Computational Statistics and Data Analysis 56 (2012) 1303–1318

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

Frailty modeling via the empirical Bayes–Hastings sampler

Richard A. Levine [a,*], Juanjuan Fan [a], Pamela Ohman Strickland [b], Shaban Demirel [c]

[a] Department of Mathematics and Statistics, 5500 Campanile Drive, San Diego State University, San Diego, CA, 92182, United States

[b] Division of Biometrics, UMDNJ, Piscataway, NJ 08854, United States

[c] Devers Eye Institute, Discoveries in Sight Research Laboratories, 1225 NE 2nd Ave, Portland, OR 97232, United States

ARTICLE INFO

Article history:
Received 27 August 2010
Received in revised form 22 July 2011
Accepted 6 September 2011
Available online 1 October 2011

Keywords:
Multivariate survival analysis
Nonparametric Pólya tree prior
Gibbs sampler
Metropolis–Hastings sampler
Goodness of fit
Glaucoma and ophthalmology data

ABSTRACT

Studies of ocular disease and analyses of time to disease onset are complicated by the correlation expected between the two eyes from a single patient. We overcome these statistical modeling challenges through a nonparametric Bayesian frailty model. While this model suggests itself as a natural one for such complex data structures, model fitting routines become overwhelmingly complicated and computationally intensive given the nonparametric form assumed for the frailty distribution and baseline hazard function. We consider empirical Bayesian methods to alleviate these difficulties through a routine that iterates between frequentist, data-driven estimation of the cumulative baseline hazard and Markov chain Monte Carlo estimation of the frailty and regression coefficients. We show both in theory and through simulation that this approach yields consistent estimators of the parameters of interest. We then apply the method to the short-wave automated perimetry (SWAP) data set to study risk factors of glaucomatous visual field deficits.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Analyses of data from studies of visual field deficits and glaucomatous progression are complicated by correlations between observed failure times from fellow eyes of a subject. Bayesian frailty models have proven to be a valuable tool for modeling this dependence through a random effect term in a proportional hazards model. However, the practitioner is left to choose from a wide array of frailty distributions, the choice of which may affect inferences drawn on parameters (hazard ratios) of interest. Not to mention, the dependence structure is unknown presenting difficulties in parameterizing a frailty model and exposing "default" models, such as a gamma frailty distribution, as seemingly arbitrary.

A nonparametric approach to frailty modeling provides a flexible alternative in which the frailty distribution is left unspecified, letting the data a posteriori drive the functional form. In such models, both the frailty distribution and the baseline hazard rate are modeled nonparametrically. The nonparametric frailty term presents no difficulties in the construction of a Markov chain Monte Carlo (MCMC) algorithm for drawing posterior inferences. Standard Gibbs samplers for fitting nonparametric Bayesian models (e.g., Walker and Mallick, 1997) may be applied for sampling full conditional distributions on the frailties and the parametric portion of the proportional hazards model. However, incorporation of the baseline hazard into the Markov chain Monte Carlo (MCMC) routine turns out to be a challenging task. A wide array of models for the baseline hazard and MCMC methods for fitting these models have been proposed in the literature (for example, see Ibrahim et al., 2001). However, as Gustafson et al. (2003) mentions in the motivation of their work, the routines are computationally and mathematically intensive and not easily automated, leaving the non-expert with a difficult task in applying such inferential procedures.

* Corresponding author. Tel.: +1 619 594 6494.
E-mail addresses: ralevine@sciences.sdsu.edu (R.A. Levine), jjfan@sciences.sdsu.edu (J. Fan), ohmanpa@umdnj.edu (P.O. Strickland), sdemirel@deverseye.org (S. Demirel).

0167-9473/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2011.09.004

In our motivating application the primary goal is to infer hazard rates, studying risk factors for glaucoma and short wavelength automated perimetry for detecting visual field defects. The baseline hazard rate is then effectively a nuisance parameter (function). An inferential routine for the baseline hazard which requires complicated mathematical derivation and substantial computational coding and implementation cost is clearly undesirable. In this paper we propose an empirical Bayes approach to alleviate difficulties in modeling the baseline hazard and subsequently incorporating it into an MCMC algorithm. The idea derives from the work of Casella (2001) in which we use the data to "estimate away" nuisance parameters, focusing computational and inferential effort on the parameters of interest. In the nonparametric Bayesian frailty model, we estimate the baseline hazard through a nonparametric frequentist estimator and then construct an MCMC algorithm to iteratively simulate posterior samples conditional on this empirical Bayes estimate. The method draws on the deep theory and vast implementation options of MCMC and EM algorithms. Furthermore, the routine is simple to automate within the construct of the Gibbs and Hastings samplers for fitting nonparametric Bayesian models. We argue that this empirical Bayes–Hastings sampler requires less coding time and computational expense than the popular piecewise hazard approaches (e.g., Walker and Mallick, 1997), requires less tweaking of tuning and model parameters in fact lending to complete automation.

The Bayesian frailty model with nonparametric specification of the frailty distribution is best suited for our study of glaucomatous progression. However, the empirical Bayes–Hastings sampler in this setting, as a general approach, lends to diagnostic tools for testing parametric forms for the frailty distribution and routines for performing model selection. We highlight these issues in our analysis of glaucomatous visual field defects. Furthermore, the proposed modeling and inferential strategies provide a flexible framework within which to mix and match nonparametric and parametric components and strategies for handling nuisance parameters.

In Section 2, we formally define the nonparametric Bayesian frailty model, expressing the frailty distribution nonparametrically through a Pólya tree process. We also define the nonparametric estimator of the cumulative baseline hazard to be incorporated into our empirical Bayes routine. In Section 3 we, primarily for notational purposes, briefly detail the Pólya tree distribution. In Section 4, we introduce the empirical Bayes–Hastings sampler for drawing inferences under the semi-parametric frailty model, estimating the baseline hazard rate in a Monte Carlo E-type step in the MCMC routine. As part of the discussion of the Hastings sampler, we derive conditions under which the random variates drawn reasonably represent a sample from the posterior distribution of interest. We also discuss issues for optimally implementing the MCMC sampling scheme in practice. Section 5 presents simulation studies to validate our proposed methods for drawing inferences under the frailty model. In Sections 6 and 7, we present routines for computing Bayes factors and traverse the regression parameter space to evaluate parametric forms of the frailty distribution and perform variable selection within our empirical Bayes–Hastings sampler framework. In Section 8, we illustrate our proposed methods in the analysis of a data set for studying glaucomatous visual field deficits. In Section 9 we conclude with a discussion of practical issues beyond the developments and applications in this paper.

2. Frailty model

Suppose that the observed data consist of clustered, and possibly censored, failure-time data represented by $Y_{ik} = \{X_{ik}, \delta_{ik}, Z_{ik}\}$, with $k = 1, \ldots, K_i$ and $i = 1, \ldots, n$. $X_{ik} = T_{ik} \wedge C_{ik}$ is the minimum of the failure time and the censoring time; $\delta_{ik} = I\{(X_{ik} = T_{ik})\}$, the failure indicator, which takes the value of 1 if $(X_{ik} = T_{ik})$ and 0 otherwise; and Z_{ik} is a p-vector of covariates. It is assumed that the failure time vector $T_i = (T_{i1}, \ldots, T_{iK_i})'$ is independent of the censoring time vector $C_i = (C_{i1}, \ldots, C_{iK_i})'$ given $Z_i = (Z_{i1}', \ldots, Z_{iK_i}')'$, $i = 1, \ldots, n$.

The proportional hazards model (Cox, 1972) has been widely applied in analyzing independent or univariate failure times. As a generalization of the Cox proportional hazards model for clustered or multivariate failure times, Clayton and Cuzick (1985) introduce the frailty model in which a random effect term (or "frailty") is assumed to have a multiplicative effect on the hazard. In terms of the hazard function, the model can be stated as follows:

$$\lambda_{ik}(t|\mathbf{Z}_{ik}, V_i) = \lambda_0(t) \exp(\boldsymbol{\beta}'\mathbf{Z}_{ik})V_i \tag{1}$$

where $\lambda_0(t)$ is an unknown baseline hazard function, $\boldsymbol{\beta}$ is a p-vector of unknown regression parameters, and V_i is the frailty, representing some common unobserved characteristics shared by all the failure times in the ith cluster. It is assumed that, given the frailty V_i, failure times within the ith cluster are independent. Note that the baseline hazard function λ_0 may be assumed to depend on k, for example, in a family study when $k = 1, 2$ refers to mothers and daughters, respectively.

Let $\theta_i = \ln V_i$ for each $i = 1, \ldots, n$ with the n-vector of log-frailties denoted by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)'$. We will avoid difficulties in specifying the frailty distribution by modeling this distribution nonparametrically. In particular, following Walker and Mallick (1997), assume

$$\begin{aligned} &\theta_1, \ldots, \theta_n \text{ i.i.d. } F \\ &F \sim PT(\boldsymbol{\alpha}, G) \\ &\boldsymbol{\beta} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \end{aligned} \tag{2}$$

where $PT(\boldsymbol{\alpha}, G)$ denotes a Pólya tree prior with prespecified parameters $\boldsymbol{\alpha}$ and G (see Section 3 for details) and $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are prespecified parameters of the normal prior distribution on the coefficients $\boldsymbol{\beta}$.

Download English Version:

https://daneshyari.com/en/article/416429

Download Persian Version:

https://daneshyari.com/article/416429

Daneshyari.com