta}^{*}, \\ 1 - \int q(\boldsymbol{\theta}^{[*]}|\boldsymbol{\theta}^{[g]})\alpha(\boldsymbol{\theta}^{[g]}, \boldsymbol{\theta}^{[*]})d\boldsymbol{\theta}^{[*]} & \text{if } \boldsymbol{\theta}^{*} \text{ is rejected such that } \boldsymbol{\theta}^{[g+1]} = \boldsymbol{\theta}^{g}, \end{cases}$$

and satisfies the detailed balance equation

$$\pi(\boldsymbol{\theta}^{[g]})\mathcal{P}(\boldsymbol{\theta}[g], \boldsymbol{\theta}[g+1]) = \pi(\boldsymbol{\theta}^{[g+1]})\mathcal{P}(\boldsymbol{\theta}[g], \boldsymbol{\theta}[g+1]),$$

[71]. This implies that $\pi(\boldsymbol{\theta})$ is the stationary distribution of the Markov chain [12, 16, 81, 82].

The irreducible, aperiodic and positive recurrent conditions, which regulate convergence of Markov chain to the stationary distribution, will be satisfied by Markov chains generated using the Metropolis-Hastings algorithm if the proposal distribution provides full support for stationary distribution $\pi(\boldsymbol{\theta})$ [70]. It should be noted that

Markov chains that are irreducible with stationary distribution $\pi(\boldsymbol{\theta})$ will be positive recurrent [71].

It is often computationally more efficient to update the components of $\boldsymbol{\theta}$ one by one, rather than all at once using a single component *Metropolis-Hastings algorithm* [39]. If the proposal distribution is symmetric, i.e. if $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$, then a special case of the Metropolis-Hastings algorithm, named as the *Metropolis algorithm*, can be used. Given a symmetric proposal distribution, the candidate value $\boldsymbol{\theta}^*$ is accepted as the next value with probability $\alpha = \min(1, \frac{\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta})})$.

Gibbs sampling is a special case of the single component Metropolis-Hastings algorithm in which the proposal distribution for updating the i^{th} component of $\boldsymbol{\theta}$ at the g^{th} iteration of the Markov chain is same as the full conditional.

A Markov chain defined by the Metropolis-Hastings algorithm will satisfy the irreducible, aperiodic and positive recurrent conditions required for convergence to a stationary distribution as long as the proposal distribution ensures that the support of $\boldsymbol{\theta}$ can be fully explored [70].

Burn-in. Early iteration in MCMC simulations are discarded to diminish the effect of the starting value [12]. The discarded iterations are referred to as *burn-in*. The length of the burn-in period will depend on the starting value and the rate of convergence of the Markov chain to the stationary distribution [39]. Geyer [36] suggests that calculating the burn-in is unnecessary and that discarding 1% to 2% of the run length should be sufficient so that extreme starting values are discarded. A more conservative rule of thumb is used by Gelman *et al.* [34], who generally discard the first half of the iterations.

Thin factor. Markov chain sequences may be thinned by keeping every k^{th} simulation draw from each sequence and discarding the rest. Thinning is useful for problems with a large number of parameters where computer storage is a problem [34]. A thinned Markov chain of length G will have less autocorrelation than a full sample of the same length [36]. As the thin factor k goes to infinity, the Markov

chain will become almost independent [36]. The amount of thinning must be weighed against the cost of sampling as the full sample will have more information.

3.4 Code Red Worm Studies

3.4.1 Introduction

The first incarnation of the Code Red worm (CRv1) began to infect hosts running unpatched versions of Microsoft's IIS webserver on July 12, 2001. The first version of the worm uses a static seed for its random number generator. Then, around 10:00 UTC (Universal Time Coordinated) in the morning of July 19th, 2001, a random seed variant of the Code Red worm (CRv2) appeared and spread. This second version shared almost all of its code with the first version, but spread much more rapidly. Finally, on August 4th, a new worm began to infect machines exploiting the same vulnerability in Microsoft's IIS webserver as the original Code Red virus. Although the new worm shared almost no code with the two versions of the original worm, it contained in its source code the string "CodeRedII" and was thus named CodeRed II. Since we model and analyse the propagation of Code Red version 2 (CRv2) worm, the characteristics of CRv2 are described below. For more details about the other versions of Code Red worm, see [59].

At approximately 10:00 UTC in the morning of July 19, 2001, a random seed variant of the Code Red worm (CRv2) began to infect hosts running unpatched versions of Microsoft's IIS webserver. The worm began to spread by probing random IP addresses and infecting all hosts vulnerable to the IIS exploit. CRv2 lacks the static seed found in the random number generator of Code Red version 1. In contrast, the CRv2 worm uses a random seed. So each infected computer attempts to infect a different list of randomly generated IP addresses. This seemingly minor change had a major impact: *more than* 359,000 *machines were infected* with CRv2 *in just fourteen hours*. CRv2 initially generated 100 threads. Each of the first 99 threads randomly chose one IP address and tried to set up connection on port 80 with the

target machine. If the connection was successful, the worm would send a copy of itself to the victim's web server to compromise it and continue to find another web server. If the victim was not a web server or the connection could not be set up, the worm thread would randomly generate another IP address to probe. The timeout of the CRv2 can exploit only Windows 2000 with IIS server installed. Figure 3.1 shows the spread of CRv2 by the number of infected hosts versus time.

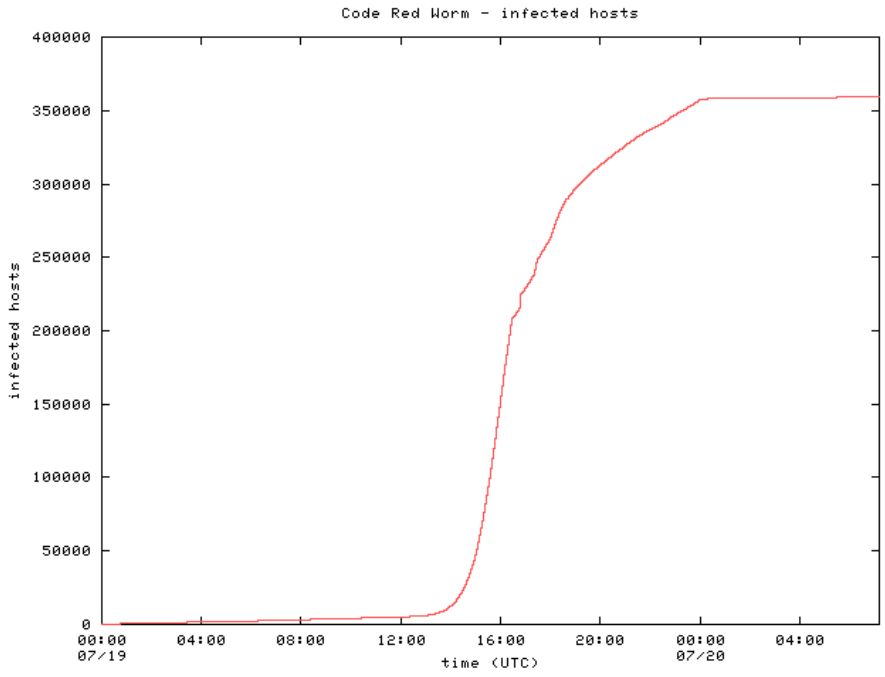

Figure 3.1 Observed Code Red worm propagation-number of infected hosts. Source: http://www.caida.org/research/security/code-red/gifs/cumulative-ts.gif

Because CRv2 is identical to Code Red version 1 in all respects except the seed for its random number generator, its only actual damage is the "Hacked by Chinese" message added to top level webpages on some hosts. However, CRv2 had a greater impact on global infrastructure due to the sheer volume of hosts infected and probes sent to infect new hosts. It also wreaked havoc on some additional devices with web interfaces, such as routers, switches, DSL modems, and printers. Although these devices were not infected with the worm, they either crashed or rebooted when an infected machine attempted to send them a copy of the worm. Even though the CRv2

is a kind of friendly worm which does not damage the computers that were infected or, leave any back doors on infected computers; the cost of this epidemic, including subsequent strains of Code Red, is estimated to be in excess of $2.6 billion [59].

3.4.2 Data and Notation

For simplicity and our purpose of modeling and analysis, the Code Red version 2 is referred to subsequently as the *Code Red worm*. The observed Code Red worm data set $\boldsymbol{D}^{obs}$ includes the start time, end time, top-level domain (TLD), country, latitude and longitude information, autonomous system (AS) number and AS name for each computer in a /8 network at UCSD (University of California at San Diego), where the start time and the end time corresponds to the infection time and the recovery time respectively. In modeling and analysis of Code Red worm propagation, a computer is characterized as being i) susceptible (S), ii) exposed (E, i.e., infected but not infectious), iii) infectious (I), or iv) recovered (R) at any given time. The number of computers in each of the susceptible, exposed, infectious and recovered compartments at time t are $S(t)$, $E(t)$, $I(t)$ and $R(t)$, respectively. N is used to denote the total number of computers in the Internet.

3.4.3 Summary of Previous Works on Code Red Worm

Both the SIS and SIR models of biological epidemics have been adapted to the study of computer virus propagation in networks [50, 51, 57, 87, 89]. However, in the SIS model the infected individual becomes re-infected as soon as it is recovered, which means that the recovered individual does not have immunity to the disease (biological, or computer malware) after it is cured. On the other hand, in the SIR model the infected individual will be healthy and will have permanent immunity to the disease after it is recovered.

Deterministic models have been used to study Code Red worm propagation [77, 90]. Zou *et al.* [90] proposed and investigated two-factor worm model based on

the classic epidemic Kermack-McKendrick model as follows:

$$\frac{dI(t)}{dt} = \beta(t)[N - I(t) - R(t) - Q(t)]I(t) - \frac{dR(t)}{dt}, \tag{3.4.1}$$

where N is total number of hosts under consideration, and $I(t)$, $R(t)$ and $Q(t)$ represent the number of infectious hosts, the number of removed hosts from the infectious population and the number of removed hosts from the susceptible population at time t, respectively. The parameter $\beta(t)$ is infection rate at time t and reflects the impact of the Internet traffic on the Code Red worm prevalence. $R(t)$ and $Q(t)$ reflect the cleaning, patching and filtering countermeasures against Code Red worm. Simulations and numerical solutions of the two-factor worm model were given, which match the observed data of Code Red worm better than previous models do.

When there is no human countermeasures and when the infection rate is constant divided by N, the two-factor worm model equation (3.4.1) is then the classic simple epidemic model used by Rohloff *et al.* [72] for the spread of Code Red worm:

$$\frac{dI(t)}{dt} = \frac{\beta}{N}[N - I(t)]I(t). \tag{3.4.2}$$

These types of model are suitable when the number of infected hosts are large. However, during the first phase of the worm propagation, the number of infected hosts are small and such deterministic models may not accurately characterize the propagation of viruses. Moreover, these models do not consider the existence of latency, which makes them unsuitable as realistic models in those situations where the existence of a random sojourn time in an exposed state is relevant to correctly describe the future propagation of infection. One of the salient points of departure from existing models in our work on computer virus propagation reported here is the explicit introduction of such latency in an exposed state via SEIR model corresponding overall finding that such a refinement significantly enhances the overall fit of the model to the actual Code Red worm propagation data.

CHAPTER 4

STOCHASTIC SEIR MODEL AND INFERENCE FOR COMPLETELY OBSERVED COMPUTER VIRUS PROPAGATION

4.1 Introduction

Stochastic epidemic models can be used to model and analyze the spreading of a computer virus. In this Chapter, the stochastic *susceptible-exposed-infectious-recovered* (SEIR) model of mathematical biology, which uses a Markovian approach is adapted for analysis of data describing a completely observed computer virus. In particular, the methods are applied to data describing the incident of Code Red worm on July 19, 2001.

4.2 Model and Methodology

Many epidemiology based models have been suggested to model virus propagation in computer networks. Recently, a dormant latent period that some viruses spend inside infected computers in a network has attracted attention of some researchers [85]. Explicitly accounting for this feature requires that we examine the so-called susceptible-exposed-infectious-recovered (SEIR) model [2] to investigate how well it can explain computer virus propagations.

Stochastic models such as SI and SIR models, which are simpler than SEIR model, have been used in previous works on computer virus propagation modeling. The point of departure in our work from previous models is to recognize the possibility that a virus may have a random latent period during which it is already infected with the virus, but is not an active agent for infecting others in the network, i.e., in reality a computer virus can be dormant for some time (referred to as in a latent period), only to become active after a certain period [85]. For example, in case of email viruses, computers are infected at the time when the users check their emails

but such infected computers cannot infect others in the network until they send out emails to other users. The latent period parameter is the 'mean *sojourn time*' spent by an infected computer in the '*exposed state*' (when it is itself infected but cannot infect others). In Markov chain models, the latency parameter can be equivalently described by the reciprocal of the mean sojourn time, which is the transition rate from the 'exposed state' to the 'infectious state'.

The SEIR model is very similar to the SIR model, but it accounts for the fact that some viruses go through a latent period before the host becomes infectious. In SEIR model, a computer may experience four states during the virus propagation: the susceptible state (S), the exposed state (E), the infectious state (I) and the recovered state (R), as shown in Figure 4.1. In state S, the computer is healthy but can be infected by viruses. In state E, the computer is infected but the virus has not been triggered yet, which means that the computer is infected but it cannot infect other computers. Then in state I, the virus has been triggered and the computer is now infectious and able to infect other computers. Finally in state R, the computer is recovered from non-healthy status and has permanent immunity.

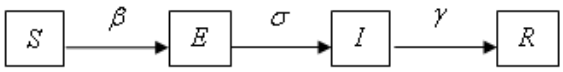

Figure 4.1 The Diagram of susceptible-exposed-infectious-recovered (SEIR) model.

In Figure 4.1, β is the *infection rate* at which a given infectious computer makes contact with other initially susceptible computers; σ is the rate at which a given infected computer that is in state E becomes infectious. This means that σ is the *transition rate* from state E to state I, or we can say that $\frac{1}{\sigma}$ is the *mean latent period*; γ is the *recovery rate* at which that an infectious computer is recovered from the non-healthy status and has immunity to the virus. The rates (β, σ, γ) are usually measured in per second time unit. $S(t)$, $E(t)$, $I(t)$ and $R(t)$ respectively denote the

number of computers that are in susceptible, exposed (infected but not infectious), infectious and recovered states at time $t \geq 0$, with $S(t)+E(t)+I(t)+R(t) = N$, where N is the total number of population. Of the N computers in the closed population (network), $N_s \leq N$ computers could potentially become infected by computer virus. At time 0, $(S(0), E(0), I(0), R(0) = (s_0, e_0, i_0, r_0)$, where $0 < s_0 \leq N_s$, $e_0 = 0$, $i_0 \geq 1$, $r_0 > 0$, denotes the initial state of the network. Since for all t, in virtue of a closed network; we have $S(t) + E(t) + I(t) + R(t) = N$, it is sufficient to keep track of $(S(t), E(t), I(t))$. Together with the assumption that the network is homogeneous, evolution of the epidemic process $(S(t) = a, E(t) = b, I(t) = c)$ can be described by the following state-to-state transition rates:

$$(a, b, c) \to (a - 1, b + 1, c) : \frac{\beta}{N} S(t)I(t) \tag{4.2.1a}$$

$$(a, b, c) \to (a, b - 1, c + 1) : \sigma E(t) \tag{4.2.1b}$$

$$(a, b, c) \to (a, b, c - 1) : \gamma I(t) \tag{4.2.1c}$$

(4.2.1a) is a schematic description of the equation

$$\begin{aligned} P((S(t + \triangle t), E(t + \triangle t), I(t + \triangle t)) = (a - 1, b + 1, c) | (S(t), E(t), I(t)) = (a, b, c)) \\ = \frac{\beta}{N} S(t)I(t) + o(\triangle t). \end{aligned}$$

The transition schematics shown by (4.2.1b) – (4.2.1c) are interpreted similarly.

4.3 Statistical Inference

4.3.1 Computer Virus Propagation Data and Notation

Assume that the epidemic is observed over the time interval $[e_1, T]$, where e_1 is the time of the first infection exposure and T is the clock time of the last observation. Our virus propagation data thus consists of all infection exposure times $\boldsymbol{e}$ observed during $[e_1, T]$; as well as the instances within this time window when some computer in the network becomes infectious and the recovery times when a computer becomes

virus free. The observed vector of infectious times instances $\boldsymbol{i}$ during $[e_1, T]$ is of the form $\boldsymbol{i} = \boldsymbol{e} + \boldsymbol{d}$, where $\boldsymbol{d}$ is the latent period spent by an infected computer in the exposed state when the infection is dormant before becoming infectious. It is reasonable to assume that latency period is small enough to guarantee that for every exposure event in $[e_1, T]$, the corresponding instances of becoming infectious falls within the observation window $[e_1, T]$ and in thus recorded. On the other hand, all computers that are recorded as having become infectious may not enter the recovered healthy state by time T, when observations end. Thus, the number of recorded recovery times $\boldsymbol{r}$ may be smaller than the number of cases of infection exposure and becoming infectious. Hence, the observed data on computer virus propagation over a time window has the form $(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r})$, where $\boldsymbol{e} = (e_1, e_2, \ldots, e_m)$, $\boldsymbol{i} = (i_1, i_2, \ldots, i_m)$ and $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)$ for some m and n such that $n \leq m \leq N$, where m and n respectively denote the number of infection/infectious cases and number of recoveries. The components of each vector are time-ordered, i.e., $e_l \leq e_{l+1}$, $i_l \leq i_{l+1}$ and $r_l \leq r_{l+1}$. The total number of computers in the network is N, with $m = n$ if the epidemic dies out and is completely observed until extinction. If $n < m$, the virus persists in the network at the time T when observations end.

For the Markovian SEIR model, with a *closed* network of computers (no 'births' or 'deaths'); the basic reproduction number is given by $R_0 = \dfrac{\beta N_s}{\gamma N}$, where N_s is the number of computers which could potentially be infected by computer virus, and determines if the epidemic will be eventually extinct, entering the 'disease-free' state $(0, 0, 0, N)$ or, persist depending on whether $R_0 \leq$ or > 1.

4.3.2 Maximum Likelihood Estimation

As a first step to estimate the parameters, the likelihood expression with data $\boldsymbol{D}^{obs} = (\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r})$ is computed. Instead of clock time, the waiting time between the different events is modeled, i.e., $t_i = t'_i - t'_{i-1}$, where t'_i denotes events in clock time of the i-th event. Likelihood of waiting times can be found using survival analysis

methodology multiple modes of failure [31]. Three events of interest are possible: infection, infectious and recovery, i.e., entering the exposed (E), infectious (I) and recovered (R) states, respectively. The corresponding hazard rate functions are

$$\lambda_{infection}(t_i \mid \beta) = \frac{\beta}{N} S(t_i) I(t_i), \tag{4.3.1a}$$

$$\lambda_{infectious}(t_i \mid \sigma) = \sigma E(t_i), \tag{4.3.1b}$$

$$\lambda_{recovery}(t_i \mid \gamma) = \gamma I(t_i), \tag{4.3.1c}$$

with $S(t_i), E(t_i)$ and $I(t_i)$ being the number of susceptible, exposed and infectious computers at time t_i, respectively. The three events of infection, becoming infectious and recovery are assumed to be mutually independent. Corresponding to the next event of 'infection', becoming 'infectious' or 'recovery', which ever occurs earlier, the overall hazard rate function λ at time t_i is thus given by

$$\lambda(t_i \mid \beta, \sigma, \gamma) = \frac{\beta}{N} S(t_i) I(t_i) + \sigma E(t_i) + \gamma I(t_i). \tag{4.3.2}$$

Between consecutive events; the hazard function $\lambda^*(t)$ expressed in *clock-time* t is a piecewise linear shift of the hazard function λ in (4.3.2). Its value at clock-time t is

$$\lambda^*(t|\beta, \sigma, \gamma) \equiv \lambda^*(t'_{i-1} + u) = \lambda(u|\beta, \sigma, \gamma),$$

if $t_i \equiv t'_i - t'_{i-1} \leq u < t'_{i+1} - t_{i-1}$ for $i \geq 1$ (setting $t'_0 = 0$); and is thus piecewise constant.

The epidemic dataset is described by $\boldsymbol{D} = (t_i, E_i)$, where $E_i \in \{$infection, infectious, recovery$\}$ denotes the event type. $\boldsymbol{D}$ consists of m infection times, m infectious times and n recovery times, where $n \leq m$. The interest here is the joint density of ($\boldsymbol{e}$, $\boldsymbol{i}$, $\boldsymbol{r}$) given the rates β, σ and γ. Consider an arbitrary single event E_i in $\boldsymbol{D}$. Counting time from the occurrence of the preceding event E_{i-1}; the density

function of the arrival time T_i to event E_i at $t > 0$, is

$$\begin{aligned} L_{E_i}(t) &= \lambda_{E_i}(t) P_{E_i}(T_i > t) \\ &= \lambda_{E_i}(t) \exp\left(- \int_0^t \lambda(u|\beta, \sigma, \gamma) du \right), \end{aligned} \tag{4.3.3}$$

where the hazard rate $\lambda_{E_i}(\cdot)$ of the i-th event E_i is given by (4.3.1a) – (4.3.1c). We have to consider however, the overall likelihood of the events in the data $\boldsymbol{D}$ at the observed clock-times t_i'; $i = 1, 2, \ldots, n+m$, when they occur. To that end, note that the inter-event arrival times $t_l = t_l' - t_{l-1}'$ are $t_l = e_l$ or, i_l if E_l is an *exposure* or *infectious* event ($E_l = e_l$, or i_l; $l = 1, 2, \ldots, m$) and $t_l = r_l$ if it is a *recovery* event ($E_l = r_l$; $l = 1, 2, \ldots, n$). Note, the survival function at the value t_i of the time to the i-th event E_i ($i = 1, 2, \ldots, m+n$) can be written as,

$$\begin{aligned} P_{E_i}(T_i > t_i) &= \exp\left(- \int_0^{t_i} \lambda(x|\beta, \sigma, \gamma) dx \right) = \exp\left(- \int_0^{t_i' - t_{i-1}'} \lambda^*(t_{i-1}' + v|\beta, \sigma, \gamma) dv \right) \\ &= \exp\left(- \int_{t_{i-1}'}^{t_i'} \lambda(u|\beta, \sigma, \gamma) du \right). \end{aligned}$$

Assuming independence of the inter-event time lengths; the overall likelihood function, considering all observed infection, infectious and recovery events, is

$$\begin{aligned} L &= \prod_{l=1}^{m+n} \lambda_{E_i}(t_i) P_{E_i}(T_i > t_i) \\ &= \prod_{l=1}^{m+n} \lambda_{E_i}(t_i) \exp(- \int_{t_{i-1}'}^{t_i'} \lambda(u|\beta, \sigma, \gamma) du) \qquad (4.3.4) \\ &= \prod_{l=1}^{m} \frac{\beta}{N} S(e_l^-) I(e_l^-) \exp(- \int_{e_{l-1}}^{e_l} \lambda(u \mid \beta, \sigma, \gamma) du) \prod_{l=1}^{m} \sigma E(i_l^-) \\ &\times \exp(- \int_{i_{l-1}}^{i_l} \lambda(u \mid \beta, \sigma, \gamma) du) \prod_{l=1}^{n} \gamma I(r_l^-) \exp(- \int_{r_{l-1}}^{r_l} \lambda(u \mid \beta, \sigma, \gamma) du) \\ &= \prod_{l=1}^{m} \frac{\beta}{N} S(e_l^-) I(e_l^-) \prod_{l=1}^{m} \sigma E(i_l^-) \prod_{i=1}^{n} \gamma I(r_l^-) \prod_{l=1}^{m} \exp(- \int_{e_{l-1}}^{e_l} \lambda(u \mid \beta, \sigma, \gamma) du) \\ &\times \prod_{l=1}^{m} \exp(- \int_{i_{l-1}}^{i_l} \lambda(u \mid \beta, \sigma, \gamma) du) \prod_{i=1}^{n} \exp(- \int_{r_{l-1}}^{r_l} \lambda(u \mid \beta, \sigma, \gamma) du), \qquad (4.3.5) \end{aligned}$$

where $S(t^-)$, $E(t^-)$ and $I(t^-)$ denote the number of susceptible, exposed and infectious computers just prior to time t, i.e., $S(t^-) = \lim_{t \to t^-} S(t)$, etc. Since the events of infection, infectiousness and recovery occur sequentially in time (see Remark 4.3.1 following (4.3.8a) – (4.3.8c) at the end of this section); the sums of integrals in the likelihood L above can be simplified into a single integral as,

$$L = \left(\frac{\beta}{N}\right)^m \prod_{l=1}^{m} S(e_l^-) I(e_l^-) \sigma^m \prod_{l=1}^{m} E(i_l^-) \gamma^n \prod_{i=1}^{n} I(r_l^-) \times \exp(-\int_{e_1}^{T} \lambda(t \mid \beta, \sigma, \gamma) dt). \tag{4.3.6}$$

With the likelihood function (4.3.6), we can obtain the log-likelihood function as

$$\begin{aligned} \ln L = {} & m \ln \beta - m \ln N + \sum_{l=1}^{m} \ln(S(e_l^-) I(e_l^-)) + m \ln \sigma + \sum_{l=1}^{m} \ln E(i_l^-) \\ & + n \ln \gamma + \sum_{l=1}^{n} \ln I(r_l^-) - \frac{\beta}{N} \int_{e_1}^{T} S(t) I(t) dt \\ & - \sigma \int_{e_1}^{T} E(t) dt - \gamma \int_{e_1}^{T} I(t) dt. \end{aligned}$$

Differentiating the above equation with respect to β, σ and γ, respectively, the following score equations are obtained

$$0 = \frac{\partial \ln L}{\partial \beta} = \frac{m}{\beta} - \frac{1}{N} \int_{e_1}^{T} S(t) I(t) dt, \tag{4.3.7a}$$

$$0 = \frac{\partial \ln L}{\partial \sigma} = \frac{m}{\sigma} - \int_{e_1}^{T} E(t) dt, \tag{4.3.7b}$$

$$0 = \frac{\partial \ln L}{\partial \gamma} = \frac{n}{\gamma} - \int_{e_1}^{T} I(t) dt; \tag{4.3.7c}$$

thus leading to the following maximum likelihood estimators (m.l.e.) of the infection, exposure and recovery rate parameters:

$$\hat{\beta} = \frac{mN}{\int_{e_1}^{T} S(t)I(t)dt}, \tag{4.3.8a}$$

$$\hat{\sigma} = \frac{m}{\int_{e_1}^{T} E(t)dt}, \tag{4.3.8b}$$

$$\hat{\gamma} = \frac{n}{\int_{e_1}^{T} I(t)dt}. \tag{4.3.8c}$$

as the unique solution of the score equations in (4.3.7).

Remark 4.3.1 In the expression (4.3.4) for the likelihood; recall that the i.i.d. random variables T_i are defined as the minimum of the times to the next event, irrespective of event type $E_i \in \{\text{infection, infectious, recovery}\}$ that occur; $i = 1, 2, \cdots$. However, the first event E_1 is necessarily an exposure to infection at clock-time $t_1' = t_1 = e_1$, and the second event E_2 is *either* the second infection *or*, the first infectious event, but cannot be a recovery (since a recovery must be preceded by an infection and a subsequent infectious event. E_2 occurs at clock-time $t_2' = t_1' + x = e_1 + x$, where $x = e_2$ or, i_1. Thus, the first two inter-event times T_1 and T_2 are not distributed as the minimum of the times to the next infection/infectious/recovery event regardless of the event type. In order to justify our maximum likelihood estimators for β, σ and γ in (4.3.8a) – (4.3.8c); we proceed to argue that (4.3.4) does indeed correctly represent the likelihood of the observed data.

In virtue of (4.3.2) and our remarks about which event types cannot occur as the first two events, together with the piecewise constancy of the overall hazard rate expressed as a function of clock-time; the ratio of the actual likelihood to its suggested

representation in (4.3.4), is easily seen to be:

$$\begin{aligned}
& \exp\left(\int_0^{t_1'} \{\sigma E(t) + \gamma I(t)\}\, dt + \int_{t_1'}^{t_2'} \gamma I(t)\, dt\right) \\
= \;& \exp\left(\int_0^{t_1'} \{\sigma E(e_1-) + \gamma I(e_1-)\}\, dt + \int_{t_1'}^{t_2'} \gamma I(i_1-)\, dt\right) \\
= \;& \exp\left\{t_1' \{\sigma E(e_1-) + \gamma I(e_1-)\} + (t_2' - t_1')\, \gamma I(i_1-)\right\} \\
\equiv \;& \exp\left\{e_1 \{\sigma E(e_1-) + \gamma I(e_1-)\} + (\min(i_1, e_2) - e_1')\gamma I(i_1-)\right\} \\
= \;& 1,
\end{aligned}$$

since, for $t \in (0, t_1')$, we have $E(t) = E(e_1-)$, $I(t) = I(e_1-)$, and prior to the first infection event, there are no computers in the exposed or, infectious state ($E(e_1-) = 0 = I(e_1-)$). Similarly, we argue that for all clock-times $t \in [t_1', t_2')$, we must also have $I(t) = I(i_1-) = 0$. This follows, since $t_1' \leq t < t_2'$ corresponds to 'time to the second event' $t_2 \in [0, t_2' - t_1') = \min(e_2, i_1)$. Then, *either* $e_2 < i_1$ (2^{nd} exposure occurs before 1^{st} infectious event), *or*, $i_1 < e_2$. In both cases, the second event occurs before the first infectious event, which implies $I(t) = I(i_1-) = 0$.

4.3.3 Bayesian Estimation

We assume the prior distributions of the unknown transition rate parameters β, σ and γ, respectively, to be *gamma* distributed as

$$\beta \sim \Gamma(\nu_\beta, \lambda_\beta), \qquad \sigma \sim \Gamma(\nu_\sigma, \lambda_\sigma), \qquad \gamma \sim \Gamma(\nu_\gamma, \lambda_\gamma),$$

where $\Gamma(\nu, \lambda)$ denotes the Gamma distribution with mean $\dfrac{\nu}{\lambda}$ and variance $\dfrac{\nu}{\lambda^2}$. In virtue of Bayes Formula, it is well known that

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}.$$

Ignoring the normalizing constants, the posterior distributions of β, σ and γ, up to a constant of proportionality, are respectively given by

$$\begin{aligned}
\pi(\beta \mid \boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}, \sigma, \gamma) &\propto f(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r} \mid \beta, \sigma, \gamma)\pi(\beta) \\
&\propto \prod_{l=1}^{m} \frac{\beta}{N} S(e_l^-) I(e_l^-) \exp\left(- \int_{e_1}^{T} \left(\frac{\beta}{N} S(t) I(t) + \sigma E(t) + \gamma I(t) \right) dt \right) \\
&\times (\beta)^{\nu_\beta - 1} \exp(-\lambda_\beta \beta) \\
&\propto (\beta)^m \exp\left(- \int_{e_1}^{T} \frac{\beta}{N} S(t) I(t) dt \right) (\beta)^{\nu_\beta - 1} \exp(-\lambda_\beta \beta) \\
&= (\beta)^{\nu_\beta + m - 1} \exp\left(-\lambda_\beta \beta - \int_{e_1}^{T} \frac{\beta}{N} S(t) I(t) dt \right) \\
&\sim \Gamma(\nu_\beta + m, \lambda_\beta + \frac{1}{N} \int_{e_1}^{T} S(t) I(t) dt), \qquad (4.3.9)
\end{aligned}$$

$$\begin{aligned}
\pi(\sigma \mid \boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}, \beta, \gamma) &\propto f(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r} \mid \beta, \sigma, \gamma)\pi(\sigma) \\
&\propto \prod_{l=1}^{m} \sigma E(i_l^-) \exp\left(- \int_{e_1}^{T} \left(\frac{\beta}{N} S(t) I(t) + \sigma E(t) + \gamma I(t) \right) dt \right) \\
&\times \sigma^{\nu_\sigma - 1} \exp(-\lambda_\sigma \sigma) \\
&\propto \sigma^m \exp\left(- \int_{e_1}^{T} \sigma E(t) dt \right) \sigma^{\nu_\sigma - 1} \exp(-\lambda_\sigma \sigma) \\
&= \sigma^{\nu_\sigma + m - 1} \exp\left(-\lambda_\sigma \sigma - \int_{e_1}^{T} \sigma E(t) dt \right) \\
&\sim \Gamma(\nu_\sigma + m, \lambda_\sigma + \int_{e_1}^{T} E(t) dt), \qquad (4.3.10)
\end{aligned}$$

and

$$\begin{aligned}
\pi(\gamma \mid \boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}, \beta, \sigma) &\propto f(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r} \mid \beta, \sigma, \gamma)\pi(\gamma) \\
&\propto \prod_{l=1}^{n} \gamma I(rl^{-}) \exp\left(- \int_{e_1}^{T} \left(\frac{\beta}{N} S(t)I(t) + \sigma E(t) + \gamma I(t) \right) dt \right) \\
&\times \gamma^{\nu_\gamma - 1} \exp(-\lambda_\gamma \gamma) \\
&\propto \gamma^{n} \exp\left(- \int_{e_1}^{T} \gamma I(t) dt \right) \gamma^{\nu_\gamma - 1} \exp(-\lambda_\gamma \gamma) \\
&= \gamma^{\nu_\gamma + n - 1} \exp\left(-\lambda_\gamma \gamma - \int_{e_1}^{T} \gamma I(t) dt \right) \\
&\sim \Gamma(\nu_\gamma + n, \lambda_\gamma + \int_{e_1}^{T} I(t) dt). \qquad (4.3.11)
\end{aligned}$$

The Bayesian estimators for β, σ and γ are then obtained as the *posterior means* of the distributions (4.3.9) – (4.3.11), as

$$\hat{\beta} = \frac{\nu_\beta + m}{\lambda_\beta + \frac{1}{N} \int_{E_1}^{T} S(t)I(t)dt}, \qquad (4.3.12a)$$

$$\hat{\sigma} = \frac{\nu_\sigma + m}{\lambda_\sigma + \int_{E_1}^{T} E(t)dt}, \qquad (4.3.12b)$$

$$\hat{\gamma} = \frac{\nu_\gamma + n}{\lambda_\gamma + \int_{E_1}^{T} I(t)dt}. \qquad (4.3.12c)$$

These estimates are applied to Code Red worm data in Section 4.4

4.4 Case Study: Code Red Worm Data

The observed Code Red worm data $\boldsymbol{D}^{obs}$ includes the start time and end time for each computer in a /8 network at UCSD, where the start time and the end time corresponds to the infection time and the recovery time respectively, i.e., $\boldsymbol{D}^{obs} = (\boldsymbol{e}, \boldsymbol{r})$. At any given time, a computer is characterized as being (1) susceptible, (2) exposed (infected but not infectious), (3) infectious, or (4) recovered. The number of computers in each

of the susceptible, exposed and infectious compartments at time t are $S(t)$, $E(t)$ and $I(t)$, respectively.

4.4.1 Parameters Estimation Results

The maximum likelihood estimators of the parameters for the Code Red worm data are computed based on the results obtained in section 4.3.2. The MLEs so computed are $\hat{\beta} = 3.2100$, $\hat{\sigma} = 0.11$ and $\hat{\gamma} = 5.625851 \times 10^{-5}$.

These *maximum likelihood estimates* lead to the following *interpretations* for the Code Red worm propagation:

1) an infectious computer makes contact with approximately three (3) susceptible computers per second;

2) an exposed computer (infected by Code Red worm but not yet infectious) in state E becomes infectious (i.e., goes to state I) with rate 0.11/sec. In other words, we can say that the average latent period is about $(0.11)^{-1} \approx 9$ seconds;

3) an infectious computer in state I, is recovered from non-healthy status and has immunity at rate 5.625851×10^{-5} computers/sec. This corresponds to an average recovery time of $(5.625851)^{-1} \times 10^5 \approx 17775.09\,\text{sec} \approx 4.94\,\text{hrs}$ per computer. Because of the relatively small value of recovery rate and the correspondingly relatively large recover time, the Code Red worm propagated rapidly (viz., m.l.e. of the basic reproduction number is $\hat{R}_0 = \dfrac{\hat{\beta} N_s}{\hat{\gamma} N} = 5.3139 > 1$).

Bayesian estimators of the parameters for the Code Red worm data also can be computed based on the results obtained in section 4.3.3. Here, $\hat{\beta} = 3.2421$, $\hat{\sigma} = 0.10$ and $\hat{\gamma} = 5.681790 \times 10^{-5}$.

The *Bayesian estimates* for the Code Red worm propagation, parameters are similarly *interpreted*; viz.,

1) contacts with approximately three (3) susceptible computers per second is made by each infectious computer;

2) an exposed computer (infected by Code Red worm but not infectious) has an average latent period of $(0.1)^{-1} \approx 9$ seconds;

3) infectious computers in the network are recovered to a healthy state at the rate $\gamma \approx 5.681790 \times 10^{-5}$ computers/sec. To a typical recovery of an infectious computer takes $\gamma^{-1} \approx 17600.09\,\text{sec} \approx 4.89\,\text{hrs}$ on average.

Table 4.1 below summarizes the model parameter estimates using maximum likelihood method and Bayesian method.

Table 4.1 Summary of Parameter Estimates for Stochastic SEIR Model using Markovian Approach

Estimates	$\hat{\beta}$	$\hat{\sigma}$	$\hat{\gamma}$
Maximum Likelihood Estimates	3.2100	0.11	5.625851×10^{-5}
Bayesian Estimates	3.2421	0.10	5.681790×10^{-5}

4.4.2 Simulation Results

4.4.2.1 Description of Simulation. The simulation program is written using Fortran under Window XP environment. In the simulation, we have $N = 2^{32}$ (the IP address space), $N_s = 400,000$ (an approximation of the size of the susceptible Code Red population), $I(0) = 1$ (the size of initial infection) for the two stochastic SEIR models with MLE and Bayesian estimates. The system in our simulation consists of $N = 2^{32}$ computers that can reach each other directly and thus there is no topology issue in our simulation. For classic simple SI model and two-factor worm model, with associated parameters used in [72] and [90], the governing ordinary differential equations are solved numerically using ODE solver in Fortran and the solutions of those two models are plotted in Figure 4.4, orange solid line and green solid line, respectively. For both stochastic SEIR models (with MLE and Bayesian estimates), we carried out 5000 simulation runs. We plot the average values of the outcome of

5000 simulation runs in Figure 4.4 for stochastic SEIR model with MLE and stochastic SEIR model with Bayesian estimates, red solid line and blue solid line, respectively.

4.4.2.2 Results. For the purpose of comparison, we plot the Figure 4.2 and Figure 4.3. We observe that our stochastic SEIR models (with MLE and with Bayesian estimates) well match the observed Code Red worm data.

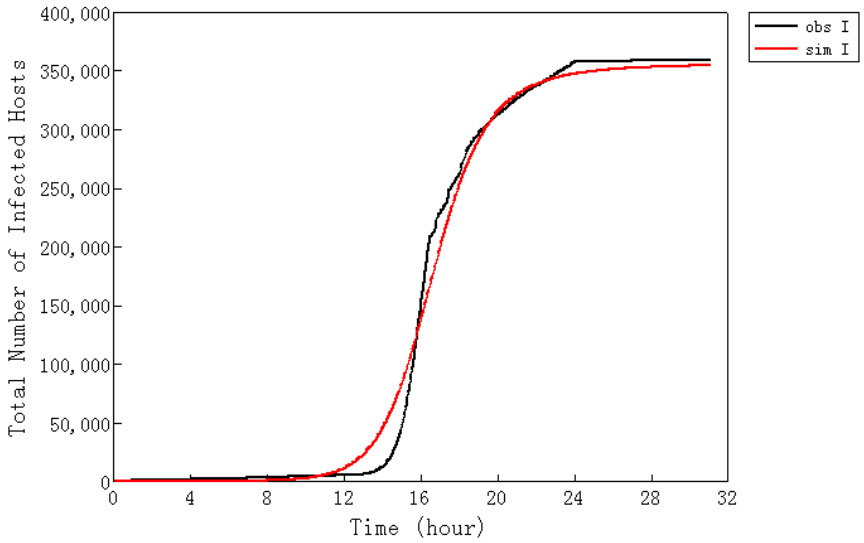

Figure 4.2 Observed data vs. Code Red worm simulation plot based on maximum likelihood estimates of parameters.

For the Code Red data, we also contrast our results of the stochastic SEIR model with the results of other models used in the literature to study Code Red worm, we do some simulation experiments. One of these is the classic SI (susceptible-infected) model presented in [72], which is deterministic and does not consider the latent period or allow recovery from a virus and acquiring a corresponding immunity. The other is the two-factor worm model presented in [90]. The two-factor worm model is also a deterministic model that considers the impacts of i) the effect of human countermeasures against worm spreading and ii) the slowing down of worm infection rate, worm's impact on Internet traffic and infrastructure. The comparative

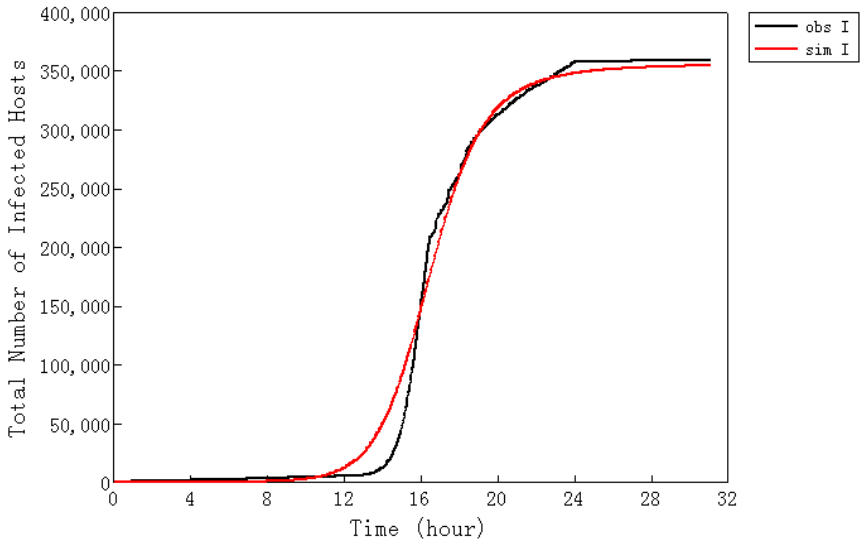

Figure 4.3 Observed data vs. Code Red worm simulation plot based on Bayesian estimates of parameters.

result of all models along with ours, are summarized in Table 4.2, with the plots of all simulated outputs shown in Figure 4.4.

Figure 4.4 shows that

1) The classic simple SI model consistently overestimates the actual infected population (observed data), represented by the black solid line. By contrast, from the overlay graphs of SEIR model simulations outputs, using both maximum likelihood (MLE) and Bayesian estimates as input parameters, is easily seen to provide a much closer fit to the observed data.

2) The two-factor worm model output, shown by green solid line in Figure 4.4, increases dramatically in the range 19/20-24 hours, which accounts for as much as 20-25% of the effective significant range of the outbreak period of the Code Red worm data (approximately 12-24 hours) and thereby increasingly overestimates the actual infections during the last 20% of the outbreak period, by the end of which the infection has already reached its saturation level.

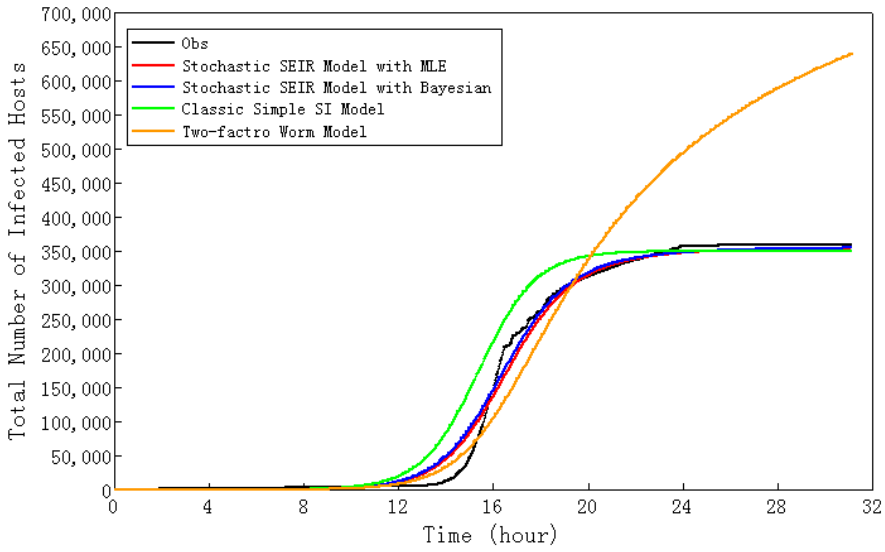

Figure 4.4 Comparison among observed data, stochastic SEIR model with maximum likelihood estimates, stochastic SEIR model with Bayesian estimates, classic simple SI model and Two-factor worm model.

To compare the four models, stochastic SEIR model with ML estimates, stochastic SEIR model with Bayesian estimates, classic simple SI model and two-factor worm model, to see which one is better, the standard deviation of the 'discrepancy' (= simulated output − actual observed) is computed for each of them. Table 4.2 summarizes the maximum and average discrepancy along with the standard deviation of the discrepancies, for each of the four models. Both MLE and Bayes estimate based simulations of the stochastic SEIR model have smaller standard deviation of the discrepancy compared to classic simple SI and two-factor worm models; a finding that is consistent with the visual plots of the simulations shown in Figure 4.4.

4.4.2.3 Simulation Results for Recovery Profiles. To examine the time profile of the number of recoveries, we also plot the observed total number of recovered hosts for the Code Red data and the simulation results of the recovered hosts against clock time using our stochastic SEIR model with the MLE estimates of β, σ and γ as input parameters in Figure 4.5 (blue solid line and green solid line respectively).

Table 4.2 Summary of Simulation Results from Stochastic SEIR Model using Markovian Approach and Comparisons with other Previous Models

	Standard Deviation	Maximum Discrepancy	Average Discrepancy
Stochastic SEIR (MLE)	13.3572	40.7684	−0.5910
Stochastic SEIR (Bayesian Estimates)	13.3128	45.9277	2.9181
Classic Simple SI Model	28.4853	100.6507	19.7426
Two-factor worm Model	41.9134	136.2675	8.9397

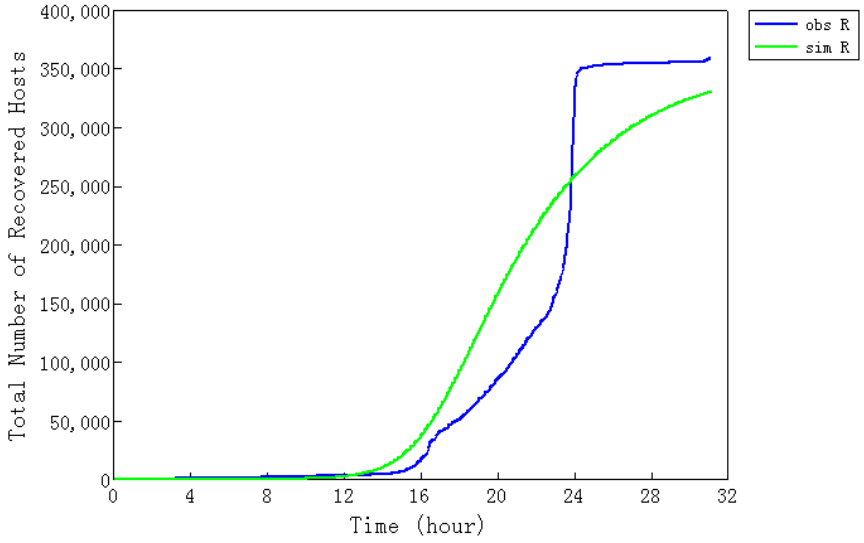

Figure 4.5 Observed and simulated total number of recovered hosts for Code Red worm data.

From Figure 4.5, it would appear that, between approximately 14-23 hours since the first instance of Code Red infection is detected and recorded, the simulated number of recoveries consistently overestimates the actual number of recovered hosts; and after which the actual number steeply climbs to overtake the simulated number of recoveries by the time the clock reads 24 hours since the first detected infection. This apparent mismatch of the number of recoveries is less serious than it seems and can be explained by noting that people are initially slow to react to a malware attack until it is fairly widespread when they actually begin to take protective countermeasures such as using antivirus software to remove/clean infected files. There is usually a time

lag between the time when an infection is detected in a computer and the time when a corresponding countermeasure is taken while the SEIR model does not include any countermeasure factors and thus effectively assumes no delay in implementing them. Such lags can explain why initially the simulated number of recoveries overestimates the actual number of recovered computers up to a certain clock time.

Simulation of the results of recoveries over time where also undertaken for other models of the SEIR *genre* and input parameter estimating methods explored in the subsequent Chapters 5 – 6. The various scenarios so considered for simulating the number of recoveries in time were: **(A)** for *homogeneous* networks, simulated recoveries with A-i) full Bayesian estimates (complete data), A-ii) MCMC based estimates (incomplete data with infection times missing), as well as **(B)** for multi-group *heterogeneous* networks (Code Red data classified into seven groups based on country location) based on B-i) Markovian model with maximum likelihood estimates (MLE), and B-ii) SDE-based least squares estimates. For the Code Red data, the profiles of actual recoveries vs. the corresponding simulation results obtained for the different cases listed above for homogeneous as well as multi-group heterogeneous networks are seen to be similar (see Figures 4.6 – 4.9) and mimic the profile of Figure 4.10, which assumes a single homogeneous network for the propagation of the 'Code Red'.

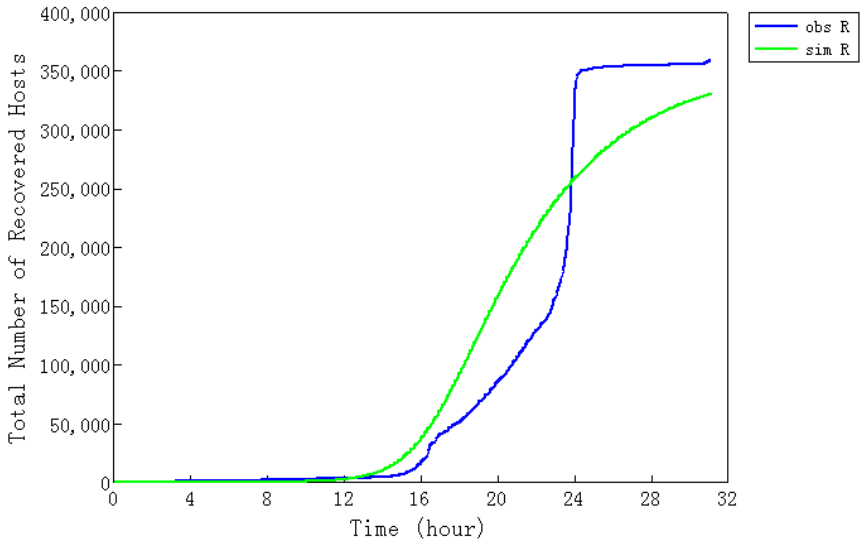

Figure 4.6 Observed and simulated total number of recovered hosts for Code Red worm data using Bayesian estimates (single homogeneous network).

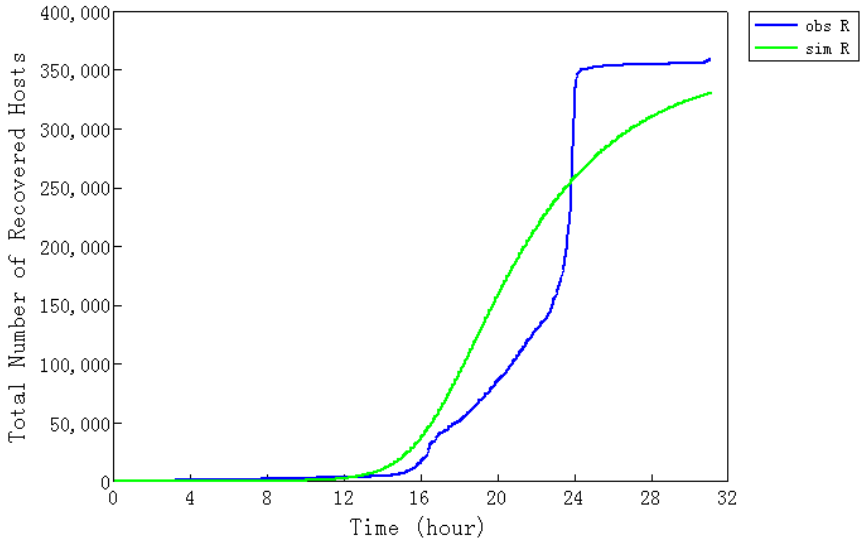

Figure 4.7 Observed and simulated total number of recovered hosts for Code Red worm data using MCMC estimates (infection times missing, single homogeneous network).

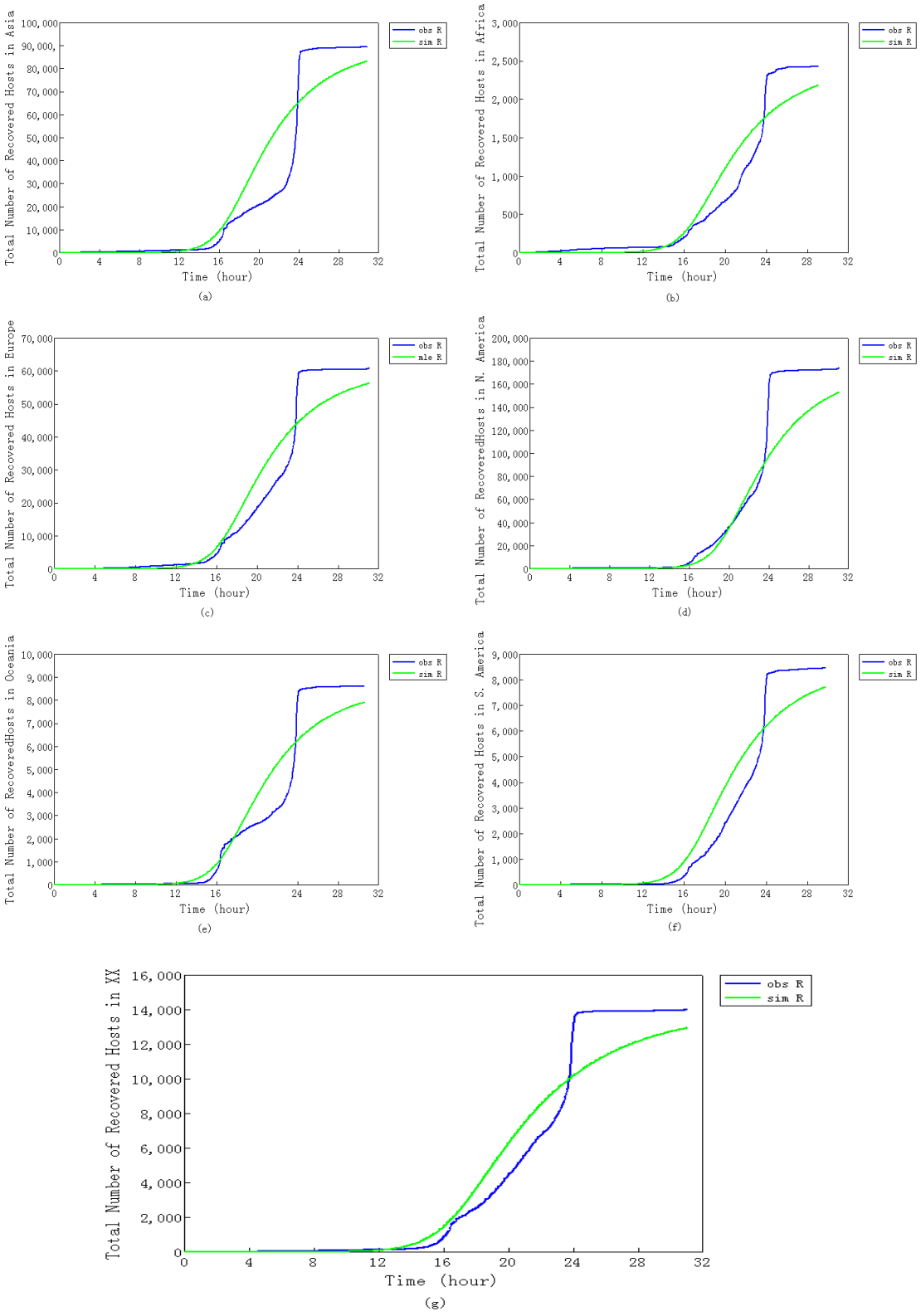

Figure 4.8 (Multi-group Markovian model) Observed and simulated number of recovered hosts by groups: (a) Asia (b) Africa (c) Europe (d) North America (e) Oceanic (f) South America and (g) XX (location unknown).

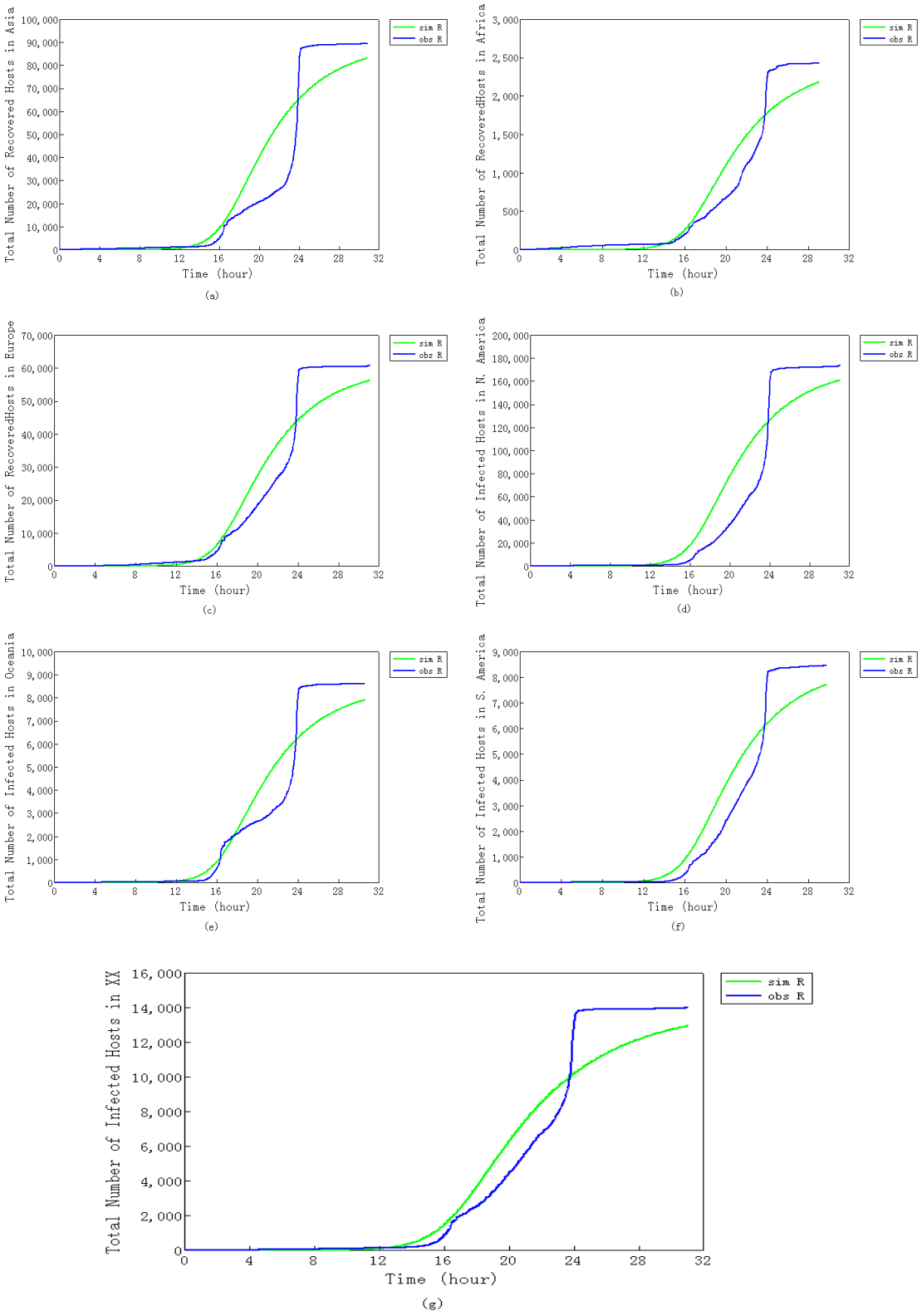

Figure 4.9 (Multi-group SDE model) Observed and simulated number of recovered hosts by groups: (a) Asia (b) Africa (c) Europe (d) North America (e) Oceanic (f) South America and (g) XX (location unknown).

As the Code Red virus has a designed active lifespan of 24 hours from the time it goes out into a network, and the first case of infection can be detected and recorded only at a time later than its initiation; it is no longer active in the network when the recording clock reads 24 hours – at which time the actual number of infected hosts also appears to have reached its saturation level; although observations on detected Code Red infections continued until about 32 hours of clock time since the first infection, with no new cases detected after 24 hours. The overlay plots in Figure 4.10 exhibits all these features in the same graph.

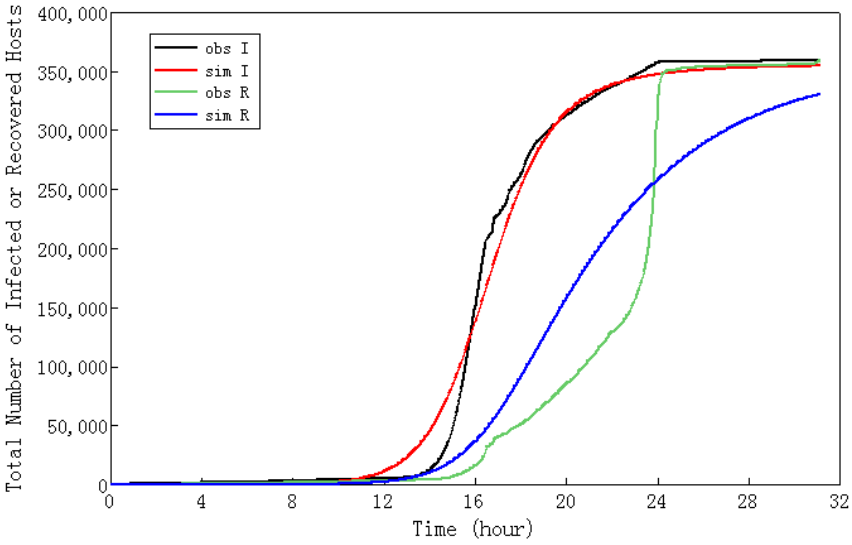

Figure 4.10 Overlay Plots of Code Red worm data vs. simulation plot for the whole network (based on maximum likelihood estimates of parameters).

From Figure 4.5 and Figure 4.10, coupled with the steeply accelerating pace of actual recoveries from about 23 hours, the gap between the actual and simulated number of recoveries continues to diminish, as clean up measures are more aggressively undertaken at an aggressive pace.

4.5 Discussion

In this Chapter, Stochastic SEIR model using a Markovian approach has been adapted to the context of computer virus propagation. Survival analysis methodology is used to express the likelihood function to estimate the transition parameters of the stochastic SEIR model using maximum likelihood and Bayesian methods. With values of the maximum likelihood estimators and Bayesian estimators as inputs, simulations have been conducted and compared with the observed Code Red worm data. It is found that the proposed stochastic SEIR model closely matches the observed data. In fact, comparing our simulation results for the stochastic SEIR model with the classic simple SI model and the two-factor worm model – both of which are deterministic; we observe that stochastic SEIR model provides a better match to the observed Code Red worm data.

CHAPTER 5

MARKOV CHAIN MONTE CARLO (MCMC) METHODS FOR PARTIALLY OBSERVED COMPUTER VIRUS PROPAGATION – STOCHASTIC SEIR MODEL

5.1 Introduction

In practice, the nature of data that are available for computer virus epidemics are typically not as exhaustive as would be desirable for modeling and inference. For example, it is unlikely that the precise times when individual computers became infected during an epidemic are known, or that other relevant details of the infection process would be observed. Consequently, for many of the epidemic models, it becomes difficult to write down the full likelihood function in the spirit of Chapter 4, where all relevant variables (exposure, infectious, recovery times) are observed. However, researchers are still interested in estimation of parameters, β, σ and γ, despite of missing infection times. Our results reported in this Chapter based on Markov Chain Monte Carlo (MCMC) methods to deal with the partially observed computer virus propagation process, using only the observed recovery times.

5.2 Model and Methodology

Stochastic SEIR model using a Markovian approach, as in Chapter 4, will be used in this Chapter, to analyze a partially observed computer virus epidemic process, where only recovery times are observed, and all infection times are missing. Bayesian inference using Markov chain Monte Carlo (MCMC) method are used to estimate the model parameters.

We use Markov chain Monte Carlo (MCMC) to generate a Markov chain whose stationary distribution is the joint posterior distribution of the parameters. Inferences

about the parameters based on the Markov chain sample are made using Monte Carlo integration.

Since the computer virus propagation is partially observed, i.e., the observed data consist only of a set of recovery times $\boldsymbol{r}$; MCMC approach here treats the unobserved infection times $\boldsymbol{e}$ as parameters of the model except the transition parameters, β, σ and γ, in the model. The following notations are adapted from Chapter 4. Let the *observed* recovery times be $\boldsymbol{r} = (r_1, r_2, \ldots, r_n)$, where $0 < r_1 \leq r_2 \leq \cdots \leq r_n = T$. The vector $\boldsymbol{e} = (e_1, e_2, \ldots, e_m)$ and $\boldsymbol{i} = (i_1, i_2, \ldots, i_m)$ respectively denote the *unobserved* infection exposure times and infectious times after latency; where $0 < e_1 \leq e_2 \leq \cdots \leq e_m$ and $e_j \leq r_j$ for $j = 1, 2, \ldots, m$ and $0 < i_1 \leq i_2 \leq \cdots \leq i_m$ and $i_j \leq r_j$ for $j = 1, 2, \ldots, m$. If the epidemic is known to have ceased, then it must be true that $m = n$; in general, $n \leq m \leq N$.

A Gibbs sampler within Metropolis sampling scheme will be used to generate random samples from the desired posterior distribution. Relevant details about how to sample from the appropriate conditional distributions are described below. First, note that the density of $(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r})$ conditionally on β, σ and γ is given by

$$\begin{aligned} L &= f(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r} | \beta, \sigma, \gamma) \\ &= \left(\frac{\beta}{N}\right)^m \prod_{j=1}^{m} S(e_j^-) I(e_j^-) \sigma^m \prod_{j=1}^{m} E(i_j^-) \gamma^n \prod_{j=1}^{n} I(r_j^-) \\ &\quad \times \exp(-\int_{e_1}^{T} \lambda(t \mid \beta, \sigma, \gamma) dt), \end{aligned} \tag{5.2.1}$$

where $\lambda(t \mid \beta, \sigma, \gamma) = \dfrac{\beta}{N} S(t) I(t) + \sigma E(t) + \gamma I(t)$, using the notations in Chapter 4, and where ${e_j}^-$ denotes the left limit. We assume that β, σ and γ have *gamma* prior distributions with parameters $(\nu_\beta, \lambda_\beta)$, $(\nu_\sigma, \lambda_\sigma)$ and $(\nu_\gamma, \lambda_\gamma)$ respectively. Then following from equation (5.2.1) and using Γ to denote the family gamma distributions,

the posterior distributions of β, σ and γ are easily shown to be:

$$\pi(\beta \mid \boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}, \sigma, \gamma) \sim \Gamma(\nu_\beta + m, \lambda_\beta + \frac{1}{N}\int_{e_1}^{T} S(t)I(t)dt), \tag{5.2.2}$$

$$\pi(\sigma \mid \boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}, \beta, \gamma) \sim \Gamma(\nu_\sigma + m, \lambda_\sigma + \int_{e_1}^{T} E(t)dt), \tag{5.2.3}$$

$$\pi(\gamma \mid \boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}, \beta, \sigma) \sim \Gamma(\nu_\gamma + n, \lambda_\gamma + \int_{e_1}^{T} I(t)dt). \tag{5.2.4}$$

For simplicity, the notation $f(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r}|\beta, \sigma, \gamma)$ for the joint density of $(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r})$ in (5.2.1) will be abbreviated as $f(\boldsymbol{e})$. Then the sampling scheme is as follow:

- **Initial state vector**

  Initial values for β, σ and γ are obtained by sampling from their respective priors, $\Gamma(\nu_\beta, \lambda_\beta)$, $\Gamma(\nu_\sigma, \lambda_\sigma)$ and $\Gamma(\nu_\gamma, \lambda_\gamma)$. By fixing the seed value of the random generator, it is possible to assure fixed initial values. Exposure times $\boldsymbol{e}$ are generated by sampling from a uniform distribution over the set of valid configurations. We draw m independent values from the uniform distribution on $(0, r_n)$, sort them in ascending order and check whether they obey the constraint of $e_j < r_j$ for $1 \leq j \leq m$. This procedure is repeated until a valid sample of infection times is obtained.

- **Generating new states by Gibbs within Metropolis**

  A Gibbs sampling algorithm within Metropolis sampling scheme is used to update the state vector $(\beta, \sigma, \gamma, \boldsymbol{e})$. The first three components, β, σ and γ, are updated by sampling from the full conditional distributions (5.2.2), (5.2.3) and (5.2.4), respectively. The Metropolis sampler is used for sampling from the uniform distribution of $\boldsymbol{e}$ given $(\boldsymbol{r}, \beta, \sigma, \gamma)$. Because the epidemic is assumed to be observed to its end, we know the size (m) of the exposure times, $\boldsymbol{e}$. The only operation is the moving of an infection event in time using the Metropolis algorithm, is as follows:

i) choose one from the existing infection times, $e_1, e_2, \ldots, e_m$, uniformly at random and denote it as s,

ii) generate a replacement time t by sampling uniformly on (e_1, r_n),

iii) accept the new infection candidate with probability

$$\min(1, \frac{f(\boldsymbol{e} - \{s\} + \{t\})}{f(\boldsymbol{e})}).$$

While implementing, the infection times vector $\boldsymbol{e}$ is represented as an array. After the new infection candidate t is written to s's position in $\boldsymbol{e}$, the array needs to be sorted before the acceptance probability can be computed.

5.3 Data and Notation

Data used in this Chapter will be partially observed, consists only of the observed recovery times, $\boldsymbol{r}$. At any given time t, a computer is characterized as being (i) susceptible, (ii) exposed (infected but not infectious), (iii) infectious, or (iv) recovered. The number of computers in each of susceptible, exposed, infectious and recovered compartments at time t are $S(t), E(t), I(t)$ and $R(t)$, respectively. β, σ and γ are infection rate, infectious rate and removal rate, respectively. Moreover, the total population is close, homogeneous and uniformly mixing such that $S(t) + E(t) + I(t) + R(t) = N$, where N is the total population size (number of computers in the network). The notations of various parameters and variables, as described in Section 4.3.1, remain in effect.

5.4 Case Study: Code Red Worm Data

In this section, results of applying MCMC method to Code Red worm data, where the infection times are treated as missing values, are summarized. Our data consists of observed recovery times only. Since the Code Red worm epidemic process is indeed

observed to its end [59], the sampling scheme described in Section 5.2 can be applied to Code Red worm data.

5.4.1 MCMC Estimation

To make inferences for transition parameters of interest in stochastic SEIR model using Markovian approach, 10,000 samples from 1,000,000 MCMC iterations with thin factor of 100 are used following a burn-in of 5000 iterations. The posterior distribution of the parameters are shown in Figure 5.1, and the corresponding point estimates (mean, median and some percentiles) are summarized in Table 5.1. The percentiles can be used to construct confidence interval of the parameters.

Table 5.1 Summary of Parameters Estimates using MCMC Methods for Code Red Worm Data with Observed Recovery Times Only

Parameters	Mean	Standard Deviation	MC Error	2.5%	Median	97.5%
β	3.206544	0.024427	0.000245	3.15866	3.20772	3.25048
σ	0.115749	0.000425	4.26×10^{-6}	0.11590	0.11575	0.11665
γ	5.61×10^{-5}	1.69308×10^{-6}	1.71×10^{-8}	5.29687×10^{-5}	5.63×10^{-5}	5.88577×10^{-5}

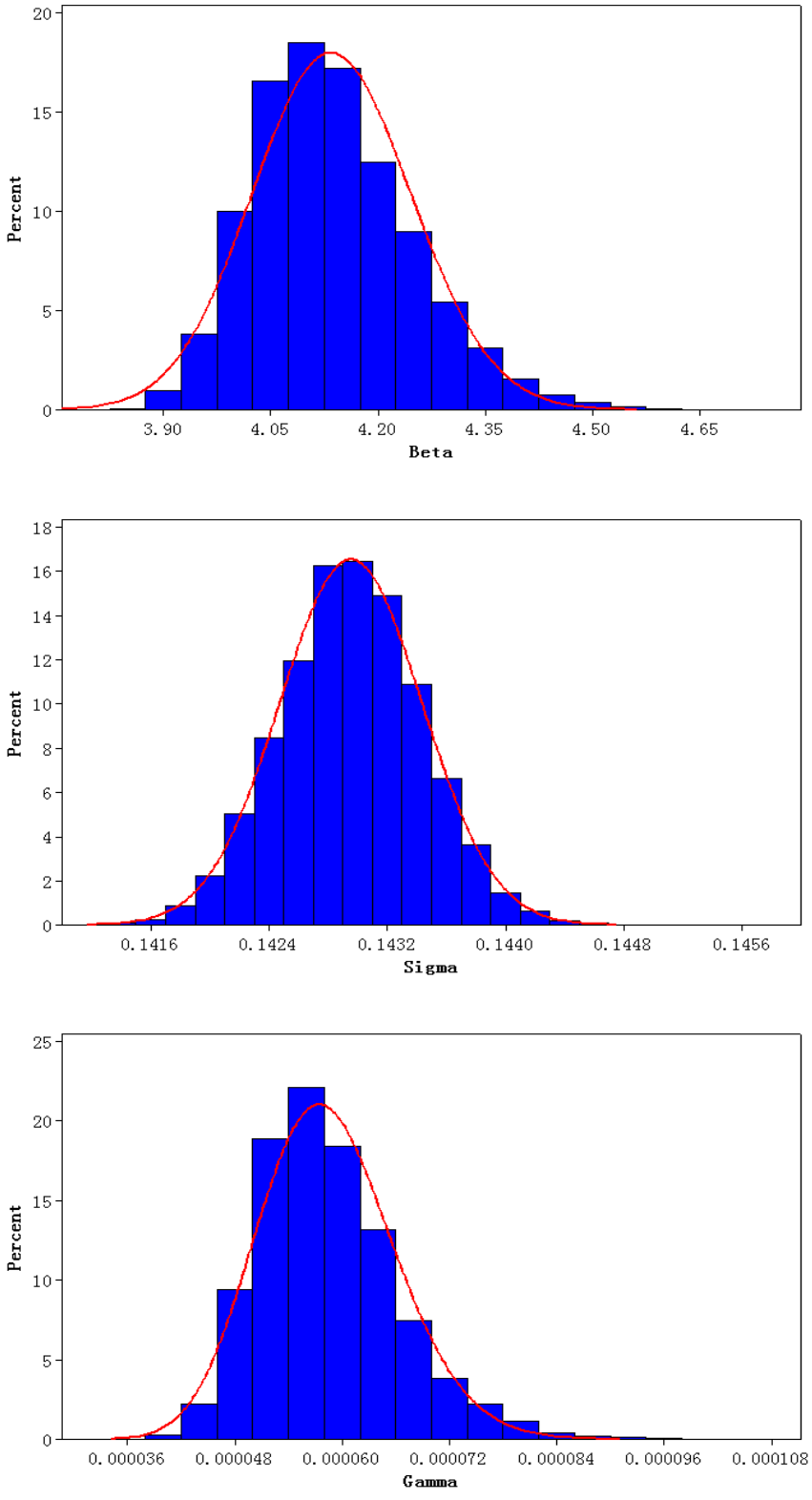

Figure 5.1 Posterior distributions of the model parameters β, σ and γ for the Markovian SEIR model with Code Red worm data.

Table 5.2 contrasts the MCMC estimate of the transition parameters for the partially observed Code Red worm data (of recovery times only) to those obtained as MLE and Bayes estimates, reported in Chapter 4 for the completely observed Code Red worm data. It shows that even with exposure and infectious times missing, MCMC-based estimates of the epidemic parameters are very close to those obtained when complete data are available.

Table 5.2 Comparison among Maximum Likelihood Estimates, Bayesian Estimates and MCMC Estimates

	β	σ	γ
Maximum Likelihood Estimates	3.2100	0.11	5.625851×10^{-5}
Bayesian Estimates	3.2421	0.10	5.681790×10^{-5}
Markov chain Monte Carlo (MCMC) Estimates	3.2065	0.1157	5.61×10^{-5}

5.4.2 Simulation Results

To further judge the extent to which MCMC-based estimates can adequately capture the Code Red virus propagation, simulations were performed using stochastic SEIR model with MCMC estimates as input values of model parameters, β, σ and γ. Figure 5.2 shows that the simulation results, which closely match the observed Code Red worm data.

For purposes of comparison, discrepancy measures are computed for each of the simulation scenarios. The results are summarized in Table 5.3. Moreover, Figure 5.3 graphically exhibits the Code Red propagation time trajectory for the different model setups of Table 5.3. The stochastic SEIR model with MLE, stochastic SEIR model with Bayesian and stochastic SEIR with MCMC estimators better fit the actual Code Red worm data much closer than the previous works, classic simple SI model and two-factor worm model reported in the literature.

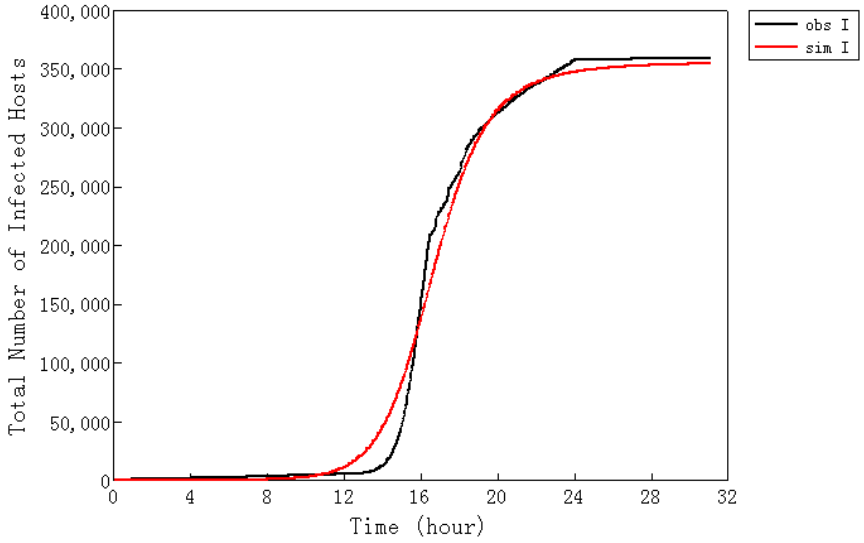

Figure 5.2 Comparison between observed Code Red worm data and simulation based on MCMC estimates for parameters.

Table 5.3 Summary of Stochastic SEIR Simulations and Comparison with other Models

	Standard Deviation of Discrepancy	Maximum Discrepancy	Average Discrepancy
Stochastic SEIR (MLE)	13.3572	40.7684	−0.5910
Stochastic SEIR (Bayesian Estimation)	13.3128	45.9277	2.9181
Stochastic SEIR (MCMC Estimation)	12.0984	39.3951	−1.0646
Classic Simple SI Model	28.4853	100.6507	19.7426
Two-factor worm Model	41.9134	136.2675	8.9397

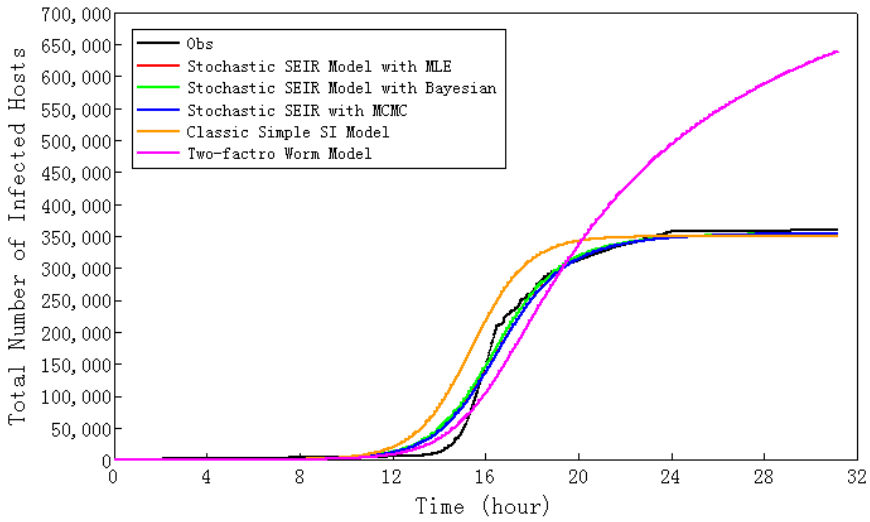

Figure 5.3 Comparison between observed Code Red worm data and simulations based on different models: Stochastic SEIR model with MLE, Stochastic SEIR model with Bayesian estimates, Stochastic SEIR model with MCMC estimates, Classic Simple SI Model and Two-factor Worm Model.

5.5 Discussion

In this Chapter, Bayesian inference using Markov chain Monte Carlo (MCMC) method is introduced to the context of computer virus propagation. The MCMC method can be used to estimate model parameters when all relevant variables of the computer virus propagation are not completely observed. The model and methodology are applied to Code Red worm data by treating the infection times as missing values. The simulation results show that MCMC method is efficient for dealing with parameter estimations with partially observed computer virus epidemic process.

CHAPTER 6

MULTI-GROUP MARKOVIAN AND SDE-BASED SEIR MODELS AND INFERENCE FOR COMPUTER VIRUS PROPAGATION

6.1 Introduction

In both Chapters 4 and 5, the stochastic SEIR model in a Markovian setup to study computer virus propagation, the underlying network is assumed to be homogeneous. It has been argued, however, that network structure has an impact on the speed and other characteristics of propagation of a computer virus [66, 67, 86], and thus the assumption of homogeneity may need to be relaxed. In this Chapter, two multi-group stochastic SEIR models will be introduced to study computer virus prevalence. Maximum likelihood estimation for model parameters of interest will be developed for multi-group stochastic SEIR model using Markovian approach. Furthermore, stochastic differential equations (SDE) will be used to construct a multi-group SEIR model, denoted as multi-group stochastic SEIR-SDE model. The method of least squares will be used to estimate parameters of interest for multi-group stochastic SEIR-SDE model.

6.2 Model and Methodology

6.2.1 Data and Notation

It is assumed that the epidemic is observed in the time interval $[e_1, T]$, where e_1 is the time of the first exposure to infection in the network; and at time T, all infected computers have been recovered, so that the epidemic is completely observed. We continue to use the *same notations and conventions to describe our data and the corresponding variables* as in Chapter 4. Our data thus consists of m observed exposure to infection times $\boldsymbol{e}$ during $[e_1, T]$ and $n(\leq m)$ recovery times $\boldsymbol{r}$ during $[e_1, T]$. Because the infected computer will stay in the exposed state for the length of a *latent*

period before it becomes infectious and is capable to infect other computers, we have another sequence of m observed infectious times $\boldsymbol{i}$ during $[e_1, T]$, where $\boldsymbol{i} = \boldsymbol{e} + \boldsymbol{d}$, and $\boldsymbol{d}$ denotes latent periods during which infected computers remain dormant.

As a way of considering the structure of network, in this dissertation the whole Internet will be divided into several subnets (groups) according to some characteristic information, i.e., if the computers who have the same characteristics will considered to be from the same group. Now suppose the Internet is divided into g subnets. We use the following notational conventions to denote the state variables and transmission parameters within and between subnets in the multi-group SEIR setup:

i) ($S_j(t)$, $E_j(t)$, $I_j(t)$ and $R_j(t)$) is the random vector representing the number of (susceptible, exposed, infectious, recovered) computers at time t satisfying $S_j(t) + E_j(t) + I_j(t) + R_j(t) = N_j$, where N_j is the total number of computers in subnet j;

ii) transmission parameters:

- β_{jk}: infection rate from subnet j to k, where $j, k = 1, 2, \ldots, g$. (Note, β_{jj} represents *intra-subnet* infection rate; i.e., the infection rate within the subnet j, $j = 1, 2, \ldots, g$.)
- σ_j: infectious rate within subnet j, where $j = 1, 2, \ldots, g$.
- γ_j: recovery rate within subnet j, where $j = 1, 2, \ldots, g$.

6.2.2 Multi-group Stochastic SEIR Model using Markovian Approach

As a first step to estimate the parameters of interest, the likelihood expression with data $(\boldsymbol{e}, \boldsymbol{i}, \boldsymbol{r})$ is computed. Instead of absolute clock time, the waiting time between the different events is modeled, i.e., $t_i = t'_i - t'_{i-1}$, where t'_i denotes the clock time of the i-th event such that $i = 1, 2, \ldots, m$ if the i-th event is of type $\boldsymbol{e}$ or $\boldsymbol{i}$ and $i = 1, 2, \ldots, n$ if it is of type $\boldsymbol{r}$ (i.e., a recovery event). Likelihood of each waiting time can be found using survival analysis methodology in a setup with multiple modes of failure [31]. The

events of interest are: infection, becoming infectious and recovery. The corresponding hazard rate functions, for a non-homogeneous network, as described in the previous section, can be expressed as

$$\lambda_{infection}(t_i \mid \beta) = \sum_{j=1}^{g}\sum_{k=1}^{g}\frac{\beta_{jk}}{N_k}S_k(t_i)I_j(t_i) \tag{6.2.1a}$$

$$\lambda_{infectious}(t_i \mid \sigma) = \sum_{j=1}^{g}\sigma_j E_j(t_i) \tag{6.2.1b}$$

$$\lambda_{recovery}(t_i \mid \gamma) = \sum_{j=1}^{g}\gamma_j I_j(t_i) \tag{6.2.1c}$$

with $S_j(t_i), E_j(t_i)$ and $I_j(t_i)$ are the number of susceptible, exposed and infectious computers at time t in the subnet j, respectively, and where $i = 1, 2, \ldots, m$ in (6.2.1a) – (6.2.1b), and $i = 1, 2, \ldots, n$ in (6.2.1c). The events of infection, infectious and recovery are assumed to be mutually independent. The overall hazard rate function is then given by

$$\lambda(t_i \mid \boldsymbol{\theta}) = \sum_{j=1}^{g}\sum_{k=1}^{g}\frac{\beta_{jk}}{N_k}S_k(t_i)I_j(t_i) + \sum_{j=1}^{g}\sigma_j E_j(t_i) + \sum_{j=1}^{g}\gamma_j I_j(t_i), \tag{6.2.2}$$

where $\boldsymbol{\theta} = (\beta_{11}, \beta_{12}, \ldots, \beta_{gg}, \sigma_1, \sigma_2, \ldots, \sigma_g, \gamma_1, \gamma_2, \ldots, \gamma_g)$ is a vector of $2g + g + g = 4g$ parameters. The hazard rate remains unchanged between events.

The epidemic dataset is described by $\boldsymbol{D} = (t_i, E_i)$, where $E_i \in \{$infection, infectious, recovery$\}$ denotes the *event type*. $\boldsymbol{D}$ consists of m infection times, m infectious times and $n(\leq m)$ recovery times. Our interest now is the density of ($\boldsymbol{e}$, $\boldsymbol{i}$, $\boldsymbol{r}$) given β, σ and γ. Consider therefore the likelihood of a single arbitrary event E_i in $\boldsymbol{D}$ as failure time in a setup with constant hazard rate and non-informative censoring. If time is zero at the occurrence of the event just prior to the event E_i, the likelihood is

$$L_{E_i} = \lambda_{E_i}(t_i \mid \boldsymbol{\theta})P(T_i \geq t_i \mid \boldsymbol{\theta}) = \lambda_{E_i}(t_i \mid \boldsymbol{\theta})\overline{F}(t_i \mid \boldsymbol{\theta}), \tag{6.2.3}$$

where T_i is a random variable denoting the clock time to event E_i, and

$$\overline{F}(t \mid \boldsymbol{\theta}) = \exp(-\int_0^t \lambda(u \mid \boldsymbol{\theta})du) \tag{6.2.4}$$

is the survival function based on the overall hazard rate function (6.2.2). Assuming independence of the event inter-arrival times; the overall likelihood is given by considering all infection, all infectious and all recovery events. Hence,

$$\begin{aligned}
L &= \prod_{l=1}^{m}\left[\sum_{j=1}^{g}\sum_{k=1}^{g}\frac{\beta_{jk}}{N_k}S_k(e_l^-)I_j(e_l^-)\exp(-\int_{e_{l-1}}^{e_l}\lambda(t \mid \boldsymbol{\theta})dt)\right] \\
&\times \prod_{l=1}^{m}\left[\sum_{j=1}^{g}\sigma_j E_j(i_l^-)\exp(-\int_{i_{l-1}}^{i_l}\lambda(t \mid \boldsymbol{\theta})dt)\right]\prod_{l=1}^{n}\left[\sum_{j=1}^{g}\gamma_j I_j(r_l^-)\exp(-\int_{r_{l-1}}^{r_l}\lambda(t \mid \boldsymbol{\theta})dt)\right] \\
&= \prod_{l=1}^{m}\left[\sum_{j=1}^{g}\sum_{k=1}^{g}\frac{\beta_{jk}}{N_k}S_k(e_l^-)I_j(e_l^-)\right]\prod_{l=1}^{m}\left[\sum_{j=1}^{g}\sigma_j E_j(i_l^-)\right]\prod_{i=1}^{n}\left[\sum_{j=1}^{g}\gamma_j I_j(r_l^-)\right] \\
&\times \prod_{l=1}^{m}\exp(-\int_{e_{l-1}}^{e_l}\lambda(t \mid \boldsymbol{\theta})dt)\prod_{l=1}^{m}\exp(-\int_{i_{l-1}}^{i_l}\lambda(t \mid \boldsymbol{\theta})dt)\prod_{i=1}^{n}\exp(-\int_{r_{l-1}}^{r_l}\lambda(t \mid \boldsymbol{\theta})dt),
\end{aligned} \tag{6.2.5}$$

where $S_j(t^-)$, $E_j(t^-)$ and $I_j(t^-)$ denote the numbers in each category just prior to time t in the subnet j, i.e., $S_j(t^-) = \lim_{t\to t^-} S_j(t)$, etc. Since the events of infection, infectious and recovery are consecutive in time, the sums of integrals in the likelihood function (6.2.5) can be simplified into a single integral to yield,

$$\begin{aligned}
L &= \prod_{l=1}^{m}\left[\sum_{j=1}^{g}\sum_{k=1}^{g}\frac{\beta_{jk}}{N_k}S_k(e_l^-)I_j(e_l^-)\right]\prod_{l=1}^{m}\left[\sum_{j=1}^{g}\sigma_j E_j(i_l^-)\right]\prod_{i=1}^{n}\left[\sum_{j=1}^{g}\gamma_j I_j(r_l^-)\right] \\
&\times \exp(-\int_{e_1}^{T}\lambda(t \mid \boldsymbol{\theta})dt).
\end{aligned} \tag{6.2.6}$$

Then, log-likelihood function is given by

$$\begin{aligned}\ln L = &\sum_{l=1}^{m} \ln \left[\sum_{j=1}^{g} \sum_{k=1}^{g} \frac{\beta_{jk}}{N_k} S_k(e_l^-) I_j(e_l^-) \right] + \sum_{l=1}^{m} \ln \left[\sum_{j=1}^{g} \sigma_j E_j(i_l^-) \right] \\ &+ \sum_{l=1}^{n} \ln \left[\sum_{j=1}^{g} \gamma_j I_j(r_l^-) \right] - \int_{e_1}^{T} \sum_{j=1}^{g} \sum_{k=1}^{g} \frac{\beta_{jk}}{N_k} S_k(e_l^-) I_j(e_l^-) dt \\ &- \int_{e_1}^{T} \sum_{j=1}^{g} \sigma_j E_j(i_l^-) dt - \int_{e_1}^{T} \sum_{j=1}^{g} \gamma_j I_j(r_l^-) dt. \end{aligned} \tag{6.2.7}$$

Differentiating (6.2.7) with respect to each component in $\boldsymbol{\theta}$, we get the following *score functions*, for $j, k = 1, 2, \ldots, g$

$$\begin{cases} \dfrac{\partial \ln L}{\partial \beta_{jk}} &= \displaystyle\sum_{l=1}^{m} \frac{S_k(e_l^-) I_j(e_l^-)/N_j}{\displaystyle\sum_{j=1}^{g} \sum_{k=1}^{g} \frac{\beta_{jk}}{N_k} S_k(e_l^-) I_j(e_l^-)} - \frac{1}{N_j} \int_{e_1}^{T} S_k(t) I_j(t) dt \\ \dfrac{\partial \ln L}{\partial \sigma_j} &= \displaystyle\sum_{l=1}^{m} \frac{E_j(i_l^-)}{\displaystyle\sum_{j=1}^{g} \sigma_j E_j(i_l^-)} - \int_{e_1}^{T} E_j(t) dt \\ \dfrac{\partial \ln L}{\partial \gamma_j} &= \displaystyle\sum_{l=1}^{n} \frac{I_j(r_l^-)}{\displaystyle\sum_{j=1}^{g} \gamma_j I_j(r_l^-)} - \int_{e_1}^{T} I_j(t) dt. \end{cases} \tag{6.2.8}$$

Setting each of the above score functions to zero, the maximum likelihood estimates (MLEs) of the parameters σ_j, γ_j and β_{jk}; $j, k = 1, 2, \ldots, g$ are obtained as the solution of the system of equations

$$\begin{cases} \displaystyle\sum_{l=1}^{m} \frac{S_k(e_l^-) I_j(e_l^-)/N_j}{\displaystyle\sum_{j=1}^{g} \sum_{k=1}^{g} \frac{\beta_{jk}}{N_k} S_k(e_l^-) I_j(e_l^-)} - \frac{1}{N_j} \int_{e_1}^{T} S_k(t) I_j(t) dt = 0 \\ \displaystyle\sum_{l=1}^{m} \frac{E_j(i_l^-)}{\displaystyle\sum_{j=1}^{g} \sigma_j E_j(i_l^-)} - \int_{e_1}^{T} E_j(t) dt = 0 \\ \displaystyle\sum_{l=1}^{n} \frac{I_j(r_l^-)}{\displaystyle\sum_{j=1}^{g} \gamma_j I_j(r_l^-)} - \int_{e_1}^{T} I_j(t) dt = 0 \end{cases} \tag{6.2.9}$$

In practice, these equations will require a numerical solution.

6.2.3 Multi-group Stochastic SEIR-SDE Model

To set up our proposed multi-group stochastic SEIR model using stochastic differential equations, let $X(t) = (S_1(t), \ldots, S_g(t), E_1(t), \ldots, E_g(t), I_1(t), \ldots, I_g(t), R_1(t), \ldots, R_g(t))^T$ denote the column vector *state of the multi-group system* at time t, where $(S_j(t), E_j(t), I_j(t), R_j(t))$ represents the corresponding distribution of N_j computers in subnet j $(j = 1, 2, \ldots, g)$ which are (susceptible, exposed, infectious, recovered) at time t.

The stochastic differential equation formulation is based on a time step, $\triangle t$. We assume that the time step, $\triangle t$, is small enough so that there can be only a change (± 1) of at most one unit among the S-E-I-R category counts within each subnet during a time step. We denote the changes in each subnet j during $\triangle t$ by $\triangle S_j(t) = S_j(t + \triangle t) - S_j(t)$, $\triangle E_j(t) = E_j(t + \triangle t) - E_j(t)$, $\triangle I_j(t) = I_j(t + \triangle t) - I_j(t)$ and $\triangle R_j(t) = R_j(t + \triangle t) - R_j(t)$. The change in the entire network is given by $\triangle X(t) = (\triangle S_1(t), \ldots, \triangle S_g(t), \triangle E_1(t), \ldots, \triangle E_g(t), \triangle I_1(t), \ldots, \triangle I_g(t), \triangle R_1(t), \ldots, \triangle R_g(t))^T$.

Now, we can compute the probabilities of the various changes occurring in each subnet. For $k = 1, 2, \ldots, g$, these transition probabilities are given by:

$$P(\triangle S_k = -1, \triangle E_k = 1 | X(t)) = \sum_{j=1}^{g} \frac{\beta_{jk}}{N_k} S_k(t) I_j(t) \triangle t + o(\triangle t) \tag{6.2.10a}$$

$$P(\triangle E_k = -1, \triangle I_k = 1 | X(t)) = \sigma_k E_k(t) \triangle t + o(\triangle t) \tag{6.2.10b}$$

$$P(\triangle I_k = -1, \triangle R_k = 1 | X(t)) = \gamma_k I_k(t) \triangle t + o(\triangle t) \tag{6.2.10c}$$

Applying these transition probabilities, the expected rate of change of the population $E(\triangle X(t))$ can be found. This expected value, after approximate algebraic computation, can be shown to be given by

$$E(\triangle X(t)) := \mu(X(t))\triangle t + \boldsymbol{o}(\triangle t)$$

$$= \begin{pmatrix} -\sum_{j=1}^{g} \frac{\beta_{j1}}{N_1} S_1(t) I_j(t) \\ \vdots \\ -\sum_{j=1}^{g} \frac{\beta_{jg}}{N_g} S_g(t) I_j(t) \\ \sum_{j=1}^{g} \frac{\beta_{j1}}{N_1} S_1(t) I_j(t) - \sigma_1 E_1(t) \\ \vdots \\ \sum_{j=1}^{g} \frac{\beta_{jg}}{N_g} S_g(t) I_j(t) - \sigma_g E_g(t) \\ \sigma_1 E_1(t) - \gamma_1 I_1(t) \\ \vdots \\ \sigma_g E_g(t) - \gamma_g I_g(t) \\ \gamma_1 I_1(t) \\ \vdots \\ \gamma_g I_g(t) \end{pmatrix} \triangle t + \boldsymbol{o}(\triangle t). \tag{6.2.11}$$

The column vector shown in (6.2.11) thus corresponds to $\mu(X(t))$.

The stochastic variance comes from the covariance matrix for the rate of change in the random variables. The covariance matrix of $\triangle X(t)$ is $V(\triangle X(t)) = E(\triangle X(t)(\triangle X(t))^T) - E(\triangle X(t))(E(\triangle X(t)))^T \approx E(\triangle X(t)(\triangle X(t))^T)$, since the elements in the second term are $o((\triangle t)^2)$. Then the $4g \times 4g$ covariance matrix of $\triangle X(t)$ up to order $\triangle t$, is given by

$$V(\triangle X(t)) = \begin{pmatrix} A & B & 0 & 0 \\ B & C & D & 0 \\ 0 & D & E & F \\ 0 & 0 & F & G \end{pmatrix} \triangle t \tag{6.2.12}$$

[1]. Here the $g \times g$ submatrices in (6.2.12) are given by

$$A = \text{diag}\left(\sum_{j=1}^{g} \frac{\beta_{jk}}{N_k} S_k(t) I_j(t)\right), \qquad B = \text{diag}\left(-\sum_{j=1}^{g} \frac{\beta_{jk}}{N_k} S_k(t) I_j(t)\right),$$

$$C = \text{diag}\left(\sum_{j=1}^{g} \frac{\beta_{jk}}{N_k} S_k(t) I_j(t) + \sigma_k E_k(t)\right), \quad D = \text{diag}\left(-\sigma_k E_k(t)\right),$$

$$E = \text{diag}\left(\sigma_k E_k(t) + \gamma_k I_k(t)\right), \qquad F = \text{diag}\left(-\gamma_k I_k(t)\right),$$

$$G = \text{diag}\left(\gamma_k I_k(t)\right),$$

where the rows and columns of the matrix in (6.2.12) shown in terms of partitioned submatrices corresponding to susceptible (S), exposed (E), infectious (I) and recovered (R) states. The null submatrices in (6.2.12) arise as a consequence of the transition equations (6.2.10) or/and due to impossibility of *direct transitions* from $S \to E$, $S \to R$, $E \to R$, $I \to S$, $R \to S$ and $R \to E$.

Considering $X(t)$ as approximating a diffusion process with small time step $\triangle t$, we have up to order $\triangle t$,

$$\triangle X(t) = E(\triangle X(t)) + \sqrt{V(\triangle X(t))}. \tag{6.2.13}$$

Let

$$H(X(t)) = \sqrt{V(\triangle X(t))} = \sqrt{\begin{pmatrix} A & B & 0 & 0 \\ B & C & D & 0 \\ 0 & D & E & F \\ 0 & 0 & F & G \end{pmatrix}},$$

where the matrices A, B, C, D, E, F, and G are defined as in (6.2.12). The matrix H is the unique because matrix $V(\triangle X(t))$ is positive definite and is given by

$$H(X(t)) = \sqrt{V(\triangle X(t))} = \begin{pmatrix} A_1 & 0 & 0 & 0 \\ B_1 & C_1 & 0 & 0 \\ 0 & D_1 & E_1 & 0 \\ 0 & 0 & 0 & F_1 \end{pmatrix}. \tag{6.2.14}$$

Here the $g \times g$ submatrices in (6.2.14) are given by

$$A_1 = \text{diag}\left(-\sqrt{\sum_{j=1}^{g} \frac{\beta_{jk}}{N_k} S_k(t) I_j(t)}\right), \quad B_1 = \text{diag}\left(\sqrt{\sum_{j=1}^{g} \frac{\beta_{jk}}{N_k} S_k(t) I_j(t)}\right),$$
$$C_1 = \text{diag}\left(-\sqrt{\sigma_k E_k(t)}\right), \qquad D_1 = \text{diag}\left(\sqrt{\sigma_k E_k(t)}\right),$$
$$E_1 = \text{diag}\left(-\sqrt{\gamma_k I_k(t)}\right), \qquad F_1 = \text{diag}\left(\sqrt{\gamma_k I_k(t)}\right).$$

Taking limit as $\triangle t \to 0$ of equation (6.2.13), the following system of Itô *stochastic differential equations* (SDE) is obtained,

$$d(X(t)) = \mu(X(t))dt + H(X(t))dW(t), \tag{6.2.15}$$

where $W(t) = (W_1(t), W_2(t), \ldots, W_{4g}(t))^T$, and each $W_i(t)$, $i = 1, 2, \ldots, 4g$ is an independent Wiener process. The stochastic differential equations (SDEs), governing the process, are:

$$\begin{cases} dS_i(t) = -\sum_{j=1}^{g} \frac{\beta_{ji}}{N_i} S_i(t) I_j(t) dt - \sqrt{\sum_{j=1}^{g} \frac{\beta_{ji}}{N_i} S_i(t) I_j(t)} dW_i(t) \\ dE_i(t) = \left[\sum_{j=1}^{g} \frac{\beta_{ji}}{N_i} S_i(t) I_j(t) - \sigma_i E_i(t)\right] dt + \sqrt{\sum_{j=1}^{g} \frac{\beta_{ji}}{N_i} S_i(t) I_j(t)} dW_i(t) \\ \qquad\qquad -\sqrt{\sigma_i E_i(t)} dW_{g+i}(t) \\ dI_i(t) = \left[\sigma_i E_i(t) - \gamma_i I_i(t)\right] dt + \sqrt{\sigma_i E_i(t)} dW_{g+i}(t) - \sqrt{\gamma_i I_i(t)} dW_{2g+i}(t) \\ dR_i = \gamma_i I_i(t) dt + \sqrt{\gamma_i I_i(t)} dW_{3g+i}(t), \end{cases} \tag{6.2.16}$$

for $i = 1, 2, \ldots, g$.

6.3 Case Study: Code Red Worm Data

The methods described in Section 6.2.2 and Section 6.2.3 will be applied to Code Red worm data.

We make the following simplifying *assumptions* for the transmission, latency and recovery parameters of the component subnets:

1) all *within* subnet infection rates are the same, i.e., $\beta_{jj} = \beta_w$; $j = 1, 2, \ldots, g$.

2) infection rates *between* any two subnets are equal, i.e., $\beta_{jk} = \beta_b$ for all distinct pairs $j \neq k$; $j, k = 1, 2, \ldots, g$.

3) infectiousness latency rate is constant across subnets, i.e., $\sigma_j = \sigma$; $j = 1, 2, \ldots, g$.

4) each subnet has the same recovery rate, i.e., $\gamma_j = \gamma$; $j = 1, 2, \ldots, g$.

Each component subnet in the network is necessarily assumed to be homogeneous within itself, since otherwise the subnet can be further decomposed into constituent homogeneous subgroups. The simplifying assumptions 1) – 4) above are often, in practice – as in the case of Code Red data to which our model is applied, due to lack of detailed information about the communication profiles of the individual nodes in the computer network that would make it feasible to choose/assign distinct but at the same time also realistic values to all different parameters.

For Code Red worm data, Figure 6.1 summarizes the total number of infected hosts according to the country information.

Based on the geographical locations (i.e., country information) for each computers in Code Red worm data set, the Internet is divided into seven groups, namely Asia, Africa, Europe, North America, Oceania, South America and XX (computers, for which no specific country information is available), which is summarized in Figure 6.2.

6.3.1 Results for Country based Multi-group Stochastic SEIR Model using Markovian Approach

6.3.1.1 Parameters Estimations Results. In the application to Code Red worm data, the whole network is divided in seven groups based on country information of each computer, namely Asia, Africa, Europe, North America, Oceania, South America and XX (no country specific information given), i.e., number of groups $g = 7$. For parameters of interest, since the reciprocal of infectious rate, σ, is the average latent period, which depends only on the computer virus itself, σ_j will be same for each

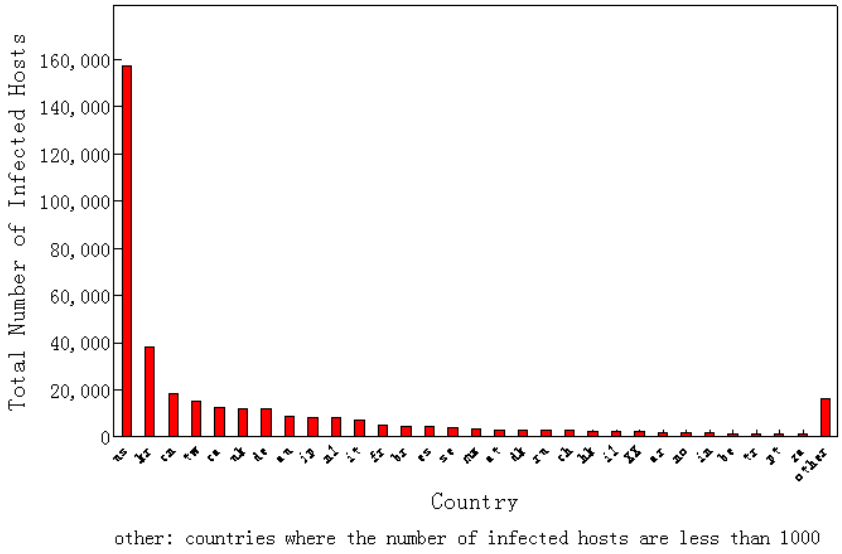

Figure 6.1 Distribution of total number of infected hosts based on location (country).

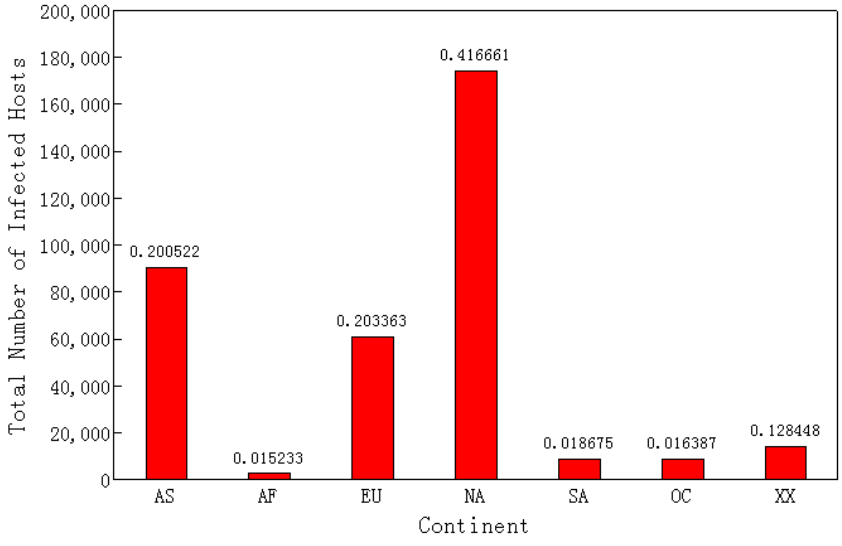

Figure 6.2 Distribution of the population of infected hosts among different groups: Asia, Africa, Europe, North America, Oceania, South America, and XX (location unknown).

of the seven groups. Also, since based on geographical location information above, there is no reason to believe that the designers/propagators of the Code Red virus have any explicit preference of targets by their location; it is reasonable to assume that all intra-subnet as well as inter-subnet transmission rates are each constant. Therefore additionally, we assume that for all seven groups, within each group the infection rates are same, i.e., $\beta_{jj} = \beta_w$ for $j = 1, 2, \ldots, 7$, and between groups the infection rate are also same, i.e., $\beta_{jk} = \beta_b$ for $j, k = 1, 2, \ldots, 7$ and $j \neq k$. Now, the likelihood function (6.2.6), the log-likelihood function (6.2.7) and the system of equations (6.2.8) can be simplified, and the maximum likelihood estimates (MLEs) of the parameters β_w, β_b, σ and γ_j; $j = 1, 2, \ldots, 7$ for each of the seven groups can be obtained as the numerical solution of the simplified system of equations corresponding to (6.2.9). Our results of the parameter estimates are summarized in Table 6.1 and Table 6.2.

Table 6.1 Summary of Parameter Estimates in Multi-group Stochastic SEIR Model using Markovian Approach (Same Recovery Rate for All Groups)

	$\hat{\beta}_w$	$\hat{\beta}_b$	$\hat{\sigma}$	$\hat{\gamma}$
Maximum Likelihood Estimates	3.3035	3.2556	0.11	5.625851×10^{-5}

If the the fourth assumption of constancy of recovery rates between groups, as assumed at the beginning of Section 6.3, is relaxed; then the maximum likelihood estimates for the possibly different recovery rates for different groups can be obtained. Table 6.2 summarizes the maximum likelihood estimates for all parameters, by suitably modifying equations (6.2.9) with this relaxation. There does not appear to by any appreciable difference in the estimates of recovery rates across groups. The estimated average recovery time across groups as measured by γ^{-1} varies only between 4.8 hours and 4.9 hours (a difference of 0.1 hour or, about 6 minutes for a recovery) across the seven groups.

Table 6.2 Summary of Parameter Estimates in Multi-group Stochastic SEIR Model using Markovian Approach (Different Recovery Rates for All Groups)

	$\hat{\beta}_w$	$\hat{\beta}_b$	$\hat{\sigma}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
MLE	3.3035	3.2556	0.11	5.6159×10^{-5}	5.7236×10^{-5}
	$\hat{\gamma}_3$	$\hat{\gamma}_4$	$\hat{\gamma}_5$	$\hat{\gamma}_6$	$\hat{\gamma}_7$
MLE	5.6359×10^{-5}	5.6259×10^{-5}	5.7528×10^{-5}	5.6288×10^{-5}	5.6269×10^{-5}

Both Table 6.1 and Table 6.2 show that the maximum likelihood estimates of β_w and β_b are close, 3.3035 and 3.2556, respectively, which are also close to the maximum likelihood estimate of β and Bayesian estimate of β (shown in Table 4.1) obtained in Chapter 4.

6.3.1.2 Simulation Results. To see how well our multi-group stochastic SEIR model using Markovian approach is, simulations using the maximum likelihood estimates for all parameters (values in Table 6.2) were conducted with 5000 simulation runs. Figure 6.3 shows the observed Code Red worm data and the corresponding simulation results separately for each of the seven different groups classified by location. By visual inspection, the simulations for each of the seven groups well match the actual observed data.

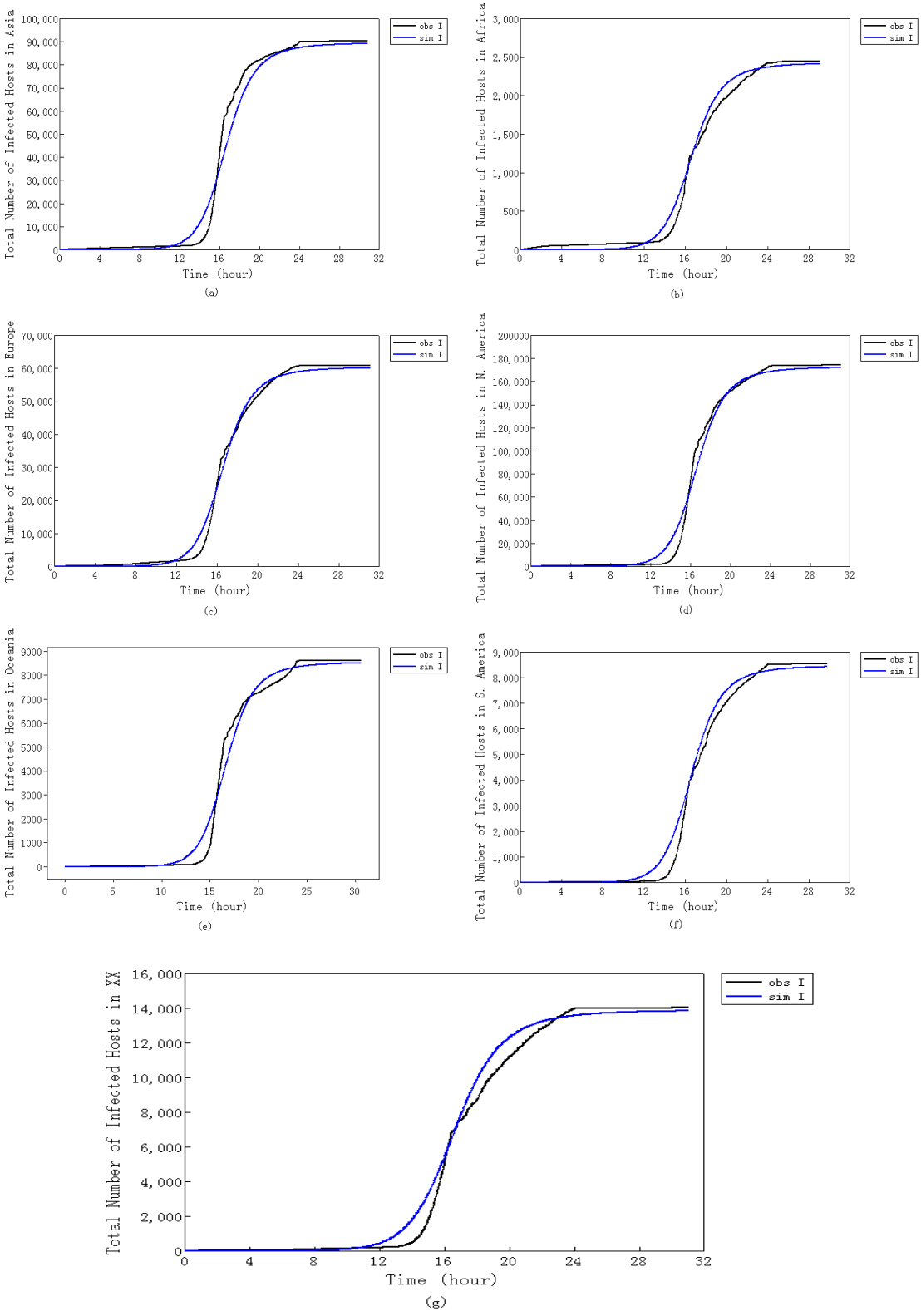

Figure 6.3 Comparison between observed data and Code Red worm simulation using multi-group stochastic SEIR with Markovian approach for the seven different groups based to geographical locations of computers: (a) Asia (b) Africa (c) Europe (d) North America (e) Oceanic (f) South America and (g) XX (location unknown).

Table 6.3 also demonstrates the close match to the observed Code Red data, which is shown in Figure 6.3, by computing the discrepancy measures.

Table 6.3 Summary of Simulations from Multi-group Stochastic SEIR Model using Markovian Approach for Code Red Worm Propagation Based on Different Groups

	Standard Deviation of Discrepancy	Maximum Discrepancy	Average Discrepancy
Asia	4.1242	10.7505	−1.1366
Africa	9.1801	24.0843	1.5849
Europe	1.6653	5.7695	0.0049
North America	0.5943	18.9769	−0.3101
Oceania	3.7204	11.6271	1.9450
South America	3.2043	11.3123	1.4794
XX (location unknown)	5.3483	15.8771	2.0977

6.3.2 Multi-group Stochastic SEIR-SDE Model

6.3.2.1 A Special Case of Multi-group Stochastic SEIR-SDE Model: One Group Only-Homogeneous Case. When the constituent groups in a multi-group stochastic SEIR-SDE model are merged together, the model reduces to a single network under homogeneity assumption. For comparison purposes, simulations are performed based on stochastic SEIR-SDE model with different parameter input values by using MLE and Bayesian estimates obtained in Chapter 4, MCMC estimates obtained in Chapter 5 and those used in previous works [72, 90]. Figure 6.4 shows the result that the simulations based on the models with MLE, Bayesian and MCMC estimates fit the observed Code Red worm data much better than those previous works. The corresponding discrepancy measures in Table 6.6 underscore this finding.

6.3.2.2 Parameters Estimations Results. In the case using stochastic differential equations, we assume that all groups have the same infectious rates, σ, and recovery rates, γ, i.e., $\sigma_j = \sigma$ and $\gamma_j = \gamma$ for $j = 1, 2, \ldots, 7$. Moreover, we assume

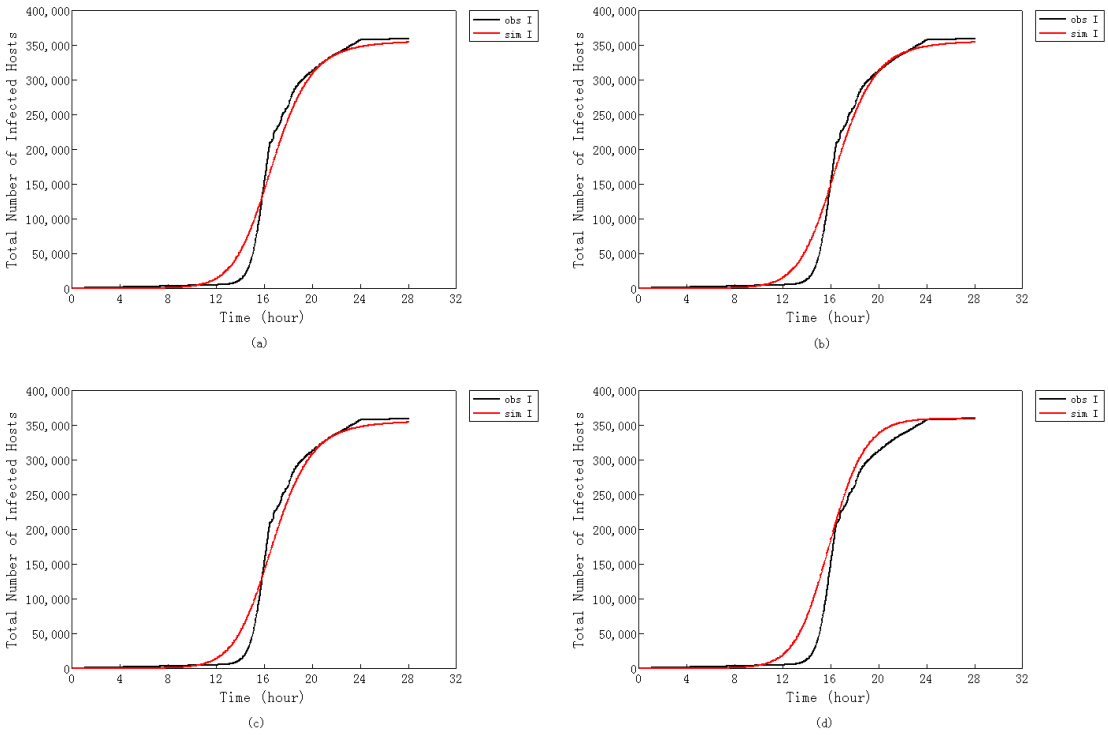

Figure 6.4 Comparison between observed data and Code Red worm simulation using SDE based on different parameter values: (a) maximum likelihood estimates ($\beta = 3.2100, \sigma = 0.11, \gamma = 5.625851 \times 10^{-5}$) (b) Bayesian estimates ($\beta = 3.2421, \sigma = 0.10, \gamma = 5.681790 \times 10^{-5}$) (c) MCMC estimates ($\beta = 3.2065, \sigma = 0.12, \gamma = 5.61 \times 10^{-5}$) and (d) previous works ($\beta = 2.83, \sigma = 0.11, \gamma = 1.39 \times 10^{-5}$).

Table 6.4 Summary of Stochastic SEIR Models Simulations for Code Red Worm Propagation in Homogeneous Network using Stochastic Epidemic Equations based on Different Scenarios

	Standard Deviation of Discrepancy	Maximum Discrepancy	Average Discrepancy
Stochastic SEIR (MLE)	15.3200	47.6329	−0.7411
Stochastic SEIR (Bayesian Estimates)	15.4543	53.7708	1.2451
Stochastic SEIR (MCMC Estimates)	15.3342	47.1415	−0.8880
Stochastic SEIR (Previous Work)	19.6102	78.3853	12.2958

that $\sigma = 0.11$ and $\gamma = 5.625851 \times 10^{-5}$, maximum likelihood estimates for σ and γ obtained in Chapter 4. Since each element in $W(t) = (W_1(t), W_2(t), \ldots, W_{4k}(t))^T$ in system of equations (6.2.15) is Wiener process, i.e., $W_i(t)$ has normal distribution with mean 0 and variance t for $i = 1, 2, \ldots, 4k$, the method of least squares can be used to estimate the parameters of interest, saying β_w and β_b. However, due to the complexity of the system of equations (6.2.15), the explicit solution for the parameter estimates can not be obtained. We obtain the numerically solutions for the least squares estimates for model parameters of interest, which are summarized in Table 6.5.

Table 6.5 Summary of Parameter Estimates in Multi-group Stochastic SEIR Model-SDE

	$\hat{\beta}_w$	$\hat{\beta}_b$
Least Squares Estimates	3.303	3.256

6.3.2.3 Simulation Results. With the least squares estimates obtained above, simulations based on the stochastic multi-group SEIR-SDE model is conducted. Figure 6.5 shows the comparison results between the observed Code Red worm data and simulation results. By visual inspection, the multi-group stochastic SEIR-SDE model well match the observed data for each of the seven groups. It also can be seen from Table 6.6.

6.3.3 Comparison between Multi-group Markovian SEIR Models and Multi-group Stochastic SEIR-SDE Model

In this Section, we compare the multi-group Markovian SEIR model and multi-group SEIR-SDE model. Table 6.7 summarizes the discrepancy measures for both two models. It shows that for each of the seven groups the two models are very close to each other and both of them well match the observed Code Red data. Figure 6.6 also visually corroborates the same view.

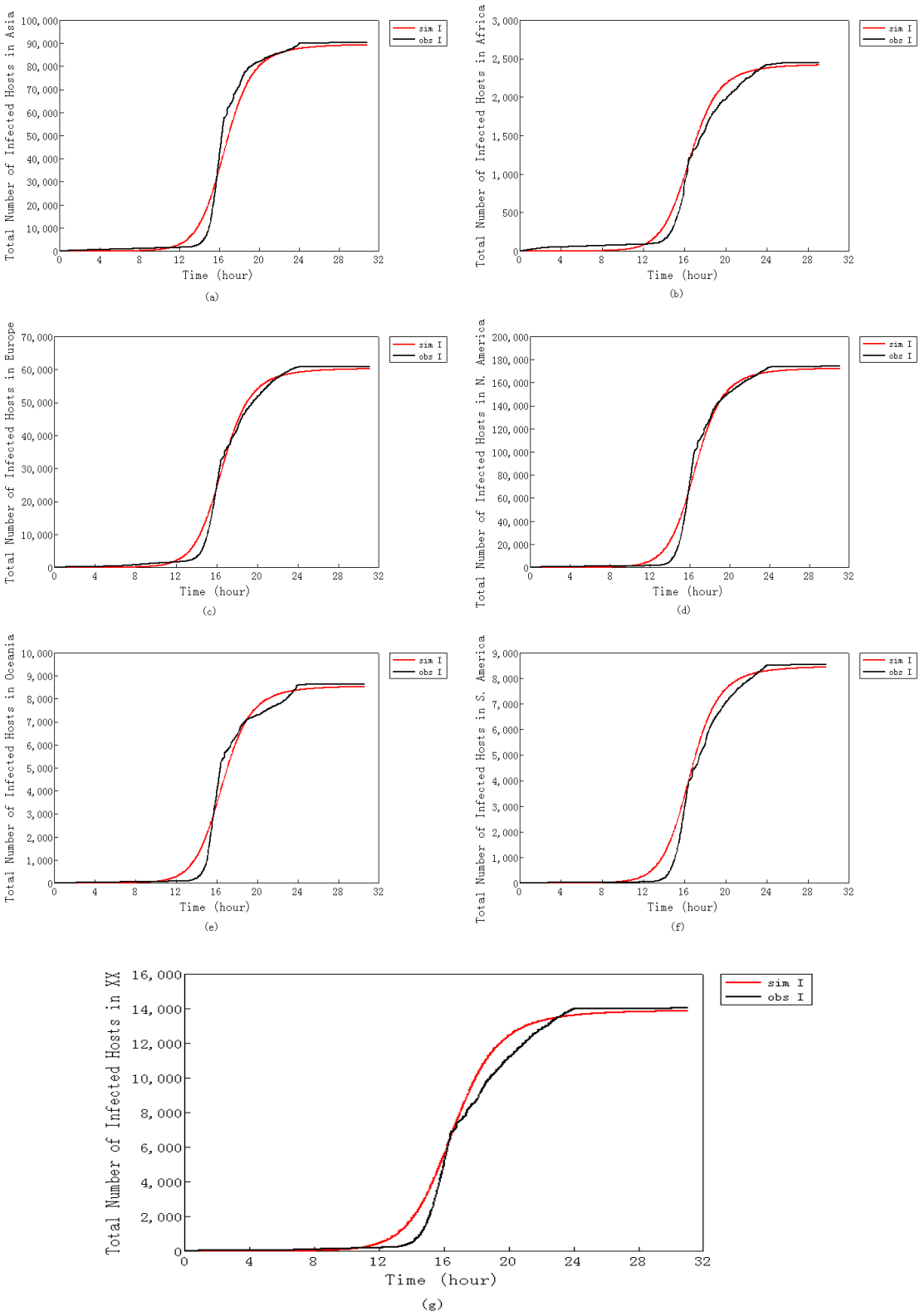

Figure 6.5 Comparison between observed data and Code Red worm simulation using multi-group stochastic SEIR-SDE model for the seven different groups based on geographical locations of computers: (a) Asia (b) Africa (c) Europe (d) North America (e) Oceanic (f) South America and (g) XX (location unknown).

Table 6.6 Summary of Multi-group Stochastic SEIR-SDE Model Simulation for Code Red Worm Propagation Based on Different Groups

	Standard Deviation of Discrepancy	Maximum Discrepancy	Average Discrepancy
Asia	3.9111	11.2867	−0.8152
Africa	9.9556	25.57	2.4000
Europe	1.7385	6.1037	2.1306
North America	0.5779	1.9992	0.3639
Oceania	3.6543	12.112	4.7506
South America	3.4150	11.81	1.7619
XX (location unknown)	5.7399	16.671	2.5190

Table 6.7 Comparison between Multi-group Stochastic SEIR Model using Markovian Approach and Multi-group Stochastic SEIR-SDE According to Different Groups

Group	SEIR Model Type	Standard Deviation of Discrepancy	Maximum Discrepancy	Average Discrepancy
Asia	Markovian	4.1242	10.7505	−1.1366
	SDE	3.9111	11.2867	−0.8152
Africa	Markovian	9.1801	24.0843	1.5849
	SDE	9.9556	25.57	2.4000
Europe	Markovian	1.6653	5.7695	0.0049
	SDE	1.7385	6.1037	2.1306
North America	Markovian	0.5943	18.9769	−0.3101
	SDE	0.5779	1.9992	0.3639
Oceania	Markovian	3.7204	11.6271	1.9450
	SDE	3.6543	12.112	4.7506
South America	Markovian	3.2043	11.3123	1.4794
	SDE	3.4150	11.81	1.7619
XX	Markovian	5.3482	15.8771	2.0977
	SDE	5.7399	16.671	2.5190

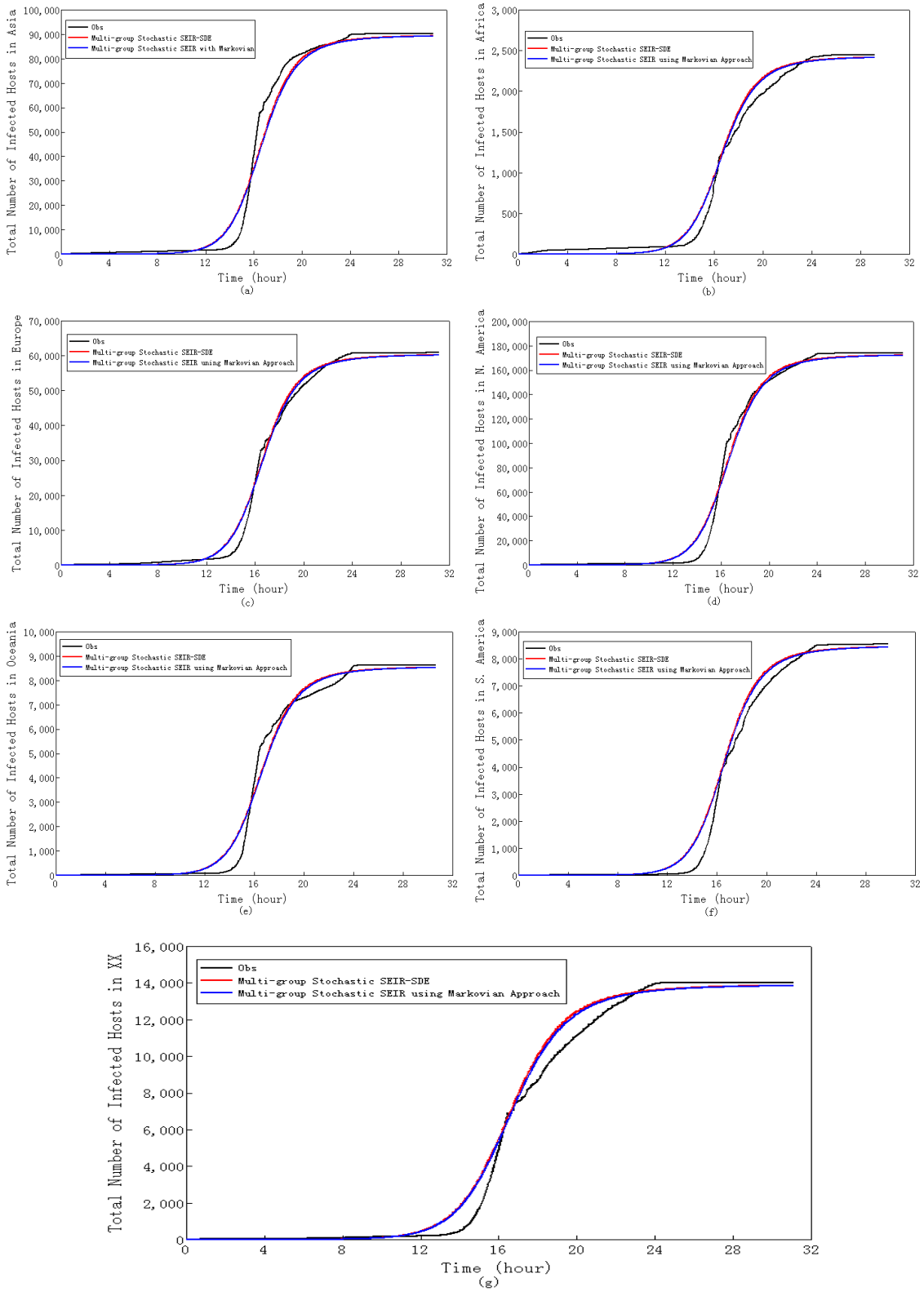

Figure 6.6 Comparison between observed data and Code Red worm simulations using multi-group stochastic SEIR with Markovian approach and multi-group stochastic SEIR-SDE for seven different groups according to geographic-al locations of computers: (a) Asia (b) Africa (c) Europe (d) North America (e) Oceanic (f) South America and (g) XX (location unknown).

Moreover, as the overall comparison between the two models, Table 6.8 and Figure 6.7 show the same result that both multi-group Markovian SEIR model and multi-group SEIR-SDE model are good in matching the observed Code Red worm data.

Table 6.8 Comparison between the Overall Simulations from Multi-group Stochastic SEIR Model using Markovian Approach and Multi-group Stochastic SEIR-SDE Model

	Multi-group SEIR-SDE Model	Multi-group Markovian SEIR Model
Standard Deviation of Discrepancy	11.7929	12.3782
Maximum Discrepancy	41.5196	39.7848
Average Discrepancy	0.2449	1.5910

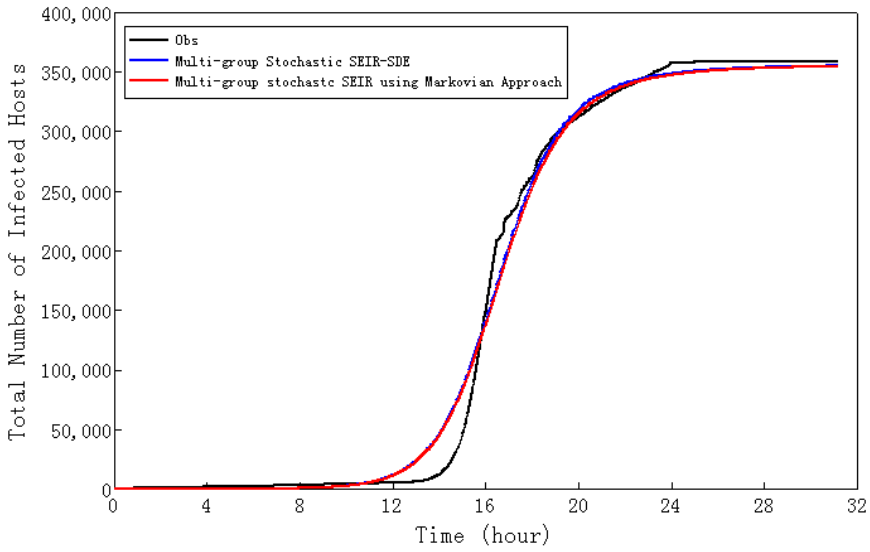

Figure 6.7 Comparison between observed Code Red worm data and simulations based on Multi-groups Markovian SEIR Model and Multi-group SEIR-SDE Model.

6.3.4 Comparison between Models with Homogeneous Assumption and Models with Heterogeneous Assumption

For the purpose of comparing the stochastic models with homogeneous assumption discussed in Chapter 4 and Chapter 5 with the multi-group stochastic SEIR model of this Chapter of a heterogeneous network, simulations based on different models are conducted. The results are shown in Figure 6.8 and Table 6.9, from which we conclude that the use of stochastic SEIR model under a homogeneous assumption or, a heterogenous assumption fit the actual Code Red worm data much better than previous models do.

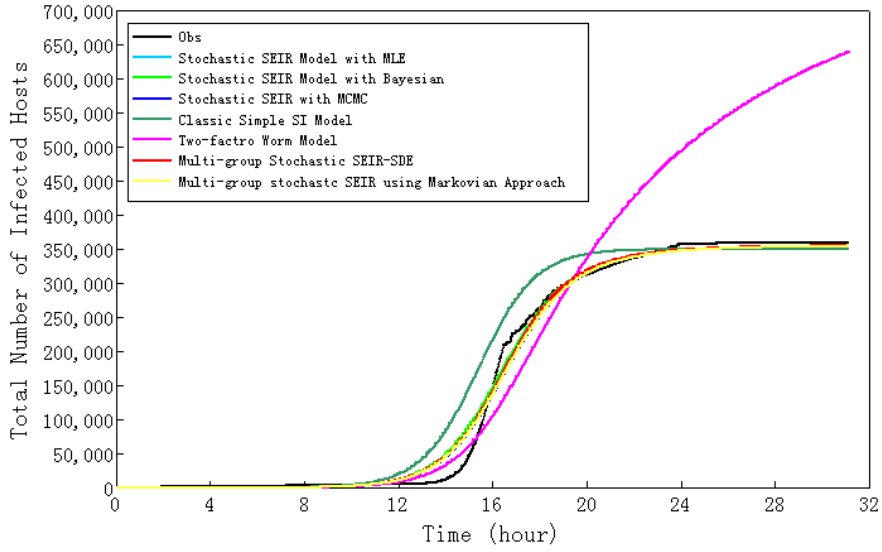

Figure 6.8 Comparison between observed data and Code Red worm simulations based on different models: Stochastic SEIR model with MLE, Stochastic SEIR model with Bayesian estimates, Stochastic SEIR model with MCMC estimates, Classic Simple SI Model, Two-factor Worm Model , Multi-groups Stochastic SEIR Model using Markovian Approach and Multi-group Stochastic SEIR-SDE Model.

Among the SEIR single homogeneous network model outputs with input parameter chosen as estimates obtained by different methods (MLE, Bayesian), the MCMC based outputs with partially observed data appear to perform best in term of average, standard deviation of discrepancies from actual data and also in term of the maximum

Table 6.9 Summary of Stochastic SEIR Models Simulations and Comparison with other Models

	Standard Deviation of Discrepancy	Maximum Discrepancy	Average Discrepancy
Stochastic SEIR (MLE)	13.3572	40.7684	−0.5910
Stochastic SEIR (Bayesian Estimation)	13.3128	45.9277	2.9181
Stochastic SEIR (MCMC Estimation)	12.0984	39.3951	−1.0646
Classic Simple SI Model	28.4853	100.6507	19.7426
Two-factor worm Model	41.9134	136.2675	8.9397
Multi-group Stochastic SEIR-SDE Model	11.7929	41.5196	0.2449
Multi-group Markovian SEIR Model	12.3782	39.7848	1.5910

deserved discrepancy (see Table 6.9). In the Markovian vs. SDE based approach to multi-group SEIR models to allow for possible network heterogeneity, the results based on parameters estimates obtained via SDE approach are again seem to perform better than the Markovian approach (see Table 6.8).

Moreover, there is no big difference between the stochastic models with homogeneous and those with heterogenous assumption for Code Red worm data. To check the possible reason, the observations (percentage of the infected hosts) Code Red worm is plotted according to the seven groups: Asia, Africa, Europe, North America, Oceania, South America and XX (location unknown) in Figure 6.9. In Figure 6.9, the total observed (in the form of percentage of the infected hosts) Code Red worm data is also plotted. It is can be seen that the feature of the propagation of Code Red worm in different groups are very similar and also similar to the total observations. This is a possible reason why there is no significant difference between the simulations using stochastic models with homogeneous assumption and with heterogeneous assumption for Code Red worm data.

6.4 Discussion

In this Chapter, two multi-group stochastic SEIR models are introduced, to study the propagation of computer virus in heterogeneous networks. Parameter estimates are

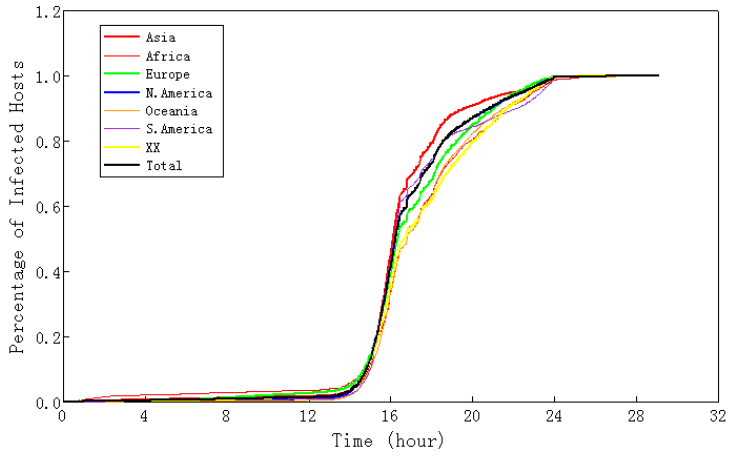

Figure 6.9 Comparison between Code Red worm observation among groups (Asia, Africa, Europe, North America, Oceania, South America and XX (location unknown)) and total observations.

developed. The models are applied to Code Red worm data. The simulation results show that the multi-group stochastic SEIR models well match the observed Code Red worm data and generally provide a much better fit to the actual data than those obtained by previous models. However, the difference between using stochastic SEIR model for a single homogeneous network vs. multi-group stochastic SEIR models to incorporate network heterogeneity does not appear to be significant when the Code Red worm data are grouped by geographical locations.

CHAPTER 7

CONCLUSION

Stochastic SEIR models and inference for computer virus propagation in Homogeneous and Heterogeneous networks using (i) a Markovian approach and (ii) an approach based on stochastic differential equations (SDE) are considered in this Part (Chapters 2 – 6) of this dissertation. The models developed and the corresponding methodology are applied to 'Code Red' outbreak data. Based on simulation results, we conclude that (a) our proposed models well match the observed 'Code Red' data; and in fact (b) provide a closer fit than offered by other models explored in the literatures to explain the observed data on 'Code Red'.

To summarize, we have two main findings:

A. Our numerical studies of the 'Code Red' data indicates an overall relative superiority of the stochastic SEIR framework to more adequately explain the propagation behavior in time of computer 'viruses' that by their intrinsic nature, remain dormant for a random 'latent period' in a computer after it is infected, until it turns infectious.

B. The impact of network structure on computer 'virus' propagation dynamics over time can be explored via multi-group variants of a stochastic SEIR framework (Markovian, or SDE based) to allow for different transmission, infection and recovery parameters between and within subnets that are homogeneous within themselves, and together constitute the overall heterogeneous network. Whether allowing for such heterogeneity improves our ability to describe a 'virus' outbreak a homogeneous model depends essentially on whether the individual nodes (computers) in the network can be suitably grouped into clusters (subnetworks) which differ substantially in communication activity profiles and in their primary

purposes (academic/administrative/govenment/industrial R&D/media/financial etc.) between clusters but are closely similar or, identical for nodes within clusters.

For the possible future work, the stochastic SEIR framework can be further modified by considering other feature. For increased realism, two such possibilities are:

(1) **Considering the human countermeasures**

Human countermeasures is a very important factor for a more realistic modeling of computer virus propagation, since appropriately including the effects of such countermeasures can substantially further improve such a models's predictive ability, and corresponding influence real life actions taken by network administrators as well as by individual users in controlling the virus outbreak and the consequent damages. Some work has been done to consider the effect of such human countermeasures, such as 'kill signals'. Wang [84] has explored the static effects of human countermeasures by considering the possibility of immunizing a fraction of the hosts before the computer virus spread becomes a serious threat, which begs the question of the timings of such immunization. Zou [90] also consider the human countermeasures as a factor while modeling Code Red worm propagation where the human countermeasures are dynamic actions and is used in a deterministic epidemic model. In this dissertation, the human countermeasure is not included as a factor. For the future work, one can attempt to include such countermeasure effects in as suitable way in the stochastic SEIR model setting as a dynamic factor, although it is not yet clear how to do this.

(2) **Considering the infection rate as a function of time**

The infection rates are assumed to be constant in the sense that it does not change over time during the whole process. In reality, it is possible that *either* (i) a large-scale computer virus propagation causes congestion in Internet traffic,

thereby slowing down the speed of infection, as seems to be the case with Code Red data [90] *or*, (ii) the infection rates increase due to the occurrence of more computer infections, i.e., due to the increased number of infectious hosts. In either case, the constant infection rates are not appropriate. Thus, for future work, the constant infection rate, β, may be generalized to be a suitable function of time.

Investigating the effects of network topology on computer malware propagation is now emerging as a new avenue of research, and is not included in the scope of our work reported here. Doing so successfully would require tying together the ideas and tools of random graph dynamics to describe the stochastic behavior of the topological structure of large computers network with important factors relevant to network traffic. Even without such sophistication, the relatively simple modification of earlier models to a SEIR framework can, as we have shown, significantly improve its explanatory ability.

Part II

ON SOME SURVIVAL DISTRIBUTION MODELS IN RELIABILITY

CHAPTER 8

INTRODUCTION

8.1 Brief Description of Research

In this Part, we explore two models of survival distribution in univariate and bivariate setups respectively. A new partial ordering called reverse star-ordering among life distributions is first introduced. Its properties including the relationship of this ordering to other partial orderings and non-parametric life distribution classes, and preservation properties of under some reliability operations are also discussed.

A new class of bivariate proportional hazard model is introduced and its different properties are discussed. The maximum likelihood estimators cannot be expressed in explicit forms in most of the cases. The expectation-maximization (EM) algorithm has been proposed to compute the maximum likelihood estimators of the unknown parameters. The model and methodology are applied to two real data sets as illustrations.

8.2 Review of Expectation-Maximization (EM) Algorithm

For data $\mathbf{D}$ generated from a model defined by parameters $\boldsymbol{\theta}$, a maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ is the value that maximizes the likelihood function $L(\boldsymbol{\theta}; \mathbf{D})$, or equivalently, the loglikelihood $l(\boldsymbol{\theta}; \mathbf{D})$. If the likelihood is differentiable, unimodal and bounded above [79], the MLE is unique and found by setting the score function $S(\boldsymbol{\theta}) = \dfrac{\partial l(\boldsymbol{\theta}; \mathbf{D})}{\partial \boldsymbol{\theta}}$ to zero and solving for $\boldsymbol{\theta}$. Numerical approaches can be used when the maximum of the likelihood can not be determined analytically. The expectation-maximization (EM) algorithm is discussed below. Other numerical approaches include the Newton-Raphson method, the Nelder-Mead simplex algorithm, quadratic optimization and the quasi-Newton method.

The expectation-maximization (EM) algorithm, first named by Dempster et al. [24], can simplify maximum likelihood estimation of the parameter vector $\boldsymbol{\theta}$ by considering a more complete and hypothetical data set. The complete data set $\mathbf{D}$ is formed by augmenting the observed data $\mathbf{D}^{\text{obs}}$ with fictitious data $\mathbf{D}^m$ which is referred to as latent or missing data. The latent data should be chosen such that the loglikelihood of the complete data is relatively straightforward. The algorithm is an iterative method which consist of two steps: the E-step (expectation step) and the M-step (maximization step). Once an initial parameter choice $\boldsymbol{\theta}^0$ is chosen, the E-step and M-step are performed repeatedly until convergence occurs, which is until the difference between successive iterates is negligible.

The EM algorithm is as follows:

E-Step. The E-step consists of computing the expected value of the complete data loglikelihood conditional on the observed data and current estimate,

$$E_{\mathbf{D}^m|\boldsymbol{\theta}^{[i]},\mathbf{D}^{\text{obs}}} l(\boldsymbol{\theta}; \mathbf{D}^{\text{obs}}\, \mathbf{D}^m).$$

M-Step. The M-step requires maximizing the expectation calculated in the E-step with respect to $\boldsymbol{\theta}$ to obtain the next iterate $\boldsymbol{\theta}^{[i]}$.

Iterates obtained using the EM algorithm converge to a turning point of the likelihood. Readers are referred to Hastie et al. [41] for an explanation as to why the EM algorithm works. A numerical example is provided in [79].

CHAPTER 9

COMPARING SURVIVAL (LIFE) DISTRIBUTIONS VIA A NEW PARTIAL ORDERING

9.1 Introduction

In statistical reliability theory, identifying the broad concept of 'aging' or degradation in any specific sense for modeling the distribution of lifelength of an equipment or system is often motivated by intuitive considerations. For example, the notion of 'increasing failure rate' (IFR) property is derived from the observation that if X denotes the lifelength of a device with cumulative distribution function (cdf) $F(\cdot)$ and probability density function (pdf) $f(\cdot)$; then given that the device has not failed at age x, the conditional probability that it does not last beyond an additional time $h > 0$, is

$$P(X \leq x + h | X > x) = hr(x) + o(h) \tag{9.1.1}$$

where $r(x) := \dfrac{f(x)}{1 - F(x)}$, as a function of device's age x can be interpreted as a measure of the equipment's proneness to fail at age x, and is called *the failure rate* function $r(x)$ at age x. Thus, if the life distribution F is such that the proneness to fail as measured by $r(x)$ increases ($\uparrow$) with age x; the device or equivalently its distribution F is said to be IFR. In studying the ramifications of this intuitive notion of aging in the sense of IFR property, it is well known that equivalence of the IFR property ($r(x) \uparrow$) to a specific stochastic partial order (viz., a convex order $>_c$, see [9]) proved to be very useful [9], [60] to discover many useful consequences of the IFR property.

Such interplay between an intuitive notion of degradation and an equivalent partial order has had a large role in the development of many nonparametric life

distribution classes (such as IFR, IFRA, DMRL, NBU, NBUE, HNBUE), useful as stochastic models of life [9], [14].

In this Chapter, a new partial ordering among life distributions on the half line $[0, \infty)$ is introduced and several of its consequences are investigated. It may be noted here that this work takes the reverse of the traditional route of going from an intuitive notion to some equivalent partial order.

9.2 Reverse Star-order among Life Distributions

For any distribution function F, which need not be continuous throughout its entire support, its generalized inverse $F^{-1} : (0,1) \rightsquigarrow (-\infty, \infty)$ is

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \qquad 0 < u < 1 \tag{9.2.1}$$

If F is a life distribution and hence its support is contained in $[0, \infty)$, then it can be obviously written $F^{-1}(u) = \inf\{x \geq 0 : F(x) \geq u\}$ for $u \in (0,1)$. Alternatively, if F is continuous on its support, then $\{x : F(x) \geq u\}$ is a singleton for each $u \in (0,1)$ so that there is a 1:1 correspondence between F and F^{-1} (i.e. F is fully invertible).

Definition 9.2.1 *A real valued function $h(x)$ on $A \subset (0, \infty)$ is star-shaped if $h(x)/x$ is nondecreasing on A; equivalently, if*

$$h(\alpha x) \leq \alpha h(x), \qquad x \in A, \quad 0 < \alpha < 1. \tag{9.2.2}$$

The idea of star-shaped functions has been used, in the context of reliability theory, to define a partial ordering among distributions on the half line $[0, \infty)$; viz.,

Definition 9.2.2 (See Barlow and Proschan [9]). *If F and G (continuous) are life distribution, then F is said to be $>_*$-dominate G if $F^{-1}G : [0, \infty) \to (0,1)$ is a star-shaped function on the support $\{x : G(x) < 1\}$ of G.*

The intuitive notion of *increasing failure rate average*(IFRA) property and the corresponding nonparametric clan of life distributions can be described via the *star-*

ordering ($>_*$) above, namely: F if IFRA if $F >_* G$ where G is the unit ($mean = 1$) exponential distribution [9].

Motivated by the above and some results of Chan, Proschan and Sethuraman [14] on comparing life distribution using a partial order they developed using notion of convexity; we introduce a new partial order defined below.

Definition 9.2.3 *Let F and G be continuous life distributions. Then F is said to be reverse star-ordered ($>^*$) relative to G, if FG^{-1} is a star-shaped function on $(0,1)$; i.e., $\dfrac{FG^{-1}(u)}{u}$ is non-decreasing on $(0,1)$.*

If F is reverse star-ordered relative to G in the sense of Definition 9.2.3; we say $F >^* G$ (F dominates G in reverse star-order, or equivalently F dominates G in $>^*$-order). Note that, while $F^{-1}G : [0,\infty) \to [0,1)$; the function $FG^{-1} : (0,1) \to (0,1)$. Clearly, *neither* the star-order ($>_*$), *nor* the reverse star-order ($>^*$) implies the other.

The following result characterizes the $>^*$-order and further shows that $>^*$-order is stronger than the usual stochastic majorization order $>_{st}$, i.e., $F >_{st} G$ ($F(x) \leq G(x)$ for all x).

Lemma 9.2.4 *Let the distribution functions F and G be continuous. Then, $F >^* G$ if and only if $\dfrac{F}{G} \uparrow$, and implies ($\Rightarrow$) $F >_{st} G$.*

Proof By Definition 9.2.3, $F >^* G$ if and only if $\dfrac{FG^{-1}(u)}{u}$ is non-decreasing in u on $(0,1)$. i.e.,

$$F >^* G \Leftrightarrow \frac{FG^{-1}(u)}{u} \equiv \frac{FG^{-1}(u)}{GG^{-1}(u)} \uparrow \quad \text{in } u \in (0,1) \tag{9.2.3}$$

$$\Leftrightarrow \frac{F(x)}{G(x)} \uparrow \quad \text{in } x > 0, \quad \text{setting } u = G(x)$$

$$\Rightarrow 0 < \frac{F(x)}{G(x)} \leq \lim_{x\to\infty} \frac{F(x)}{G(x)} = \frac{F(\infty)}{G(\infty)} = 1, \quad x > 0 \tag{9.2.4}$$

$$\Leftrightarrow 1 - G(x) \leq 1 - F(x)$$

$$\Leftrightarrow F >_{st} G$$

Note:

(i) G is continuous on $[0,\infty)$ implies $GG^{-1}(u) = u$ in (9.2.3). If G is an arbitrary cdf, then $GG^{-1}(u) \geq u$ on $(0,1)$ with strictly inequality if u is a discontinuity point of G. On the other hand $G^{-1}G(u) = u$ on $(0,1)$ for any cdf G, continuous or not.

(ii) It is clearly possible to extend the domain of $>^*$ ordered life distribution by allowing a possible jump discontinuity (infant mortality effect) at zero. If X, Y are lifetimes with cdfs F, G continuous on $(0,\infty)$ but with possible discontinuous at zero; note that right continuity of cdfs and (9.2.4) together implies

$$P(X > 0) = 1 - F(0) = 1 - \lim_{x\to 0+} F(x) \geq 1 - \lim_{x\to 0+} G(x) = 1 - G(0) = P(Y > 0).$$

Hence if G is continuous at 0 ($\Leftrightarrow P(Y > 0) = 1$), then so must be F.

Lemma 9.2.5 *The relation $>^*$ is a partial order.*

Proof (i) Reflexive: $FF^{-1}(u) = u$ is trivially star-shaped, thus $F >^* F$, i.e., F is $>^*$-ordered relative to itself.

(ii) Transitive: $F >^* G$ and $G >^* H$ imply $F >^* H$, since $FH^{-1}(u) = FG^{-1}GH^{-1}(u)$ is star-shaped. viz., $a(u)$, $b(u)$ both are star-shaped and $b(u) \uparrow$ in $u \Rightarrow a \circ b(u) \equiv a(b(u))$ is star-shaped, since $\dfrac{a \circ b(u)}{u} = \dfrac{a(b(u))}{b(u)} \cdot \dfrac{b(u)}{u}$ is $\uparrow$ in u.

(iii) Antisymmetric: $F >^* G$ and $G >^* F$ imply $F = G$, since by Lemma 9.2.4, $F >^* G \Rightarrow F >_{st} G$ and $G >^* F \Rightarrow G >_{st} F$, together, $F = G$ pointwise is obtained.

9.3 Results for Reverse Star-order

The idea and our formulation of the reverse star-order ($>^*$-order) as a concept has been motivated by some work of Chan, Proschan and Sethuraman [14]. They define

a convex order ($>^c$) as: life distributions $F >^c G$ (F is *more convex than* G in $>^c$-ordered) if $FG^{-1}(u)$ is convex for $0 < u < 1$, and investigate its reliability theoretic consequences. It may be noted here that the convex partial order $>^c$ is *different* from the convex order $>_c$ described in Section 9.1 which is an equivalent description of the well known IFR (increasing failure rate) property as a non-parametric notion of an equipment's degradation with age.

The notion of reverse star-order ($>^*$) can be extended to real valued right-continuous monotone non-decreasing ($\uparrow$) and non-increasing ($\downarrow$) functions or the half-line by suitably re-defining their inverse functions as:

(a) if f is monotone $\downarrow$ on $[0, \infty)$, then $f^{-1}(z) := \sup\{x : f(x) \leq z\}$,

(b) if f is monotone $\uparrow$ on $(0, \infty)$, then $f^{-1}(z) := \inf\{x : f(x) \geq z\}$, as defined earlier for cdfs.

If f is continuous and strictly monotone, then $f(x)$ is of course fully invertible.

In the spirit of Definition 9.2.3; for such real valued function f and g, it will be said that $f >^* g$ (f is more $>^*$-star shaped than g) if $f \circ g^{-1}(x) := f(g^{-1}(x))$ is star-shaped on its domain. A corresponding convex ordering ($>^c$) among continuous functions f and g is similarly defined, viz., $f >^c g$ (f is more $>^c$-convex than g) if $f \circ g^{-1}(x) := f(g^{-1}(x))$ is convex on its domain.

A question of natural interest in comparing life distributions F and G is: if it is known that the d.f.s F and G are suitably ordered what can be said about possible ordering of the corresponding survival probabilities (tail functions) $\overline{F}$ and $\overline{G}$.

Lemma 9.3.1 *If F and G are absolutely continuous life distributions, then*

$$\overline{G} >^c \overline{F} \Rightarrow \overline{G} >^* \overline{F} \tag{9.3.1}$$

i.e., survival function of G is more $>^c$-convex than that of F implies that survival function of G is more $>^*$-star-shaped than the survival function of F.

Proof Follows from the facts that,

(i) F and G are continuous on $[0,\infty)$, which requires $F(0) = G(0) = 0$,

(ii) for any real valued convex function $h(x)$ on $[0,\infty)$, $h(0) \leq 0$ implies $h(x)$ is star-shaped on $(0,\infty)$, viz., for $x > 0$ and $0 < \alpha < 1$,

$$h(\alpha x) \equiv h(\alpha x + (1-\alpha)\cdot 0) \leq \alpha h(x) + (1-\alpha)h(0), \quad \text{by convexity}$$
$$\leq \alpha h(x)$$

Hence, $\overline{G} >^c \overline{F}$ and continuous at zero together imply $\overline{G} >^* \overline{F}$.

Remark 9.3.2 We may note that the conclusion of Lemma 9.3.1 also follows from a corresponding result of Chan, Proschan and Sethnraman [14] under a stronger $>^c$-ordering which they define (viz., $F >^c G$ if FG^{-1} is convex on $(0,1)$ where d.f. G is continuous and $\dfrac{dF}{dG}$ exists) states

$$F >^c G \text{ if and only if } \overline{G} >^c \overline{F}$$

(see their theorem 2.8, p.125 in [14]), and further that $F >^c G$ implies both the ratios $\dfrac{F}{G}$ and $\dfrac{\overline{F}}{\overline{G}}$ of d.f.s and corresponding survival functions are nondecreasing. We note however that in Lemma 9.3.1, the implication holds only one way.

For a lifetime X with cdf F, and a finite mean, let

$$\nu_F(t) = E(X - t | X > t) = \frac{\int_t^\infty \overline{F}(x)\,dx}{\overline{F}(t)}, \quad t > 0 \tag{9.3.2}$$

be its mean residual life (MRL) function at age t. For simplicity, it is assumed that the support of life distributions is the entire half line $[0,\infty)$. If the right end point of support of F is finite, i.e., $\sup\{x \geq 0 : F(x) < 1\} < \infty$, the argument and result below can be suitably modified.

Our next result shows the relationship between reverse star-order ($>^*$) and the mean residual life (MRL) order (for life distribution function F and G, it is said that $G \geq_{MRL} F$ if $\nu_G(t) \geq \nu_F(t)$) pointwise.

Corollary 9.3.3 $\overline{G} >^* \overline{F} \Rightarrow \nu_G(t) \geq \nu_F(t)$ *for all* $t > 0$.

Proof Note that,

$$\begin{aligned}
\overline{G} >^* \overline{F} &\Rightarrow \frac{\overline{G}(x)}{\overline{F}(x)} \quad \uparrow \quad \text{(from an obvious analog of Lemma 9.2.4)} \qquad (9.3.3)\\
&\Rightarrow \frac{\overline{G}(x)}{\overline{F}(x)} \geq \frac{\overline{G}(t)}{\overline{F}(t)}, \quad \text{i.e.,} \quad \overline{F}(t)\overline{G}(x) \geq \overline{G}(t)\overline{F}(x), \quad \text{all } x \geq t\\
&\Rightarrow \overline{F}(t) \int_t^\infty \overline{G}(x)\, dx \geq \overline{G}(t) \int_t^\infty \overline{F}(x)\, dx > 0\\
&\Leftrightarrow \nu_G(t) = \frac{\int_t^\infty \overline{G}(x)\, dx}{\overline{G}(t)} \geq \frac{\int_t^\infty \overline{F}(x)\, dx}{\overline{F}(t)} = \nu_F(t), \quad t > 0
\end{aligned}$$

Remark 9.3.4 The first implication (9.3.3) above follows from arguments paralleling to those used in the proof of Lemma 9.2.4.

Our next result shows the relationship between our proposed partial order $>^*$ and the well known hazard rate order $>_{hr}$, which is defined [10] as: for life distribution function F and G, it is said that $G >_{hr} F$ if $r_G(t) \geq r_F(t)$, where $r(\cdot)$ is defined as in Section 9.1.

Corollary 9.3.5 $\overline{G} >^* \overline{F} \Leftrightarrow G >_{hr} F$.

Proof Suppose the survival function $\overline{G}$ $>^*$-dominates the survival function $\overline{F}$. Hence,

$$\begin{aligned}
\overline{G} >^* \overline{F} &\Leftrightarrow \frac{\overline{G}\,\overline{F}^{-1}(u)}{u} \quad \uparrow \quad \text{(from Definition 9.2.3)}\\
&\Leftrightarrow u^{-2}\left[u\frac{g(\overline{F}^{-1}(u))}{f(\overline{F}^{-1}(u))} - \overline{G}\,\overline{F}^{-1}(u)\right] \geq 0\\
&\Leftrightarrow \frac{g(\overline{F}^{-1}(u))}{f(\overline{F}^{-1}(u))} \geq \frac{\overline{G}\,\overline{F}^{-1}(u)}{u} = \frac{\overline{G}\,(\overline{F}^{-1}(u))}{\overline{F}\,(\overline{F}^{-1}(u))}\\
&\text{i.e.,} \ \frac{g(x)}{f(x)} \geq \frac{\overline{G}(x)}{\overline{F}(x)}\\
&\Leftrightarrow \frac{g(x)}{\overline{G}(x)} \geq \frac{f(x)}{\overline{F}(x)} \quad \text{i.e., } r_G(x) \geq r_F(x)\\
&\Leftrightarrow G >_{hr} F
\end{aligned}$$

In the spirit of (9.1.1), if X is a lifetime with a density $f(\cdot)$ and cdf $F(\cdot)$; then given an equipment with lifetime X had failed by clock-time x, the conditional probability that it actually failed shortly before time x within a small left-neighborhood of x is:

$$P(X > x - h | X \leq x) = h\lambda(x) + o(h) \tag{9.3.4}$$

where $\lambda(x) := \dfrac{f(x)}{F(x)}$ is the so called *reversed hazard rate* (see [10]). Continuous life distributions F and G on $[0, \infty)$ with densities f and g respectively are said to be *reversed hazard ordered* in the sense of following definition.

Definition 9.3.6 *$F >^{rh} G$ (F has more reversed hazard rate than G) if $(f/F) \geq (g/G)$ pointwise.*

As an apparently surprising implication of the reverse star-order ($>^*$), we show that a subset of $>^*$-ordered life distributions coincides with the set of *reversed hazard ordered* distributions. This result appears to be new and provides an alternative description of the reversed hazard order that was originally motivated intuitively by Block, Savits and Singh [10].

Consider the subset of all $>^*$-ordered (and hence continuous) life distributions which are *absolutely continuous* and thus have probability density functions (pdfs). The result below shows that within this subclass, $>^*$ and $>^{rh}$ ordering are equivalent.

Theorem 9.3.7 *If F and G are absolutely continuous life distributions with densities f and g respectively, then $F >^* G$ if and only if $(f/F) \geq (g/G)$ pointwise.*

Proof Using Definition 9.2.3, the following can be having

$$\begin{aligned} F >^* G &\Leftrightarrow FG^{-1}(u) \text{ is star-shaped for } 0 < u < 1 \\ &\Leftrightarrow h(u) := \frac{FG^{-1}(u)}{u} \uparrow \text{ in } u \text{ on } (0,1) \\ &\Leftrightarrow h'(u) \geq 0 \quad \text{for } 0 < u < 1, \text{ if } F \text{ and } G \text{ have densities.} \end{aligned}$$

Let $G^{-1}(u) = x \in (0, \infty)$, then $u = G(x)$ and,

$$\frac{du}{dx} = G'(x) \equiv g(x) \Leftrightarrow \frac{dx}{du} = \frac{1}{g(x)} = \frac{1}{g(G^{-1}(u))} \equiv \frac{1}{g(x)}.$$

Now considering $h'(u)$,

$$\begin{aligned} h'(u) &= u^{-2}\left[u\, f(G^{-1}(u))\, \frac{dG^{-1}(u)}{du} - FG^{-1}(u)\right] \\ &= u^{-1}\left[f(G^{-1}(u))\, \frac{dG^{-1}(u)}{du} - \frac{FG^{-1}(u)}{u}\right], \quad 0 < u < 1. \end{aligned}$$

Since $sgn(h'(u)) = sgn\left[f(G^{-1}(u))\, \frac{dG^{-1}(u)}{du} - \frac{FG^{-1}(u)}{u}\right]$ for $0 < u < 1$,

$$\begin{aligned} 0 \le h'(u) &\Leftrightarrow 0 \le f(G^{-1}(u))\, \frac{dG^{-1}(u)}{du} - \frac{FG^{-1}(u)}{u} \\ &= f(x)\, \frac{dx}{du} - \frac{F(x)}{G(x)} \\ &= \frac{f(x)}{g(x)} - \frac{F(x)}{G(x)} \\ &= \frac{F(x)}{g(x)}\left[\frac{f(x)}{F(x)} - \frac{g(x)}{G(x)}\right] \\ &\Leftrightarrow \frac{f(x)}{F(x)} \ge \frac{g(x)}{G(x)}, \quad i.e., F >^{rh} G \quad \text{for} \quad x \in S(F) \cap S(G) \end{aligned}$$

where $S(F)$ and $S(G)$ is the support of F and G respectively.

If F is a life distribution with a finite mean μ, let

$$\widetilde{F}(x) := \mu^{-1} \int_0^x \overline{F}(t)dt, \qquad x > 0,$$

denote the induced distribution, whose interpretations in the context of renewal theory, as the initial distribution of a stationary renewal process and as the asymptotic distribution of the *age* or, the *remaining life of an item in use* at time t as $t \to \infty$, are well known. The following are some additional consequences of the *reverse star-order* ($>^*$) property.

Corollary 9.3.8 *Let F be a life distribution with a finite mean. Then, $F >^* \widetilde{F}$ ($\widetilde{F} >^* F$) implies F is DMRL (IMRL respectively).*

[DMRL (IMRL) ≡ Decreasing (Increasing) Mean Residual Life].

Proof By Lemma 9.2.4,

$$F >^* \widetilde{F} \Rightarrow \frac{1-F(x)}{1-\widetilde{F}(x)} = \frac{\mu\overline{F}(x)}{\int_x^\infty \overline{F}(t)dt} = \frac{\nu_F(0)}{\nu_F(x)} \text{ is } \uparrow \text{ in } x > 0,$$

where $\mu_F(\cdot)$ is the MRL function of F, defined in (9.3.2). The duel result follows analogously.

Corollary 9.3.9 *Let F, G be life distributions with finite means. Then,*

$$F >^* G \text{ if and only if } \widetilde{F} >^* \widetilde{G}.$$

Proof If λ denotes the Lebesgue measure on the real line; simply note that,

$$\widetilde{F} >^* \widetilde{G} \Leftrightarrow \frac{d\widetilde{F}}{d\widetilde{G}} = \left(\frac{d\widetilde{F}}{d\lambda}\right) / \left(\frac{d\widetilde{G}}{d\lambda}\right) = c\left(\frac{\overline{F}}{\overline{G}}\right),$$

where $c = \dfrac{\mu_G}{\mu_F}$, the ratio of the means. Using Lemma 9.2.4 again, completes the argument for the claim in the statement of Corollary 9.3.9.

A final aging property characterization result, analogous to the preceding corollary is:

Corollary 9.3.10 *F is IFR (DFR) if and only if $F >^c \widetilde{F}$ ($\widetilde{F} >^c F$, respectively).*
[IFR (DFR) ≡ Increasing (Decreasing) Failure Rate].

Proof It is standard [9] that F is IFR implies, it is absolutely continuous and thus has a density (f) with respect to Lebesgue measure. The induced distribution $\widetilde{F}$ has a density $\widetilde{f} := \dfrac{f}{\overline{F}}$ by the former's definition. Now, by the definition of $>^c$-order (see [14] and remarks preceding (9.3.2),

$$F >^c \widetilde{F} \Leftrightarrow \frac{dF}{d\widetilde{F}}(x) = \frac{f}{\widetilde{f}} \equiv \frac{f}{(\overline{F}/\mu)} = \mu r(x),$$

where $r(x)$ is the *failure rate* function of F. Now use the characterization of $>^c$-order (Theorem 2.3, p.124 in [14] which shows that $\frac{dF}{dG} \uparrow$ is necessary and sufficient for $F >^c G$ such that $\frac{dF}{dG}$ exists.

Let A be an index set where elements are ordered; and let $\{F_\alpha : \alpha \in A\}$ be a family of distributions indexed by $\alpha \in A$.

Definition 9.3.11 *$\{F_\alpha : \alpha \in A\}$ is a $>^*$-ordered family if $\alpha_2 > \alpha_1$ implies $F_{\alpha_2} >^* F_{\alpha_1}$.*

Theorem 9.3.12 *A family $\{F_\alpha : \alpha \in A\}$ of life distributions is $>^*$-ordered if $F_\alpha(x) : A \times [0, \infty) \to [0, 1]$ is TP_2 (Totally Positive of Order-2) in (α, x).*

Proof For any α_1, α_2 in A such that $\alpha_2 > \alpha_1$,

$$\begin{aligned} F_{\alpha_2} >^* F_{\alpha_1} &\Leftrightarrow F_{\alpha_2} F_{\alpha_1}^{-1}(u) \text{ is star-shaped} \\ &\Leftrightarrow \frac{F_{\alpha_2} F_{\alpha_1}^{-1}(u_2)}{u_2} \geq \frac{F_{\alpha_2} F_{\alpha_1}^{-1}(u_1)}{u_1} \end{aligned}$$

for $0 < u_1 \leq u_2 < 1$, or equivalently, if and only if

$$0 \leq \begin{vmatrix} u_1 & u_2 \\ F_{\alpha_2} F_{\alpha_1}^{-1}(u_1) & F_{\alpha_2} F_{\alpha_1}^{-1}(u_2) \end{vmatrix} = \begin{vmatrix} F_{\alpha_1}(x_1) & F_{\alpha_1}(x_2) \\ F_{\alpha_2}(x_1) & F_{\alpha_2}(x_2) \end{vmatrix}, \tag{9.3.5}$$

setting $0 < x_i := F_{\alpha_i}(u_i)$; $i = 1, 2$. The condition (9.3.5) is of course the TP_2-property for the family of cdfs $F_\alpha(x)$.

9.4 Preservation Results for Reverse Star-order

Theorem 9.4.1 *If $F_\alpha >^* G$ for each α, then $\int F_\alpha d\mu(\alpha) >^* G$ for any mixing distribution μ.*

Proof From the proof of Lemma 9.2.4, it follows that,

$$F_\alpha >^* G \Leftrightarrow \frac{F_\alpha}{G} \uparrow, \qquad \text{for each} \alpha,$$
$$\Rightarrow \frac{\int F_\alpha d\mu(\alpha)}{G} \uparrow$$
$$\Leftrightarrow \int F_\alpha d\mu(\alpha) >^* G.$$

A similar proof holds for our next result.

Theorem 9.4.2 *If $F >^* G_\alpha$ for each α, then $F >^* \int G_\alpha d\nu(\alpha)$ for any mixing distribution ν.*

Theorems 9.4.1 – 9.4.2 together implies

Theorem 9.4.3 *If $F_\alpha >^* G_\beta$ for each pair (α, β), then $\int F_\alpha d\mu(\alpha) >^* \int G_\beta d\nu(\beta)$ for any mixing distributions μ and ν.*

Next, we will show that the reverse star-ordered is preserved under formation of parallel systems of independent components. The life distribution of a parallel system of n independent components with possibly different life distributions F_i $(i = 1, 2, \ldots, n)$ of the components, is given by the product $\prod_{i=1}^{n} F_i(t)$. The following theorem shows that the reverse star-order is preserved under formation of parallel system.

Theorem 9.4.4 *Suppose $F_i >^* G_j$ for each pair (i, j). Then a parallel system of n independent components with life d.f.s $F_1, \ldots, F_n$ is more reliable than a comparable system with life d.f.s $G_1, \ldots, G_n$.*

Proof It is sufficient to prove the theorem for $n = 2$. By Lemma 9.2.4, for each pair (i, j) with $i \neq j$, we have

$$F_i >^* G_j \Leftrightarrow \frac{F_i}{G_j} \uparrow, \text{ and hence } \frac{F_1}{G_2} \uparrow \text{ and } \frac{F_2}{G_1} \uparrow$$
$$\Rightarrow \frac{F_1}{G_2} \cdot \frac{F_2}{G_1} \uparrow \text{ i.e., } \frac{F_1 F_2}{G_1 G_2} \uparrow.$$

Therefore, $F_1F_2 >^* G_1G_2$, which then implies $F_1F_2 >_{st} G_1G_2$, by Lemma 9.2.4. Since the survival function of a parallel system of two independent components with life d.f.s F_1 and F_2 is $\overline{S}_{F_1,F_2}(t) := 1 - F_1(t)F_2(t)$; it follows that a corresponding system with component life d.f.s G_1, G_2 is less reliable, i.e.,

$$\overline{S}_{F_1,F_2}(t) \geq \overline{S}_{G_1,G_2}(t), \qquad \text{pointwise in } t \geq 0.$$

CHAPTER 10

A FRAMEWORK FOR BIVARIATE PROPORTIONAL HAZARD MODELS

10.1 Introduction

Suppose X is an absolute continuous positive random variable with the probability density function (p.d.f.) and cumulative distribution function (d.f.) as $f(\cdot)$ and $F(\cdot)$ respectively. Then the hazard rate function at time t is defined as

$$r(t) = \frac{f(t)}{\overline{F}(t)}, \qquad t > 0, \tag{10.1.1}$$

where the tail function $\overline{F} := 1 - F$ is usually referred to as the *survival function* in reliability theory.

The class of proportional hazard model can described as follows; A life distribution F on the non-negative half line is referred to as a *proportional hazard model* if

$$F(t) = 1 - \left[\overline{F}_0(t)\right]^{\alpha}, \quad t > 0, \tag{10.1.2}$$

for some $\alpha > 0$, and some (baseline) life distribution F_0 with survival function $\overline{F}_0 = 1 - F_0$. In such a case, the *hazard functions* $\Lambda_F := -\ln \overline{F}$ and $\Lambda_{F_0} := -\ln \overline{F}_0$ satisfy

$$-\ln \overline{F}(t) = -\alpha \ln \overline{F}_0(t), \quad t > 0.$$

If further F_0 is also absolutely continuous so that F_0 and F both have density functions, then the failure (hazard) rate $r_F(\cdot)$ is proportional to $r_{F_0}(\cdot)$, as

$$r_F(t) = \alpha r_{F_0}(t), \quad t > 0, \tag{10.1.3}$$

whence the name proportional hazards. In practice, theoretical developments in the proportional hazard model literature in the univariate set up within the basic

framework exemplified by (10.1.2) and (10.1.3) are in conjunction with a vector $\boldsymbol{X}$ of covariates [20], [45], [46], [30], [43]. The underlying idea behind including a vector $\boldsymbol{X}$ of covariates is to account for possible influence of some environmental but typically unobserved variables as the hazard rate $r_F(t)$ formulated as

$$r_F(t) = r_{F_0}(t)\, e^{\mathbf{X}\beta},$$

referred to in the literature as the Cox-regression model. The vast majority of its fruitful applications with real life data sets and associated methodological refinements via censoring etc. are to be found mostly in biostatistics and engineering literature too numerous to mention, and in other fields such as economics and political science as well (e.g., [83], [11]).

10.1.1 Notions of Bivariate Proportional Hazards in Literature

For modeling the joint distribution of possibly mutually dependent lifetimes that may be considered as multivariate version(s) of the proportional hazards idea, however there is no universally agreed natural way to formulate such a notion. Several researchers have considered this modeling problem from different perspectives.

Clayton and Cuzick [17] proposed and investigated a bivariate generalization of the univariate proportional hazard (their Theorem 1, p. 85 in [17]). Their proposed class of bivariate survival functions is the solution (unique, under some regularity conditions) of a functional equation requiring a suitably defined bivariate hazard rate (called by Clayton and Cuzick as 'mortality potential') to be equal to a scalar multiple of the product of a version of the corresponding univariate conditional hazard rates given the other lifetime exceeds a threshold. Beyond this formulation, however, the main accent of their work is on statistical inference for estimation and testing of an 'association parameter' between the component lifetimes from right censored sample data using only rank-order information.

Hougaard [46] proposed a class of multivariate lifetime models that account for possible heterogeneity across individuals or units within a group via unobserved covariates with a *positive stable distribution*. His models have the property that if the hazards conditional on the covariates are proportional, then so are the hazards of marginal distributions (with different constants of proportionality). Additionally, for Hougaard model the hazard of the minimal lifetime in a group is proportional to the sum of the marginal hazards.

The idea of dependence among component lifetimes induced by an unobserved covariate, known as the *frailty* effect has also been used by Oakes [61]. He considers a class of bivariate survival distributions, where the two component lifetimes are conditionally independent univariate proportional hazard models given the frailty. The unobserved frailty variable is shown to induce a negative association between the observed survival times. In the class of models investigated by Oakes, the observable bivariate distribution determines the unobserved frailty distribution up to scale equivalence, so that there are no identifiability issues.

The main aim of this Chapter is to formulate a suitable new notion of *bivariate proportional hazard models* (BPHM) and study its consequence and applicability. This is done by defining BPHMs in such a way that implies their marginal distributions follow *univariate* proportional hazard model (PHM) distributions. The proposed model is shown to have a structure that has a singular part, a feature often shared by multivariate distributions (see Marshall and Olkin [55]). It is observed that, as expected, the maximum likelihood estimators of the unknown parameters cannot be obtained in explicit forms. Non-linear equations need to be solved to compute the estimators of the unknown parameters. An expectation-maximization (EM) algorithm is used for computing the MLEs of parameters. For the purpose of illustration, the proposed bivariate proportional hazard models and the method of parameter estimation will be applied to two real data sets.

10.2 A Framework for Bivariate Proportional Hazard Models

The point of departure in our framework, compared to the formulation by Clayton and Cuzick [17], Hougaard [46] and Oakes [61] in the sense of, possibly unobserved, latent variables that inject dependence between the components of the bivariate lifetime in a way that guarantees all marginal distributions to have univariate proportional hazards.

Our formulation is motivated by a recent work of Kundu and Gupta [53] with proportional reversed hazards. The proposed BPHMs do not include covariates (frailties), since the emphasis here is on developing a basic formulation and investigate its corresponding distribution theoretic consequences. From an application point of view, inclusion of such unobserved frailties may indeed be important and can be a topic of future research, not included within the scope of the work reported here.

If $\{\overline{F}_0(t;\theta) : \theta \in \Theta\}$ is a family of *baseline* survival functions with parameters $\theta \in \Theta \subset \mathcal{R}^k$, the k-dimensional Euclidian space for same $k \geq 1$; in the spirit of (10.1.2), a univariate proportional hazard model (PHM) is defined as a *parametric family* of lifetime cdfs ($\equiv$ d.f.)

$$F_{PHM}(x;\alpha,\theta) = 1 - [\overline{F}_0(x;\theta)]^\alpha, \quad x > 0, \quad \alpha > 0 \tag{10.2.1}$$

with parameters $\alpha > 0$ and θ (possibly vector valued). If the baseline d.f. F_0 admits a density f_0, then the PHM family admits a density

$$f_{PHM}(x;\alpha,\theta) = \alpha[\overline{F}_0(x;\theta)]^{\alpha-1} f_0(x;\theta), \quad x > 0. \tag{10.2.2}$$

10.3 A New Class of Bivariate Proportional Hazard Models

The approach to define the notion of bivariate proportional hazard is as follows.

Definition 10.3.1 *A pair of lifetimes* (X_1, X_2) *has a bivariate proportional hazard model (BPHM) distribution iff*

$$X_1 = \min(T_1, T_3), \quad X_2 = \min(T_2, T_3)$$

for some (T_1, T_2, T_3) *which are independent and each with a univariate proportional hazard model (PHM) distribution.*

An example that illustrates the genesis and rationale from Definition 10.3.1 can be thought of as follows. Suppose T_1, T_2 denote the lifelengths of heating/cooling systems in two independently owned homes. Let T_3 denote the time to a catastrophic power failure, which occurs independently of the failures of the home heating/cooling systems, and which would automatically shut down the home systems by cutting off the power supply (referred to as a 'common cause failure'). Let X_i, $i = 1, 2$ be the *observed* failure time of the home heating/cooling systems, then (X_1, X_2) are as in Definition 10.3.1.

It is important to note here that while the component lifetimes (X_1, X_2) are observable, the background latent variables T_1, T_2, and T_3 defining them are not *covariates* in the sense as usually understood in the context of survival analysis.

If each T_i, $i = 1, 2, 3$ has a d.f. in a parametric PHM-family $\{F(\cdot\,; \alpha, \theta), \alpha > 0, \theta \in \Theta\}$ as in (10.2.1), i.e.,

$$T_i \sim PHM(\alpha_i, \theta), \quad i = 1, 2, 3 \tag{10.3.1}$$

then following Definition 10.3.1 it can be said that the random vector

$$(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta) \tag{10.3.2}$$

i.e., (X_1, X_2) follows a bivariate proportional hazard model with survival function

$$\overline{F}_{X_1,X_2}(x_1, x_2) := P(X_1 > x_1, X_2 > x_2) \tag{10.3.3}$$

given by,

Proposition 10.3.2 *If* $(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta)$, *then the joint survival function for* $x_1 > 0$ *and* $x_2 > 0$ *is*

$$\overline{F}_{X_1,X_2}(x_1, x_2) = [\overline{F}_0(x_1; \theta)]^{\alpha_1} [\overline{F}_0(x_2; \theta)]^{\alpha_2} [\overline{F}_0(z; \theta)]^{\alpha_3} \tag{10.3.4}$$

where $z = \max(x_1, x_2)$.

Proof Follows as a consequence of (10.3.1) – (10.3.2) and independence of T_1, T_2, T_3; observing that,

$$\begin{aligned}\overline{F}_{X_1,X_2}(x_1,x_2) &= P(\min(T_1,T_3) > x_1, \min(T_2,T_3) > x_2) \\ &= P(T_1 > x_1, T_2 > x_2, T_3 > \max(x_1,x_2))\end{aligned}$$

and then using (10.3.1) and independence of T_1, T_2, T_3.

Note that (10.3.4) can be written as

$$\overline{F}_{X_1,X_2}(x_1,x_2) = \begin{cases} [\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3}\,[\overline{F}_0(x_2;\theta)]^{\alpha_2}, & \text{if } 0 < x_2 < x_1 < \infty \\ [\overline{F}_0(x_1;\theta)]^{\alpha_1}\,[\overline{F}_0(x_2;\theta)]^{\alpha_2+\alpha_3}, & \text{if } 0 < x_1 < x_2 < \infty \\ [\overline{F}_0(x;\theta)]^{\alpha_1+\alpha_2+\alpha_3}. & \text{if } 0 < x_1 = x_2 = x < \infty \end{cases} \tag{10.3.5}$$

If the baseline univariate PHM d.f. F_0 is absolutely continuous, then the BPHM survival time (X_1, X_2) has a joint probability density function (p.d.f.) given by the following result.

Proposition 10.3.3 *If the baseline d.f.* F_0 *in (10.3.1) is absolutely continuous so that the univariate PHM family in (10.2.1) admits a p.d.f.* $f_{PHM}(x;\alpha,\theta)$ *given by (10.2.2), then the joint pdf of* $(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta)$ *is given by*

$$\begin{aligned} &f_{X_1,X_2}(x_1,x_2) \\ &= \begin{cases} f_{PHM}(x_1;\alpha_1+\alpha_3,\theta) \times f_{PHM}(x_2;\alpha_2,\theta), & \text{if } 0 < x_2 < x_1 < \infty \\ f_{PHM}(x_1;\alpha_1,\theta) \times f_{PHM}(x_2;\alpha_2+\alpha_3,\theta), & \text{if } 0 < x_1 < x_2 < \infty \\ \dfrac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} f_{PHM}(x;\alpha_1+\alpha_2+\alpha_3,\theta). & \text{if } 0 < x_1 = x_2 = x < \infty \end{cases} \end{aligned} \tag{10.3.6}$$

Proof By taking $\dfrac{\partial^2}{\partial x_1 \partial x_2}\overline{F}_{X_1,X_2}(x_1,x_2)$ for $x_2 < x_1$ and $x_1 < x_2$ respectively, the first two expressions in (10.3.6) can be obtained. For the remaining case when $x_1 = x_2 = x$;

use

$$1 = \int_0^\infty \int_0^\infty f_{X_1,X_2}(x_1,x_2) = \int_0^\infty \int_0^{x_1} f_{PHM}(x_1;\alpha_1+\alpha_3,\theta) \times f_{PHM}(x_2;\alpha_2,\theta)\, dx_2\, dx_1 + \int_0^\infty \int_0^{x_2} f_{PHM}(x_1;\alpha_1,\theta) \times f_{PHM}(x_2;\alpha_2+\alpha_3,\theta)\, dx_1\, dx_2 + \int_0^\infty f_{X_1,X_2}(x,x)\, dx,$$

together with the identities

$$\int_0^\infty \int_0^{x_1} f_{PHM}(x_1;\alpha_1+\alpha_3,\theta) \times f_{PHM}(x_2;\alpha_2,\theta)\, dx_2\, dx_1 = \frac{\alpha_2}{\alpha_1+\alpha_2+\alpha_3};$$

$$\int_0^\infty \int_0^{x_2} f_{PHM}(x_1;\alpha_1,\theta) \times f_{PHM}(x_2;\alpha_2+\alpha_3,\theta)\, dx_1\, dx_2 = \frac{\alpha_1}{\alpha_1+\alpha_2+\alpha_3};$$

thus leading us to,

$$\int_0^\infty f_{X_1,X_2}(x,x)\, dx = 1 - \frac{\alpha_1+\alpha_2}{\alpha_1+\alpha_2+\alpha_3} = \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}.$$

Since the joint survival function along the diagonal (x,x) given by (10.3.5) has the density function

$$-\frac{d}{dx}[\overline{F}_0(x;\theta)]^{\alpha_1+\alpha_2+\alpha_3} = f_{PHM}(x;\alpha_1+\alpha_2+\alpha_3,\theta),$$

which is a proper univariate density on $(0,\infty)$, it follows that

$$f_{X_1,X_2}(x,x) := \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} f_{PHM}(x;\alpha_1+\alpha_2+\alpha_3,\theta)$$

is a version of the joint density $f_{X_1,X_2}(x_1,x_2)$ along the diagonal (x,x) on the positive quadrant.

Remark 10.3.4 Proposition 10.3.3 shows that even when the baseline PHM is absolutely continuous, the bivariate density of (X_1,X_2) has a *singular component* (in the case when $x_1 = x_2$), a feature often shared by other multivariate model in the

literature, e.g. by the well known multivariate exponential distribution of Marshall and Olkin [55]. An item of future work plan in this dissertation is to provide a decomposition of the joint survival function $\overline{F}_{X_1,X_2}(x_1,x_2)$ into its absolutely continuous and singular parts.

The next result, which easily follow again from (10.3.1) and (10.3.2) and independence of T_i, $i = 1,2,3$, justifies the formulation of BPHM in Definition 10.3.1 by demonstrating that its marginals indeed belong to a family of univariate proportional hazard models.

Proposition 10.3.5 *If* $(X_1,X_2) \sim BPHM(\alpha_1,\alpha_2,\alpha_3,\theta)$, *then*

$$X_1 \sim PHM(\alpha_1+\alpha_3,\theta); \qquad X_2 \sim PHM(\alpha_2+\alpha_3,\theta);$$

$$\min(X_1,X_2) \sim PHM(\alpha_1+\alpha_2+\alpha_3,\theta).$$

Proof Follow from (10.3.5); viz., for the case $x_2 < x_1$ in (10.3.5), it can be obtained that

$$\begin{aligned} P(X_1 > x_1) = \overline{F}_{X_1,X_2}(x_1,0+) &= \lim_{x_2\to 0+} [\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3}\,[\overline{F}_0(x_2;\theta)]^{\alpha_2} \\ &= [\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3}. \end{aligned}$$

The remaining claims follow analogously.

If the bivariate hazard rate $h(\cdot,\cdot)$ of (X_1,X_2) defined by

$$\begin{aligned} h_{X_1,X_2}(x_1,x_2) :&= \frac{f_{X_1,X_2}(x_1,x_2)}{\overline{F}_{X_1,X_2}(x_1,x_2)} \\ &= \frac{\partial^2}{\partial x_1 \partial x_2} \ln \overline{F}_{X_1,X_2}(x_1,x_2) \end{aligned} \tag{10.3.7}$$

is used as the obvious generalization of the univariate hazard (failure) rate, as a measure of joint failure proneness of (X_1,X_2), the following result also can be obtained.

Proposition 10.3.6 *If* $(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta)$ *admits a density on* $\mathcal{R}^2$, *then the bivariate hazard rate as defined by (10.3.7) satisfies*

$$h_{X_1,X_2}(x_1,x_2) = \begin{cases} (\alpha_1+\alpha_3)\,\alpha_2\, r(x_1;\theta)\, r(x_2;\theta), & \text{if } 0 < x_2 < x_1 < \infty \\ \alpha_1\,(\alpha_2+\alpha_3)\, r(x_1;\theta)\, r(x_2;\theta), & \text{if } 0 < x_1 < x_2 < \infty \\ \alpha_3\, r(x;\theta), & \text{if } 0 < x_1 = x_2 = x < \infty \end{cases} \tag{10.3.8}$$

where $r(\cdot;\theta)$ *is the hazard (failure) rate of the univariate baseline d.f.* F_0 *in (10.2.1).*

Proof Since $\overline{F}_{X_1,X_2}(x_1,x_2) = [\overline{F}_0(x_1;\theta)]^{\alpha_1}\,[\overline{F}_0(x_2;\theta)]^{\alpha_2}\,[\overline{F}_0(z;\theta)]^{\alpha_3}$, where $z = \max(x_1,x_2)$, the bivariate hazard rate at (x_1,x_2), $h_{X_1,X_2}(x_1,x_2) = \dfrac{f_{X_1,X_2}(x_1,x_2)}{\overline{F}_{X_1,X_2}(x_1,x_2)}$ is computed as follows:

For $0 < x_2 < x_1$:

$$\begin{aligned} h_{X_1,X_2}(x_1,x_2) &= \frac{f_{PHM}(x_1;\alpha_1+\alpha_3,\theta)\times f_{PHM}(x_2;\alpha_2,\theta)}{[\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3}\,[\overline{F}_0(x_2;\theta)]^{\alpha_2}} \\ &\equiv \frac{(\alpha_1+\alpha_3)[\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3-1} f_0(x_1;\theta)\alpha_2[\overline{F}_0(x_2;\theta)]^{\alpha_2-1} f_0(x_2;\theta)}{[\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3}\,[\overline{F}_0(x_2;\theta)]^{\alpha_2}} \\ &= (\alpha_1+\alpha_3)\frac{f_0(x_1;\theta)}{\overline{F}_0(x_1;\theta)}\cdot\alpha_2\frac{f_0(x_2;\theta)}{\overline{F}_0(x_2;\theta)} \\ &= (\alpha_1+\alpha_3)\alpha_2 r(x_1;\theta)\, r(x_2;\theta). \end{aligned}$$

For $0 < x_1 < x_2$: analogous computations yield,

$$\begin{aligned} h_{X_1,X_2}(x_1,x_2) &= \frac{f_{PHM}(x_1;\alpha_1,\theta)\times f_{PHM}(x_2;\alpha_2+\alpha_3,\theta)}{[\overline{F}_0(x_1;\theta)]^{\alpha_1}[\overline{F}_0(x_2;\theta)]^{\alpha_2+\alpha_3}} \\ &= \alpha_1\frac{f_0(x_1;\theta)}{\overline{F}_0(x_1;\theta)}(\alpha_2+\alpha_3)\frac{f_0(x_2;\theta)}{\overline{F}_0(x_2;\theta)} \\ &= \alpha_1(\alpha_2+\alpha_3) r(x_1;\theta)\, r(x_2;\theta). \end{aligned}$$

For $0 < x_1 = x_2 \equiv x$: it is got,

$$\begin{aligned} h_{X_1,X_2}(x,x) &= \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} \cdot \frac{f_{PHM}(x;\alpha_1+\alpha_2+\alpha_3,\theta)}{[\overline{F}_0(x;\theta)]^{\alpha_1+\alpha_2+\alpha_3}} \\ &= \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} \cdot \frac{(\alpha_1+\alpha_2+\alpha_3)[\overline{F}_0(x;\theta)]^{\alpha_1+\alpha_2+\alpha_3-1} f_0(x;\theta)}{[\overline{F}_0(x;\theta)]^{\alpha_1+\alpha_2+\alpha_3}} \\ &= \alpha_3 \frac{f_0(x;\theta)}{\overline{F}_0(x;\theta)} \\ &\equiv \alpha_3 r(x;\theta). \end{aligned}$$

Remark 10.3.7 In addition to Proposition 10.3.5, Proposition 10.3.6 provides a further justification of the framework of BPHMs via Definition 10.3.1.

It should be mentioned that the BPHM has both an absolute continuous part and a singular part similar to the Marshall-Olkin bivariate exponential or bivariate Weibull model. In Proposition 10.3.3, the function $f_{X_1,X_2}(\cdot,\cdot)$ is considered to be the joint p.d.f. of the BPHM, if it is understood that the first two terms are densities with respect to the two dimensional Lebesgue measure and the third term is a density function with respect to the one dimensional Lebesgue measure. The next result shows the decomposition of the proposed bivariate proportional hazard model survival function into its constituent absolutely continuous and singular parts.

Proposition 10.3.8 *If* $(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta)$ *with an absolute continuous base-line d.f., then*

$$\overline{F}_{X_1,X_2}(x_1,x_2) = \frac{\alpha_1+\alpha_2}{\alpha_1+\alpha_2+\alpha_3}\,\overline{F}_a(x_1,x_2) + \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}\,\overline{F}_s(x_1,x_2), \quad (10.3.9)$$

where,

$$\begin{aligned} \overline{F}_a(x_1,x_2) &= \frac{\alpha_1+\alpha_2+\alpha_3}{\alpha_1+\alpha_2}[\overline{F}_0(x_1,\theta)]^{\alpha_1}[\overline{F}_0(x_2,\theta)]^{\alpha_2}[\overline{F}_0(z,\theta)]^{\alpha_3} \\ &\quad - \frac{\alpha_3}{\alpha_1+\alpha_2}[\overline{F}_0(z,\theta)]^{\alpha_1+\alpha_2+\alpha_3} \end{aligned} \quad (10.3.10)$$

and

$$\overline{F}_s(x_1,x_2) = [\overline{F}_0(\max(x_1,x_2),\theta)]^{\alpha_1+\alpha_2+\alpha_3} \quad (10.3.11)$$

respectively denote the joint survival function's absolutely continuous and singular parts.

Proof Set $A := \{T_1 \geq T_3\} \cap \{T_2 \geq T_3\}$, where T_1, T_2, T_3 are as in Definition 10.3.1. Then $P(A) = \dfrac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3}$. Therefore,

$$\begin{aligned} \overline{F}_{X_1,X_2}(x_1, x_2) &= P(X_1 > x_1, X_2 > x_2|A)P(A) \\ &\quad + P(X_1 > x_1, X_2 > x_2|A')P(A'). \end{aligned} \tag{10.3.12}$$

Writing $z \equiv= \max(x_1, x_2)$ note that

$$\begin{aligned} A \cap \{X_1 > x_1, X_2 > x_2\} &= \{T_1 \geq T_3, T_2 \geq T_3, \min(T_1, T_3) > x_1, \min(T_2, T_3) > x_2\} \\ &= \{T_1 \geq T_3, T_2 \geq T_3, T_1 > x_1, T_2 > x_2, T_3 > \max(x_1, x_2)\} \\ &= \{T_1 \geq T_3, T_2 \geq T_3, T_3 > z\}. \end{aligned}$$

Since $T_i \geq T_3 > z$ implies $T_i > x_i$ for $i = 1, 2$. Also, since independence of T_1, T_2 and (10.3.1) together implies

$$P(\min(T_1, T_2) > t) = [\overline{F}_0(t; \theta)]^{\alpha_1+\alpha_2}, \quad t > 0. \tag{10.3.13}$$

Then the following is obtained

$$\begin{aligned} P(A \cap \{X_1 > x_1, X_2 > x_2\}) &= P(T_1 \geq T_3, T_2 \geq T_3, T_3 > z) \\ &= P(\min(T_1, T_2) \geq T_3 > z) \\ &= \int_z^\infty P(\min(T_1, T_2) \geq t) f_{T_3}(t) dt \end{aligned} \tag{10.3.14}$$

by independence of T_3 from T_1 and T_2. Using (10.3.12) and the PHM assumption in (10.3.1) for the pdf of T_3, the last step above equals

$$\begin{aligned}\alpha_3 \int_z^\infty [\overline{F}_0(t;\theta)]^{\alpha_1+\alpha_2+\alpha_3-1} f_0(t;\theta)dt &= \alpha_3 \int_0^{\overline{F}_0(z;\theta)} u^{\alpha_1+\alpha_2+\alpha_3-1}du \\ &= \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}[\overline{F}_0(z;\theta)]^{\alpha_1+\alpha_2+\alpha_3} \\ &= P(A)[\overline{F}_0(z;\theta)]^{\alpha_1+\alpha_2+\alpha_3}. \qquad (10.3.15)\end{aligned}$$

Combining (10.3.14) – (10.3.15), the following is obtained

$$\begin{aligned}P(X_1 > x_1, X_2 > x_2|A) &= \frac{P(A \cap \{X_1 > x_1, X_2 > x_2\})}{P(A)} \\ &= [\overline{F}_0(z,\theta)]^{\alpha_1+\alpha_2+\alpha_3}.\end{aligned}$$

This is obviously the singular part of the joint survival function. The other factor $P(X_1 > x_1, X_2 > x_2|A)$ in the decomposition (10.3.12) can be obtained by subtracting the second term in (10.3.12) from its left hand side, and corresponds to the absolutely continuous part of $\overline{F}_{X_1,X_2}(x_1, x_2)$.

The joint p.d.f. of (X_1, X_2) now follows using Proposition 10.3.8 and can be decomposed as a mixture of absolutely continuous and singular components;

$$f_{X_1,X_2}(x_1,x_2) = \frac{\alpha_1+\alpha_2}{\alpha_1+\alpha_2+\alpha_3} f_a(x_1,x_2) + \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3} f_s(z), \qquad (10.3.16)$$

where $z = \max(x_1, x_2)$,

$$\begin{aligned}f_a(x_1,x_2) = &\frac{\alpha_1+\alpha_2+\alpha_3}{\alpha_1+\alpha_2} \\ &\times \begin{cases} f_{PHM}(x_1;\alpha_1+\alpha_3,\theta) \times f_{PHM}(x_2;\alpha_2,\theta), & \text{if } 0 < x_2 < x_1 < \infty \\ f_{PHM}(x_1;\alpha_1,\theta) \times f_{PHM}(x_2;\alpha_2+\alpha_3,\theta), & \text{if } 0 < x_1 < x_2 < \infty \end{cases}\end{aligned} \qquad (10.3.17)$$

and

$$f_s(x_1,x_2) = f_{PHM}(z;\alpha_1+\alpha_2+\alpha_3,\theta). \qquad (10.3.18)$$

Some further additional useful consequences of BPHM (Proposition 10.3.9 and Proposition 10.3.10), which easily follow, are as follows.

Proposition 10.3.9 *If* $(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta)$, *then*

$$P(X_1 < X_2) = \frac{\alpha_2}{\alpha_1 + \alpha_2 + \alpha_3}, \qquad P(X_2 < X_1) = \frac{\alpha_1}{\alpha_1 + \alpha_2 + \alpha_3},$$
$$P(X_1 = X_2) = \frac{\alpha_3}{\alpha_1 + \alpha_2 + \alpha_3}.$$

Proposition 10.3.10 *Let* $(X_1, X_2) \sim BPHM(\alpha_1, \alpha_2, \alpha_3, \theta)$. *Suppose the baseline d.f.* F_0 *is absolutely continuous. Then the conditional survival function of* X_1 *given* $X_2 > x_2$, *say* $\overline{F}_{X_1|X_2>x_2}(x_1)$ *is an absolute continuous survival function as follows:*

$$\begin{aligned}\overline{F}_{X_1|X_2>x_2}(x_1) &\equiv P(X_1 > x_1 | X_2 > x_2) \\ &= \begin{cases} [\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3} \cdot [\overline{F}_0(x_2;\theta)]^{-\alpha_3}, & \text{if } 0 < x_2 < x_1 < \infty \\ [\overline{F}_0(x_1;\theta)]^{\alpha_1}. & \text{if } 0 < x_2 < x_1 < \infty \end{cases}\end{aligned} \tag{10.3.19}$$

Note, by right continuity of cdfs, we can have the following,

$$\overline{F}_{X_1|X_2>x_2}(x_2) := \overline{F}_{X_1|X_2>x_2}(x_2+) = [\overline{F}_0(x_2;\theta)]^{\alpha} = \overline{F}_{X_1|X_2>x_2}(x_2-),$$

so that $\overline{F}_{X_1|X_2>x_2}(\cdot)$ is continuous at x_2. Along that the absolutely continuity of the baseline d. f. F_0, this proves $\overline{F}_{X_1|X_2>x_2}(\cdot)$ is also absolutely continuous for each $x_2 > 0$ and hence admits a

The conditional distributions $F_{X_1|X_2>x_2}(\cdot)$ **vs.** $F_{X_1|X_2}(\cdot|x_2)$

It is institutive to construct the conditional distribution of X_1 given $\{X_2 > x_2\}$ vs. the conditional distribution of X_1 when X_2 has an observed value x_2. Denote the corresponding c.d.fs by $F_{X_1|X_2>x_2}(\cdot)$ and $F_{X_1|X_2}(\cdot|x_2)$ respectively. The former if the complement of the one-dimensional survival (tail) probability $\overline{F}_{X_1|X_2.x_2}$, developed in Theorems and , which the latter is the *traditional* conditional cdf of X_1 when the value x_2 assumed by the other associated future time X_2 is known.

For fixed $x_2 > 0$,

$$\overline{F}_{X_1|X_2}(x_1|x_2) = \int_{x_1}^{\infty} f_{X_1|X_2}(u|x_2)du, \qquad x_1 > 0,$$

where, the corresponding conditional density (pdf)

$$f_{X_1|X_2}(x_1|x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_2}(x_2)}, \tag{10.3.20}$$

can be computed using Proposition 10.3.6 and Proposition 10.3.5, to get:

$$f_{X_1|X_2}(x_1|x_2) = \begin{cases} \dfrac{f_{PHM}(x_1;\alpha_1+\alpha_3,\theta)f_{PHM}(x_2;\alpha_2,\theta)}{f_{PHM}(x_2;\alpha_2+\alpha_3,\theta)}, & \text{if } 0 < x_2 < x_1 < \infty \\ \dfrac{f_{PHM}(x_1;\alpha_1,\theta)f_{PHM}(x_2;\alpha_2+\alpha_3,\theta)}{f_{PHM}(x_2;\alpha_2+\alpha_3,\theta)}, & \text{if } 0 < x_1 < x_2 < \infty \end{cases}$$

$$= \begin{cases} \dfrac{(\alpha_1+\alpha_3)[\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3-1}f_0(x_1;\theta)\alpha_2[\overline{F}_0(x_2;\theta)]^{\alpha_2-1}}{(\alpha_2+\alpha_3)[\overline{F}_0(x_2;\theta)]^{\alpha_2+\alpha_3-1}}, & \text{if } 0 < x_2 < x_1 < \infty \\ \alpha_1[\overline{F}_0(x_1;\theta)]^{\alpha_1-1}f_0(x_1;\theta), & \text{if } 0 < x_1 < x_2 < \infty \end{cases}$$

Hence, the conditional density of X_1 when X_2 has value x_2, can be expressed

$$f_{X_1|X_2}(x_1|x_2)$$

$$= \begin{cases} \alpha_2\dfrac{\alpha_1+\alpha_3}{\alpha_2+\alpha_3}[\overline{F}_0(x_1;\theta)]^{\alpha_1+\alpha_3-1}[\overline{F}_0(x_2;\theta)]^{-\alpha_3}f_0(x_1;\theta), & \text{if } 0 < x_2 < x_1 < \infty \\ \alpha_1[\overline{F}_0(x_1;\theta)]^{\alpha_1-1}f_0(x_1;\theta), & \text{if } 0 < x_1 < x_2 < \infty \\ 0, & \text{if } 0 < x_1 = x_2 < \infty \end{cases}$$

Remark 10.3.11 Note that along the diagonal $x_1 = x_2$, the conditional density, according to (10.3.20), can be strictly positive; viz., if $x_1 = x_2 \equiv x > 0$,

$$\begin{aligned} f_{X_1|X_2}(x|x) &= \frac{f_{X_1,X_2}(x,x)}{f_{X_2}(x)} \\ &= \frac{\dfrac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}f_{PHM}(x;\alpha_1+\alpha_2+\alpha_3,\theta)}{f_{PHM}(x;\alpha_2+\alpha_3,\theta)} \\ &= \frac{\alpha_3}{\alpha_2+\alpha_3}[\overline{F}_0(x;\theta)]^{\alpha_1}, \qquad \text{using (10.2.2)} \end{aligned}$$

can be strictly positive if $F_0(x;\theta) < 1$, all x, i.e., if the baseline d.f. F_0 has support unbounded to the right, but contributes no mass to the conditional pdf, since

$$\int_{\{x_1:x_1=x_2\}} f_{X_1|X_2}(x_1|x_2)dx_1 = \int_{\{x_2\}} f_{X_1|X_2}(x_1|x_2)dx_1 = 0.$$

Hence, for $x > 0$, $f_{X_1|X_2}(x|x)$ can be defined arbitrarily and hence can be set to zero.

10.4 Parameter Estimators: Maximum Likelihood

Suppose $\boldsymbol{x} = ((x_{11}, x_{21}), \cdots, (x_{1n}, x_{2n}))$ are the observations in a random sample from BPHM$(\alpha_1, \alpha_2, \alpha_3, \theta)$, the distribution of (X_1, X_2). To compute the maximum likelihood estimators (MLEs) of the parameters, define:

$$J_0 = \{j; x_{1j} = x_{2j} = x_j\},\ J_1 = \{j; x_{1j} < x_{2j}\},\ J_2 = \{j; x_{1j} > x_{2j}\},\ J = J_0 \cup J_1 \cup J_2.$$

The cardinalities of the sets are, respectively,

$$|J_i| = n_i, \quad \text{for} \quad i = 0, 1, 2 \quad \text{and} \quad n = n_0 + n_1 + n_2.$$

Using (10.2.2) and (10.3.6), the log-likelihood function can be expressed as

$$\begin{aligned}
&l(\alpha_1, \alpha_2, \alpha_3, \theta|((x_{11}, x_{12}), \cdots, (x_{1n}, x_{2n}))) \\
&= n_1 \ln \alpha_1 + (\alpha_1 - 1) \sum_{j\in J_1} \ln \overline{F}_0(x_{1j};\theta) + \sum_{j\in J_1} \ln f_0(x_{1j};\theta) \\
&+ n_1 \ln(\alpha_2 + \alpha_3) + (\alpha_2 + \alpha_3 - 1) \sum_{j\in J_1} \ln \overline{F}_0(x_{2j};\theta) + \sum_{j\in J_1} \ln f_0(x_{2j};\theta) \\
&+ n_2 \ln(\alpha_1 + \alpha_3) + (\alpha_1 + \alpha_3 - 1) \sum_{j\in J_2} \ln \overline{F}_0(x_{1j};\theta) + \sum_{j\in J_2} \ln f_0(x_{1j};\theta) \\
&+ n_2 \ln \alpha_2 + (\alpha_2 - 1) \sum_{j\in J_2} \ln \overline{F}_0(x_{2j};\theta) + \sum_{j\in J_2} \ln f_0(x_{2j};\theta) \\
&+ n_0 \ln \alpha_3 + (\alpha_1 + \alpha_2 + \alpha_3 - 1) \sum_{j\in J_0} \ln \overline{F}_0(x_j;\theta) + \sum_{j\in J_0} \ln f_0(x_j;\theta). \qquad (10.4.1)
\end{aligned}$$

For the sake of brevity, write $l(\alpha_1, \alpha_2, \alpha_3, \theta|(x_{11}, x_{12}), \cdots, (x_{1n}, x_{2n}))$ as $l(\alpha_1, \alpha_2, \alpha_3, \theta|data)$. Now consider two cases.

Case (1). θ is known: In this case the baseline distribution function F_0 is completely known and hence the baseline survival function $\overline{F}_0 = 1 - F_0$ is completely known. The only unknown parameters are α_1, α_2 and α_3. By differentiating the log-likelihood function (10.4.1) with respect to α_1, α_2 and α_3 respectively, the following can be obtained:

$$\begin{pmatrix} \dfrac{\partial l(\alpha_1,\alpha_2,\alpha_3,\theta|data)}{\partial \alpha_1} \\ \dfrac{\partial l(\alpha_1,\alpha_2,\alpha_3,\theta|data)}{\partial \alpha_2} \\ \dfrac{\partial l(\alpha_1,\alpha_2,\alpha_3,\theta|data)}{\partial \alpha_3} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{n_1}{\alpha_1} + \sum\limits_{j\in J_1} \ln \overline{F}_0(x_{1j};\theta) + \dfrac{n_2}{\alpha_1+\alpha_3} + \sum\limits_{j\in J_2} \ln \overline{F}_0(x_{1j};\theta) + \sum\limits_{j\in J_0} \ln \overline{F}_0(x_j;\theta) \\ \dfrac{n_1}{\alpha_2+\alpha_3} + \sum\limits_{j\in J_1} \ln \overline{F}_0(x_{2j};\theta) + \dfrac{n_2}{\alpha_2} + \sum\limits_{j\in J_2} \ln \overline{F}_0(x_{2j};\theta) + \sum\limits_{j\in J_0} \ln \overline{F}_0(x_j;\theta) \\ \dfrac{n_1}{\alpha_2+\alpha_3} + \sum\limits_{j\in J_1} \ln \overline{F}_0(x_{2j};\theta) + \dfrac{n_2}{\alpha_1+\alpha_3} + \sum\limits_{j\in J_2} \ln \overline{F}_0(x_{1j};\theta) + \sum\limits_{j\in J_0} \ln \overline{F}_0(x_j;\theta) \end{pmatrix}. \tag{10.4.2}$$

Set the left hand side of (10.4.2) to $\mathbf{0} = (0,0,0)^T$ to obtain the three normal equations as:

$$\frac{n_1}{\alpha_1} + \frac{n_2}{\alpha_1+\alpha_3} = -\sum_{J_1\cup J_2} \ln \overline{F}_0(x_{1j};\theta) - \sum_{J_0} \ln \overline{F}_0(x_j;\theta), \tag{10.4.3a}$$

$$\frac{n_2}{\alpha_2} + \frac{n_1}{\alpha_2+\alpha_3} = -\sum_{J_1\cup J_2} \ln \overline{F}_0(x_{2j};\theta) - \sum_{J_0} \ln \overline{F}_0(x_j;\theta), \tag{10.4.3b}$$

$$\frac{n_0}{\alpha_3} + \frac{n_1}{\alpha_2+\alpha_3} + \frac{n_2}{\alpha_1+\alpha_3} = -\sum_{J_1} \ln \overline{F}_0(x_{2j};\theta) - \sum_{J_2} \ln \overline{F}_0(x_{1j};\theta) - \sum_{J_0} \ln \overline{F}_0(x_j;\theta). \tag{10.4.3c}$$

It follows that for fixed $\alpha_3 > 0$, the MLEs of $\hat{\alpha}_1(\alpha_3)$ and $\hat{\alpha}_2(\alpha_3)$ are:

$$\hat{\alpha}_1(\alpha_3) = \frac{(n_1+n_2-a_1\alpha_3) + \sqrt{(a_1\alpha_3-n_1-n_2)^2 + 4a_1n_1\alpha_3}}{2a_1}, \tag{10.4.4}$$

$$\hat{\alpha}_2(\alpha_3) = \frac{(n_1+n_2-a_2\alpha_3) + \sqrt{(a_2\alpha_3-n_1-n_2)^2 + 4a_2n_2\alpha_3}}{2a_2}, \tag{10.4.5}$$

where,

$$a_1 := -\sum_{J_1 \cup J_2} \ln \overline{F}_0(x_{1j};\theta) - \sum_{J_0} \ln \overline{F}_0(x_j;\theta),$$
$$a_2 := -\sum_{J_1 \cup J_2} \ln \overline{F}_0(x_{2j};\theta) - \sum_{J_0} \ln \overline{F}_0(x_j;\theta).$$

By plugging in (10.4.4) – (10.4.5) into (10.4.3c), the MLE of α_3 can be obtained, which can be seen to be a solution of equation

$$h(\alpha_3) = \alpha_3, \tag{10.4.6}$$

where,

$h(\alpha_3) :=$

$$-\frac{n_0}{\sum_{J_1} \ln \overline{F}_0(x_{2j};\theta) + \sum_{J_2} \ln \overline{F}_0(x_{1j};\theta) + \sum_{J_0} \ln \overline{F}_0(x_j;\theta) + \frac{n_1}{\hat{\alpha_2}(\alpha_3)+\alpha_3} + \frac{n_2}{\hat{\alpha_1}(\alpha_3)+\alpha_3}}, \tag{10.4.7}$$

i.e., α_3 is the fixed point of the real valued function $h(\cdot)$. Any suitable numerical procedure can be used to solve (10.4.7) iteratively.

Case (2). θ is unknown: Here instead of computing the MLEs directly, EM algorithm will be used. Treating this as a missing value problem.

Corresponding to the bivariate survival times (X_1, X_2) which follows a BPHM model (10.3.2), define an associated random vector (Δ_1, Δ_2) by:

$$\Delta_1 = \begin{cases} 1, & \text{if } T_1 < T_3 \\ 3, & \text{if } T_1 > T_3 \end{cases}, \qquad \Delta_2 = \begin{cases} 2, & \text{if } T_2 < T_3 \\ 3, & \text{if } T_2 > T_3 \end{cases}.$$

Given (X_1, X_2), it is clear that the associated (Δ_1, Δ_2) is not always known. Only when $X_1 = X_2$ (and hence $\Delta_1 = \Delta_2 = 3$), (Δ_1, Δ_2) is completely known. When $X_1 < X_2$ or $X_1 > X_2$; the associated (Δ_1, Δ_2) is missing.

The 'E'-step and 'M'-step of EM algorithm can now be described.

E-step: The observations belong to J_0 are treated as the complete observations, and the observation is treated as a missing observation if (x_1, x_2) belongs to J_1 or J_2. If $(x_1, x_2) \in J_1$, similarly as in [29], the 'pseudo observation' is formed by fractionizing (x_1, x_2) to two partially complete observation of the form $(x_1, x_2, u_1(\gamma))$ and $(x_1, x_2, u_2(\gamma))$. Here $u_1(\gamma) = P((\Delta_1, \Delta_2) = (1,2)|X_1 < X_2)$ and $u_2(\gamma) = P((\Delta_1, \Delta_2) = (1,3)|X_1 < X_2)$. Similarly, for $(x_1, x_2) \in J_2$, the 'pseudo observation' of the form $(x_1, x_2, v_1(\gamma))$ and $(x_1, x_2, v_2(\gamma))$ are formed. Here $v_1(\gamma) = P((\Delta_1, \Delta_2) = (1,2)|X_1 > X_2)$ and $v_2(\gamma) = P((\Delta_1, \Delta_2) = (3,2)|X_1 > X_2)$. Since

$$P(T_1 < T_2 < T_3) = \frac{\alpha_1\alpha_2}{(\alpha_2+\alpha_3)(\alpha_1+\alpha_2+\alpha_3)},$$
$$P(T_1 < T_3 < T_2) = \frac{\alpha_1\alpha_3}{(\alpha_2+\alpha_3)(\alpha_1+\alpha_2+\alpha_3)},$$

the following can be had:

$$\begin{aligned} u_1(\gamma) &= P((\Delta_1, \Delta_2) = (1,2)|X_1 < X_2) = P(T_1 < T_3, T_2 < T_3|X_1 < X_2) \\ &= \frac{P(T_1 < T_2 < T_3)}{P(X_1 < X_2)} \\ &= \frac{\alpha_1}{\alpha_2+\alpha_3}, \end{aligned}$$

and

$$\begin{aligned} u_2(\gamma) &= P((\Delta_1, \Delta_2) = (1,3)|X_1 < X_2) = P(T_1 < T_3, T_3 < T_2|X_1 < X_2) \\ &= \frac{P(T_1 < T_3 < T_2)}{P(X_1 < X_2)} \\ &= \frac{\alpha_1\alpha_3}{\alpha_2(\alpha_2+\alpha_3)}. \end{aligned}$$

Similarly, since

$$P(T_2 < T_1 < T_3) = \frac{\alpha_1\alpha_2}{(\alpha_1+\alpha_3)(\alpha_1+\alpha_2+\alpha_3)},$$
$$P(T_2 < T_3 < T_1) = \frac{\alpha_2\alpha_3}{(\alpha_1+\alpha_3)(\alpha_1+\alpha_2+\alpha_3)},$$

the following can be hold:

$$\begin{aligned} v_1(\gamma) &= P((\Delta_1,\Delta_2)=(1,2)|X_1>X_2) = P(T_1<T_3, T_2<T_3|X_1>X_2) \\ &= \frac{P(T_2<T_1<T_3)}{P(X_1>X_2)} \\ &= \frac{\alpha_2}{\alpha_1+\alpha_3}, \end{aligned}$$

and

$$\begin{aligned} v_2(\gamma) &= P((\Delta_1,\Delta_2)=(3,2)|X_1>X_2) = P(T_3<T_1, T_2<T_3|X_1>X_2) \\ &= \frac{P(T_2<T_3<T_1)}{P(X_1>X_2)} \\ &= \frac{\alpha_2\alpha_3}{\alpha_1(\alpha_1+\alpha_3)}. \end{aligned}$$

Accordingly, the log-likelihood function of the 'pseudo data' is,

$$\begin{aligned} & l_{pseudo}(\alpha_1,\alpha_2,\alpha_3,\theta) \\ & = n_0 \ln\alpha_3 + (\alpha_1+\alpha_2+\alpha_3-1)\sum_{j\in J_0}\ln\overline{F}_0(x_j;\theta) + \sum_{j\in J_0}\ln f_0(x_j;\theta) \\ & + u_1(\gamma)[n_1\ln\alpha_2 + (\alpha_2+\alpha_3-1)\sum_{j\in J_1}\ln\overline{F}_0(x_{2j};\theta) + \sum_{j\in J_1}\ln f_0(x_{2j};\theta)] \\ & + u_2(\gamma)[n_1\ln\alpha_3 + (\alpha_2+\alpha_3-1)\sum_{j\in J_1}\ln\overline{F}_0(x_{2j};\theta) + \sum_{j\in J_1}\ln f_0(x_{2j};\theta)] \\ & + n_1\ln\alpha_1 + (\alpha_1-1)\sum_{j\in J_1}\ln\overline{F}_0(x_{1j};\theta) + \sum_{j\in J_1}\ln f_0(x_{1j};\theta) \\ & + v_1(\gamma)[n_2\ln\alpha_1 + (\alpha_1+\alpha_3-1)\sum_{j\in J_2}\ln\overline{F}_0(x_{1j};\theta) + \sum_{j\in J_2}\ln f_0(x_{1j};\theta)] \\ & + v_2(\gamma)[n_2\ln\alpha_3 + (\alpha_1+\alpha_3-1)\sum_{j\in J_2}\ln\overline{F}_0(x_{1j};\theta) + \sum_{j\in J_2}\ln f_0(x_{1j};\theta)] \\ & + n_2\ln\alpha_2 + (\alpha_2-1)\sum_{j\in J_2}\ln\overline{F}_0(x_{2j};\theta) + \sum_{j\in J_2}\ln f_0(x_{2j};\theta). \end{aligned}$$

M-step: Maximize $l_{pseudo}(\alpha_1,\alpha_2,\alpha_3,\theta)$ with respect to α_1, α_2, α_3 and θ at each step.

For fixed θ, it is hold that $\arg\max_{(\alpha_1,\alpha_2,\alpha_3)} l_{pseudo}(\alpha_1,\alpha_2,\alpha_3,\theta) = (\hat{\alpha}_1(\theta),\hat{\alpha}_2(\theta),\hat{\alpha}_3(\theta))$, where

$$\hat{\alpha}_1(\theta) = -\frac{n_1 + n_2 v_1(\gamma)}{\sum_{j\in J_0} \ln \overline{F}_0(x_j;\theta) + \sum_{j\in J_1} \ln \overline{F}_0(x_{1j};\theta) + (v_1+v_2)\sum_{j\in J_2} \ln \overline{F}_0(x_{1j};\theta)},$$

$$\hat{\alpha}_2(\theta) = -\frac{n_2 + n_1 u_1(\gamma)}{\sum_{j\in J_0} \ln \overline{F}_0(x_j;\theta) + (u_1+u_2)\sum_{j\in J_1} \ln \overline{F}_0(x_{2j};\theta) + \sum_{j\in J_2} \ln \overline{F}_0(x_{2j};\theta)},$$

$$\hat{\alpha}_3(\theta) = -\frac{n_0 + n_1 u_2(\gamma) + n_2 v_2(\gamma)}{\sum_{j\in J_0} \ln \overline{F}_0(x_j;\theta) + (u_1+u_2)\sum_{j\in J_1} \ln \overline{F}_0(x_{2j};\theta) + (v_1+v_2)\sum_{j\in J_2} \ln \overline{F}_0(x_{1j};\theta)}.$$

Finally the maximized $\hat{\theta}$ is obtained as $\hat{\theta} = \arg\max_{\theta} l_{pseudo}(\hat{\alpha}_1,\hat{\alpha}_2,\hat{\alpha}_3,\theta)$.

To describe the iterative step of the EM algorithm, suppose that $({\alpha_1}^{(i)}, {\alpha_2}^{(i)}, {\alpha_3}^{(i)}, \theta^{(i)})$ is the estimates of $(\alpha_1,\alpha_2,\alpha_3,\theta)$ at the i-th step. Then the estimate in the $(i+1)$-th step is obtained using the following version of the EM algorithm known as Expectation-Conditional-Maximization (ECM) algorithm:

Step 1: Compute $u_1(\gamma)$, $u_2(\gamma)$, $v_1(\gamma)$ and $v_2(\gamma)$ using ${\alpha_1}^{(i)}$, ${\alpha_2}^{(i)}$ and ${\alpha_3}^{(i)}$.

Step 2: Compute $\theta^{(i+1)}$ by maximizing $l_{pseudo}(\hat{\alpha}_1(\theta),\hat{\alpha}_2(\theta),\hat{\alpha}_3(\theta),\theta)$.

Step 3: Compute ${\alpha_1}^{(i+1)} = \hat{\alpha}_1(\theta^{(i+1)})$, ${\alpha_2}^{(i+1)} = \hat{\alpha}_2(\theta^{(i+1)})$ and ${\alpha_3}^{(i+1)} = \hat{\alpha}_3(\theta^{(i+1)})$ with obtained $\theta^{(i+1)}$.

Iterations continue until the convergence criterion is satisfied.

10.5 Illustrative Applications of the Bivariate Proportional Hazard Models

10.5.1 Examples of Bivariate Proportional Hazard Model (BPHM) Families

We will examine the following models as examples of parametric families of BPHMs:

Bivariate Generalized Exponential (BGE),

Bivariate Exponentiated Weibull (BEW),

Bivariate Linear Failure Rate (BLFR);

which are respectively specified via the joint survival function representation in Proposition 10.3.2, using *univariate* Exponential (E), Weibull (W) and Linear Failure Rate (LFR) baseline distributions of the latent variables T_i, $i = 1, 2, 3$ of Definition 10.3.1. The *univariate* survival distribution of these baseline distributions, are of course:

Exponential $E(\lambda) : \overline{F}_0(x; \lambda) = e^{-\lambda x}$; $\quad \lambda > 0$

Weibull $W(\lambda, \beta) : \overline{F}_0(x; \lambda, \beta) = e^{-\lambda x^\beta}$; $\quad \lambda > 0, \beta > 0$

Linear Failure Rate $LFR(\lambda, \beta) : \overline{F}_0(x; \lambda, \beta) = e^{-\lambda(x+\beta x^2)}$; $\quad \lambda > 0, \beta > 0$

If we denote the corresponding *univariate* proportional hazard model (PHM) families by 'Generalized Exponential' (GE), 'Exponentiated Weibull' (EW), and 'Generalized Linear Failure Rate' (GLFR) distributions, then from their respective survival functions expressed via (10.2.1); it is clear that

$GE(\alpha, \lambda) \equiv E(\alpha\lambda)$,

$EW(\alpha, \lambda, \beta) \equiv W(\alpha\lambda, \beta)$,

$GLFR(\alpha, \lambda, \beta) \equiv LFR(\alpha\lambda, \beta)$;

so that in the univariate case, the above PHM families do *not* lead to new distribution models, and are *not identifiable* in the sense different combinations (α_i, λ_i) can lead to the same model, *viz.*,

$GE(\alpha_1, \lambda_1) = GE(\alpha_2, \lambda_2) \equiv E(\lambda^*)$,

$EW(\alpha_1, \lambda_1, \beta) = EW(\alpha_2, \lambda_2, \beta) \equiv W(\lambda^*, \beta)$,

$GLFR(\alpha_1, \lambda_1, \beta) = GLFR(\alpha_2, \lambda_2, \beta) \equiv LFR(\lambda^*, \beta)$;

where $\alpha_1\lambda_1 = \alpha_2\lambda_2 \equiv \lambda^*$. However, the identifiability issue ceases to exist for the bivariate proportional hazard model (BPHM) families that we can construct using the above baseline univariate distributions, as can be seen by writing out the corresponding joint survival functions. This leads us to define the following BPHM parametric families; in accordance with Definition 10.3.1:

- $BGE(\alpha_1, \alpha_2, \alpha_3, \lambda)$, *Bivariate Generalized Exponential*, defined by

$$(X_1, X_2) \stackrel{d}{=} (T_1 \wedge T_3, T_2 \wedge T_3), \text{ where}$$
$$T_i \sim E(\lambda\alpha_i); \qquad i = 1, 2, 3.$$

Here, and below, $\stackrel{d}{=}$ denotes equivalence in distribution, and $\sim$ stands for 'distributed as'.

- $BEW(\alpha_1, \alpha_2, \alpha_3, \lambda, \beta)$, *Bivariate Exponentiated Weibull*, defined by

$$(X_1, X_2) \stackrel{d}{=} (T_1 \wedge T_3, T_2 \wedge T_3), \text{ where}$$
$$T_i \sim W(\lambda\alpha_i, \beta); \qquad i = 1, 2, 3.$$

- $BLFR(\alpha_1, \alpha_2, \alpha_3, \lambda, \beta)$, *Bivariate Linear Failure Rate*, defined by

$$(X_1, X_2) \stackrel{d}{=} (T_1 \wedge T_3, T_2 \wedge T_3), \text{ where}$$
$$T_i \sim LFR(\lambda\alpha_i, \beta); \qquad i = 1, 2, 3.$$

For example, the family $BEW(\alpha_1, \alpha_2, \alpha_3, \lambda, \beta)$ distributions, in virtue of Proposition 10.3.3, specified by the joint density function

$$f_{X_1,X_2}(x_1, x_2) =$$
$$\begin{cases} (\alpha_1 + \alpha_3)\alpha_2(\lambda\beta)^2(x_1x_2)^{\beta-1}e^{-\lambda((\alpha_1+\alpha_3)x_1^\beta + \alpha_2 x_2^\beta)}, & \text{if } x_2 < x_1 \\ \alpha_1(\alpha + 2 + \alpha_3)(\lambda\beta)^2(x_1x_2)^{\beta-1}e^{-\lambda(\alpha_1 x_1^\beta + (\alpha_2+\alpha_3)x_2^\beta)}, & \text{if } x_1 < x_2 \\ \alpha_3\lambda\beta x^{\beta-1}e^{-\lambda(\alpha_1+\alpha_2+\alpha_3)x^\beta}, & \text{if } x_1 = x_2 = x \end{cases} \tag{10.5.1}$$

a five-parameters family on $(0, \infty)^2$.

10.5.2 Application to Data Sets

For the purposes of illustration, we analyze two data sets:

Data Set 1. This data set (Table 10.1) was first published in 'Washington Post' and available in [21], which describes goal timings of the National Football League (NFL) matches in three consecutive weekends. In this bivariate data set for (X_1, X_2), the variable X_1 denotes the game time (in minutes and seconds) to the first points scored by kicking the ball between goal posts; similarly, X_2 denotes the game time by moving (in minutes and seconds) the ball into the end zone. The events described via X_1 and X_2 have the following meaning: (i) $X_1 < X_2$ means that the first score is a field goal, (ii) $X_1 = X_2$ means the first score is a converted touchdown and (iii) $X_1 > X_2$ means that the first score is an unconverted touchdown.

Table 10.1 National Football League (NFL) Data Obtained from the Matches on Three Consecutive Weekend in 1986

X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2	X_1	X_2
2:03	3:59	5:47	25:59	10:24	14:15	14:32	20:34	8:52	8:52	2:54	2:54
9:03	9:03	13:48	49:45	2:59	2:59	10:51	38:04	17:50	17:50	7:14	9:41
0:51	0:51	7:15	7:15	3:53	6:26	31:08	49:53	10:09	10:09	14:35	14:35
3:26	3:26	4:15	4:15	0:45	0:45	8:32	14:34	8:59	8:59	11:49	11:49
7:47	7:47	1:39	1:39	11:38	17:22	19:39	10:42	5:31	11:16	12:08	12:08
10:34	14:17	6:25	15:05	1:23	1:23	32:27	42:21	6:25	6:25	15:32	15:32
7:03	7:03	4:13	9:29	10:21	10:21	6:51	34:35	7:01	7:01	2:35	2:35

Data Set 2. This data set (Table 10.2) is from the record of the IX FINA World Cup diving competition, held in Atlanta, Georgia in 1995. The data were obtained from [25, 54] and show the *scores given by the seven judges* who were from the seven different countries. In this bivariate data set for (X_1, X_2), the variables X_1 (X_2, respectively) represents the maximum score given by a judge from Asia or Caucasus (from the Western countries, respectively). Here, (i) $X_1 < X_2$ means that the maximum score for the diver is given by a Western judge, (ii) $X_1 = X_2$ means

that the maximum score for the diver is given by a judge from Asia/Caucasus and the Western – is a tie, while (iii) $X_1 > X_2$ means that maximum score was assigned by a judge from Asia/Caucasus.

Table 10.2 Dive Data Obtained from the IX FINA World Cup diving competition, held in Atlanta, Georgia in 1995

Diver	Country	X_1	X_2	Diver	Country	X_1	X_2
Sun Shuwei	China	19	19	Sun Shuwei	China	15	16
David Pichler	USA	15	15	David Pichler	USA	15	15
Jan Hempel	Germany	13	14	Jan Hempel	Germany	17	18
Roman Volodkuv	Ukraine	11	12	Roman Volodkuv	Ukraine	16	16
Sergei Kudrevich	Belarus	14	14	Sergei Kudrevich	Belarus	12	13
Patrick Jeffrey	USA	15	14	Patrick Jeffrey	USA	14	14
Valdimir Timoshinin	Russia	13	16	Valdimir Timoshinin	Russia	12	13
Dimitry Sautin	Russia	7	5	Dimitry Sautin	Russia	17	18
Xiao Hailiang	China	13	13	Xiao Hailiang	China	9	10
Sun Shuwei	China	18	18				

10.5.3 Numerical Results

We fit three different bivariate proportional hazard models, namely (i) bivariate generalized exponential (BGE), (ii) bivariate exponentiated Weibull (BEW) and (iii) bivariate linear failure rate (BLFR), to each of the two data sets described in Section 10.5.2. The maximum likelihood estimates of model parameters for all three bivariate proportional hazard models were obtained using EM algorithm.

10.5.3.1 Data Set 1: NFL Data. We analyze the NFL data set using the three proposed bivariate proportional hazard models. First the data (shown in Table 10.1) is reset with minutes as unit of time as in [21, 53], e.g., 2:03 is converted to 2.05 minutes. Although in reality, the possible scoring times are restricted by the duration

of the game, it is ignored for modeling purposes as in [21, 53], by considering time to goals a non-negative random variable.

The maximum likelihood estimates of the parameters for three different bivariate proportional hazard models are summarized in Table 10.3, from which we can conclude that BEW model fits NFL data set better than both BGE and BLFR models.

10.5.3.2 Data Set 2: Dive Data. We analyze the Dive data set (shown in Table 10.2), also using the three proposed bivariate proportional hazard models listed above. The maximum likelihood estimates of the parameters for three models are summarized in Table 10.4. From the log-likelihood (LL) values, as well as from the values of the AIC, BIC criteria of model fit shown in Table 10.4, it is seen that the BEW as a bivariate proportional hazard model fits the dive data best among the three candidates.

Table 10.3 The Maximum Likelihood Estimators (MLE) and Corresponding 95% Confidence Intervals, the Corresponding Log-likelihood (LL) Values, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for Three Different Bivariate Proportional Hazard Models for *NFL Data Set*

		M O D E L P A R A M E T E R S							
Models		λ	β	α_1	α_2	α_3	LL	AIC	BIC
BGE	MLE	8.3479	$--$	0.0893	0.7658	1.8996	-29.88	67.76	74.71
	95% CI	(5.9069, 10.7889)	$--$	(0.0758, 0.1028)	(0.6147, 0.9169)	(1.3468, 2.4324)			
BEW*	MLE	6.5339	0.5638	0.4877	4.6872	7.9636	-18.28	46.56	55.25
	95% CI	(4.8623, 8.2055)	(0.2866, 0.8410)	(0.2566, 0.7188)	(1.6984, 7.6760)	(5.5563, 10.3709)			
BLFR	MLE	5.5966	1.2876	0.0644	0.5866	1.5635	-24.89	59.78	68.47
	95% CI	(2.2677, 8.9255)	(0.6836, 1.8916)	(0.0489, 0.0799)	(0.4048, 0.7684)	(1.1921, 1.9349)			

* The BEW model provides the best fit to NFL data set among the BGE, BEW and BLFR models.

Table 10.4 The Maximum Likelihood Estimators (MLE) and Corresponding Associates 95% Confidence Intervals, the Corresponding Log-likelihood (LL) Values, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for Three Different Bivariate Proportional Hazard Models for *Dive Data Set*

		M O D E L P A R A M E T E R S							
Models		λ	β	α_1	α_2	α_3	LL	AIC	BIC
BGE	MLE	5.6349	$--$	0.0634	0.8944	1.5633	-19.88	47.76	51.54
	95% CI	(3.1934, 8.3479)	$--$	(0.0458, 0.0810)	(0.6566, 1.1322)	(1.1988, 1.9278)			
BEW*	MLE	5.8876	0.3489	0.6955	5.5633	9.5488	-17.59	45.18	49.90
	95% CI	(3.6422, 8.1330)	(0.2367, 0.4611)	(0.4866, 0.9044)	(2.0642, 9.0624)	(7.1643, 11.9333)			
BLFR	MLE	6.8934	1.7855	0.1126	0.8622	2.0164	-20.56	51.12	55.84
	95% CI	(3.7688, 10.0180)	(1.2066, 2.3644)	(0.0755, 0.1497)	(0.5988, 1.1256)	(1.3674, 2.6654)			

* The BEW model provides the best fit to dive data set among the BGE, BEW and BLFR models.

With the maximum likelihood estimates for model parameters shown in Table 10.4, we can also compute the plug-in estimates of probabilities: (i) $P(X_1 > X_2)$, (ii) $P(X_1 < X_2)$ and (iii) $P(X_1 = X_2)$, *the probability of tied scores*, for three different bivariate proportional hazard models, according to Proposition 10.3.9. The results are summarized in Table 10.5, which also shows the estimates of these probabilities obtained by Li [54] using bivariate geometric distribution as a model to describe the data in Table 10.2. We further compute the AIC and BIC criteria, to judge the fit of the bivariate geometric model in [54] to the dive data set and obtained the values: $AIC = 226.9937$ and $BIC = 229.827$ for the discrete bivariate geometric distribution, as a model for the dive data.

Table 10.5 Summary of Probabilities using Three Different Bivariate Proportional Hazard Models and Bivariate Geometric Distribution (BGD) with Maximum Likelihood Estimations for Model Parameters

	BGE	BEW	BLFR	BGD[1]
$P(X_1 > X_2)$	0.0251	0.0440	0.0376	0.1204
$P(X_1 < X_2)$	0.3548	0.3519	0.2882	0.3263
$P(X_1 = X_2)$	0.6201	0.6041	0.6741	0.5533

Comparing these AIC and BIC values to the corresponding AIC and BIC values in Table 10.4; it appears that all three bivariate proportional hazard models provide a better fit to the dive data set. This conclusion is however based on the premise that assigned scores are continuous variables rather than discrete integers, which is the assumed framework for using the bivariate geometric model [54].

10.6 Discussion

In this Chapter, we introduce a framework for bivariate proportional hazard models and investigate its properties. Expectation-Maximization (EM) algorithm is discussed

[1]All probabilities based on BGD are obtained from [54].

to estimate model parameters. The proposed models and methodology are applied to two real data sets for the illustrative purpose.

CHAPTER 11

CONCLUSION

In Part II, reverse star-ordered, a new partial ordering, and its properties including the relationship of this partial order to other partial orders and non-parametric life distribution classes, and preservation properties of under some reliability operations are discussed.

In addition, we introduce a new class of bivariate proportional hazard models and explore its properties. EM algorithm is used to obtain maximum likelihood estimations for model parameters of interest. The model and methodology are applied to real data set for illustrative purposes. The performance are quite satisfactory. To illustrate the applicability of our proposed class of BPHM models, we consider two different real data sets and three different parametric families of bivariate proportional hazard models, and show one of the parametric class of BPHMs can explain both data set very well in terms of log-likelihood and other well known model-fit criterions.

APPENDIX A

FORTRAN PROGRAM CODES

A.1 Fortran 90 Program Code for MLE of Bivariate Generalized Exponential Model Parameters using EM Algorithm with NFL Data Set

```
program football_BGE
integer, parameter :: n00=24, n11=17, n22=1
real*8, parameter :: n0=24, n1=17, n2=1
real*8, parameter :: small_value = 1e-20
real*8  :: j0x(n00) = (/9.05, 0.85, 3.43, 7.78, 7.05, 2.58,
                        8.88, 17.83, 10.15, 8.98, 6.42, 7.02,
                        11.82, 14.58, 2.9, 15.53, 12.13,
                        10.35, 1.38, 1.65, 0.75, 4.25, 7.25,
                        2.98/)
real*8  :: j1x1(n11) = (/2.05, 10.57, 7.23, 6.85, 32.45,
                         8.53, 31.13, 14.58, 10.85, 5.52,
                         4.22, 6.42, 11.63, 3.98, 13.8, 5.78,
                         10.4/)
real*8  :: j1x2(n11) = (/3.98, 14.28, 9.68, 34.58, 42.35,
                         14.27, 49.88, 20.57, 38.07, 11.27,
                         9.48, 15.08, 17.37, 6.43, 49.75,
                         25.98, 14.25/)
real*8  :: j2x1(n22) = (/19.65/)
real*8  :: j2x2(n22) = (/10.7/)
integer(4)  :: i, k
real*8      :: N
```

```
real*8       :: u1, u2, v1, v2
real*8       :: a1, a2, a3, lamda0, lamda1
real*8       :: diff
real*8       :: sumj0x, sumj1x1, sumj1x2, sumj2x1, sumj2x2,
                sumj12x1, sumj12x2
! Initialization
N = n0 + n1 + n2
a1 = 0.1d0
a2 = 0.2d0
a3 = 0.05d0
lamda1 = -100.d0
diff = 100d0
i = 0
sumj0x  = sum(j0x)
sumj1x1 = sum(j1x1)
sumj1x2 = sum(j1x2)
sumj2x1 = sum(j2x1)
sumj2x2 = sum(j2x2)
open(unit = 20, file = 'output.txt',status = 'unknown')
! Main loop
do while ( diff > small_value .and. i < 10000)
   i = i +1
! Calculate  u1, u2, v1, v2
u1 = a1 / (a2+a3)
u2 = a1*a3 / (a2* (a2+a3) )
v1 = a2 / (a1+a3)
v2 = a2*a3 / (a1* (a1+a3) )
! Calculate lamda1
```

```
lamda0 = lamda1
lamda1 = ( N + (u1+u2)*n1 + (v1+v2)*n2 ) / &
         ((a1+a2+a3)*sumj0x + (1 + a3/a2)*a1*sumj1x2 + &
         a1*sumj1x1 + (1+a3/a1)*a2*sumj2x1 + a2*sumj2x2)
! Calculate a1 a2 a3
a1 = (n1 + n2*v1) / &
     (lamda1*(sumj0x +sumj1x1 + (v1+v2)*sumj2x1))
a2 = (n2 + n1*u1) / &
     (lamda1*(sumj0x + (u1+u2)*sumj1x2 + sumj2x2))
a3 = (n0 + n1*u2 + n2*v2)/ &
     (lamda1*(sumj0x +(u1+u2)*sumj1x2 + (v1+v2)*sumj2x1))
diff = abs ((lamda1 - lamda0)/lamda0)
write(20,*) '----------------'
write(20,*) 'diff       = ', diff
write(20,*) 'loop       = ', i
write(20,*) 'lamda      = ', lamda1
write(20,*) 'a1         = ', a1
write(20,*) 'a2         = ', a2
write(20,*) 'a3         = ', a3
write(20,*) 'u1         = ', u1
write(20,*) 'u2         = ', u2
write(20,*) 'v1         = ', v1
write(20,*) 'v2         = ', v2
end do
! End main loop
close(20)
write(*,*) 'diff       = ', diff
write(*,*) 'loop       = ', i
```

```
write(*,*) 'lamda       = ', lamda1
write(*,*) 'a1          = ', a1
write(*,*) 'a2          = ', a2
write(*,*) 'a3          = ', a3
end program
```

A.2 Fortran 90 Program Code for MLE of Bivariate Exponentiated Weibull Model Parameters using EM Algorithm with NFL Data Set

```
program football_BEW
integer, parameter :: n00=24, n11=17, n22=1
real*8, parameter :: n0=24, n1=17, n2=1
real*8, parameter :: small_value = 1e-10,
                     small_value2 = 1e-10
real*8  :: j0x(n00) = (/9.05, 0.85, 3.43, 7.78, 7.05,
                        2.58, 8.88, 17.83, 10.15, 8.98,
                        6.42, 7.02, 11.82, 14.58, 2.9,
                        15.53, 12.13, 10.35, 1.38, 1.65,
                        0.75, 4.25, 7.25, 2.98/)
real*8  :: j1x1(n11) = (/2.05, 10.57, 7.23, 6.85, 32.45,
                         8.53, 31.13, 14.58, 10.85, 5.52,
                         4.22, 6.42, 11.63, 3.98, 13.8,
                         5.78, 10.4/)
real*8  :: j1x2(n11) = (/3.98, 14.28, 9.68, 34.58, 42.35,
                         14.27, 49.88, 20.57, 38.07, 11.27,
                         9.48, 15.08, 17.37, 6.43, 49.75,
                         25.98, 14.25/)
real*8  :: j2x1(n22) = (/19.65/)
real*8  :: j2x2(n22) = (/10.7/)
```

```
integer(4)   :: i, k
real*8       :: N
real*8       :: u1, u2, v1, v2
real*8       :: a1, a2, a3, lamda0, lamda1
real*8       :: diff
real*8       :: sumj0x, sumj1x1, sumj1x2, sumj2x1, sumj2x2
real*8       :: lnj0x(n00), lnj1x1(n11), lnj1x2(n11),
                lnj2x1(n22), lnj2x2(n22)
real*8       :: j0xb(n00), j1x1b(n11), j1x2b(n11),
                j2x1b(n22), j2x2b(n22)
real*8       :: sumlnj0x, sumlnj1x1, sumlnj1x2, sumlnj2x1,
                sumlnj2x2
real*8       :: sum_tmp1, term1, suma123, suma1, suma2
logical      :: do_loop = .true.
! Initialization
N = n0 + n1 + n2
a1 = 0.1d0
a2 = 0.2d0
a3 = 0.05d0
lamda1 = -100.d0
beta0 = 0.9
beta1 = 0.9
diff_lamda = 100d0
diff_beta  = 100d0
i =0
lnj0x   = log(j0x)
lnj1x1  = log(j1x1)
lnj1x2  = log(j1x2)
```

```
lnj2x1  = log(j2x1)
lnj2x2  = log(j2x2)
sumlnj0x        = sum(lnj0x)
sumlnj1x1       = sum(lnj1x1)
sumlnj1x2       = sum(lnj1x2)
sumlnj2x1       = sum(lnj2x1)
sumlnj2x2       = sum(lnj2x2)
sum_tmp1        = sumlnj0x + sumlnj1x1 + sumlnj2x2
open(unit = 20, file = 'output.txt',status = 'unknown')
! Main loop
do while ( do_loop )
   i = i +1
! Calculate  u1, u2, v1, v2
u1 = a1 / (a2+a3)
u2 = a1*a3 / (a2* (a2+a3) )
v1 = a2 / (a1+a3)
v2 = a2*a3 / (a1* (a1+a3) )
! Calculate sum
j0xb            = j0x**beta1
j1x1b   = j1x1**beta1
j1x2b   = j1x2**beta1
j2x1b   = j2x1**beta1
j2x2b   = j2x2**beta1
sumj0x  = sum(j0xb)
sumj1x1 = sum(j1x1b)
sumj1x2 = sum(j1x2b)
sumj2x1 = sum(j2x1b)
sumj2x2 = sum(j2x2b)
```

```
! Calculate lamda1
lamda0 = lamda1
term1  = N + (u1+u2)*n1 + (v1+v2)*n2
suma123 = a1 + a2 + a3
suma1   =  (1 + a3/a2)*a1
suma2   =  (1 + a3/a1)*a2
lamda1 = term1 / &
         ( suma123*sumj0x + suma1*sumj1x2 + a1*sumj1x1 + &
         suma2*sumj2x1 + a2*sumj2x2 )
! Calculate beta
beta0 = beta1
beta1 = term1 / &
        (lamda1 * ( suma123*sum(j0xb *lnj0x) + suma1* &
        sum(j1x2b*lnj1x2) + a1*sum(j1x1b*lnj1x1)+ suma2* &
        sum(j2x1b*lnj2x1) + a2*sum(j2x2b*lnj2x2) ) - &
        (sum_tmp1+(u1+u2)*sumlnj1x2+(v1+v2)*sumlnj2x1))
! Calculate a1 a2 a3
a1 = (n1 + n2*v1) / &
     (lamda1*(sumj0x +sumj1x1 + (v1+v2)*sumj2x1))
a2 = (n2 + n1*u1) / &
     (lamda1*(sumj0x + (u1+u2)*sumj1x2 + sumj2x2))
a3 = (n0 + n1*u2 + n2*v2)/ &
     (lamda1*(sumj0x +(u1+u2)*sumj1x2 + (v1+v2)*sumj2x1))
diff_lamda = abs ((lamda1 - lamda0)/lamda0)
diff_beta = abs ((beta1 - beta0)/beta0)
write(20,*) '----------------'
write(20,*) 'diff_lamda = ', diff_lamda
write(20,*) 'diff_beta  = ', diff_beta
```

```
write(20,*) 'loop        = ', i
write(20,*) 'lamda       = ', lamda1
write(20,*) 'beta        = ', beta1
write(20,*) 'a1          = ', a1
write(20,*) 'a2          = ', a2
write(20,*) 'a3          = ', a3
write(20,*) 'u1          = ', u1
write(20,*) 'u2          = ', u2
write(20,*) 'v1          = ', v1
write(20,*) 'v2          = ', v2
do_loop = diff_lamda > small_value .or. &
          diff_beta > small_value2
do_loop = do_loop  .and. i < 10000
if(mod(i,100) .eq. 0) then
    write(*,*)i
end if
end do
! End main loop
close(20)
write(*,*) 'diff_lamda  = ', diff_lamda
write(*,*) 'diff_beta   = ', diff_beta
write(*,*) 'loop        = ', i
write(*,*) 'lamda       = ', lamda1
write(*,*) 'beta        = ', beta1
write(*,*) 'a1          = ', a1
write(*,*) 'a2          = ', a2
write(*,*) 'a3          = ', a3
end program
```

A.3 Fortran 90 Program Code for MLE of Bivariate Linear Failure Rate Model Parameters using EM Algorithm with NFL Data Set

```
program football_BLFR
integer, parameter :: n00=24, n11=17, n22=1
real*8, parameter :: n0=24, n1=17, n2=1
real*8, parameter :: small_value = 1e-6,
                     small_value2 = 1e-6
real*8  :: j0x(n00) = (/9.05, 0.85, 3.43, 7.78, 7.05,
                        2.58, 8.88, 17.83, 10.15, 8.98,
                        6.42, 7.02, 11.82, 14.58, 2.9,
                        15.53, 12.13, 10.35, 1.38, 1.65,
                        0.75, 4.25, 7.25, 2.98/)
real*8  :: j1x1(n11) = (/2.05, 10.57, 7.23, 6.85, 32.45,
                         8.53, 31.13, 14.58, 10.85, 5.52,
                         4.22, 6.42, 11.63, 3.98, 13.8,
                         5.78, 10.4/)
real*8  :: j1x2(n11) = (/3.98, 14.28, 9.68, 34.58, 42.35,
                         14.27, 49.88, 20.57, 38.07, 11.27,
                         9.48, 15.08, 17.37, 6.43, 49.75,
                         25.98, 14.25/)
real*8  :: j2x1(n22) = (/19.65/)
real*8  :: j2x2(n22) = (/10.7/)
integer(4)  :: i
real*8      :: N
real*8      :: u1, u2, v1, v2
real*8      :: a1, a2, a3, lamda0, lamda1, beta0, beta1
real*8      :: diff_lamda, diff_beta
```

```
real*8       :: sumj0x, sumj1x1, sumj1x2, sumj2x1, sumj2x2
real*8       :: sumj0xs, sumj1x1s, sumj1x2s, sumj2x1s,
                sumj2x2s
real*8       :: sum_tmp1, term1, suma123, suma1, suma2,
                tmp1, tmp2, tmp3
logical      :: do_loop = .true.
! Initialization
N = n0 + n1 + n2
a1 = 0.1d0
a2 = 0.2d0
a3 = 0.05d0
lamda1 = -100.d0
beta0 = 0.9
beta1 = 0.9
diff_lamda = 100d0
diff_beta  = 100d0
i =0
sumj0x  = sum(j0x)
sumj1x1 = sum(j1x1)
sumj1x2 = sum(j1x2)
sumj2x1 = sum(j2x1)
sumj2x2 = sum(j2x2)
sumj0xs  = sum(j0x **2)
sumj1x1s = sum(j1x1**2)
sumj1x2s = sum(j1x2**2)
sumj2x1s = sum(j2x1**2)
sumj2x2s = sum(j2x2**2)
open(unit = 20, file = 'output.txt',status = 'unknown')
```

```
! Main loop
do while ( do_loop )
   i = i +1
! Calculate  u1, u2, v1, v2
u1 = a1 / (a2+a3)
u2 = a1*a3 / (a2* (a2+a3) )
v1 = a2 / (a1+a3)
v2 = a2*a3 / (a1* (a1+a3) )
! Calculate lamda1
lamda0 = lamda1
term1  = N + (u1+u2)*n1 + (v1+v2)*n2
suma1   =  (1 + a3/a2)*a1
suma2   =  (1 + a3/a1)*a2
suma123 = a1 + a2 + a3
tmp1  = sum( 2*j0x/(1+2*beta1*j0x))+(u1+u2)*sum(2*j1x2 &
      /(1+2*beta1*j1x2))+sum(2*j1x1/(1+2*beta1*j1x1)) &
      +(v1+v2)*sum(2*j2x1/(1+2*beta1*j2x1))+sum(2*j2x2 &
      /(1+2*beta1*j2x2))
tmp2  = suma123*sumj0xs + suma1*sumj1x2s + a1*sumj1x1s &
      + suma2*sumj2x1s + a2*sumj2x2s
lamda1 = tmp1 / tmp2
! Calculate beta
beta0 = beta1
tmp3  = suma123*sumj0x + suma1*sumj1x2 + a1*sumj1x1 &
      + suma2*sumj2x1 + a2*sumj2x2
beta1 = (term1 / lamda1 - tmp3 ) / tmp2
! Calculate a1 a2 a3
a1 = (n1 + n2*v1) / &
```

```
      (lamda1*(sumj0x+sumj1x1+(v1+v2)*sumj2x1)+lamda1 &
      *beta1*(sumj0xs + sumj1x1s + (v1+v2)*sumj2x1s))
a2 = (n2 + n1*u1) / &
      (lamda1*(sumj0x+(u1+u2*sumj1x2+sumj2x2)+lamda1 &
      *beta1*(sumj0xs + (u1+u2)*sumj1x2s + sumj2x2s))
a3 = (n0 + n1*u2 + n2*v2)/ &
      (lamda1*(sumj0x+(u1+u2)*sumj1x2+v1+v2)*sumj2x1) &
      +lamda1*beta1*(sumj0xs+(u1+u2)*sumj1x2s+(v1+v2) &
      *sumj2x1s))
diff_lamda = abs ((lamda1 - lamda0)/lamda0)
diff_beta = abs ((beta1 - beta0)/beta0)
write(20,*) '----------------'
write(20,*) 'diff_lamda = ', diff_lamda
write(20,*) 'diff_beta  = ', diff_beta
write(20,*) 'loop       = ', i
write(20,*) 'lamda      = ', lamda1
write(20,*) 'beta       = ', beta1
write(20,*) 'a1         = ', a1
write(20,*) 'a2         = ', a2
write(20,*) 'a3         = ', a3
write(20,*) 'u1         = ', u1
write(20,*) 'u2         = ', u2
write(20,*) 'v1         = ', v1
write(20,*) 'v2         = ', v2
do_loop = diff_lamda > small_value .or. &
          diff_beta > small_value2
do_loop = do_loop  .and. i < 20000
if(mod(i,100) .eq. 0) then
```

```
    write(*,*)i
end if
end do
! End main loop
close(20)
write(*,*) 'diff_lamda  = ', diff_lamda
write(*,*) 'diff_beta   = ', diff_beta
write(*,*) 'loop        = ', i
write(*,*) 'lamda       = ', lamda1
write(*,*) 'beta        = ', beta1
write(*,*) 'a1          = ', a1
write(*,*) 'a2          = ', a2
write(*,*) 'a3          = ', a3
end program
```

REFERENCES

[1] L. J. S. Allen. *Mathematical Epidemiology*, chapter An introduction to stochastic epidemic models, pages 81–130. Springer, 2008.

[2] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1991.

[3] H. Andersson. Epidemic models and social networks. *Math Scientist*, **24**:128–147, 1999.

[4] H. Andersson and T. Britton. *Stochastic Epidemic and Their Statistical Analysis*, volume 151. New York: Springer-Verlag, 2000.

[5] N. T. J. Bailey. *The Mathematical Theory of Infectious Diseases and Its application*. London: Griffin, second edition edition, 1975.

[6] F. Ball. A unified approach to the distribution of total size and total area under the trajectory. *Advances in Applied Probability*, **18**:289–310, 1986.

[7] F. Ball, D. Mollison, and G. Scalia-Tomba. Epidemics with two levels of mixing. *Annals of Applied Probability*, **7**(1), 1997.

[8] F. Ball and P. Neal. The great circle epidemic model. *Stochastic Processes and their Applications*, **107**(2), 2003.

[9] R. E. Barlow and F. Proschan. *Statistical Theory of Reliability and Life Testing: Probability Models*. To Begin with: Silber Springs, 1981.

[10] H. W. Block, T. H. Savits, and H. Singh. The reversed hazard rate function. *Probability in the Engeneering and Informational Sciences*, **12**:69–90, 1998.

[11] J. M. Box-Steffensmeier and C. J. W. Zorn. Duration models and proportional hazards in political science. *American Journal of Political Science*, **45**(4):972–988, 2001.

[12] S. P. Brooks. Markov chain monte carlo and its application. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **47**(1):69–100, 1998.

[13] G. Casella and E. I. Georage. Explaining the gibbs sampler. *The American Statistician*, **46**(3):167–174, 1992.

[14] W. Chan, F. Proschan, and J. Sethuraman. Convex-ordering among functions, with applications to reliability and mathematical statistics. *Lecture Notes-Monograph Series: Topics in Statistics Dependence*, **16**:121–134, 1990.

[15] L. Chen and K. M. Carlen. The impact of countermeasure propagation on the prevalence of computer viruses. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Bybernetics*, **34**(2):823–833, 2004.

[16] S. Chib and E. Greenberg. Understanding the metropolis-hastings algorithm. *The American Statistician*, **49**(4):327–335, 1995.

[17] D. Clayton and J. Cuzick. Multivariate generalizations of the proportional hazards model. *Journal of the Royal Statistical Society. Series A (General)*, **148**(2), 1985.

[18] F. Cohen. *Computer Viruses*. University of Southern California, 1985.

[19] F. Cohen. Computer viruses: theory and experiments. *Computer and Security*, **6**:22–35, 1987.

[20] D. R. Cox. Regresion models and life-tables. *Journal of the Royal Statistical Society. Series B (Methdological)*, **34**(2), 1985.

[21] S. Csorgo and A. H. Welsh. Testing for exponential and marshall-olkin distribution. *Journal of Statistical Planning and Inference*, **23**:287–300, 1989.

[22] D. J. Daley. *Epidemic modeling: An Introduction*. Cambridge University Press, 2001.

[23] M. C. M. de Jong, O. Diekmann, and H. Heesterbeek. *Epidemic Models: Their Structure and Relation to Data*, chapter How does transmission of infection depend on population size?, pages 84–94. Cambridge University Press, 1995.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**(1):1–38, 1977.

[25] S. K. Dhar. *Advances on Methodological and Applied Aspects of Probability and Statistics*, chapter Modeling with a bivariate geometric distribution, pages 101–109. CRC Press, 2003.

[26] O. Diekmann, H. Heesterbeek, and H. Metz. *Epidemic Models: Their Structure and Relation to Data*, chapter The legacy of Kermack and McKendrick, pages 95–113. Cambridge University Press, 1995.

[27] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley, England, 2000.

[28] K. Dietz and D. Schenzle. *A Celebration of Statistics: the ISI Centenary Volume*, chapter Mathematical models for infectious disease statistics, pages 167–204. Springer-Verlag, New York, 1985.

[29] G. E. Dinse. Non-parametric estimation of partially imcomplete time and types of failure data. *Biometrics*, **38**:417–431, 1982.

[30] C. Elbers and G. Ridder. True and spruious duration dependence: the identifiability of the proportional hazard model. *The Review of Economic Studies*, **49**(3), 1982.

[31] L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer-Verlag, 1994.

[32] M. Garetto, W. Gong, and D. Towsley. Modeling malware spreading dynamics. In *The Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies*, pages 1869–1876, 2003.

[33] A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**:398–409, 1990.

[34] A. Gelman, J. B. Carlin, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 1995.

[35] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *Journal of Applied Statistics*, **20**, 1984.

[36] C. Geyer. Practical markov chain monte carlo. *Statistical Science*, **7**, 1992.

[37] G. J. Gibson and E. Renshaw. Estimating parameters in stochastic compartmental models using markov chain methods. *Ima Journal of Mathematics Applied in Medicine and Biology*, **15**:19–40, 1998.

[38] G. J. Gibson and E. Renshaw. Likelihood estimation for stochastic compartment models using markov chain methods. *Statistics and Computing*, **11**:347–358, 2001.

[39] W. R. Gilks, S. Richardson, and D. J. Siegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London, 1996.

[40] R. Grimes. *Malicious Mobile Code: Virus Protection for Windows*. Sebastopol, CA: O'Reilly and Associates, 2001.

[41] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, 2011.

[42] W. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, **57**, 1970.

[43] J. Heckman and B. Singer. The identifiability of the proportional hazard model. *The Review of Economic Studies*, **51**(2), 1984.

[44] H. Hethcote. *Frontiers in Mathematical Biology (lecture notes in Biomathematics)*, chapter A thousand and one epidemic models, pages 504–515. Berlin: Springer, 1994.

[45] P. Hougaard. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*, **71**(1), 1984.

[46] P. Hougaard. A class of multivariate failure time distributions. *Biometrika*, **73**(3), 1986.

[47] J. Jacquez. *Compartmental Analysis in Biology and Medicine*. University of Michigan Press, third edition edition, 1996.

[48] R. E. Kass, B. P. Carlin, and R. M. Gelman, A. Neal. Markov chain monte carlo in practice: a roundtable discussion. *The American Statistician*, **52**, 1998.

[49] J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. In *Proceedings of the Computer Society Symposium on Research in Security and Privacy*, pages 343–359, 1991.

[50] J. O. Kephart and S. R. White. Measuring and modeling computer virus prevalence. In *Proceedings of the Computer Society Symposium on Research in Security and Privacy*, pages 2–15, 1993.

[51] J. O. Kephart, S. R. White, and D. M. Chess. Computers and epidemiology. *IEEE Spectrum*, **30**:2–15, 1993.

[52] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. In *Proceeding of the Royal Society London Series A*, volume **115**, pages 700–721, 1927.

[53] D. Kundu and R. D. Gupta. A class of bivariate models with proportional reversed hazard maginals. *Sankhyā*, to appear.

[54] J. Li. *Modeling with bivariate geometric distributions.* PhD thesis, New Jersey Institute of Technology, 2010.

[55] A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of the American Statistical Association*, **62**(317), 1967.

[56] A. W. Metropolis, N. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1953.

[57] B. K. Mishra and N. Jha. Fixed period of temporary immunity after run of anti-malicious software on computer nodes. *Applied Mathematics and Computation*, **190**, 2007.

[58] D. Mollison, V. Isham, and B. Grenfell. Epidemics: models and data. *Journal of the Royal Statistical Society (Series A): Statistics in Society*, **157**:115–149, 1994.

[59] D. Moore, C. Shannon, and J. Brown. Code-red: a case study on the spread and victims of an internet worm. In *Proceedings of the ACM SIGCOMM/USENIX Internet Measurement Workshop*, pages 273–284, 2002.

[60] A. Müller and D. Stoyan. *Comparison Method for Stochastic Models and Risks.* Wiley: New York, 2002.

[61] D. Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, **84**(406), 1989.

[62] H. Okamura, H. Kobayashi, and T. Dohi. Dependence of computer virus prevalence on network structure-stochastic modeling approach. In *Proceedings of the Fourth International Workshop on Advanced Reliability Modeling*, pages 379–386, 2004.

[63] H. Okamura, H. Kobayashi, and T. Dohi. Markovian modeling and analysis of internet worm propagation. In *Proceedings of the 16th IEEE International Symposium on Software Reliability Engineering*, pages 149–158, 2005.

[64] P. G. O'Neill. A tutorial introduction to bayesian inference for stochastic epidemic models uning markov chain marte carlo. *Mathematical Biosciences*, **180**:103–114, 2002.

[65] P. G. O'Neill and G. O. Roberts. Bayesian inference for partionally observed stochastic epidemics. *Journal of the Royal Statistical Society (Series A): Statistics in Society*, **162**:121–129, 1999.

[66] R. Pastor-Satorras and A. Vespigani. Epidemic dynamics and endemic states in complex network. *Physical Review E*, **63**:066117–066124, 2001.

[67] R. Pastor-Satorras and A. Vespigani. Epidemic spreading in scale-free networks. *Physical Review Letters*, **86**, 2001.

[68] E. Renshaw. *Advanced Ecological Theory: Principles and Applications*, chapter Stochastic effects in population models, pages 23–63. Blackwell Science, Malden, Mass, 1994.

[69] C. J. Rhodes and R. M. Anderson. Contact rate calculation for a basic epidemic model. *Mathematical Biosciences*, **216**:56–62, 2008.

[70] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. New York: Springer, 2004.

[71] G. O. Roberts. *Markov Chain Monte Carlo in Practice*, chapter Markov chain concepts related to sampling algorithm, pages 45–58. Chapman and Hall, London, 1996.

[72] K. R. Rohloff and T. Basar. Stochastic behavior of random constant scanning worms. In *Proceedings of the 14th International Conference on Computer Communications and Networks*, pages 339–344, 2005.

[73] K. R. Rohloff and T. Basar. Deterministic and stochastic models for the detection of random constant scanning worms. *ACM Transactions on Modeling and Computer Simulation*, **18**:1–24, 2008.

[74] R. B. Schinazi. On the role of social clusters in the transmission of infectious diseases. *Theoretical Population Biology*, **61**(2):163–169, 2002.

[75] G. Serazzi and S. Zanero. *Performance Tools and Applications to Networked Systems*, chapter Computer virus propagation models, pages 26–50. Springer, 2004.

[76] E. H. Spafford. The internet worm: crisis and aftermath. *Communications of the ACM*, **32**:678–687, 1989.

[77] S. Staniford, V. Paxson, and N. Weaver. How to own the internet in your spare time. In *Proceedings of the 11th USENIX Security Symposium*, pages 149–167, 2002.

[78] G. Streftaris and G. J. Gibson. Bayesian inference for stochastic epidemics in closed populations. *Statistical Modelling*, **4**:63–75, 2004.

[79] M. A. Tanner. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer-Verlag, New York, 1996.

[80] M. A. Tanner and W. H. Wong. The calculation of posterior distribution by data augmentation. *Journal of the American Statistical Association*, **82**, 1987.

[81] L. Tierney. Markov-chains for exploring posterior distributions. *Annals of Statistics*, **22**, 1994.

[82] L. Tierney. Rejoinder: Markov-chains for exploring posterior distributions. *Annals of Statistics*, **22**, 1994.

[83] G. J. Van den Berg. Duration models: specification, identification and multiple durations. *Handbook of Econometrics*, **5**:3381–3460, 2001.

[84] C. Wang, J. C. Knight, and E. C. Elder. On computer viral infection and the effect of immunization. In *The 16th Annual Conference of Computer Security Applications*, pages 246–256, 2000.

[85] Y. Wang and C. Wang. Modeling the effects of timeing parameters on virus propagation. In *The 2003 ACM Workshop on Rapid Malcode*, pages 61–66, 2003.

[86] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, **393**:440–442, 1998.

[87] J. C. Wierman. Probabilistic analysis of a computer virus epidemic model. In *Proceedings of the Workshop on Statistical and Machine Learning Techniques in Computer Intrustion Detection*, 2002.

[88] J. C. Wierman and D. J. Marchette. Modeling computer virus prevalence with a susceptible-infected-susceptible model with reintroduction. *Computational Statistics and Data Analysis*, **45**:3–23, 2004.

[89] C. C. Zou, L. Gao, W. Gong, and D. Towsley. Monitoring and early warning for internet worms. In *Proceedings of the 10th ACM Conference on Computer and Communications Security*, pages 190–199, 2003.

[90] C. C. Zou, W. Gong, and D. Towsley. Code red worm propagation modeling and analysis. In *Proceedings of the 9th ACM Conference on Computer and Communications Security*, pages 138–147, 2002.

[91] C. C. Zou, D. Towsley, and W. Gong. Email worm modeling and defense. In *Proceedings of the 13th International Conference on Computer Communications and Networks*, pages 409–414, 2004.