Expert Systems with Applications 36 (2009) 11341–11346

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Comparison of Bayesian survival analysis and Cox regression analysis in simulated and breast cancer data sets

Imran Kurt Omurlu [a,*], Kazim Ozdamar [b], Mevlut Ture [c]

[a] Trakya University Medical Faculty, Department of Biostatistics, 22030 Edirne, Turkey
[b] Eskişehir Osmangazi University Medical Faculty, Department of Biostatistics, Eskişehir, Turkey
[c] Adnan Menderes University Medical Faculty, Department of Biostatistics, Aydın, Turkey

ARTICLE INFO

Keywords:
Cox regression
Bayesian survival
Survival
Breast cancer
Markov Chain Monte Carlo
Simulation

ABSTRACT

We aimed to compare the performance of Cox regression analysis (CRA) and Bayesian survival analysis (BSA) by using simulations and breast cancer data.

Simulation study was carried out with two different algorithms that were informative and noninformative priors. Moreover, in a real data set application, breast cancer data set related to disease-free survival (DFS) that was obtained from 423 breast cancer patients diagnosed between 1998 and 2007 was used.

In the simulation application, it was observed that BSA with noninformative priors and CRA methods showed similar performances in point of convergence to simulation parameter. In the informative priors' simulation application, BSA with proper informative prior showed a good performance with too little bias. It was found out that the bias of BSA increased while priors were becoming distant from reliability in all sample sizes. In addition, BSA obtained predictions with more little bias and standard error than the CRA in both small and big samples in the light of proper priors.

In the breast cancer data set, age, tumor size, hormonal therapy, and axillary nodal status were found statistically significant prognostic factors for DFS in stepwise CRA and BSA with informative and noninformative priors. Furthermore, standard errors of predictions in BSA with informative priors were observed slightly.

As a result, BSA showed better performance than CRA, when subjective data analysis was performed by considering expert opinions and historical knowledge about parameters. Consequently, BSA should be preferred in existence of reliable informative priors, in the contrast cases, CRA should be preferred.

1. Introduction

Survival analysis is a family of statistical procedures for data analysis for which the outcome variable of interest is time until an event occurs. Most popular of survival procedures is Cox regression analysis (CRA). Because it is a semiparametric and a method for investigating the effect of several variables upon the time a specified event takes to happen (Kleinbaum & Klein, 1996). But over the last few years there has been increased interest shown in the application of survival analysis based on Bayesian methodology. Researchers did not use Bayesian analysis frequently in medical studies because it has a complex theory. Bayesian analysis of survival data has received much recent attention due to advances in computational and modeling techniques (Ibrahim, Chen, & Sinha, 2001).

Bayesian survival analysis (BSA) provides inferences that are exact, while CRA bases maximum likelihood estimations of parameters on asymptotic considerations (Calle, Hough, Curia, & Gómez, 2006; SAS Institute, 2006).

BSA consists of data and prior information. It generates conclusions based on the synthesis of new information from an observed data and historical knowledge or expert opinion. Historical knowledge from past similar studies can be very helpful in interpreting the results of the current study. Therefore, BSA reflects researches' subjective beliefs. Prior elicitation plays the most crucial role in BSA. BSA cannot be used for any modeling without using a prior distribution (Ibrahim et al., 2001; SAS Institute, 2006).

Recently, few works have been published on the BSA method. Yin and Ibrahim (2006) analyzed a simulation study using BSA for varying sample sizes, 1000 replications, 5000 Gibbs samples and 200 burn-in samples and a real data set from a melanoma clinical trial. Calle et al. (2006) analyzed data from sensory shelf-life studies. Wong, Lam, and Lo (2005) used BSA to investigate the effectiveness of silver diamine fluoride and sodium fluoride varnish in arresting active dentin caries in Chinese pre-school children.

The purpose of this study was to compare performances of CRA and BSA under varying sample sizes using Monte Carlo simulation and to apply CRA and BSA for disease-free survival (DFS) in breast cancer patients.

* Corresponding author. Tel.: +90 284 2357641/1633; fax: +90 284 2357652.
E-mail address: ikurt@trakya.edu.tr (I.K. Omurlu).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.03.058

2. Material and methods

2.1. Cox regression analysis

The CRA is the most general of the regression models because it is not based on any assumptions concerning the nature or shape of the underlying survival distribution (Ahmed, Vos, & Holbert, 2007). The CRA is the most widely used method of survival analysis.

Survival analysis typically examines the relationship of the survival distribution to covariates. Most commonly, this examination entails the specification of a linear-like model for the log hazard. The Cox model may be written as

$$h(t,x) = h_0(t)e^{\beta' x}$$

where x is the covariate vector, β is the unknown parameter vector and $h_0(t)$ is called the baseline hazard (it is the hazard for the respective individual when all independent variable values are equal to zero). $h(t,x)$ denotes the resultant hazard, given the values of the covariates for the respective case and the respective survival time (t). This method uses the partial likelihood to estimate the parameters, and parameter estimates in the method are obtained by maximizing partial likelihood function. The partial likelihood is given by

$$L(\beta) = \prod_{i=1}^{k} \frac{\exp(\beta' x_{(i)})}{\sum_{l \in R_{(t_{(i)})}} \exp(\beta' x_1)}$$

where the summation in the denominator is the over all subjects in the risk set at time $t_{(i)}$, denoted by $R(t_{(i)})$, the product is over the k distinct ordered survival times and $x_{(i)}$ denotes the value of the covariate for the subject with ordered survival time $t_{(i)}$ (Hosmer & Lemeshow, 1999; Kleinbaum & Klein, 1996).

The CRA has two assumptions, while no assumptions are made about the shape of the underlying hazard function. First, they specify a multiplicative relationship between the underlying hazard function and the log-linear function of the covariates. The second assumption is that there is a log-linear relationship between the independent variables and the underlying hazard function (Hosmer & Lemeshow, 1999; Kleinbaum & Klein, 1996).

2.2. Bayesian survival analysis

Bayesian analysis generates conclusions based on the synthesis of new information from the observed data and previous knowledge or external evidence (Wong et al., 2005).

In classical approaches such as maximum likelihood, inference is based on the likelihood of the data alone. In Bayesian models, the likelihood of the observed data x given parameters β, denoted as $p(x|\beta)$ or equivalently $L(\beta)$, is used to modify the prior beliefs $\pi(\beta)$, with the updated knowledge summarized in a posterior density, $p(\beta|x)$. The relationship between these densities is:

$$p(\beta|x) \propto L(\beta)\pi(\beta)$$

Thus, updated beliefs are a function of prior knowledge and the sample data evidence. From the Bayesian perspective the likelihood is viewed as a function of β given fixed data x, and so elements in the likelihood which are not functions of β become part of the proportionality in this equation. $L(\beta)$ is the partial likelihood function with regression coefficients β as parameters (Congdon, 2003, 2006; Ibrahim et al., 2001; SAS Institute, 2006).

In complex models, posterior densities can often be too difficult to work with directly. To update knowledge about the parameters requires that one can sample from the posterior density. With Markov Chain Monte Carlo (MCMC) method, it is possible to generate samples from a posterior density and to use these samples to approximate expectations of quantities of interest. MCMC method samples successively from a target distribution, with each sample drawn depending on the previous one. Gibbs sampler is a MCMC method, and a very powerful simulation algorithm. Gibbs sampler can be efficient when the parameters are not highly dependent on each other and the full conditional distributions are easy to sample from SAS Institute (2006) and Robert and Casella (2004).

Gibbs sampler works as follows (Ibrahim et al., 2001; Robert & Casella, 2004; SAS Institute, 2006):

1. Set $m = 0 (m = 1,2,\ldots,M)$, and choose an arbitrary initial value of $\boldsymbol{\beta}^{(0)} = \{\beta_1^{(0)}, \beta_2^{(0)}, \ldots \beta_p^{(0)}\}'$.
2. Generate each component of $\boldsymbol{\beta}^{(m+1)} = \{\beta_1^{(m+1)}, \beta_2^{(m+1)}, \ldots \beta_p^{(m+1)}\}'$ as follows:
   - Draw $\beta_1^{(m+1)}$ from $\pi(\beta_1|\beta_2^{(m)}, \ldots, \beta_p^{(m)}, \boldsymbol{x})$
   - Draw $\beta_2^{(m+1)}$ from $\pi(\beta_2|\beta_2^{(m+1)}, \beta_3^{(m)} \ldots, \beta_p^{(m)}, \boldsymbol{x})$
   - ……
   - Draw $\beta_p^{(m+1)}$ from $\pi(\beta_p|\beta_1^{(m+1)}, \beta_2^{(m+1)} \ldots, \beta_{p-1}^{(m+1)}, \boldsymbol{x})$
3. Set $m = m + 1$ and go to step 1.

Convergence diagnostics help to resolve whether the Markov chain has reached its stationary. Many diagnostic tests (Gelman–Rubin, Geweke, autocorrelation and so forth) are designed to verify a necessary but not sufficient condition for convergence. With some models, certain parameters can appear to have very good convergence behavior, but that could be misleading due to the slow convergence of other parameters. If some of the parameters have bad mixing, posterior inference for parameters is failed (Congdon, 2003; SAS Institute, 2006).

In Bayesian analysis, prior elicitation plays the most crucial role. Bayesian analysis cannot be used for any modeling without using a prior distribution. Bayesian analysis is used to noninformative (objective) or informative (subjective) prior in inference. Informative prior is obtained from previous studies, past experiences or expert opinions. It is not dominated by the likelihood and has an impact on the posterior distribution. Sometimes there is no prior information about any of the model parameters, and what is often referred to as noninformative prior density is used. Noninformative prior has minimal impact on the posterior distribution of β, and can lead to improper posteriors. However, while noninformative prior is very popular in some applications, it is not always easy to construct (Gelman, 2002a; Gelman, 2002b; SAS Institute, 2006).

2.3. Simulation algorithms

Our interest in this study was to compare the parameter estimates from CRA and BSA in different conditions. The models developed here have the same multiplicative structure as the Cox regression model. We used two different simulation algorithms for analyses. The probability models with one explanatory variable were used in simulations and the following steps were applied to carry out the simulations.

Algorithm I:

We compared CRA and BSA with noninformative prior in this algorithm.

(1) Set up a value of the parameter β.
(2) Set up a value of the sample size.
(3) Set up a value of the baseline hazard function ($h_0(t)$).
(4) The variable E was generated from exponential distribution, $E \sim$ Exponential (1).
(5) The explanatory variable was generated from uniform distribution with (0,1) parameters.
(6) Survival time ($t = E/h_0(t)e^{\beta' x}$) (Bender, Augustin, & Blettner, 2005) was generated by using values obtained in steps 1–5.