Computational Statistics and Data Analysis 71 (2014) 859–867

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Optimal experimental designs for partial likelihood information

J. López-Fidalgo[a,*], M.J. Rivas-López[b]

[a] Department of Mathematics, University of Castilla-La Mancha, 13071-Ciudad Real, Spain
[b] Department of Statistics, University of Salamanca, 37008-Salamanca, Spain

ARTICLE INFO

Article history:
Received 16 February 2012
Received in revised form 4 October 2012
Accepted 6 October 2012
Available online 13 October 2012

Keywords:
Censoring
Cox regression
Proportional hazards model

ABSTRACT

In a follow-up study the time-to-event may be censored either because of dropouts or the end of study is earlier. This situation is frequently modeled through a Cox-proportional hazards model including covariates, some of which are under the control of the experimenter. When the model is to be fitted to n observed times these are known and for each of them it is also known whether that time is censored or not. When the experiment is to be designed neither the observed times nor the information about whether a particular unit will be censored are known. For censoring some additional prior probability distribution has to be assumed. Thus, the design problem faces two sources of imprecision when the experiment is to be scheduled. On the one hand, the censored times are not known. On the other hand, there is uncertainty about occurrence of censoring. A prior probability distribution is needed for this. Moreover, the Cox partial likelihood instead of the full likelihood is usually considered for these models. A partial information matrix is built in this case and optimal designs are computed and compared with the optimal designs for the full likelihood information. The usual tools for computing optimal designs with full likelihood are no longer valid for partial information. Some general results are provided in order to deal with this new approach. An application to a simple case with two possible treatments is used to illustrate it. The partial information matrix depends on the parameters and therefore a sensitivity analysis is conducted in order to check the robustness of the designs for the choice of the nominal values of the parameters.

1. Introduction

In survival analysis, one is interested in exploring the possible relationship between a survival time T and covariates X. For example, when a decision is to be made about the proportion of patients receiving an aggressive (in the sense of immediate side effects) treatment and the proportion to be put on a control group. The survival time is usually censored in a number of cases. Optimal experimental designs for models with incomplete data or potentially censored variables have been considered in the literature. Hackl (1995) proposed a procedure for computing D-optimal designs, assuming a monotonic probability distribution for missing responses. Imhof et al. (2002) suggested a general method for computing optimal designs when the probability of a missing response is positive. Garcet-Rodríguez et al. (2008) computed potentially censored designs, not in the survival sense. Balakrishnan and Han (2007) considered designs for a simple stepstress model under progressive type-II censoring. Recently, Balakrishnana and Kundub (2013) have provided a hybrid censoring scheme, which is a mixture of Type-I and Type-II censoring schemes.

López Fidalgo et al. (2009) considered a simple model for two parameters. In that case working with the full likelihood information was feasible from the estimation and the experimental design point of view. The problem was equivalent

* Corresponding author. Tel.: +34 926 295212; fax: +34 926 295361.
E-mail addresses: jesus.lopezfidalgo@uclm.es (J. López-Fidalgo), chusrl@usal.es (M.J. Rivas-López).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2012.10.009

to marginally restricted optimality (López Fidalgo and Garcet-Rodríguez, 2004; Martín-Martín et al., 2007). The approach considered in this paper is a continuation of that paper, considering now a Cox-proportional hazards model. Estimation for this model is usually based on the so called partial likelihood (Cox, 1975) rather than on the full likelihood. Thus, the main novelty of this work is the consideration of proportional hazards models and the computation of optimal designs through partial information as well as the inclusion of possible dropout. The concept of partial likelihood generalizes the ideas of conditional and marginal likelihood introduced by Cox (1972). For right censored and left truncated data Andersen et al. (1993) have shown that in most cases the properties of the maximum partial likelihood estimator (MPLE) are similar to those of the maximum full likelihood estimator (MLE). That is, the estimators of θ based on partial likelihood are consistent and asymptotically normal with a covariance matrix estimated consistently by the inverse of the observed information matrix. When there is more than one event at a given time several partial likelihoods have been proposed (see for instance Johnson and Klein, 2000).

Let $f(t|x)$, $S(t|x)$, $h(t|x)$ and $H(t|x)$ denote the conditional density, survival, hazard and cumulative hazard functions of T given the covariate X takes a particular value x. Additionally $H(t|x) = -\log(S(t|x))$.

A *Proportional hazards model* describing a covariate effect on survival time is traditionally used in this context and it is given by

$$h(t|x) = h_0(t)\exp\{\theta^T x\},$$

where $h_0(t)$ is the *baseline hazard function*, that is, the conditional hazard function of T at baseline conditions, and θ is a m-vector of regression parameters. For this model, survival and density functions can be written as

$$S(t|x) = S_0(t)^{\exp\{\theta^T x\}}, \qquad f(t|x) = S_0(t)^{\exp\{\theta^T x\}} h_0(t)\exp\{\theta^T x\},$$

where $S_0(t) = \exp\{-H_0(t)\}$ is the *baseline survival function* with $H_0(t) = \int_0^t h_0(u)du$.

In most applications the main interest is on the estimation of the regression coefficients θ and thus the baseline hazard function is left unspecified and estimated with nonparametric methods.

Let us assume the experiment is carried out during the time interval $[0, d]$, with all n individuals starting the experiment at time 0. In trials with time-to-event outcomes, some observations may be right censored either due to dropout during the study or because the event has not occurred before the end of the study.

Thus we have a data set with n observed times, some of them being distinct failure times (events) T, some of them being dropouts at time W and the rest being right censored times at time d. We assume that W is independent of failure time T.

For the ith individual in the experiment there is a pair (y_i, δ_i), where y_i is the observed time, event or censored time, and $\delta_i = I_{\{\text{event}\}}$ is an indicator of the event of interest for $i = 1, \ldots, n$.

Therefore the variable of the observed times will be

$$Y = \begin{cases} T & \text{if } \delta = 1, \\ W & \text{if } \delta = 0 \text{ and } W < d, \\ d & \text{if } \delta = 0 \text{ and } W > d, \end{cases} = \begin{cases} T & \text{if } T < \min\{W, d\}, \\ W & \text{if } W < \min\{T, d\}, \\ d & \text{if } d < \min\{T, W\}. \end{cases}$$

An introduction to optimal experimental design is given in Section 1.1. Then Sections 2 and 3 provide results for full and partial likelihoods respectively. An example is used to illustrate the procedure in both cases. Section 3 ends with a sensitivity analysis in respect of the choice of the nominal values of the parameters and a comparison of the designs computed for both cases. Finally a Discussion section with some comments is provided.

1.1. Optimal experimental design background

The optimal design theory for non-linear models is usually based on the Fisher information matrix (FIM), which depends on the parameters. Under mild conditions the inverse of this matrix is asymptotically proportional to the covariance matrix of the MLEs. As mentioned above in this paper partial information in the Cox sense will be used instead. Locally optimal designs will be computed for some nominal values of the parameters. Some basic ideas about the theory of optimal experimental design are provided in what follows.

A general nonlinear model where the response y is modeled by a probability distribution with probability density function (pdf) $f(y; x, \theta)$ is considered. The vector of variables x will take values on some compact space χ. Any probability measure ξ on χ will be referred to as an *approximate design. Exact designs of size n* concentrate masses $\xi(x_i)$ at different points x_i, $i = 1, \ldots,$ subject to the restriction that $n\xi(x_i) = n_i$ is an integer for all i. An exact design specifies that the experimenter has to take n_i observations at x_i, $\sum_i n_i = n$.

If uncorrelated observations are assumed the resulting covariance matrix of the maximum likelihood estimator (MLE) of θ is $\sigma^2 n^{-1} I^{-1}(\xi)$, where

$$I(\xi) = \int_\chi \mathrm{E}\left(-\frac{\partial^2 LL(x,\theta)}{\partial\theta^2}\right)\xi(dx)$$

is the FIM and $LL(x, \theta)$ is the log-likelihood function of an observation y at point x. In this paper, the partial likelihood is used instead. The decision about what treatment should be applied to each unit is made at the time all units enter the study.