

Optimal experimental designs for accelerated failure time with Type I and random censoring

María J. Rivas-López^{*,1}, Jesús López-Fidalgo², and Rodrigo del Campo³

¹ Department of Statistics, University of Salamanca. Plaza de los Caídos, 37008-Salamanca, Spain

² Department of Mathematics, University of Castilla-La Mancha. Avda. Camilo José Cela 3, 13071-Ciudad Real, Spain

³ Department of History and Fundamentals of Economics. University of Salamanca, 37008-Salamanca, Spain

Received zzz, revised zzz, accepted zzz

Summary

Proportional Hazards models have been widely used to analyze survival data. In many cases survival data do not verify the assumption of proportional hazards. An alternative to the PH models with more relaxed conditions are Accelerated Failure Time models. These models are fairly commonly used in the field of manufacturing, but they are more and more frequent for modeling clinical trial data. They focus on the direct effect of the explanatory variables on the survival function allowing an easier interpretation of the effect of the corresponding covariates on the survival time. Optimal experimental designs are computed in this framework for Type I and random arrival. The results are applied to clinical models used to prevent Tuberculosis in Ugandan adults infected with HIV.

Key words: Acceleration Factor; Log-Logistic Distribution; Optimal Experimental Design; Proportional Hazards; Survival Models.

1 Introduction

Proportional Hazards (PH) models have been widely used to analyze survival data. López Fidalgo et al. [2009] obtained optimal designs for a simple survival regression model. Later on Müller [2013] used a similar framework and gave some theoretical results. Cox [1975] in a seminal work advised the use of partial likelihood for dealing with the estimation of PH models and it has been used widely since then. Optimal experimental designs for partial likelihood information were computed by López Fidalgo and Rivas-López [2013]. The use of partial likelihood for fitting the parameters in a Cox model is rather simple. Nevertheless, the procedure to obtain optimal designs for maximum partial likelihood is a difficult task. Konstantinou et al. [2014, 2013] provided more insight about this in some particular cases.

If the evidence of non-proportionality for one or more covariates is demonstrated or at least suspected, two popular methods are frequently used to deal with this, either the stratified Cox model or the Cox regression model with time-dependent variables. Accelerated failure time (AFT) models could be an alternative to the PH models with more relaxed conditions. The AFT models focus on the direct effect of the explanatory variables on the survival function instead of the hazard function in the PH models. This property allows for an easier interpretation measuring directly the effect of the corresponding covariate on the survival time T . AFT models have not been used much for analyzing clinical trial data, although they are fairly common in the field of manufacturing. Nevertheless, there is an increasing interest in using AFT models for clinical trials as some literature reflects: Collett [2003], Lawless [1982], Cox and Oakes [1984] or Orbe and Núñez Antón [2006]. Bai and Chung [1991] considered optimal designs for accelerated life

* Corresponding author: e-mail: chusr@usal.es, Tel. 923 295300

tests in which two levels of stress were constantly applied and the failed test items were replaced with new ones using an exponential distribution for a log–linear function of stress.

Wei [1992] proposed AFT models as a useful alternative to the Cox regression model in survival analysis and Keiding et al. [1997] considered that, in survival analysis, deviations from proportional hazards may sometimes be explained by unaccounted random heterogeneity, or frailty. They provided an overview of the literature on omitted covariates in survival analysis and showed in a case study how unstable frailty models might behave when asked to account for unobserved heterogeneity in standard survival analysis with no replications per heterogeneity unit. Then, they showed that accelerated failure time modeling seems to avoid these difficulties. Bagdonavicius and Nikulin [2002] provided an interesting introduction to these models in chapter 2.3.

Kay and Kinnersley [2002] dealt with a case study in influenza and stated that the AFT model should be considered as an alternative to the proportional hazards model in the analysis of time to event data, especially in applications where the effects of treatment are to accelerate (or delay) the event of interest with no permanent effect in the context of the follow-up period. Lambert et al. [2004] applied parametric AFT models with random effects and decided these models seem to be a good fit to the kidney transplant survival data.

Qi [2009] presented the details of a randomized placebo-controlled trial to prevent Tuberculosis (TB) in Ugandan adults infected with HIV. They fitted the data set using non-parametric methods, Cox PH model, Cox model with time-dependent variables, piecewise exponential model (Breslow, 1974) and the AFT model. After fitting all of the models and the assessment of goodness-of-fit, they found that the Log-Logistic AFT model is the best fit for their data set. Thus, in this paper optimal experimental designs will be computed for the Log-Logistic AFT model. The results will be applied to the example considered by Qi [2009] (Section 6). The procedures developed here can be extended to other models in a similar way. In Section 3 the Fisher Information matrix is derived for Type I censoring. Sections 3, 4 and 5 will show general computations for different assumptions of AFT models. Section 1.1 will give an introduction to the optimal experimental design theory used in this paper and Section 1.2 will introduce the general AFT model.

1.1 Optimal experimental design background

Let x be the vector of covariates, say an experimental condition, which can be chosen from a compact *design space*, χ , typically a product of intervals. At this point a response variable t is observed. This is considered as a random variable from a parametric family of distributions, where β is the vector of parameters. An *exact design* of size n is defined by a collection of experimental conditions x_1, x_2, \dots, x_n in χ , where some of them may be repeated. Thus, a probability measure can be defined with support on the different points and weights proportional to the number of repetitions (replicates). This leads to an extension of this definition to any probability measure, ξ , the so called *approximate design*, which will be used in this paper.

The Fisher Information Matrix (FIM) associated with this design will be

$$M(\beta; \xi) = \int_{\chi} I(\beta; x) \xi(dx), \quad I(\beta; x) = \mathbf{E} \left(- \frac{\partial^2 LL(\beta; t, x)}{\partial \beta \partial \beta'} \right),$$

where $LL(\beta; t, x)$ is the log-likelihood function of the observation t for a particular experiment x .

In particular, it is well known that restricting to approximate designs with finite support preserves the whole space of information matrices. This is a consequence of the application of Caratheodory's theorem. Thus, we may confine ourselves to these designs. Since the covariance matrix of the estimators is asymptotically proportional to the inverse of the FIM (Silvey, 1980), an optimality criterion will be based on a function of (the inverse of) the FIM. Two criteria will be considered in this paper.

The most popular is D–optimality, which maximizes the determinant of the FIM. The statistical interpretation of this criterion is the minimization of the volume of the confidence ellipsoid of the parameters

of the model. It is easy to check that if the support of the D-optimal design is the number of parameters of the model, which is rather frequent, then the determinant is the product of the weights of the design and a function depending just on the support points. Then the product of the weights is maximized for equal weights. Thus, the best of this type of designs may be computed in an easy way and then the so called General Equivalence Theorem (GET) can be used to check whether this is actually the D-optimal design or not. The GET for D-optimality states that a design ξ_D is D-optimal if and only if

$$d(\xi_D, \beta; x) = \text{tr} (M^{-1}(\beta; \xi_D)I(\beta; x)) \leq m, x \in \chi,$$

where m is the number of parameters. Moreover, the equality is reached at the support points of the design.

When the interest is in estimating a function of the parameters, e.g. one of the parameters, c-optimality is used. This criterion focuses on estimating a linear combination of the parameters, $c'\beta$, where vector c defines the linear combination. For instance, $c = (1, 0, \dots, 0)$ means the interest is just in the first parameter. For this criterion Elfving [1952] provided a graphical method to construct c-optimal designs based on the convex hull of the union of the curve defined by the regressors and its reflection through the origin (see e.g. Figures 2 and 4). Then, the point where the line defined by vector c cuts the boundary determines the c-optimal design. The method is quite simple for dimension two, but for higher dimension it is not easy to be applied in its present form (López Fidalgo and Rodríguez-Díaz, 2004).

In what follows the Elfving method is described for two parameters in a design space $\chi = [a, b]$. Let consider a model with FIM at point $s \in \chi$ of the type

$$I(s) = g(s)g'(s), g'(s) = (g_1(s), g_2(s)).$$

This is the usual case of the general theory. The function $g(s)$ may depend on the parameters for a non-linear model. In such a case nominal values of the parameters have to be assumed in order to compute locally c-optimal designs. The Elfving locus is the convex hull of the set formed by the curve $g(\chi)$ and its reflection $-g(\chi)$. Curve $g(\chi)$ can be depicted using the parametric equations,

$$\begin{aligned} X(s) &= g_1(s), \\ Y(s) &= g_2(s), \end{aligned}$$

for any value of $s \in \chi$. Figures 2 and 4 show typical cases.

The boundary of the convex hull is part of the curves and segments linking two points of the curve, which may include extreme and tangential points. The simplest case is joining the extremes of the curves given by the extremes of the design space, e.g. $(g_1(a), g_2(a))$ from the original curve and $(-g_1(b), -g_2(b))$ from the reflected one. If a segment in the boundary has to link a extreme point, say $(-g_1(a), -g_2(a))$, with a tangential point, say $(g_1(s_t), g_2(s_t))$, then the following equation gives the value of s_t ,

$$\frac{g_1(s) + g_1(a)}{g_1'(s)} = \frac{g_2(s) + g_2(a)}{g_2'(s)}.$$

There are more complex situations which can be solved with basic geometry.

The c-optimal design is given by the boundary point meeting the straight line defined by vector $c' = (c_1, c_2)$, say (x^*, y^*) . This point is then a convex combination of the corresponding two points of the curves, e.g. $(g_1(s_1), g_2(s_1))$ and $(-g_1(s_2), -g_2(s_2))$. The c-optimal design is then a design,

$$\xi_c = \left\{ \begin{array}{cc} s_1 & s_2 \\ 1 - p^* & p^* \end{array} \right\},$$

where p^* comes from the convex combination of the two points. For that it is enough to solve the equation given by one of the components, e.g. $x^* = (1 - p)g_1(s_1) - pg_1(s_2)$, to obtain the optimal weight in this case,

$$p^* = \frac{g_1(s_1) - x^*}{g_1(s_1) + g_1(s_2)}.$$

Detailed introductions to the theory of optimal experimental design are provided, e.g., by Pázman [1986], Pukelsheim [1993], Fedorov and Hackl [1997], Walter and Pronzato [1997], Atkinson et al. [2007], Berger and Wong [2009], Goos and Jones [2011], among others.

1.2 Accelerated Failure Time models background

An AFT model is written mathematically through the survival function,

$$S(t|x, \beta) = S_0(t/\eta(x, \beta)),$$

where S_0 is the *baseline survival function* and $\eta(x, \beta)$ is an *acceleration factor*, which speeds up the effect of the survival time in the survival function. The acceleration factor is typically the exponential of a linear relationship,

$$\eta(x, \beta) = \exp(\beta'x).$$

Under AFT models the covariates effects are assumed to be constant and multiplicative on the time scale, that is, the covariates impact on survival through the acceleration factor, just mentioned. According to the relationship between the survival and the hazard functions

$$h(t|x, \beta) = \frac{1}{\eta(x, \beta)} h_0(t/\eta(x, \beta)).$$

The corresponding log-linear form of the AFT model with respect to time may be reduced to (e.g. Orbe and Núñez Antón, 2006)

$$\log(T) = \mu + \theta'x + \sigma\varepsilon,$$

where $\mu \equiv E(\log T_0)$ is the intercept, $\sigma \equiv \text{var}(\log T_0)$ the scale parameter and ε is a random variable with a particular distribution and mean 0. The random variable T_0 follows the distribution given by the survival baseline. This form is adopted by most software packages for the AFT model.

For each distribution of ε , there is a corresponding distribution for T . As mentioned above the AFT models are discussed in detail in the textbooks by Collett [2003], Cox and Oakes [1984] and Lawless [1982]. The AFT models are usually named by the distribution of T rather than the distribution of ε or $\log(T)$. The survival function of T can be expressed by the survival function of ε ,

$$S(t|x; \mu, \theta) = S_\varepsilon \left(\frac{\log(t) - \mu - \theta'x}{\sigma} \right).$$

Thus, the probability density function (pdf) is

$$f(t|x; \mu, \theta) = f_\varepsilon \left(\frac{\log(t) - \mu - \theta'x}{\sigma} \right) \cdot \frac{1}{\sigma t}.$$

The time ratio comparing two levels of a covariate (e.g. $x = 1$ vs $x = 0$), after controlling all the other covariates, is then $\exp(\theta)$. This is interpreted as the estimated ratio of the expected survival times for two groups. A ratio above 1 means that the time is accelerated by a constant factor, while a ratio below 1 means that the time is slowed down by a constant factor. Therefore, AFT models can be interpreted in terms of the rate of progression of a disease. The effect of the covariates in an accelerated failure time model is equivalent to changing the scale, but not the location of a baseline distribution of survival times.

Next sections will be devoted to compute the FIM for different situations. For that some general results on information matrices for censored observations will be useful. For instance Efron and Johstone [1990], Gertsbakh [1995], Gertsbakh and Kogan [1999], Lawless [1982] or Fedorov and Leonov, 2013 in Ch.1.6 and Ch.5.2 offered interesting results (analytic formulas) on the use of elemental (single observation) matrices in design of experiments.

2 The Fisher Information Matrix for Type I censoring

In this section all the individuals in the study are recruited at time 0 and there is a follow-up until time d . The end of study is the only censoring considered here. Let $\delta \equiv 1_{\{T < d\}}$ be an indicator variable being 1 if the observation corresponds to an event and 0 if it corresponds to a censored time. Thus, the log-likelihood function for just one observation at value x is then

$$\begin{aligned} LL(\mu, \theta; t, x) &= \delta \log(f(t|x; \mu, \theta)) + (1 - \delta) \log(S(t|x; \mu, \theta)) \\ &= -\delta \log(\sigma t) + \delta \log(f_\varepsilon(z_t)) + (1 - \delta) \log(S_\varepsilon(z_t)), \end{aligned}$$

where

$$z_t = \frac{\log(t) - \mu - \theta'x}{\sigma}.$$

In order to find the FIM the log-likelihood is differentiated twice and the expectation is computed (the details can be found in the Appendix). Then the FIM for one particular observation is

$$I(\mu, \theta, \sigma^2; x) = \begin{pmatrix} A(x) & xA(x) & B(x) \\ xA(x) & x^2A(x) & xB(x) \\ B(x) & xB(x) & C(x) \end{pmatrix},$$

where

$$\begin{aligned} A(x) &= \mathbb{E} \left(-\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\ &= \frac{1}{\sigma^2} \left([f'_\varepsilon(z)]_{(-\infty)} + \frac{f_\varepsilon(z_d)^2}{S_\varepsilon(z_d)} + \int_{-\infty}^{z_d} \frac{f'_\varepsilon(z)^2}{f_\varepsilon(z)} dz \right), \\ B(x) &= \mathbb{E} \left(-\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu \partial (\sigma^2)} \right) \\ &= \frac{1}{2\sigma^3} \left([zf'_\varepsilon(z)]_{(-\infty)} + f_\varepsilon(z_d) + z_d \frac{f_\varepsilon(z_d)^2}{S_\varepsilon(z_d)} + \int_{-\infty}^{z_d} z \frac{f'_\varepsilon(z)^2}{f_\varepsilon(z)} dz \right), \\ C(x) &= \mathbb{E} \left(-\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial (\sigma^2)^2} \right) \\ &= \frac{1}{4\sigma^4} \left([zf_\varepsilon(z) + z^2 f'_\varepsilon(z)]_{(-\infty)} + S_\varepsilon(z_d) - 1 + 2z_d f_\varepsilon(z_d) + z_d^2 \frac{f_\varepsilon(z_d)^2}{S_\varepsilon(z_d)} + \int_{-\infty}^{z_d} z^2 \frac{f'_\varepsilon(z)^2}{f_\varepsilon(z)} dz \right). \end{aligned}$$

The following lemma may help to reduce last formulas in some typical cases. The proof can be made using L'Hopital rule.

Lemma 2.1 *Let Z be a random variable with pdf f_ε .*

- i) *If there exists the expectation $E(Z^r)$ then $[z^r f_\varepsilon(z)]_{(\pm\infty)} \equiv \lim_{z \rightarrow \pm\infty} z^r f_\varepsilon(z) = 0$, $r = 1, 2, \dots$*
- ii) *If f_ε is differentiable and the following limit exists, then $[f'_\varepsilon(z)]_{(\pm\infty)} \equiv \lim_{z \rightarrow \pm\infty} f'_\varepsilon(z) = 0$.*
- iii) *If f_ε is differentiable and the following limit exists, then $[zf'_\varepsilon(z)]_{(\pm\infty)} \equiv \lim_{z \rightarrow \pm\infty} z f'_\varepsilon(z) = 0$.*
- ii) *If f_ε is differentiable, there exists $E(Z)$ and the following limit exists, then $[z^2 f'_\varepsilon(z)]_{(\pm\infty)} = \lim_{z \rightarrow \pm\infty} z^2 f'_\varepsilon(z) = 0$.*

For an approximate design

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \cdots & x_k \\ p_1 & p_2 & \cdots & p_k \end{array} \right\}$$

the FIM is

$$M(\mu, \theta, \sigma^2; \xi) = \begin{pmatrix} \sum_{i=1}^k A(x_i)p_i & \sum_{i=1}^k A(x_i)p_i x_i & \sum_{i=1}^k B(x_i)p_i \\ \sum_{i=1}^k A(x_i)p_i x_i & \sum_{i=1}^k A(x_i)p_i x_i^2 & \sum_{i=1}^k B(x_i)p_i x_i \\ \sum_{i=1}^k B(x_i)p_i & \sum_{i=1}^k B(x_i)p_i x_i & \sum_{i=1}^k C(x_i)p_i \end{pmatrix}. \quad (1)$$

3 Type I censoring and known variance

In this section the variance σ^2 is assumed known for simplicity. Then the FIM will be given by the first two columns and rows of (1). If x is one-dimensional, and so θ , the FIM is of the form $Q(\tau)$ times the simple linear regression matrix, with $\tau = \mu + \theta x$. Konstantinou et al. [2014] considered a class of information matrices of this form. Checking whether $Q(\tau)$ satisfies their conditions may be used to compute easily D-optimal and c-optimal designs. Moreover, the complete class of results given by Yang [2010], Yang and Stufken [2012], Dette and Melas [2011], Dette and Schorning [2013] can be applied to prove that the optimal designs have 2 support points, one of which is a boundary point for given models with two parameters. Otherwise the equivalence theorem will be checked for each particular case. If it is not satisfied then three-point designs with any possible weights have to be considered.

3.1 Log-Logistic survival time

A Log-Logistic survival time is considered here in order to be applied later to the real example mentioned in the introduction. Thus, the variable time-to-event, T , is assumed to follow a Log-Logistic distribution, i.e. ε follows a Logistic distribution with pdf

$$f_\varepsilon(z) = \frac{e^z}{(1 + e^z)^2}, \quad -\infty < z < \infty$$

and survival function

$$S_\varepsilon(z) = \frac{1}{1 + e^z}, \quad -\infty < z < \infty,$$

where σ is again assumed known.

The FIM for a single point x of a one-dimensional covariate is

$$I(\mu, \theta; x) = \frac{1}{3\sigma^2} \left(1 - \frac{1}{(1 + e^{za})^3} \right) \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}, \quad x \in [a, b].$$

This is actually the FIM of a linear model with regressors

$$g(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \end{pmatrix} = \sqrt{\frac{1}{3\sigma^2} \left(1 - \frac{1}{(1 + e^{za})^3} \right)} \begin{pmatrix} 1 \\ x \end{pmatrix}, \quad x \in [a, b]. \quad (2)$$

Exchanging x with $\Delta = \frac{\log d - \mu - \theta x}{\sigma}$, the D- and c-optimal designs for this information matrix are the same as those for the linear models with regressors

$$\sqrt{1 - \frac{1}{(1 + e^\Delta)^3}} \begin{pmatrix} 1 \\ \Delta \end{pmatrix}, \quad \Delta \in \chi^\Delta.$$

The design interval is now $\chi^\Delta = [\Delta_b, \Delta_a]$ if $\theta > 0$ and $\chi^\Delta = [\Delta_a, \Delta_b]$ if $\theta < 0$, where $\Delta_a = \frac{\log d - \mu - \theta a}{\sigma}$ and $\Delta_b = \frac{\log d - \mu - \theta b}{\sigma}$. The FIM for a design ξ will be

$$M(\beta; \xi) = \sum_{\Delta} \xi(\Delta) \left(1 - \frac{1}{(1 + e^\Delta)^3} \right) \begin{pmatrix} 1 & \Delta \\ \Delta & \Delta^2 \end{pmatrix}, \Delta \in \chi^\Delta.$$

Thus the D- and c-optimal designs depend on the parameters only through the design space χ^Δ .

D-optimal designs

If the D-optimal design is supported at two points, say Δ_1 and Δ_2 , then the weights should be the same, 1/2. The determinant of the FIM for these designs is proportional to

$$\frac{(e^{\Delta_1} (e^{\Delta_1} + 3) + 3) (e^{\Delta_2} (e^{\Delta_2} + 3) + 3) e^{\Delta_1 + \Delta_2} (\Delta_1 - \Delta_2)^2}{(e^{\Delta_1} + 1)^3 (e^{\Delta_2} + 1)^3}.$$

Figure 1 shows the shape of this function. It is easy to check analytically that the derivatives cannot vanish

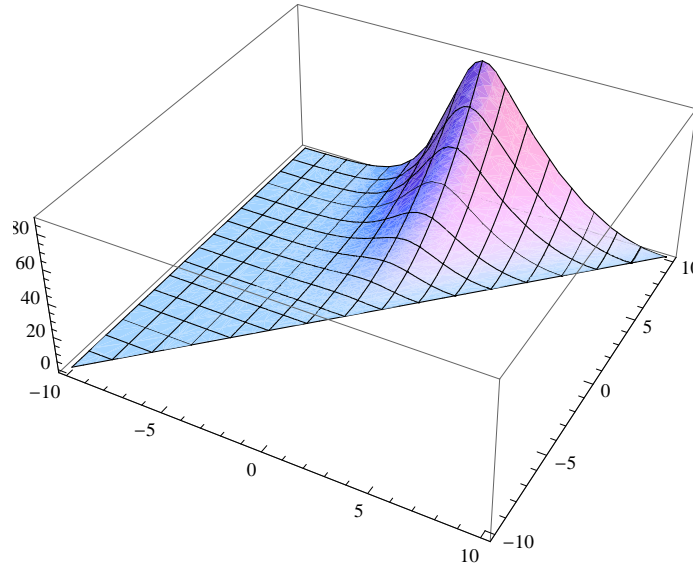


Fig. 1 Shape of the determinant of the FIM for a two-point equally weighted design for $\Delta_1 \leq \Delta_2$ for the Log-Logistic model with known variance and Type I censoring

simultaneously for any values of Δ_1 and Δ_2 . Therefore, the maximum must be reached at the boundary. Fixing $\Delta_2 = \Delta_{max}$ (where $\Delta_{max} = \Delta_a$, for $\theta > 0$ and $\Delta_{max} = \Delta_b$, for $\theta < 0$), the derivative with respect to Δ_1 vanishes at the unique root, $\Delta_D(max)$, of

$$3(\Delta - \Delta_{max} + 2) + 2e^\Delta (e^\Delta (e^\Delta + 4) + 6) \tag{3}$$

in $[-\infty, \Delta_{max}]$. The maximum is reached at point $(\Delta_D(max), \Delta_{max})$. Thus, if $\Delta_D(max) < \Delta_{min}$ then the two-point equally weighted design maximizing the determinant is supported at a and b , otherwise it will be supported at a and $x_D(a) = \frac{\log d - \mu - \sigma \Delta_D(a)}{\theta}$, for $\theta > 0$, or at b and $x_D(b) = \frac{\log d - \mu - \sigma \Delta_D(b)}{\theta}$, for $\theta < 0$.

In order to check whether this design, say ξ_D , is actually D-optimal the equivalence theorem can be easily checked numerically,

$$d(\xi; x) = \text{tr} (M^{-1}(\mu, \theta; \xi_D) I(\mu, \theta; x)) \leq 2, x \in [a, b].$$

c-optimal designs

The Elfving method described in the introduction will be used to compute c -optimal designs for the model defined by (2). Figure 2 shows the Elfving locus for a complete case according to the possible extremes of the interval $[a, b]$. The curves may be reduced by any of the to extremes shortening the interval $[a, b]$. According to this there are different possible Elfving locus, which are reduced to two possibilities. Fixing a and considering b large enough, let $x_c(a)$ be the tangential point to the reflected curve of a straight line passing through the point defined by a . Notice there is a symmetric situation, which is equivalent. The boundary of the Elfving locus is then constructed with a segment joining the points defined by a and $x_c(a)$ if $b \geq x_c(a)$. Otherwise the boundary is formed by a segment joining the points defined by a and b . Therefore, there are two types of c -optimal designs supported either at the two points just mentioned or else at one-point if the line defined by vector c cuts the boundary in the curves. In particular, for estimating parameter θ a two-point design is always needed. The c_μ -optimal design for estimating μ will be a two-point design if $a > 0$ or $b < 0$. Otherwise, a one-point design will be c_μ -optimal, supported at 0, because the vector $c = (1, 0)$ will cut the Elfving locus at a point with $g_2(x) = \sqrt{A(x)}x = 0$ corresponding to $x = 0$.

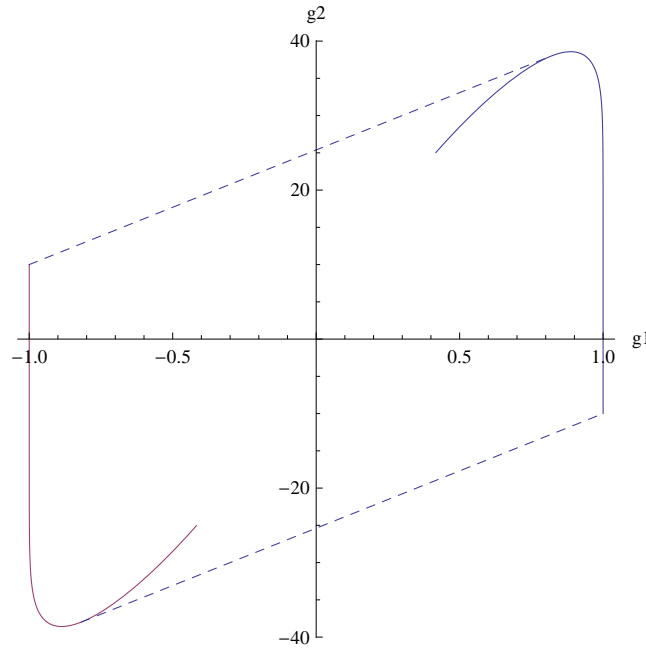


Fig. 2 Elfving locus for the Log-Logistic model with known variance and Type I censoring

4 Random censoring and known variance

Let us assume now a patient enters the study randomly at time W independently from others with a known probability distribution on the interval $[0, d]$. Let $r(c)$ be the pdf of the random variable $C = d - W$ (possible censoring time for this patient). Given a particular value $C = c$ the conditional pdf for the time-to-event T is the one given in the previous section replacing d by c . For simplicity of notation let us call again the conditional distribution of ε , $f_\varepsilon(z|c)$. The joint pdf of (ε, c) will be the product $r(c)f_\varepsilon(z|c)$. Thus exchanging d with c in (1) and integrating we obtain the conditional expectation. Then computing

the marginal expectation with respect to c we get the FIM,

$$I(\mu, \theta; x) = \frac{1}{\sigma^2} \int_0^d \left(\frac{f_\varepsilon(z_c)^2}{S_\varepsilon(z_c)} + \int_{-\infty}^{z_c} \frac{f'_\varepsilon(z)^2}{f_\varepsilon(z)} dz \right) r(c) dc \quad xx'.$$

4.1 Log-Logistic survival time

Assuming uniform random censoring, $r(c) = 1/d$, $c \in [0, d]$, the FIM is

$$I(\mu, \theta; x) = \frac{1}{3\sigma^2 d} \int_0^d \left(1 - \frac{1}{(1 + e^{z_c})^3} \right) dc \begin{pmatrix} 1 & x \\ x & x^2 \end{pmatrix}.$$

In this case an analytic expression cannot be obtained, but the procedure is similar both for D- and c-optimality. Now,

$$g(x) = \begin{pmatrix} g_1(x) \\ g_1(x) \end{pmatrix} = \sqrt{\frac{1}{3\sigma^2 d} \int_0^d \left(1 - \frac{1}{(1 + e^{z_c})^3} \right) dc} \begin{pmatrix} 1 \\ x \end{pmatrix}, \quad x \in [a, b].$$

The same procedure applies for any possible distribution of the arrivals.

5 Unknown variance

So far the variance has been assumed known in order to show the methodology in a simple case. In a more realistic setup all the parameters can be considered unknown, including more covariates. If σ^2 is unknown and needs to be estimated the FIM has been computed in Section 3. The determinant of the FIM given by (1) is then

$$\det(M) = \sum_{i=1}^n C(x_i) p_i \sum_{i=1}^n A(x_i) p_i \sum_{j>i}^n A(x_j) p_j (x_j - x_i)^2 - \sum_{i=1}^n A(x_i) p_i \left(\sum_{j \neq i}^n B(x_j) p_j (x_j - x_i) \right)^2.$$

It will be specified now for the Log-Logistic model.

5.1 Log-Logistic survival time

$$\begin{aligned}
A(x) &= \frac{1}{3\sigma^2} \left(1 - \frac{1}{(1 + e^{z_d})^3} \right) = \frac{1}{3\sigma^2} (1 - S_\varepsilon(z_d)^3), \\
B(x) &= \frac{1}{6\sigma^3} \left(z_d \left(1 - \frac{1}{(1 + e^{z_d})^3} \right) + \frac{e^{z_d}}{(1 + e^{z_d})^2} - \log(1 + e^{z_d}) \right) \\
&= \frac{1}{6\sigma^3} (z_d (1 - S_\varepsilon(z_d)^3) + f_\varepsilon(z_d) + \log(S_\varepsilon(z_d))), \\
C(x) &= \frac{1}{12\sigma^4} \left(z_d^2 \left(1 - \frac{1}{(1 + e^{z_d})^3} \right) + 2z_d \left(\frac{e^{z_d}}{(1 + e^{z_d})^2} - \log(1 + e^{z_d}) \right) + \frac{e^{z_d}}{(1 + e^{z_d})} \right. \\
&\quad \left. - 2 \int_{-e^{z_d}}^0 \frac{\log(1-t)}{t} dt \right) \\
&= \frac{1}{12\sigma^4} (z_d^2 (1 - S_\varepsilon(z_d)^3) + 2z_d (f_\varepsilon(z_d) + \log(S_\varepsilon(z_d))) + 1 - S_\varepsilon(z_d) \\
&\quad - 2 \int_{-e^{z_d}}^0 \frac{\log(1-t)}{t} dt).
\end{aligned}$$

As in the case of known variance, the D-optimal design is a two-point design. And again, for a design interval $[a, b]$, the D-optimal design is supported at an extreme point and a critical point ($\{a, x(a)\}$ for $\theta > 0$ and $\{x(b), b\}$ for $\theta < 0$). The weight for the extreme point is the largest. If the critical point is not in $[a, b]$ then the D-optimal design will be supported at a and b . Then, the further the critical point is from the interval the more similar the two weights are.

6 Experimental designs for prevention of Tuberculosis in patients infected with HIV

As mentioned in the introduction Qi [2009] considered a randomized placebo-controlled trial to prevent Tuberculosis (TB) in Ugandan adults infected with HIV. They fitted the dataset using different procedures and assessed the goodness-of-fit. They found that the Log-Logistic AFT model was the best fit for their data set. In what follows optimal experimental designs are computed for this model.

Let $x \in \mathcal{X} = [0, 1]$ be a covariable (dose) to be designed. Nominal values of the parameters and other constants were taken from this example, in particular, $d = 3$, $\sigma = 0.71$, $\mu = -2.97$ and $\theta = 0.1$.

6.1 Type I censoring and known variance

The root of (3) is then $\Delta_D(a) = -0.413706$, which corresponds to a value $x_D(a) = 43.6234 \notin [0, 1]$. Therefore the best two-point equally weighted design is supported at the extremes of the interval.

The Equivalence Theorem states that if

$$d(\mu, \theta; x) = \frac{1}{3\sigma^2} \left(1 - \frac{1}{(1 + e^{z_d})^3} \right) (1, x) M^{-1}(\mu, \theta; \xi_D) \begin{pmatrix} 1 \\ x \end{pmatrix} \leq 2$$

for any x in the design space, then the design ξ_D is actually a D-optimal design and the equality is reached at the support points of the design. Figure 3 (left) shows how the sensitivity function satisfies the Equivalence Theorem for this design and then it is D-optimal. Therefore, the equally weighted two-point design at 0 and 1, ξ_D , is actually D-optimal.

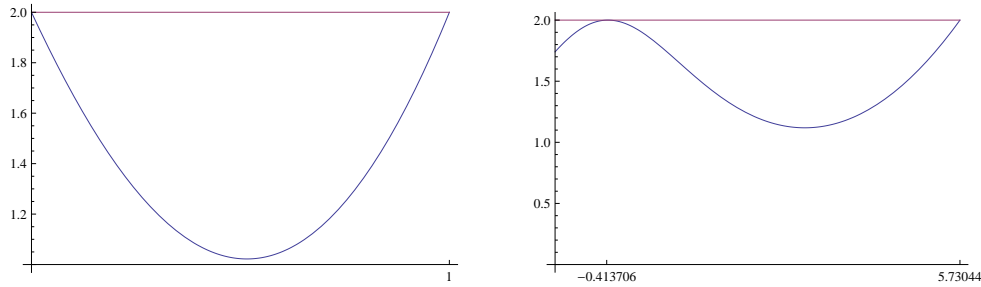


Fig. 3 Sensitivity function for the D-optimal design for the Log-Logistic with known variance and Type I censoring for $\chi = [0, 1]$ (left) and $\chi = [0, 50]$ (right). The later as function of Δ

In practice there is usually more interest in estimating parameter θ than the auxiliary parameter μ . Therefore a c_θ -optimal design for estimating optimally θ is computed using the Elfving graphical procedure. For this example the c_θ -optimal design is again ξ_D . On the other hand, the c_μ -optimal design for estimating μ puts all the mass at the lower limit of the interval.

Enlarging the design space to $\chi = [0, 50]$ the D-optimal design is the equally weighted design at points 0 ($\Delta_a = 5.73044$) and 43.6234 ($\Delta_D(a) = -0.413706$). This is actually the D-optimal design since the sensitivity function satisfies the Equivalence Theorem (Figure 3, right). Then the D-optimal equally weighted two-point design is

$$\xi_{I,K} = \begin{Bmatrix} 0 & 43.6234 \\ 1/2 & 1/2 \end{Bmatrix}.$$

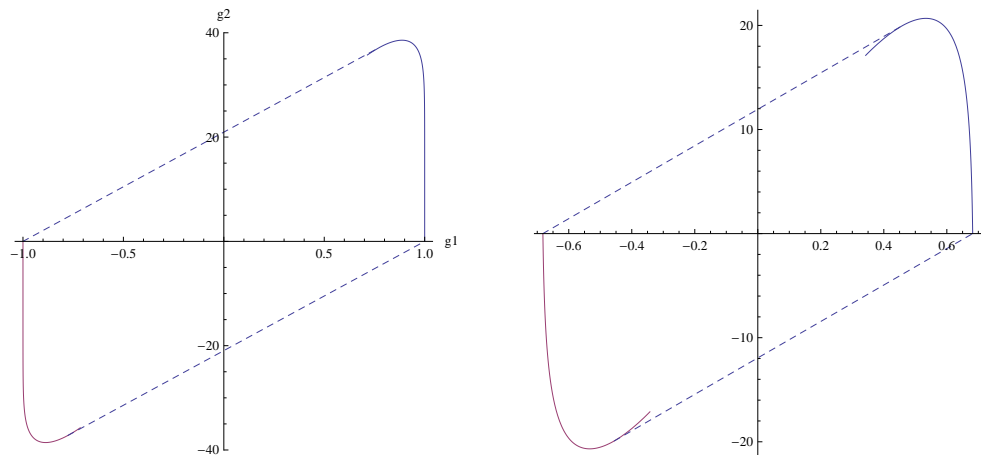


Fig. 4 Elfving locus for $\chi = [0, 50]$ for the Log-Logistic model with known variance and Type I censoring (left) and random censoring (right)

In this case the c -optimal designs have also a point in the interior of the interval (Figure 4, left). The tangential point in the first quadrant of the graphic comes from solving the equation

$$\frac{g_1(s) + 1}{g'_1(s)} = \frac{g_2(s)}{g'_2(s)},$$

which gives the solution $s_c = 47.97$. The c_θ -optimal design is given by the boundary point meeting the axis OY, say $(0, s_c)$, which is a convex combination of $(-1, 0)$ and $(g_1(s_c), g_2(s_c))$. It is enough to solve

Design space	Type I censoring		Random censoring	
	D-optimal	c_θ -optimal	D-optimal	c_θ -optimal
$[0, 1]$	1 (1/2)	1 (1/2)	1 (1/2)	1 (0.50012)
$[0, 50]$	43.6234 (1/2)	47.9735 (0.52025)	38.7614 (1/2)	43.7551 (0.59985)

Table 1 Optimal design second point and weight between brackets for the TB case. The first point is always 0

the equation $0 = -(1 - p) + pg_1(s^*)$ to obtain the optimal weight $p^* = 0.52025$. Thus the c_θ -optimal design in this case is

$$\xi_\theta = \left\{ \begin{array}{cc} 0 & 47.9735 \\ 0.47975 & 0.52025 \end{array} \right\}.$$

Notice the c_μ -optimal design for estimating μ puts again all the mass at the lower limit of the interval.

6.2 Random censoring and known variance

Assuming now there is random uniform arrival during the period of follow-up, $[0, d]$, similar results were obtained with slight differences. In particular, for $\chi = [0, 1]$ both the D-optimal and the c_θ -optimal design are supported at the extreme points with equal weights for the D-optimal design and weight $p = 0.50012$ for the c_θ -optimal design. If the design space is $\chi = [0, 50]$ then the D-optimal design, say $\xi_{R,K}$, is supported at 0 and 38.7614. On the other hand the c_θ -optimal design is supported at 0 and 43.7551 with weight $p = 0.59985$ for the second one.

Similar graphics are obtained for this situation. Figure 4 (right) shows the Elfving locus for this situation. All the results for this example are summarized in Table 1, where the second point and its weight, between brackets, are shown. The first point is always 0. It is remarkable that the critical points are a little bit lower for random censoring (see the Discussion for more details). The c_μ -optimal design is always concentrated at 0. The optimal designs for any design space $[0, b]$ with b after or before the critical points are those of the table. If the interval is more general, say $[a, b]$, then the D-optimal designs are those of the table depending of whether the critical point is in the interval. For c-optimality the critical (tangential) point depends on a and then it must be checked whether it is in the interval or not. Then the c_μ -optimal design can be a two-point design if $0 \notin [a, b]$.

6.3 Type I censoring and unknown variance

If $\chi = [0, 1]$ the corresponding design space for Δ will be $\chi^\Delta = [5.5896, 5.7304]$. The D-optimal design is the equally weighted two-point design at the extreme points, which it is actually the D-optimal design for known variance. Figure 5 (left) shows the fulfillment of the equivalence theorem for this design.

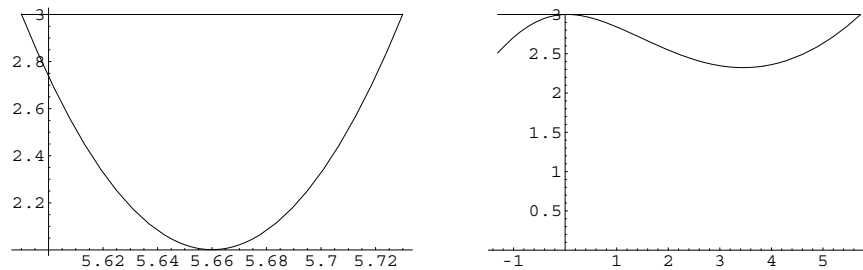


Fig. 5 Sensitivity function for $\chi = [0, 1]$ (left) and $\chi = [0, 50]$ (right)

	σ^2 unknown	
	Type I censoring	Random censoring
$\xi_{I,K}$	0.9191	0.7817
$\xi_{R,K}$	0.9699	0.9248
$\xi_{I,U}$	1	0.9440
$\xi_{R,U}$	0.9397	1

Table 2 Relative efficiencies of the D-optimal designs under different scenarios for $\chi = [0, 50]$

If $\chi = [0, 50]$ the corresponding design space for Δ will be $\chi^\Delta = [-1.3118, 5.7304]$. The D-optimal design is now

$$\xi_{I,U}(\Delta) = \left\{ \begin{array}{cc} 0.04297 & 5.7304 \\ 0.4156 & 0.5844 \end{array} \right\},$$

that is

$$\xi_{I,U}(x) = \left\{ \begin{array}{cc} 0 & 40.381 \\ 0.5844 & 0.4156 \end{array} \right\}.$$

Figure 5 (right) shows the equivalence theorem.

6.4 Random censoring with unknown variance

If $\chi = [0, 1]$ the D-optimal design is again the equally weighted two-point design at the extreme points. If $\chi = [0, 50]$ the D-optimal design is

$$\xi_{R,U}(x) = \left\{ \begin{array}{cc} 0 & 35.9493 \\ 0.595 & 0.405 \end{array} \right\}.$$

Table 2 shows relative efficiencies of the four optimal designs, $\xi_{I,K}$, $\xi_{R,K}$, $\xi_{I,U}$ and $\xi_{R,U}$, for estimating the three parameters when $\chi = [0, 50]$ and the two considered censoring schemes. Almost all of them are over 90%, what suggests that in this case the designs computed for known variance may not be so bad in practice for other censoring schemes. This may be used for saving a lot of computational burden, specially when more variables are introduced in the model. In particular, c-optimal designs for three parameters cannot be computed using directly the graphical Elfving method and significant higher computational cost is needed for that (López Fidalgo and Rodríguez-Díaz, 2004). Nevertheless, the efficiency of the optimal design for known variance and Type I censoring is less than 80% for the scenario with unknown variance and random censoring. This means some care has to be taken where random arrival is considered.

7 Discussion

Although PH models have been widely used to analyze survival data they are rather restrictive and more flexible models are required. For example, proportional hazards cannot be assumed, e.g. if the hazard ratio for two individuals is not constant over time. The AFT models seem to be an alternative more and more used in practice, not only in engineering, where they were first introduced, but also in the medical field.

Optimal experimental designs have been rarely computed in each of the two classes of models. This may be a consequence of the difficulties in outlining the problem and computing the designs. In particular, an additional distribution (model) for censoring has to be considered, while it will not be used for the final fitting. The use of partial likelihood for fitting the parameters in a Cox model is rather simple. Nevertheless, the procedure to obtain optimal designs for maximum partial likelihood is a difficult task.

In this paper a methodology has been developed to obtain optimal designs for an AFT model. The simpler case of known variance has been considered to show the details. Then the more realistic case of unknown variance, which makes the computations still tractable for D-optimality was investigated. The Log-Logistic distribution was assumed since it was the most appropriate for the example. Other distributions may not lead to analytic solutions. This actually happens in the case of random censoring and unknown variance considered in the paper. Numerical computations are feasible when no more analytic solutions are at hand. The censoring schemes used here are rather common in practical use. Assuming random arrivals of the patients is equivalent to consider possible dropouts at random.

It is noteworthy that, for both known and unknown variance, the support points of the D-optimal design are always an extreme point of the design space and an interior critical point (when it belongs to the design space). Thus, if the design space is the interval $[a, b]$ then the D-optimal design is supported at $\{a, x(a)\}$, for $\theta > 0$, or $\{x(b), b\}$, for $\theta < 0$. Moreover, for $\theta > 0$ the critical point is smaller for random censoring than for Type I censoring, and for $\theta < 0$ the critical point is larger for random censoring than for Type I censoring. This result might be intuitive. For random censoring the expected time a patient is followed up is shorter than for Type I censoring, so maybe the optimal design in this case needs to use a critical value where the probability of observing an event is greater (an event gives “more information” than a censored observation). For $\theta > 0$, the larger the covariate is the larger the expected survival time will be, so the probability of observing an event is lower, so in this case the critical value is smaller for random censoring than for Type I censoring. Reciprocally for $\theta < 0$.

In particular, a uniform distribution was assumed. At this respect there are further interesting considerations to be taken into account,

- i) For random arrivals a sequential design is quite natural in a survival study. When a particular patient enters the study at time w_{j+1} , j patients with their respective treatments have been so far in the study. At that moment it is known they have entered the study at times w_1, \dots, w_j . Therefore, at time w_{j+1} , the observed time for the i th patient ($i = 1, 2, \dots, j$) is $t_i \in (w_i, w_{j+1}]$ if he has already experienced the event, otherwise it is a censored time, $w_{j+1} - w_i$. Let assume, re-naming the subscripts without loss of generality, that the $n_j \leq j$ events occurred at that moment are t_1, \dots, t_{n_j} . With all this information an optimal treatment has to be assigned to the $(j+1)$ th patient at that moment. The log-likelihood function at the end of the study computed at time w_{j+1} will be,

$$\begin{aligned} LL(\mu, \theta; t, x) &= \log f(t_1) + \dots + \log f(t_{n_j}) \\ &+ \delta (\log P(T_{n_j+1} > t_{n_j+1}) + \dots + \log P(T_j > t_j) + \log P(T_{j+1} > t_{j+1})) \\ &+ (1 - \delta) (\log P(T_{n_j+1} > d - w_{n_j+1} | T_{n_j+1} > w_{j+1} - w_{n_j+1}) + \dots \\ &+ \log P(T_j > d - w_j | T_j > w_{j+1} - w_j) + \log P(T_{j+1} > d - w_{j+1})). \end{aligned}$$

From here the FIM has to be constructed, depending just on the possible treatment for the new patient, and then optimized. An arrival distribution for those who will enter the study in the future may be taken into account as well.

- ii) Other distributions can be considered. In particular, dropouts may be related to the treatment assigned to the patient. This influences very much the optimal design.
- iii) Optimizing the stopping time for the follow-up is also of much interest.
- iv) Incremental windows of increasing stress can also be optimized to produce faster experiments minimizing toxicity. This is typical for AFT models.

Appendix: Derivation of the FIM for Type I censoring

These are the derivatives computed for building the FIM of Section 3,

$$\begin{aligned}
\frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} &= \frac{-1}{\sigma} \left(\delta \frac{f'_\varepsilon(z_t)}{f_\varepsilon(z_t)} - (1 - \delta) \frac{f_\varepsilon(z_t)}{S_\varepsilon(z_t)} \right), \\
\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} &= \frac{-1}{\sigma^2} \left(-\delta \frac{f''_\varepsilon(z_t) f_\varepsilon(z_t) - f'_\varepsilon(z_t)^2}{f_\varepsilon(z_t)^2} + (1 - \delta) \frac{f'_\varepsilon(z_t) S_\varepsilon(z_t) + f_\varepsilon(z_t)^2}{S_\varepsilon(z_t)^2} \right), \\
\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu \partial \theta} &= x \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2}, \\
\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu \partial (\sigma^2)} &= \frac{-1}{2 \sigma^2} \frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} + \frac{z_t}{2 \sigma} \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2}, \\
\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \theta^2} &= x^2 \frac{\partial LL(\mu^2, \theta, \sigma; t, x)}{\partial \mu^2}, \\
\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \theta \partial (\sigma^2)} &= x \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu \partial (\sigma^2)}, \\
\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial (\sigma^2)^2} &= \frac{\delta}{2 \sigma^4} - \frac{3 z_t}{4 \sigma^3} \frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} + \frac{z_t^2}{4 \sigma^2} \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2}.
\end{aligned}$$

The expectations involved in the FIM are

$$\begin{aligned}
& \mathbb{E} \left(\frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} \right) \\
&= \frac{-1}{\sigma} \left(\int_{-\infty}^{z_d} f'_\varepsilon(z) dz - \frac{f_\varepsilon(z_d)}{S_\varepsilon(z_d)} \int_{z_d}^{\infty} f_\varepsilon(z) dz \right) = 0,
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E} \left(\frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\
&= \frac{-1}{\sigma^2} \left(- \int_{-\infty}^{z_d} \left(f''_\varepsilon(z) - \frac{f'_\varepsilon(z)^2}{f_\varepsilon(z)} \right) dz + \frac{f'_\varepsilon(z_d) S_\varepsilon(z_d) + f_\varepsilon(z_d)^2}{S_\varepsilon(z_d)^2} \int_{z_d}^{\infty} f_\varepsilon(z) dz \right) \\
&= \frac{-1}{\sigma^2} \left([f'_\varepsilon(z)]_{(-\infty)} + \frac{f_\varepsilon(z_d)^2}{S_\varepsilon(z_d)} + \int_{-\infty}^{z_d} \frac{f'_\varepsilon(z)^2}{f_\varepsilon(z)} dz \right),
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E} \left(z \frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} \right) \\
&= \frac{-1}{\sigma} \left(\int_{-\infty}^{z_d} z f'_\varepsilon(z) dz - \frac{z_d f_\varepsilon(z_d)}{S_\varepsilon(z_d)} \int_{z_d}^{\infty} f_\varepsilon(z) dz \right) \\
&= \frac{-1}{\sigma} \left(z_d f_\varepsilon(z_d) - [z f_\varepsilon(z)]_{(-\infty)} - \int_{-\infty}^{z_d} f_\varepsilon(z) dz - z_d f_\varepsilon(z_d) \right) \\
&= \frac{1}{\sigma} \left([z f_\varepsilon(z)]_{(-\infty)} + (1 - S_\varepsilon(z_d)) \right),
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E} \left(z \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\
&= \frac{-1}{\sigma^2} \left(- \int_{-\infty}^{z_d} \left(z f_{\varepsilon}''(z) - z \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} \right) dz + z_d \frac{f_{\varepsilon}'(z_d) S_{\varepsilon}(z_d) + f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)^2} \int_{z_d}^{\infty} f_{\varepsilon}(z) dz \right) \\
&= \frac{-1}{\sigma^2} \left([z f_{\varepsilon}'(z)]_{(-\infty)} - z_d f_{\varepsilon}'(z_d) + \int_{-\infty}^{z_d} f_{\varepsilon}'(z) dz + \int_{-\infty}^{z_d} z \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz + z_d f_{\varepsilon}'(z_d) + z_d \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} \right) \\
&= \frac{-1}{\sigma^2} \left([z f_{\varepsilon}'(z) - f_{\varepsilon}(z)]_{(-\infty)} + f_{\varepsilon}(z_d) + z_d \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} + \int_{-\infty}^{z_d} z \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz \right),
\end{aligned}$$

$$\begin{aligned}
& \mathbb{E} \left(z^2 \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\
&= \frac{-1}{\sigma^2} \left(- \int_{-\infty}^{z_d} \left(z^2 f_{\varepsilon}''(z) - z^2 \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} \right) dz + z_d^2 \frac{f_{\varepsilon}'(z_d) S_{\varepsilon}(z_d) + f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)^2} \int_{z_d}^{\infty} f_{\varepsilon}(z) dz \right) \\
&= \frac{-1}{\sigma^2} \left([z^2 f_{\varepsilon}'(z)]_{(-\infty)} - z_d^2 f_{\varepsilon}'(z_d) + 2 \int_{-\infty}^{z_d} z f_{\varepsilon}'(z) dz + \int_{-\infty}^{z_d} z^2 \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz + z_d^2 f_{\varepsilon}'(z_d) + z_d^2 \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} \right) \\
&= \frac{-1}{\sigma^2} \left([z^2 f_{\varepsilon}'(z) - 2z f_{\varepsilon}(z)]_{(-\infty)} + 2z_d f_{\varepsilon}(z_d) - 2(1 - S_{\varepsilon}(z_d)) + z_d^2 \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} + \int_{-\infty}^{z_d} z^2 \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz \right).
\end{aligned}$$

$$\mathbb{E} \left(\frac{\delta}{2 \sigma^4} \right) = \frac{1}{2 \sigma^4} \int_{-\infty}^{z_d} f_{\varepsilon}(z) dz = \frac{1}{2 \sigma^4} (1 - S_{\varepsilon}(z_d)).$$

Therefore

$$\begin{aligned}
A(x) &= \mathbb{E} \left(- \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\
&= \frac{1}{\sigma^2} \left([f_{\varepsilon}'(z)]_{(-\infty)} + \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} + \int_{-\infty}^{z_d} \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz \right),
\end{aligned}$$

$$\begin{aligned}
B(x) &= \mathbb{E} \left(- \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu \partial (\sigma^2)} \right) \\
&= \frac{1}{2 \sigma^2} \mathbb{E} \left(\frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} \right) - \frac{1}{2 \sigma} \mathbb{E} \left(z_t \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\
&= \frac{1}{2 \sigma^3} \left([z f_{\varepsilon}'(z)]_{(-\infty)} + f_{\varepsilon}(z_d) + z_d \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} + \int_{-\infty}^{z_d} z \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz \right),
\end{aligned}$$

$$\begin{aligned}
C(x) &= \mathbb{E} \left(- \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial (\sigma^2)^2} \right) \\
&= -\mathbb{E} \left(\frac{\delta}{2 \sigma^4} \right) + \frac{3}{4 \sigma^3} \mathbb{E} \left(z \frac{\partial LL(\mu, \theta, \sigma; t, x)}{\partial \mu} \right) - \frac{1}{4 \sigma^2} \mathbb{E} \left(z_t^2 \frac{\partial^2 LL(\mu, \theta, \sigma; t, x)}{\partial \mu^2} \right) \\
&= \frac{1}{4 \sigma^4} \left([z f_{\varepsilon}(z) + z^2 f_{\varepsilon}'(z)]_{(-\infty)} + S_{\varepsilon}(z_d) - 1 + 2z_d f_{\varepsilon}(z_d) + z_d^2 \frac{f_{\varepsilon}(z_d)^2}{S_{\varepsilon}(z_d)} + \int_{-\infty}^{z_d} z^2 \frac{f_{\varepsilon}'(z)^2}{f_{\varepsilon}(z)} dz \right).
\end{aligned}$$

Acknowledgements The authors have been sponsored by Ministerio de Ciencia e Innovación and fondos FEDER MTM2010-20774-C03-01 and -03, Junta de Comunidades de Castilla-La Mancha PEI10-0291-1850 and Fondo Social Europeo FSE 2007-2013. The comments from the Associate Editor and two anonymous referees contributed very much to improve the paper, the authors want to thank them for their thorough suggestions.

References

- A.C. Atkinson, A.N. Donev, and R.D. Tobias. *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford, 2007.
- V. Bagdonavicius and M. Nikulin. *Accelerated Life Models, Modeling and Statistical Analysis*. Boca Raton: Chapman & Hall/CRC, 2002.
- D.S. Bai and S.W. Chung. An optimal design of accelerated life test for exponential distribution. *Reliability Engineering and System Safety*, 31:57–64, 1991.
- M.P.F. Berger and W.K. Wong. *An Introduction to Optimal Designs for Social and Biomedical Research*. John Wiley & Sons, 2009.
- N.E. Breslow. Covariance analysis of censored survival data. *Biometrika*, 30:89–100, 1974.
- D. Collett. *Modelling Survival Data in Medical Research*. Chapman & Hall, London, 2003.
- D.R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.
- D.R. Cox and D. Oakes. *Analysis of Survival Data*. Chapman & Hall, London, 1984.
- H. Dette and V. B. Melas. A note on the de la garza phenomenon for locally optimal designs. *The Annals of Statistics*, 39(2):1266–1281, 2011.
- H. Dette and K. Schorning. Complete classes of designs for nonlinear regression models and principal representations of moment spaces. *The Annals of Statistics*, 41(3):1260–1267, 2013.
- B. Efron and I. Johstone. Fisher’s information in terms of the hazard rate. *The Annals of Statistics*, 18:38–62, 1990.
- G. Elfving. Optimum allocation in linear regression theory. *Ann. Math. Stat.*, 23:255–262, 1952.
- V. Fedorov and S. Leonov. *Optimal Design for Nonlinear Response Models*. CRC Press, 2013.
- V. V. Fedorov and P. Hackl. *Model-Oriented Design of Experiments*. Lecture Notes in Statistics. Springer, New York, 1997.
- I. Gertsbakh. On the fisher information in type-i censored and quantal response data. *Statistics and Probability Letters*, 23:297–306, 1995.
- I. Gertsbakh and A. Kogan. Characterization of the weibull distribution by properties of the fisher information under type-i censoring. *Statistics and Probability Letters*, 42:99–105, 1999.
- P. Goos and B. Jones. *Optimal Design of Experiments: A Case-Study Approach*. Wiley, New York, 2011.
- R. Kay and N. Kinnersley. On the use of the accelerated failure time model as an alternative to the proportional hazards model in the treatment of time to event data: A case study in influenza. *Drug Information Journal*, 36:571–579, 2002.
- N. Keiding, P.K. Andersen, and J.P. Klein. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, 16:215–224, 1997.
- M. Konstantinou, S. Biedermann, and A.C. Kimber. Optimal designs for full and partial likelihood information - with application to survival models. *Under review*, 2013.
- M. Konstantinou, S. Biedermann, and A.C. Kimber. Optimal designs for two-parameter nonlinear models with application to survival models. *Statistica Sinica*, 24(4):415–428, 2014.
- P. Lambert, D. Collett, A. Kimber, and R. Johnson. Parametric accelerated failure time models with random effects and an application to kidney transplant survival. *Statistics in Medicine*, 23:3177–3192, 2004.
- J.F. Lawless. *Statistical Models and Methods for Lifetime Data Analysis*. Wiley, New York, 1982.
- J. López Fidalgo and M.J. Rivas-López. Optimal experimental designs for partial likelihood information. *CSDA (in press)*, 2013.
- J. López Fidalgo and J.M. Rodríguez-Díaz. Elfving method for m-dimensional models. *Metrika*, 59:235–244, 2004.

- J. López Fidalgo, M.J. Rivas-López, and R. del Campo. Optimal designs for cox regression. *Statistica Neerlandica*, 63:135–148, 2009.
- Ch.H. Müller. D-optimal designs for lifetime experiments with exponential distribution and censoring. In A.C. Atkinson D. Uciński and M. Patan, editors, *mODa10: Advances on Model-Oriented Design and Analysis*, pages 161–168, Heidelberg, 2013. Physica-Verlag.
- J. Orbe and V. Núñez Antón. Alternative approaches to study lifetime data under different scenarios: from the ph to the modified semiparametric aft model. *Computational Statistics & Data Analysis*, 50: 1565–1582, 2006.
- A. Pázman. *Foundations of Optimum Experimental Design*. D. Reidel publishing company, Dordrecht, 1986.
- F. Pukelsheim. *Optimal Design of Experiments*. John Wiley and Sons, New York, 1993.
- J. Qi. *Comparison of Proportional Hazards and Accelerated Failure Time Models*. PhD thesis, Department of Mathematics and Statistics, University of Saskatchewan, 2009.
- S. D. Silvey. *Optimal Design*. Chapman & Hall, London, 1980.
- E. Walter and L. Pronzato. *Identification of parametric models from experimental data*. Springer, London, 1997.
- L. J. Wei. The accelerated failure time model: A useful alternative to the cox regression model in survival analysis. *Statistics in Medicine*, 11:1871–1879, 1992.
- M. Yang. On the de la garza phenomenon. *Annals of Statistics*, 38(4):2499–2524, 2010.
- M. Yang and J. Stufken. Identifying locally optimal designs for nonlinear models: a simple extension with profound consequences. *Annals of Statistics*, 40(3):1665–1681, 2012.