Journal of Computational and Applied Mathematics 291 (2016) 85–93

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Bayesian prediction for flowgraph models with covariates. An application to bladder carcinoma

B. García-Mora[a,*], C. Santamaría[a], G. Rubio[a], J.L. Pontones[b]

[a] Instituto de Matemática Multidisciplinar, Universitat Politècnica de València, Spain
[b] Departamento de Urología, Hospital Universitario La Fe de Valencia, Spain

ARTICLE INFO

Article history:
Received 16 October 2014
Received in revised form 13 March 2015

Keywords:
Flowgraph model
Erlang distribution
Phase-type distribution
Bayesian approach
Bladder carcinoma

ABSTRACT

Statistical Flowgraph Models are an efficient tool to model multi-state stochastic processes. They support both frequentist and Bayesian approaches. Inclusion of covariates is also available. In this paper we propose an easy way to perform a Bayesian approach with covariates. Results are presented with an application to bladder carcinoma data.

Contents

1. Introduction

Statistical Flowgraph Models are an efficient tool to model multi-state stochastic processes. A Flowgraph is a graphical representation of a stochastic system, where there are nodes that represent the system states and branches that represent transitions between states. A useful methodology is developed to efficiently compute magnitudes of interest, such as the probability distribution of time to reach some state. It supports both frequentist and Bayesian approaches, and in fact flowgraph methods for Bayesian analysis arise from the beginning of the subject [1].

* Corresponding author.
E-mail addresses: magarmo5@imm.upv.es (B. García-Mora), crisanna@imm.upv.es (C. Santamaría), grubio@imm.upv.es (G. Rubio), pontones_jos@gva.es (J.L. Pontones).

http://dx.doi.org/10.1016/j.cam.2015.03.045
0377-0427/

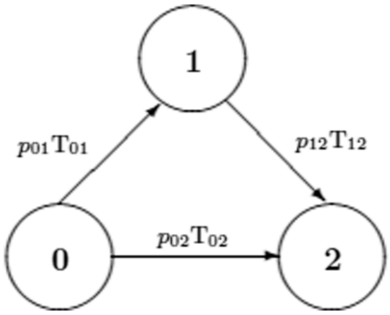

Fig. 1. Three state flowgraph model.

In practical applications, these models were successfully used to model a variety of time-to-event data arising in multistate stochastic networks such as prediction of service degradation and failure in cellular telephone networks [2], prediction of cumulative seismic damage [3], maintenance and repair of aircraft [4] and various biomedical applications [5]. See [6] for an introductory book on Statistical Flowgraphs.

The incorporation of covariates in this methodology have been recent, just the work [7] extends previous flowgraph modeling to incorporate covariates. Our aim in this paper is to propose an easy way to perform a Bayesian approach with covariates. Bayesian methods are reaching an increasing importance. An important feature is that they allow a fuller treatment of the information available. Therefore it is very interesting to develop efficient Bayesian techniques. For an introduction to Bayesian methods see, for instance, [8,9].

The paper is organized as follows: in Section 2 we describe the construction of a Flowgraph Model, step by step. We also analyze a simple flowgraph for a three states process, on which our model is based, and we introduce the *phase-type* and *Erlang* distributions, needed to build the model. Section 3 is a short introduction to our approach to incorporate covariates, and then we explain it in detail within the Bayesian framework in Section 4, modeling bladder carcinoma data. Finally, in Section 5 some discussion is given.

2. Statistical flowgraph models

A *Statistical Flowgraph Model* represents a multistate model by means of directed line segments or transitions (*branches*) connecting the states (*nodes*). Fig. 1 represents a three state flowgraph model where, for example, state 0 is the start of the process, state 1 could represent a partial failure and state 2 the total failure.

The step-by-step process of setting up a *Statistical Flowgraph Model* and solving for quantities of interest consists of:

1. **Establishing states graph and transitions of the model**
   Nodes or states of a flowgraph model are connected by directed line segments called *branches* or transitions. Every transition $i \to j$ has:
   (a) A transition probability $\mathbf{p_{ij}}$: the probability, on entry to state i, that the next transition will be to state j.
   (b) A waiting time distribution $\mathbf{F_{ij}(x)}$: the cumulative distribution function of the time x spent in state i, given that a transition to j occurs.
   Both the $\mathbf{p_{ij}}$ and the $\mathbf{F_{ij}(x)}$ may be based on data analysis if sufficient data on transitions are available. The $\mathbf{F_{ij}(x)}$ would more precisely be written as $\mathbf{F_{ij}(x|\theta_{ij})}$, where θ_{ij} represents the parameters of the distribution.
2. **Decide on a probability model family $\mathbf{F_{ij}(x|\theta_{ij})}$ for each waiting time distribution in each *branch* $i \to j$.**
   In this regard for the $\mathbf{F_{ij}(x|\theta_{ij})}$ distribution, many authors have considered mixture of distributions as a procedure to obtain new distributions with suitable properties from a computational point of view. Often the analytical expression of this mixture is not mathematically manageable. A way to include distributions with manageable expressions is to introduce the *phase-type* (PH) distributions [10]. Such distributions present some advantages that make them interesting to be included in statistical problems e.g. they form a class weakly dense in the class of general distributions defined on the positive real line; so, it is possible to approximate any lifetime distribution by a *phase-type*-distribution.
   In the present work we use a linear combination of three Erlang distributions given in [11]. The *Erlang distribution* is a particular case of a *PH*-distribution. Some works showing the application of the *phase-type* and *Erlang* distributions are the ones of [12,13] among others. Next we describe both distributions:
   (a) The distribution $F(\cdot)$ on $[0, \infty[$ is a *PH*-distribution with representation (α, T) if it is the distribution of time until absorption time in a Markov process on the states $\{1, \dots, m, m+1\}$ where $\{1, \dots, m\}$ are the transition states and $m+1$ is the absorption state. The generator (matrix of transition rates between states) of this Markov process is

$$\begin{pmatrix} T & T^0 \\ 0 & 0 \end{pmatrix}.$$

   T represents transition rates between transient states and T^0 transition rates from each transient state to the absorption state. T represents transition rates between transient states. This matrix is upper triangular, because the only