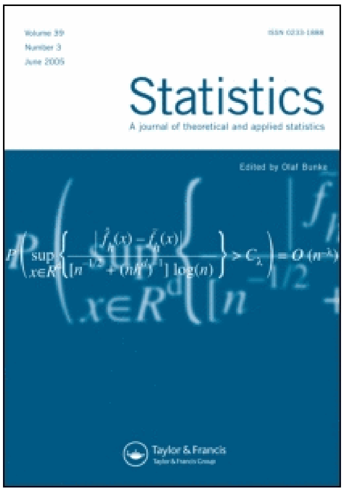

Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713682269

Optimal experimental designs when an independent variable is potentially censored

Jesús López-Fidalgo [a]; Sandra Garcet-Rodríguez [b]
[a] Department of Mathematics, Institute of Mathematics Applied to Science and Engineering, E.T.S.I. Industriales, University of Castilla-La Mancha, Ciudad Real, Spain [b] Department of History and Fundamentals of Economics, University of Salamanca, Salamanca, Spain

First published on: 22 February 2010

To link to this Article: DOI: 10.1080/02331880903546357
URL: http://dx.doi.org/10.1080/02331880903546357

Statistics, Vol. 00, No. 0, Month 2010, 1–12

Optimal experimental designs when an independent variable is potentially censored

Jesús López-Fidalgo[a]* and Sandra Garcet-Rodríguez[b]

[a] *Department of Mathematics, Institute of Mathematics Applied to Science and Engineering, E.T.S.I. Industriales, University of Castilla-La Mancha, Avda. Camilo José Cela 3, Ciudad Real 13071, Spain;* [b] *Department of History and Fundamentals of Economics, University of Salamanca, Plaza de los Caídos s/n, Salamanca 37008, Spain*

(*Received 24 May 2007; final version received 21 November 2009*)

This paper considers the problem of constructing optimal approximate designs when an independent variable might be censored. The problem is which design should be applied in practice to obtain the best approximate design when a censoring distribution is assumed known in advance. The approach for finite or continuous design spaces deserves different attention. In both cases, equivalent theorems and algorithms are provided in order to calculate optimal designs. Some examples illustrate this approach for D-optimality.

Keywords: approximate designs; censoring distribution; D-optimality; information matrix

AMS 2000 Mathematics Subject Classification: 62K05; 62N01

1. Introduction

In the classical theory of optimal experimental design for linear models, it is usually assumed that the experimenter can perform an experiment at any precise value of the independent variables within the whole design space. Nevertheless, in many areas of application it is common to find an independent variable that may be censored during the process. For example, Varela *et al.* [1] applied an exercise test to obtain more information to predict surgery complications in the treatment of lung cancer. The test consists in riding a static bicycle using a medical protocol. There are several independent variables, but the exercise time in minutes is the only one that can be controlled by the experimenter. In practice, a target exercise time is assigned to a patient, but he or she may stop the exercise before the time is over. Here, it is supposed that the stopping reasons are non-informative for the prediction. It will be assumed that the censoring of this variable has a known distribution. This means that the experimental design for this variable will be 'censored'. In this case the experiment must be designed taking into account the information given by the censoring

*Corresponding author. Email: jesus.lopezfidalgo@uclm.es

ISSN 0233-1888 print/ISSN 1029-4910 online

DOI: 10.1080/02331880903546357
http://www.informaworld.com

distribution. This particular problem deserves more theoretical and practical research since apart from the exercise time there is a mixture of two types of uncontrolled variables. For instance, the physician may take into account the 'expired volume of air in 1 s (respiratory function)', measured as a percentage of the expected values for sex, age and height. Values under 25th percentile are usually considered pathological for the investigation purposes. Another type of variable to be considered is the 'oxygen desaturation during the test'. Both variables are assumed binary in the medical context. López-Fidalgo and Garcet-Rodríguez [2] computed optimal designs assuming the patient always reaches the target time cycling. They used a design with two possible times: 12 and 18 min. A logistic model for the cardio-respiratory morbidity as the response variable predicts the complications after surgery. Because of all these types of variables, this problem exceeds the scope of the paper, and needs further research. In this paper, a simpler case will be considered to introduce the theoretical results.

Another real motivating problem is described in what follows. The influence of temperature on the rate of kinetic processes is usually interpreted in terms of what is known as the Arrhenius equation $k(t) = A\,\mathrm{e}^{B/t}$. Such a relation represents the temperature dependence of the rates of chemical reactions, conductivity coefficients and viscosity or diffusion phenomena. Experimental designs may be computed for this model [3]. In practice, it is not uncommon that a particular fixed temperature for an experiment cannot be reached exactly, that is, there is potential censoring. A typical use of the Arrhenius equation is a chemical reaction of atmospheric interest: $NO + O_3 \rightarrow NO_2 + O_2$. The temperature considered in the laboratory for this experiment typically goes from 212 to 422 with nominal values of the parameters $A = 3.0 \times 10^{-12}$ and $B = 1500$ [4]. This case needs a slight correction to the approach of this paper since trying temperature $t_1 = 212$ means getting in practice a temperature before the design interval.

In both problems the censoring is from above in the sense that the target experimental condition is never overtaken. This is the case considered in this paper.

Some work has been done in optimal experimental design for some kind of potentially failing in the response. Hackl [5] provided a criterion based on D-optimality and obtained optimal exact uniform designs for possible missing observations in the quadratic model. Imhof *et al.* [6,7] provided general procedures to compute approximate designs. Recently Ortega-Azurduy *et al.* [8] provided an insight into the effect of dropout on D-optimal designs for linear mixed models. Garcet-Rodrìguez *et al.* [9] dealt with the problem of obtaining optimal approximate designs for a linear model when the values of some independent variables are potentially censored according to a known probability distribution function. They considered the problem for a discrete space. The continuous case needs a different approach, which is treated in this paper. Appropriate algorithms are provided for computing optimal designs in both the discrete and the continuous cases.

Let a linear model be $E[y] = \beta^{\mathrm{T}}\eta(t)$, $\mathrm{var}(y) = \sigma^2$, $t \in \chi$, with $\eta^{\mathrm{T}}(t) = (\eta_1(t), \ldots, \eta_m(t))$ continuous linearly independent known functions defined in a compact design space χ and $\beta^{\mathrm{T}} = (\beta_1, \ldots, \beta_m)$, the unknown parameters to be estimated.

If the conditions of the experiment t are under the control of the practitioner, a design of the experiment can be made in advance. An *exact design* will be a sequence of experimental conditions $t_1, \ldots, t_N$ from a design space χ, while an *approximate design* [10] will be any probability distribution ξ. The information matrix associated to a design ξ is defined as follows:

$$M(\xi) = \int_{\chi} \eta(t)\eta^{\mathrm{T}}(t)\xi(\mathrm{d}t).$$

The set of approximate designs, $\Im$, and the set of the information matrices, $\mathcal{M}$, are convex. Moreover, if χ is compact then $\mathcal{M}$ is also compact. The inverse of the information matrix is proportional to the covariance matrix of the least-square estimates. Thus, an experimental design 'minimizing', in some sense, the inverse of the information matrix should be found.

A function Φ defined on $\mathcal{M}$ will define an optimality criterion if it is non-decreasing in the Loewner sense. The most popular criterion is D-optimality, which looks for the maximum determinant of the information matrix, $\Phi_D(M) = \det(M)^{-1/m}$. If a Φ-optimal design is denoted by $\xi^\star$, a measure of the goodness of a particular design ξ with respect to this criterion is the efficiency,

$$\text{eff}_\Phi[M(\xi)] = \frac{\Phi[M(\xi^\star)]}{\Phi[M(\xi)]}.$$

For a more detailed introduction of the theory of optimal design, there are a number of books, for instance [11,12].

2. Approach to the problem

Throughout this paper, a time variable is used for the explanatory variable as an illustrative and typical case for this problem, but all the results may be applied to any kind of variable. Let t be a particular time to be completed during the experiment. Assume this time; is not always reached and let t' be the actual time attained. In this approach, only censoring from above, is considered, so the target time t is never overtaken ($t' \leq t$). Thus, t' may be considered as a random variable depending on the target time t. Let T be a theoretical random variable, which measures the time; a chosen experimental unit is going to stop given no prior limitation in time from above. In the real case mentioned above, T would be the time a patient stops if he or she starts to ride the bicycle without any time limit imposed in advance. Note that this is just a theoretical time. In this approach, only the probability of a patient to reach the target time or stopping before matters. Let $f(t)$ and $F(t)$ be the probability density function (pdf) and the cumulative distribution function (cdf) of T, respectively. A particular, but typical, case may be a censoring distribution on $[0, \infty)$ with the design space contained in it, for example, $\chi = [0, b]$. Summarizing, if a target time t is fixed for the experiment the final time reached will be $t' = t$ if $T \geq t$ or $t' = T$ if $T \leq t$ and an observation $y = y(t')$ will be available at this censored time.

There are always three associated designs: the design to be tried in practice ($\breve{\xi}$), the actual censored design realized at the end of the experimentation (ξ) which is not known when designing the experiment and has to be estimated by the expected censored design ($\hat{\xi}$). Therefore, an approximate design $\breve{\xi}$ with finite support should be found such that the expected censored design $\hat{\xi}$ will be optimal among all possible expected designs. Note that the set of all possible expected designs is only a subset of $\Im$,

$$\Im_R = \{\hat{\xi} \mid \hat{\xi} \text{ is the expected design of some finite design } \breve{\xi}\}.$$

We will call an optimal design $\breve{\xi}_R^\star$ with this constraint, a censoring restricted (CER) optimal design. The associated expected design, $\hat{\xi}_R^\star$, will be called the expected censoring restricted (ECER) optimal design. Sometimes it is possible to find $\breve{\xi}$ such that the expected design $\hat{\xi}$ will be optimum according to the criterion without censoring, but usually this is not the case and a restricted search has to be performed.

Whittle [13] provided a general equivalence theorem for the general theory. In a similar way, Theorem 1 of López-Fidalgo and Garcet-Rodríguez [2] stated an equivalence theorem for restricted optimality. For that, the definition of a directional derivative of Φ at M in the direction of N is needed,

$$\partial\Phi(M, N) = \lim_{\epsilon \to 0} \frac{\Phi[(1-\epsilon)M + \epsilon N] - \Phi(M)}{\epsilon}.$$

THEOREM 1 (Equivalence theorem) *If Φ is a convex function, then the following statements are equivalent:*

(1) $\Phi[M(\hat{\xi}_R^\star)] = \inf_{\hat{\xi}\in\Im_R}\Phi[M(\hat{\xi})]$, *where $\hat{\xi}_R^\star$ is the ECER Φ-optimal design.*

(2) $\inf_{N\in\mathcal{M}_R}\partial\Phi[M(\hat{\xi}_R^\star), N] = \sup_{\hat{\xi}\in\Im_R^+}\inf_{N\in\mathcal{M}_R}\partial\Phi[M(\hat{\xi}), N]$, *where $\mathcal{M}_R = \{M(\hat{\xi}) \mid \hat{\xi}\in\Im_R\}$ and $\Im_R^+$ is the set of the designs with non-singular information matrices.*

(3) $\inf_{N\in\mathcal{M}_R}\partial\Phi[M(\hat{\xi}_R^\star), N] = 0$.

The objective is then to find

$$\hat{\xi}_R^\star = \arg\min\{\Phi[M(\hat{\xi})] \mid \hat{\xi}\in\Im_R\}.$$

Only in a few cases this programming problem can be solved analytically. Usually, an algorithm will be needed.

From the definition of the directional derivative, the inequality $\partial\Phi(M, N) \le \Phi(N) - \Phi(M)$ holds. Using this inequality and the formula $\partial\Phi(M, N) = \mathrm{tr}[\nabla\Phi(M)(N - M)]$ for a differentiable criterion Φ, the following result provides an important lower bound for the efficiency.

COROLLARY 1 *Under the above conditions*

$$\mathrm{eff}_\Phi(M) \ge 1 + \frac{\inf_N \partial\Phi(M, N)}{\Phi(M)}.$$

For more details in what has been setup in this section see [9], where an optimal restricted design was computed directly in a simple example. As pointed above, this is usually not possible, for instance, when the design space is large, and some kind of algorithm for the computation of optimal designs is needed.

In what follows, the two possible cases with either a finite (Section 2.1) or a continuous (Section 3) design space will be considered. In Sections 2.1 and 4, algorithms for the computation of these designs will be provided.

2.1. *Finite design space*

This case has been considered in detail in [9]. A brief review is provided in this section. Let a finite design space be $\chi = \{t_1, t_2, \ldots, t_n\}$, where $t_1 < \cdots < t_k < \cdots < t_n$. The censoring distribution of T will be considered as a discrete distribution on χ. The cdf will be $F(t) = \sum_{i=1}^k f(t_i)$, $t_k \le t < t_{k+1}, k = 1, \ldots, n-1$, where $f(t_i) = P(T = t_i),\ i = 1, \ldots, n$. Then, $F(t) = 0, t < t_1$ and $F(t) = 1, t \ge t_n$. If there would not be any fixed time to stop the experiment, then $f(t_k)$ would be the probability to stop exactly at time t_k, that is, completing all the stages until k and then stop. But if the experiment has to be performed by a particular individual just until t_k, then it may stop the experiment at either $t_1, t_2, \ldots$ or t_k according to the censoring distribution. Then the probability to stop exactly at t_k is $f(t_k) + \cdots + f(t_n)$.

If a proportion $\breve{\xi}(t_k)$ of experiments are tried at t_k, then some of them will reach the target t_k while the rest will be distributed at $t_{k-1}, \ldots$ and t_1. Thus, when a design $\breve{\xi}$ is tried in practice an expected censored design $\hat{\xi}$ will be performed [9]:

$$\hat{\xi}(t_k) = (1 - F_{k-1})\breve{\xi}(t_k) + f(t_k)(1 - \breve{\Xi}_k), \quad k = 1, \ldots, n, \tag{1}$$

where $F_k \equiv F(t_k)$; $\breve{\Xi}_k \equiv \sum_{i=1}^k \breve{\xi}(t_i)$, $k = 1, \ldots, n$; $F_0 \equiv F(t_0) = 0$ and $\breve{\Xi}_0 \equiv 0$, being t_0 any point less than t_1, just for consistency of notation.

Working $\breve{\xi}$ out from the previous formula is possible if and only if the ratio $r_i = (1 - \hat{\Xi}_i)/(1 - F_i)$ is non-increasing for $i = 0, \ldots, n-1$. Thus, the convex set of all the possible expected designs is then $\Im_F = \{\hat{\xi} | r_i \text{ is non-increasing}, i = 0, \ldots, n-1\}$. Let $\mathcal{M}_F$ be the set of the information matrices associated with these designs.

The information matrix associated with an expected design $\hat{\xi}$, obtained from Equation (1), is

$$M(\hat{\xi}) = \sum_{k=1}^{n} \breve{\xi}(t_k) \left[\sum_{i=0}^{k-1} f(t_i)\eta(t_i)\eta^{\mathrm{T}}(t_i) + (1 - F_{k-1})\eta(t_k)\eta^{\mathrm{T}}(t_k) \right]. \tag{2}$$

If the criterion function is differentiable, the directional derivative is linear on the second argument and therefore the equivalence theorem has to be checked only in the direction of the generators of the set of the information matrices of the ECER designs,

$$\sum_{i=0}^{k-1} f(t_i)\eta(t_i)\eta^{\mathrm{T}}(t_i) + (1 - F_{k-1})\eta(t_k)\eta^{\mathrm{T}}(t_k), \quad k = 1, \ldots, n.$$

These matrices come from the CER designs

$$\begin{Bmatrix} t_1 & \cdots & t_{k-1} & t_k \\ f(t_1) & \cdots & f(t_{k-1}) & 1 - F(t_k) \end{Bmatrix}, \quad k = 1, \ldots, n.$$

Therefore, a condition to check the optimality for differentiable criteria may be given as follows.

Corollary 2 *Under the hypotheses of Theorem* 1, *if* Φ *is a differentiable function then* $\hat{\xi}_R^{\star}$ *is an ECER* Φ*-optimal design if and only if*

$$\psi(\hat{\xi}_R^{\star}, t) \geq 0, \quad t \in \chi,$$

where for $k = 1, \ldots, n$,

$$\psi(\hat{\xi}, t_k) = \mathrm{tr} \left\{ \nabla\Phi[M(\hat{\xi})] \left[\sum_{i=0}^{k-1} f(t_i)\eta(t_i)\eta^{\mathrm{T}}(t_i) + (1 - F_{k-1})\eta(t_k)\eta^{\mathrm{T}}(t_k) - M(\hat{\xi}) \right] \right\}.$$

Based on the equivalence theorem and the results given above, the following general algorithm may be stated for a differentiable criterion Φ:

(1) Set a non-singular initial design $\breve{\xi}^{(0)}$ to be tried in practice.
(2) Let $\breve{\xi}^{(r-1)}$ be the design obtained at step $r-1$ and $\hat{\xi}^{(r-1)}$ the corresponding expected design after the experimentation. Compute the point

$$t^{(r)} = \arg \min_{k=1,\ldots,n} \mathrm{tr} \left\{ \nabla\Phi[M(\hat{\xi}^{(r-1)})] \left[\sum_{i=0}^{k-1} f(t_i)\eta(t_i)\eta^{\mathrm{T}}(t_i) + (1 - F_{k-1})\eta(t_k)\eta^{\mathrm{T}}(t_k) \right] \right\}.$$

(3) The new design at step r will be

$$\breve{\xi}^{(r)} = (1 - \alpha_r)\breve{\xi}^{(r-1)} + \alpha_r \hat{\xi}_{t^{(r)}},$$

where either $\alpha_r = \min_{\alpha \in [0,1]} \Phi\{M[(1-\alpha)\breve{\xi}^{(r-1)} + \alpha\hat{\xi}_{t^{(r)}}]\}$ or $\alpha_r = 1/(r+1)$.

(4) The algorithm will stop as soon as an appropriate bound for the efficiency will be reached:

$$1+\frac{\min_{k=1,\dots,n}\operatorname{tr}\left\{\nabla\Phi[M(\hat{\xi}^{(r)})]\left[\sum_{i=0}^{k-1}f(t_i)\eta(t_i)\eta^{\mathrm{T}}(t_i)+(1-F_{k-1})\eta(t_k)\eta^{\mathrm{T}}(t_k)\right]\right\}}{\Phi[M(\hat{\xi}^{(r)})]}$$
$$-\frac{\operatorname{tr}\left\{\nabla\Phi[M(\hat{\xi}^{(r)})]M(\hat{\xi}^{(r)})\right\}}{\Phi[M(\hat{\xi}^{(r)})]}.$$

Last formula comes from Corollary 1 after some algebra. Taking into account that $\nabla\Phi_D[M(\hat{\xi})]=-1/(m|M(\hat{\xi})|^{1/m})M^{-1}(\hat{\xi})$ for D-optimality, then

$$t^{(r)}=\arg\max_{k=1,\dots,n}\left[\sum_{i=0}^{k-1}f(t_i)\eta^{\mathrm{T}}(t_i)M^{-1}(\hat{\xi}^{(r-1)})\eta(t_i)+(1-F_{k-1})\eta^{\mathrm{T}}(t_k)M^{-1}(\hat{\xi}^{(r-1)})\eta(t_k)\right]$$

and the bound for the D-efficiency will be

$$2-\frac{\max_{k=1,\dots,n}\left[\sum_{i=0}^{k-1}f(t_i)\eta^{\mathrm{T}}(t_i)M^{-1}(\hat{\xi}^{(r-1)})\eta(t_i)+(1-F_{k-1})\eta^{\mathrm{T}}(t_k)M^{-1}(\hat{\xi}^{(r-1)})\eta(t_k)\right]}{m}.$$

An example is considered now in order to clarify and motivate the procedure.

Example 1 In a company the commercial travellers are going to take 10 group therapy sessions during a quarter in order to evaluate whether the number of sessions will influence proportionally the sales or not. The model considered is a simple linear regression model,

$$\mathrm{E}(y)=\beta_1+\beta_2 t,\quad \mathrm{Var}(y)=\sigma^2,\ \ t\in\chi=\{0,1,2,\dots,10\}.$$

The sessions are independent and equivalent since they deal with the situation of the participants at that particular moment. A fix number of sessions is assigned to each commercial. For different reasons it is likely that not all of them are going to complete their numbers. Assume there is a geometric censoring distribution on χ with a probability of failure (in this case) $\theta=0.2$, that is,

$$f(k)=0.8^k 0.2,\quad k=0,1,\dots,9;\quad f(10)=1-f(0)-f(1)-\cdots-f(9).$$

The algorithm was tried using the D-optimal unrestricted design as initial design,

$$\xi_D\equiv\left\{\begin{matrix}0 & 10\\ \frac{1}{2} & \frac{1}{2}\end{matrix}\right\},$$

and the design

$$\breve{\xi}_R\equiv\left\{\begin{matrix}0 & 10\\ \frac{3}{38} & \frac{35}{38}\end{matrix}\right\}$$

was obtained with efficiency near 100%. This means that about $100(41/58)=70.7\%$ of the commercials would have to attend all the sessions and the rest would attend none of them. If the unrestricted D-optimal design ξ_D is tried, the expected design will not be optimal as is the case of the expected design from $\breve{\xi}_R$. Comparing both determinants the restricted D-efficiency of the expected design from ξ_D is in this case

$$\left(\frac{\det M[\hat{\xi}_D]}{\det M[\hat{\xi}_R]}\right)^{1/2} = 0.889,$$

meaning an efficiency of more than 10% is lost if the unrestricted optimal design is used instead of the restricted one.

3. Continuous design space

Let $\chi = [0, b]$ be the design space. In this section, a continuous censoring distribution defined on the set $[0, \infty)$ is considered. Let $f(t)$ be an upper continuous version of the pdf and $F(t)$ the corresponding cdf. Let $\breve{\xi}$ be the design that is going to be applied in practice, supported at points $t_1 < \cdots < t_k < \cdots < t_n$ with weights $\breve{\xi}(t_i),\ i = 1, \ldots, n$. Let $t_0 = 0$ for the convenience of notation. Eventually, $t_1 = t_0 = 0$.

The expected design at the end of the experimentation will be a mixture of a continuous distribution pdf and discrete probabilities ('weights') at points $t_1, \ldots, t_n$. It is computed in a similar way as in the discrete case [9]:

(1) A proportion of $\breve{\xi}(t_1)$ experiments are tried at time t_1. Taking into account the censoring distribution, this proportion of times will be distributed on the interval $[0, t_1)$ according to density $\breve{\xi}(t_1) f(t),\ t \in [0, t_1)$ and the rest $\breve{\xi}(t_1)[1 - F(t_1)]$ will reach the challenged time t_1. Note that if $t_1 = 0$, then $\breve{\xi}(0)[1 - F(0)] = \breve{\xi}(0)$.
(2) Similarly, a proportion of $\breve{\xi}(t_2)$ experiments are tried at time t_2. The actual times attained will be distributed on the interval $[0, t_2)$ according to density $\breve{\xi}(t_2) f(t),\ t \in [0, t_2)$ and the rest $\breve{\xi}(t_2)[1 - F(t_2)]$ will actually reach t_2.

Following a similar argument for the rest of the points, the final expected design will be

$$\hat{\xi}(t) = \begin{cases} \sum_{i=k}^{n} \breve{\xi}(t_i) f(t), & \text{if } t_{k-1} < t < t_k, \\ \breve{\xi}(t_k)[1 - F(t_k)], & \text{if } t = t_k, \end{cases} \quad k = 1, \ldots, n. \tag{3}$$

The information matrix of the expected design $\hat{\xi}$ is then

$$\begin{aligned} M(\hat{\xi}) &= \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \hat{\xi}(t)\eta(t)\eta^{\mathrm{T}}(t)\,\mathrm{d}t + \sum_{k=1}^{n} \hat{\xi}(t_k)\eta(t_k)\eta^{\mathrm{T}}(t_k), \\ &= \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \sum_{i=k}^{n} \breve{\xi}(t_i) f(t)\eta(t)\eta^{\mathrm{T}}(t)\,\mathrm{d}t + \sum_{k=1}^{n} \breve{\xi}(t_k)[1 - F(t_k)]\eta(t_k)\eta^{\mathrm{T}}(t_k), \\ &= \sum_{k=1}^{n} \breve{\xi}(t_k)[G(t_k) + \eta(t_k)\eta^{\mathrm{T}}(t_k)[1 - F(t_k)]], \end{aligned}$$

where $G(t) = \int_0^t f(s)\eta(s)\eta^{\mathrm{T}}(s)\,\mathrm{d}s$. A criterion function of the information matrix should be used to obtain the optimal design $\breve{\xi}$.

Let $\Im_F$ be the set of all the expected designs when a design with finite support on a continuous design space χ is tried. A finite support is assumed since this is a practical design to be tried in practice. It can be easily proved that the set $\Im_F$ is convex and Theorem 1 can be used here again. If the criterion function is differentiable, then the directional derivative is linear in the second argument and therefore the equivalence theorem needs to be checked only in the direction of the generators of the set of the information matrices of ECER designs,

$$G(t) + [1 - F(t)]\eta(t)\eta^{\mathrm{T}}(t), \quad t \in [0, b].$$

COROLLARY 3 *If the criterion function Φ is differentiable, then a design $\breve{\xi}_R^\star$ is CER Φ-optimal if and only if*

$$\psi(\hat{\xi}_R^\star, t) \geq 0, \quad t \in \chi = [0, b],$$

where $\hat{\xi}_R^\star$ is the expected design of $\breve{\xi}_R^\star$ and

$$\psi(\hat{\xi}, t) = \mathrm{tr}\left\{\nabla\Phi[M(\hat{\xi})]\left[G(t) + [1 - F(t)]\eta(t)\eta^{\mathrm{T}}(t) - M(\hat{\xi})\right]\right\}, \quad t \in \chi = [0, b].$$

The proof of this corollary is based on the well-known formula for the directional derivative of a differentiable function, $\partial\Phi(M, N) = \mathrm{tr}[\nabla\Phi(M)\,(N - M)]$.

For D-optimality, the condition to be checked is

$$\psi(\hat{\xi}, t) = m - \mathrm{tr}\{M^{-1}(\hat{\xi})\,[G(t) + [1 - F(t)]\eta(t)\eta^{\mathrm{T}}(t)]\} \geq 0, \quad t \in \chi = [0, b].$$

The following example will illustrate the procedure.

Example 2 The efficacy of a mixture of an additive in the gasoline wants to be measured. Let t be the percentage of the additive in the mixture and let y be the kilometres per litre of gasoline. The design space is from 0% to 1% of the additive. The model to be checked is then

$$\mathrm{E}(y) = \beta_1 + \beta_2 t, \quad \mathrm{Var}(y) = \sigma^2, \;\; t \in \chi = [0, 1].$$

There is always some proportion of the additive which is not absorbed. This means the fixed proportion for a particular experiment is going to be censored. An exponential censoring distribution was assumed on χ, $f(t) = \mathrm{e}^{-t}$, $F(t) = 1 - \mathrm{e}^{-t}$, $t > 0$. Then,

$$G(t) = \int_0^t \eta(s)\eta^{\mathrm{T}}(s) f(s)\,ds = \int_0^t \begin{pmatrix} 1 & s \\ s & s^2 \end{pmatrix} \mathrm{e}^{-s}\,\mathrm{d}s,$$
$$= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} - \mathrm{e}^{-t} \begin{pmatrix} 1 & 1+t \\ 1+t & 2+2t+t^2 \end{pmatrix}.$$

The expected information matrix is

$$M(\hat{\xi}) = \sum_{k=1}^{n} \breve{\xi}(t_k)[G(t_k) + \mathrm{e}^{-t_k}\eta(t_k)\eta^{\mathrm{T}}(t_k)].$$

Note that under this framework a one-point design is possible if the expected design will have more points in its support. It is worth to mention that a singular design is also possible for the finite case. The one-point D-optimal design is concentrated at $t_1 = 1$. If the one-point design is tried, the following theoretical experimental design is expected:

$$\hat{\xi}(t) = \mathrm{e}^{-t}, \quad 0 \leq t \leq 1.$$

Thus, if for instance $N = 30$ experiments are tried at $t_1 = 1$, then about 11 ($\approx \mathrm{e}^{-1} \times 30$) experiments are expected to succeed while the 19 remaining will be distributed according to the distribution given by e^{-t} on the interval [0, 1].

The determinant of the information matrix associated with the expected design is 0.129.

The two-point D-optimal design is

$$\breve{\xi}_R = \begin{Bmatrix} 0 & 1 \\ 0.339 & 0.661 \end{Bmatrix}.$$

This design means that 33.9% of the experiments have no additive at all and the rest 1% of additive. In order to verify that this is a CER optimal design, the following condition has to be satisfied (Corollary 3):

$$\mathrm{e}^t - 1 - t(\mathrm{e} - 1) \leq 0, \quad t \in \chi = [0, 1].$$

This function is convex and attains the value 0 at the boundaries of the interval [0, 1], therefore the inequality is satisfied.

When this design is tried in practice, for example, for $N = 30$ experiments, then 10 ($\approx 0.339 \times 30$) will be made at $t_1 = 0$, about 7 ($\approx 0.661 \times [1 - F(1)] \times 30$) at $t_2 = 1$ and the rest will be located on (0, 1) according to the distribution density $0.661\,\mathrm{e}^{-t}$.

The efficiency of the expected design from the unrestricted D-optimal design

$$\xi_D = \begin{Bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{Bmatrix}$$

with respect to the expected censored D-optimal design is now 97.0%, which is very high in this particular case.

4. Algorithm for the continuous case

Based on the equivalence theorem, an algorithm may be stated for a differentiable criterion Φ in a similar way as in Section 2.1 exchanging $\sum_{i=0}^{k-1} f(t_i)\eta(t_i)\eta^{\mathrm{T}}(t_i) + (1 - F_{k-1})\eta(t_k)\eta^{\mathrm{T}}(t_k)$ with $G(t) + [1 - F(t)]\eta(t)\eta^{\mathrm{T}}(t)$ appropriately:

(1) Set a non-singular initial design $\breve{\xi}^{(0)}$ to be tried in practice.

(2) Let $\breve{\xi}^{(r-1)}$ be the design obtained at step $r - 1$ and $\hat{\xi}^{(r-1)}$ be the corresponding expected design after the experimentation. Let

$$t^{(r)} = \arg \min_{t \in [0,b]} \operatorname{tr} \left\{ \nabla\Phi[M(\hat{\xi}^{(r-1)})] \left[G(t) + [1 - F(t)]\eta(t)\eta^{\mathrm{T}}(t) \right] \right\}.$$

(3) The new design at step r will be

$$\breve{\xi}^{(r)} = (1 - \alpha_r)\breve{\xi}^{(r-1)} + \alpha_r \hat{\xi}_{t^{(r)}},$$

where either $\alpha_r = \min_{\alpha \in [0,1]} \Phi\{M[(1 - \alpha)\breve{\xi}^{(r-1)} + \alpha \hat{\xi}_{t^{(r)}}]\}$ or $\alpha_r = 1/(r + 1)$.

(4) The algorithm will stop as soon as an appropriate bound for the efficiency will be reached,

$$1 + \frac{\min_{t \in [0,b]} \operatorname{tr} \left\{ \nabla\Phi[M(\hat{\xi}^{(r-1)})] \left[G(t) + [1 - F(t)]\eta(t)\eta^{\mathrm{T}}(t) \right] \right\}}{\Phi[M(\hat{\xi}^{(r)})]} - \frac{\operatorname{tr} \left\{ \nabla\Phi[M(\hat{\xi}^{(r)})] M(\hat{\xi}^{(r)}) \right\}}{\Phi[M(\hat{\xi}^{(r)})]}.$$

Last formula comes from Corollary 1 after some algebra.

For D-optimality,

$$t^{(r)} = \arg\max_{t\in[0,b]} \operatorname{tr}\left\{M^{-1}(\hat{\xi}^{(r-1)})\left[G(t)+[1-F(t)]\eta(t)\eta^{\mathrm{T}}(t)\right]\right\},$$

and the bound for the D-efficiency will be

$$2-\frac{\max_{t\in[0,b]}\operatorname{tr}\left\{M^{-1}(\hat{\xi}^{(r-1)})\left[G(t)+[1-F(t)]\eta(t)\eta^{\mathrm{T}}(t)\right]\right\}}{m}.$$

Example 3 There is a debate of what should be the proportion of discussions with exercises and theoretical lectures in a first course of Algebraic Geometry. It is supposed there is a maximum at some middle point. An experiment is going to be conducted in several universities in order to find this optimal proportion. Let t be the proportion of discussions and y be the performance of the students. A quadratic model is assumed,

$$\mathrm{E}(y)=\beta_1+\beta_2 t+\beta_3 t^2,\quad \mathrm{Var}(y)=\sigma^2,\ \ t\in\chi=[0,1].$$

When a professor is asked to adjust the classes to a particular proportion of discussions, in practice this proportion is not always reached. An exponential censoring distribution is assumed on χ, $f(t)=\theta\mathrm{e}^{-\theta t}$, $F(t)=1-\mathrm{e}^{-\theta t}$, $t>0$. Then,

$$G(t)=\int_0^t \eta(s)\eta^{\mathrm{T}}(s)f(s)\,\mathrm{d}s=\int_0^t\begin{pmatrix}1 & s & s^2\\ s & s^2 & s^3\\ s^2 & s^3 & s^4\end{pmatrix}\theta\,\mathrm{e}^{-\theta s}\,\mathrm{d}s,$$

$$=\begin{pmatrix}1 & \frac{1}{\theta} & \frac{2}{\theta^2}\\ \frac{1}{\theta} & \frac{2}{\theta^2} & \frac{6}{\theta^3}\\ \frac{2}{\theta^2} & \frac{6}{\theta^3} & \frac{24}{\theta^4}\end{pmatrix}-\mathrm{e}^{-\theta t}\begin{pmatrix}1 & \frac{1+\theta t}{\theta} & \frac{2+2\theta t+\theta^2t^2}{\theta^2}\\ \frac{1+\theta t}{\theta} & \frac{2+2\theta t+\theta^2t^2}{\theta^2} & \frac{6+6\theta t+3\theta^2t^2+\theta^3t^3}{\theta^3}\\ \frac{2+2\theta t+\theta^2t^2}{\theta^2} & \frac{6+6\theta t+3\theta^2t^2+\theta^3t^3}{\theta^3} & \frac{24+24\theta t+12\theta^2t^2+4\theta^3t^3+\theta^4t^4}{\theta^4}\end{pmatrix}.$$

Applying the algorithm for CER D-optimality ($\theta=0.1$), the design obtained is

$$\breve{\xi}_R=\begin{Bmatrix}0 & 0.512 & 1\\ 0.361 & 0.149 & 0.590\end{Bmatrix}$$

with efficiency near 100%. When this design is tried in practice the following design is expected to happen:

$$\hat{\xi}_R(t)=\begin{cases}0.361, & \text{if } t=0,\\ 0.665\,\mathrm{e}^{-0.9t}, & \text{if } 0<t<0.512,\\ 0.149\,\mathrm{e}^{-0.461}, & \text{if } t=0.512,\\ 0.531\,\mathrm{e}^{-0.9t}, & \text{if } 0.512<t<1,\\ 0.590\,\mathrm{e}^{-0.9}, & \text{if } t=1.\end{cases}$$

The efficiency of the expected design from the unrestricted D-optimal design equally distributed on {0, 0.5, 1} with respect to ECER D-optimal design is now 94.3%.

5. Discussion

Two real problems of possible censoring on the variable that has been designed motivated this paper. The censoring rule is assumed known under a probability distribution and therefore an expected experimental design may be computed from the design tried in practice. If the censoring distribution is continuous the expected censored design happens to be a distribution with a continuous part and a discrete part. This work shows how the general theory has to be adjusted to this approach. In particular, Caratheodory's theorem may still be used in this context to limit the support points of the design to be tried in practice to $m(m+1)/2+1$ points.

Since at the moment of designing the experiment the actual censored design realized in practice is unknown, there is not any other chance than estimating it. Thus, assuming a censoring distribution is needed. This is the traditional problem of assuming prior distributions. The typical life distributions may be used in this context. Instead of using a particular distribution as done in this paper, a parametric family may be used including new parameters to be added to the model. This deserves non-trivial further research. Note that the expected design is an estimate of the actual design that is happening in practice, which is the design actually used to estimate the parameters.

In the examples, the efficiency of the expected unrestricted optimal design was computed. Depending on the severity of the censoring much lower efficiencies may be obtained.

Only approximate designs have been considered in this paper. These ideas may be developed for exact designs in a similar way, except that the equivalence theorem is not true and new algorithms have to be developed. Moreover, with exact designs a different approach can be given applying the expectation to the criterion instead of to the design tried in practice. In particular, instead of optimizing $\Phi[M(\hat{\xi})] = \Phi\{M[\mathrm{E}(\breve{\xi})]\} = \Phi\{\mathrm{E}[M(\breve{\xi})]\}$, a different approach is considering the expectation of the criterion,

$$\mathrm{E}\{\Phi[M(\breve{\xi})]\}.$$

This is an interesting idea for new research.

Under the approach considered in this paper, a singular optimal design to be tried in practice may appear. But the actual resulting design is expected to be non-singular. For example, in the cycling exercise a singular design with all tries at 18 min may be considered. Then, because of the censoring some patients will not reach this target and the resulting actual design will allow the estimation of the parameters. It is true there is a chance that all of the tries reach the target of 16 min, but it is very unlikely since an optimal singular design is usually obtained when the probability of censoring is high. In any case this may be avoided using a regularization procedure in the sense of Fedorov and Hackl [12] introducing in the algorithm the criterion function $\Phi_\gamma(\breve{\xi}) = \Phi((1-\gamma)\breve{\xi} + \gamma\tilde{\breve{\xi}})$ for an appropriate design $\tilde{\breve{\xi}}$ and $\gamma \in (0, 1)$.

The algorithms provided in this paper converge to the optimal designs under the conditions given in the general theory [14] adapted to the restricted case [9, Theorem 3]. In particular, the convergence applies for D-optimality, which is the criterion considered in this paper. As regards the choice of α_r, it has to satisfy the conditions $0 < \alpha_r < 1$, $\lim_{r\to\infty} \alpha_r = 0$, $\sum_{r=0}^{\infty} \alpha_r = \infty$ and $\sum_{r=0}^{\infty} \alpha_r^2 < \infty$ stated in [9] as is the case of the two possible choices suggested in the algorithms.

A particular design space, $[0, b]$, and a particular survival distribution domain, $[0, \infty)$, have been used in the outline of the problem. This has been done for a justification of the real problem considered in this paper. Apart from the difficulties introduced in optimal design with more than one dimension, a generalization, in a theoretical sense, of this particular approach to more general spaces may be more straightforward.

Acknowledgements

This paper has been written when the second author was doing a postdoc in the University of Salamanca and it was partially sponsored by the Ministerio de Educación y Ciencia and Fondos FEDER MTM2007 67211 C03-01, Junta de Comunidades de Castilla-La Mancha PAI07-0019-2036 and University of Castilla-La Mancha. The authors want to thank the referees for their helpful their comments.

References

[1] G. Varela, R. Cordovilla, M.F. Jiménez, and N. Novoa, *Utility of standardized exercise oximetry to predict cardiopulmonary morbidity after lung resection*, Eur. J. Cardiothoracic Surg. 19 (2001), pp. 351–354.
[2] J. López-Fidalgo and S. Garcet-Rodríguez, *Optimal experimental designs when some independent variables are not subject to control*, J. Am. Stat. Assoc. 99(468) (2004), pp. 1190–1199.
[3] L.J. Rodríguez-Aragón and J. López-Fidalgo, *Optimal designs for the Arrhenius equation*, Chemometr. Intell. Lab. Syst. 77 (2005), pp. 131–138.
[4] Jet Propulsion Laboratory, *Chemical kinetics and photochemical data for use in atmospheric studies*, Tech. Rep. Evaluation Number 14, California Institute of Technology, NASA, Pasadena, CA, February 2003.
[5] P. Hackl, *Optimal Designs for Experiments with Potentially Failing Trials*, MODA 4: Advances in Model-Oriented Data Analysis, Physica Verlag, Heidelberg, 1995, pp. 117–124.
[6] L. Imhof, D. Song, and W.K. Wong, *Optimal design of experiments with possibly failing trials*, Stat. Sin. 12(4) (2002), pp. 1145–1155.
[7] L. Imhof, D. Song, and W.K. Wong, *Optimal design of experiments with anticipated pattern of missing observations*, J. Theor. Biol. 228(2) (2004), pp. 251–260.
[8] S.A. Ortega-Azurduy, F.E.S. Tan, and M.P.F. Berger, *The effect of dropout on the effciency of d-optimal designs of linear mixed models*, Stat. Med. 27 (2008), pp. 2601–2617.
[9] S. Garcet-Rodríguez, J. López-Fidalgo, and R. Martín-Martín, *Some complexities in optimal experimental designs introduced by real life problems*, Tatra Mt. Math. Publ. 39 (2008), pp. 135–143.
[10] J. Kiefer and J. Wolfowitz, *Optimum design in regression problems*, Ann. Math. Stat. 30 (1959), pp. 271–294.
[11] A.C. Atkinson and A.N. Donev, *Optimum Experimental Designs*, Oxford University Press, New York, 1992.
[12] V.V. Fedorov and P. Hackl, *Model-Oriented Design of Experiments*, Lecture Notes in Statistics, Spinger, New York, 1997.
[13] P. Whittle, *Some general points in the theory of optimal experimental design*, J. R. Stat. Soc. Ser. B 35 (1973), pp. 123–130.
[14] C.-F. Wu and H.P. Wynn, *The convergence of general step-length algorithms for regular optimum design criteria*, Ann. Stat. 6(6) (1978), 1273–1285.