ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Model robust designs for survival trials

Maria Konstantinou [a], Stefanie Biedermann [b,*], Alan Kimber [b]

[a] Ruhr-Universität Bochum, Fakultät für Mathematik, Germany

[b] University of Southampton, Statistical Sciences Research Institute, UK

ARTICLE INFO

Article history:
Received 28 February 2016
Received in revised form 25 July 2016
Accepted 14 October 2016
Available online xxxx

Keywords:
Proportional hazards models
Minimax optimal designs
D-optimality
c-optimality
Type-I censoring

ABSTRACT

The exponential-based proportional hazards model is often assumed in time-to-event experiments but may only approximately hold. Deviations in different neighbourhoods of this model are considered that include other widely used parametric proportional hazards models and the data are assumed to be subject to censoring. Minimax designs are then found explicitly, based on criteria corresponding to classical *c*- and *D*-optimality. Analytical characterisations of optimal designs are provided which, unlike optimal designs for related problems in the literature, have finite support and thus avoid the issues of implementing a density-based design in practice. Finally, the proposed designs are compared with the balanced design that is traditionally used in practice, and recommendations for practitioners are given.

1. Introduction

Optimal experimental designs are often constructed assuming that the model generating the data is known, up to the values of the parameters involved. In many practical situations, however, the proposed parametric model may only be approximately true and thus may cause the vector of parameter estimators to be biased. As illustrated by Box and Draper (1959) for the case of a linear regression model, the advantages of using an optimal design that minimises just the variance are lost even if the deviations from the assumed model are small.

Following Box and Draper (1959), robust designs for approximately linear regression have been constructed by Wiens (1992) based on classical optimality criteria but involving the mean squared error matrix. He finds minimax designs which are optimal in that they minimise the criteria functions for the worst possible deviation from the linear regression model. Prediction and extrapolation problems with possible heteroscedasticity are studied by Wiens (1998) and Fang and Wiens (1999) respectively among others. Sinha and Wiens (2002) consider the construction of sequential designs which are robust against model uncertainty for nonlinear models. Further results on misspecified nonlinear regression include Woods et al. (2006), Wiens and Xu (2008) and Xu (2009a) for prediction and extrapolation problems.

However, none of these authors considers the case where the data are subject to censoring. This arises in many time-to-event experiments when a particular event of interest is not observed for some of the subjects utilised in the experiment. Censoring is often a result of the fact that the experiments are not run as long as necessary in order to obtain complete data, that is, event times for all the subjects, because of time and cost limitations. Therefore, it is of interest to find optimal designs which are robust to misspecifications of the assumed model and which allow for the possibility of the data being censored.

* Correspondence to: University of Southampton, Highfield Campus, University Road, School of Mathematics, Building 54, Room 9027, SO17 1BJ Southampton, UK.

E-mail address: S.Biedermann@soton.ac.uk (S. Biedermann).

http://dx.doi.org/10.1016/j.csda.2016.10.013
0167-9473/© 2016 Published by Elsevier B.V.

Please cite this article in press as: Konstantinou, M., et al., Model robust designs for survival trials. Computational Statistics and Data Analysis (2016), http://dx.doi.org/10.1016/j.csda.2016.10.013

The available literature on model robust designs for time-to-event data is focused on accelerated life tests for which the subjects are put under extreme conditions in order for the event of interest to occur sooner than under normal circumstances. In this case, extrapolation to lower covariate values and prediction problems is often of interest; see, for example, Pascual and Montepiedra (2003), Xu (2009b) and McGree and Eccleston (2010).

An alternative class of models used for the modelling of time-to-event data is studied, namely that of proportional hazards models. Such models satisfy the proportional hazards assumption of constant hazard ratio over time and are frequently used in practice because of the simple interpretation of the regression coefficients in terms of hazard ratios. When a specific distribution is assumed for the event times, the resulting parametric models are referred to as distribution-based proportional hazards models. Cox's proportional hazards model, on the other hand, leaves the underlying distribution unspecified and therefore inference is based on the partial likelihood function (see Collett, 2003 for further details).

Konstantinou et al. (2015) consider Cox's model and show that in the presence of Type-I censoring an exponential distribution can be assumed without greatly affecting the optimal choice of design for partial likelihood estimation. They also find that the full and partial likelihood approaches result in very similar designs for the same assumed model.

Following these findings, small deviations in a neighbourhood of the exponential-based proportional hazards model are considered. The model uncertainty is formulated via a contamination function and the data are assumed to be subject to Type-I censoring. Then following along the same lines as in Xu (2009b), both censoring and model uncertainty are incorporated to obtain the asymptotic properties of the maximum likelihood estimator. Based on the asymptotic mean squared error matrix, minimax optimal designs for full likelihood estimation are constructed which protect against the worst possible misspecification of the assumed exponential model. Note that Xu (2009b) considers various prediction and extrapolation problems for normally distributed data and investigates the construction of designs that are continuous with respect to the Lebesgue measure. However, the focus of the present paper is on designs with finite support. This allows for explicit solutions to be obtained and then compared with the corresponding results of Konstantinou et al. (2014) for the case of the assumed model being true.

In Section 2 the assumed and true models considered are introduced and two different classes of contamination functions are defined to account for the various forms of the true distribution for the data. Then in Section 3 the asymptotic properties of the maximum likelihood estimator for the parameter vector are derived under model uncertainty and Type-I censoring. Analytical characterisations of minimax c- and D-optimal designs are given in Section 4. These designs are found using criteria corresponding to the classical c- and D-optimality criteria but are based on the mean squared error matrix rather than just the information matrix. In Section 5 the behaviour of the proposed designs is illustrated and they are compared with the balanced design traditionally used in practice. Finally, the main conclusions are discussed in Section 6.

2. Models and contamination functions

Time-to-event experiments are usually conducted in order to evaluate a particular intervention or treatment. Therefore, in what follows the focus is on models that involve one explanatory variable x. The mean squared error matrix for general designs is derived, and then design search is illustrated for the situation in which x takes values in the binary design space denoted by $\mathcal{X} = \{0, 1\}$, corresponding, for example, to a placebo and an active treatment in a clinical trial.

The aim of the experiment is assumed to be the estimation of one or both of the two model parameters. Let c be the predetermined duration of the experiment at which point the observations of subjects for which the event of interest has not occurred are said to be right-censored. Possibly censored data are summarised mainly using the hazard function which expresses the risk of the event of interest occurring at any time after the commencement of the experiment (Collett, 2003).

Consider the situation where the experimenter assumes the exponential-based proportional hazards model specified by the hazard function

$$h_1(t) = \exp\{\alpha + \beta x\}, \quad t > 0,\ x \in \mathcal{X} \subseteq \mathbb{R}, \tag{2.1}$$

where α and β are real parameters, when in fact this is only an approximation to the true underlying model. Denote the hazard function of the unknown true model by

$$h_2(t) = \exp\left\{\alpha + \beta x + \frac{g(t)}{\sqrt{n}}\right\}, \quad t > 0,\ x \in \mathcal{X} \subseteq \mathbb{R},\ g(t) \in \mathcal{G}, \tag{2.2}$$

where n denotes the sample size. The function $g(t)$ represents uncertainty about the exact form of the underlying distribution for the data and, following the literature, it is called the contamination function or just the contaminant. It is assumed that $g(t)$ is unknown and ranges in a neighbourhood specified by the class $\mathcal{G}$.

The parametrisation in (2.2) allows one to remain within a proportional hazards framework and ensures that the model parameters are well defined. In particular, unlike the existing literature, see, for example, Wiens (1992), the contamination function is independent of the covariate value x. Therefore, the parameter β corresponds to the effect of the explanatory variable. For identifiability reasons it is further required that $g(t)$ does not involve an additive constant. If this were not the case, the constant term would be absorbed in the quantity $\exp\{\alpha\}$ that represents the baseline hazard for model (2.1), that is, the hazard function for a subject with $x = 0$.

The factor $n^{-1/2}$ is included so that the deviations are of the order $O(1/\sqrt{n})$, resulting in models that are in a neighbourhood of the exponential model (2.1). At the same time, the dependence of g on the time t ensures that the general