

# Bayesian Analysis for Semi-Markov Processes using Flowgraph Models

Aparna V. Huzurbazar

Department of Mathematics and Statistics

University of New Mexico

Albuquerque, NM 87131-1141

USA

*aparna@stat.unm.edu*

## Abstract

Flowgraph models are useful in a wide variety of systems engineering and survival analysis problems. They are especially useful for analyzing time to event data and constructing corresponding Bayes predictive distributions. When a continuous time semi-Markov process defines transition times between a finite number of states and interest focuses on estimating densities, survival/reliability and hazard functions, or predictive distributions, flowgraph models provide a way of presenting the model and an associated method for data analysis. I will introduce flowgraph models and related saddlepoint methods for problems in systems engineering and multistate modelling as arises in survival analysis. An important advantage of flowgraph / saddlepoint methods is the ability to construct likelihoods for incomplete data. I will focus on modelling using data from a study on diabetic retinopathy. I will discuss advantages over direct simulation and also situations where data are missing but known to the investigator. Much of this work has been applied to cellular telephone networks and, time permitting, I will discuss that as well.

## 1 Introduction

Flowgraph models are useful for modeling time to event data that result from a stochastic process. A flowgraph models potential outcomes, probabilities of outcomes, and waiting times for those outcomes to occur. Flowgraphs model semi-Markov processes and provide a practical alternative methodology for data analysis. They are useful when a continuous time semi-Markov process defines the transition times between disease states and interest focuses on estimating the density, reliability or survival function, or hazard function of the process. Given a stochastic process with conditionally independent states, a flowgraph model allows for the use of most standard waiting time distributions to model the different states. It provides a method for accessing the waiting time distribution for any partial or total waiting time. Flowgraph models operate on moment generating functions (MGFs) and use saddlepoint approximations (cf. Daniels (1954)) to convert the MGFs to waiting time probability density functions (pdfs), cumulative distribution functions (CDFs), reliability or survival functions, and hazard functions. As a data analytic method, flowgraphs have distinct advantages over other methods for semi-Markov processes. Flowgraphs handle censoring and can be used in either a frequentist or a Bayesian framework.

Block diagrams and signal flowgraphs are widely used to represent engineering systems, especially in circuit analysis. Basic flowgraph ideas were developed in engineering, but they never incorporated probabilities, waiting times, or data analysis. The literature on flowgraph methods is vast. Introductions to flowgraph methods are contained in most circuit analysis or control systems textbooks such as Dorf and Bishop (1995), Gajic and Lelic (1996), and Whitehouse (1977). Statistical flowgraph models are based on flowgraph ideas but unlike their antecedents, flowgraph models can also be used to model and analyze data from complex stochastic systems. For literature on statistical flowgraph models see Butler and Huzurbazar (1997), Huzurbazar (1999), and Huzurbazar (2000).

## 2 Flowgraph Models for Engineering Systems

Figure 1 shows a complex system consisting of outcomes in series and cascaded in parallel with feedback loops. The system is an assembly line for a manufacturing process for car stereos. State 0 represents an initial detection of a problem with a stereo. The problem is categorized into one of two types of severity. If the severity is of type I, the system is in state 1 for repair of the item. Eventually, the problem is fixed and the item moves to state 3 where it is specifically inspected to make sure that the type I problem is fixed. If the problem is not fixed, the item is returned to state 1, otherwise, it passes inspection and moves to state 5. Similarly, if the severity is of type II, the system is in state 2 for repair. Eventually, the problem is fixed and the item moves to state 4 where it is specifically inspected to make sure that the type II problem is fixed. If the problem is not fixed, the item is returned to state 2, otherwise, it passes inspection and moves to state 5.

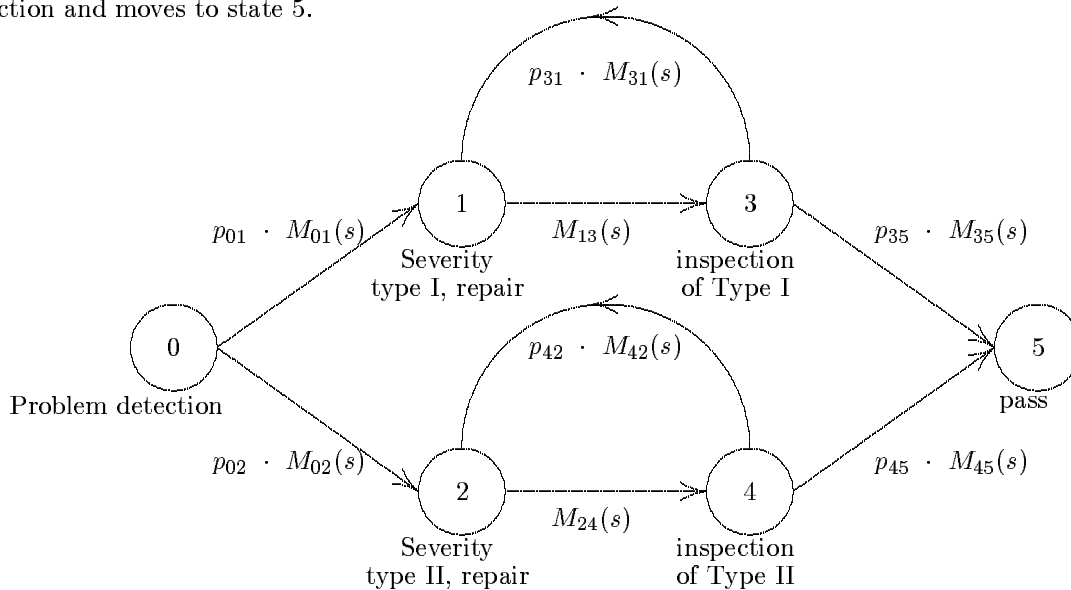


Figure 1: Flowgraph Model for Manufacturing System.

In a flowgraph model, the states or nodes represent outcomes. The nodes are connected by directed line segments called *branches*. These branches are labeled with *transmittances*. These transmittances are labelled with the “transition probability  $\times$  moment generating function (MGF) the of waiting time distribution in the previous state” which is a quantity called the *branch transmittance*. In the figure, probabilities and moment generating functions of the waiting time distributions are shown as branch transmittances. The waiting times on the branches can be any parametric distributions that admit moment generating functions. Hence, the model is quite general in that Markovian assumptions are not made. We use the branch transmittances of a flowgraph model to solve for the MGF of the distribution of the waiting time of interest. This MGF is converted to a density, reliability, or hazard function using a saddlepoint approximation. Quantities of interest include predicting the distribution of the total time until the item is up to standards,  $0 \rightarrow 5$ ; predicting the waiting time in repair, say,  $0 \rightarrow 3$  or  $0 \rightarrow 4$ , or  $0 \rightarrow 3$  or  $4$ ; or predicting the total number of times an item fails inspection from a given state such as 3, i.e. the total number of times the transition  $3 \rightarrow 1$  is made.

More generally, flowgraph models can be used to assess system reliability and quantities such as the time to total and partial failure of a system. Huzurbazar (2000) gives details on using flowgraph models for computing the total and partial system failure for cellular telephone networks. This article also deals with data analysis for general queueing systems using flowgraph models and considers M/M/q queues as a special case, although the methods presented are for general semi-Markov processes, including M/G/1 queues.

### 3 Flowgraph Models for Survival Analysis

Flowgraph models also provide an innovative way to analyze time-to-event data arising from problems survival analysis. Flowgraph models for stochastic networks were introduced by Butler and Huzurbazar (1997). The data analytic examples presented there are limited to Markov models for birth processes. Butler and Huzurbazar (1997) focus on Bayesian prediction using flowgraphs and present methodology for models of disease progression for kidney failure, cancer, and HIV/AIDS. Situations involving left, right, and interval censored data are discussed using AIDS data from the San Francisco Men's Health Study. Flowgraph models also provide useful extensions to phase type (PH) distributions. PH distributions are defined to be distributions of absorption times in Markov processes. In survival analysis, these distributions are used to model stages of disease progression that lead to an absorbing end state, usually death. Aalen (1995) presents many different PH distributions for modeling survival times in situations where the overall survival time involves progression through several stages. Analysis of these models is quite complicated and requires a number of simplifying assumptions. The most common of these is the Markovian assumption which necessitates exponential waiting times between stages of the disease. While this may be a reasonable assumption for many situations, it is also quite restrictive. Flowgraph modeling extends these to generalized PH distributions by allowing the use of any waiting time distribution with a tractable MGF. Huzurbazar (1999) discusses such extensions of PH distributions using flowgraph models.

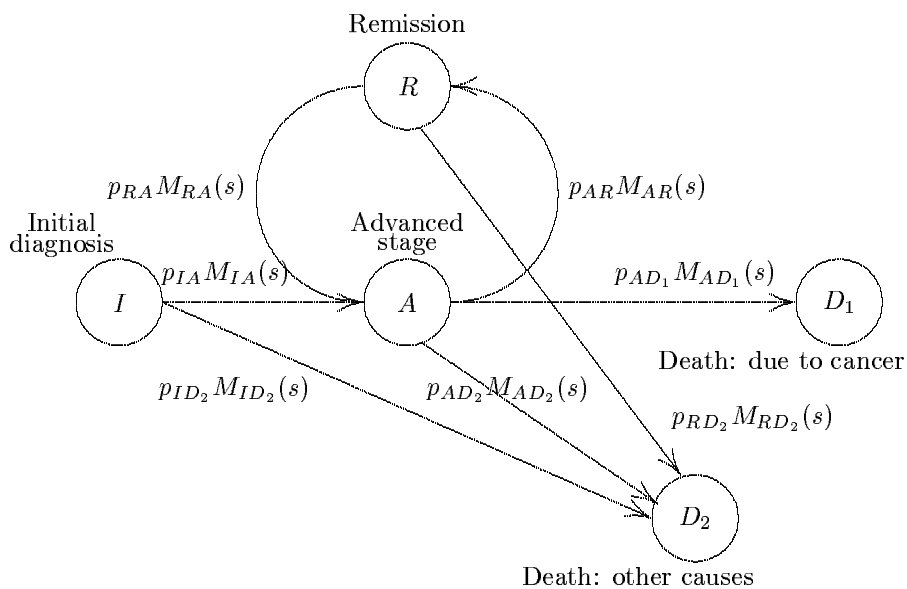


Figure 1 : Flowgraph Model for Cancer.

Figure 2 presents a flowgraph model for the progression of cancer. In this figure, the outcomes represent various stages of cancer or death. For example, a patient with an initial diagnosis of cancer is considered to be in state  $I$  in Figure 2. The patient can have two possible outcomes: progress to an advanced stage,  $A$ , of the disease or die from a different cause  $D_2$ . Once in the advanced stage, the person has 3 possible outcomes: she can go into remission,  $R$ , she can die from cancer,  $D_1$ , or she can die from other causes,  $D_2$ . If she goes into remission, she can subsequently return to an advanced stage or she can die from other causes. There is interest in the probabilities of being in any stage and in how long a person remains in a given stage. In particular, interest centers on remission of cancer and death. Flowgraphs model this entire complex stochastic network.

The current work is concerned with modeling disease progression using flowgraph models for censored and incomplete survival data for semi-Markov processes. Censored observations contribute to the likelihood function in terms of survival functions. When data are incomplete, this involves piecewise likelihood

construction. The methods presented involve modeling a large data set and developing methods for data that are missing in transitions. We illustrate the flowgraph modeling with application to a diabetic retinopathy data set (cf. Marshall and Jones (1995)). One potential problem with any sampling scheme is having unrecognized incomplete data, i.e. data that are missing but unknown to the investigator. If data are known to be missing, we have several methods for imputing missing values, for example, EM, MCMC, etc. Our aim is to investigate how flowgraph models and standard models perform when handling data that are missing but not known to be missing. For example, if we observe a patient making a transition  $I \rightarrow R$ , we recognize that the data are incomplete since we are missing the transition  $I \rightarrow A$  and  $A \rightarrow R$ . However, if in a well designed study, we observe a transition  $I \rightarrow A \rightarrow R$ , we typically assume that it is complete data. Unfortunately, that need not be the case. The patient could have transitioned as  $I \rightarrow A \rightarrow R \rightarrow A \rightarrow R$  without our observing the second transition to state  $A$  and one of the transitions to state  $R$  but not knowing which transition to state  $R$  was actually observed. For example, in  $I \rightarrow A \rightarrow R \rightarrow A \rightarrow R$ , the initial loop  $A \rightarrow R \rightarrow A$  could be missed and reported as all being time in state  $A$ , or the loop  $R \rightarrow A \rightarrow R$  could be missed and treated as time only in state  $R$ . In good studies, this should not occur, but with anything less than continuous monitoring, it is possible.

## References

- Aalen, O.O. (1995). Phase type distributions in survival analysis. *Scandinavian Journal of Statistics*, 22, 447-463.
- Butler, R.W. and Huzurbazar, A.V. (1997). Stochastic network models for survival analysis. *Journal of the American Statistical Association*, 92, 246-257.
- Daniels, H. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics*, 25, 631-650.
- Dorf, R. and Bishop, R. (1995). *Modern control systems*. Reading, Mass. : Addison-Wesley.
- Gajic, Z., and Lelic, M. (1996). *Modern control systems engineering*. New York : Prentice Hall.
- Huzurbazar, A. V. (1999). Flowgraph models for generalized phase type distributions with non-exponential waiting times. *Scandinavian Journal of Statistics*, 26, 145-157.
- Huzurbazar, A. V. (2000). Modeling and analysis of engineering systems data using flowgraph models. *Technometrics*, 42, 300-306.
- Marshall, G. and Jones, R. (1995). Multistate models and diabetic retinopathy. *Statistics in Medicine*, 14, 1975-1983.
- Whitehouse, G. (1977). *Systems analysis and design using network techniques*. New Jersey: Prentice-Hall.