Computational Statistics and Data Analysis 71 (2014) 1103–1112

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

'Nearly' universally optimal designs for models with correlated observations

Holger Dette [a], Andrey Pepelyshev [b], Anatoly Zhigljavsky [c,*]

[a] Fakultät für Mathematik, Ruhr-Universität Bochum, Bochum, 44780, Germany

[b] Faculty of Mathematics, St. Petersburg State University, Universitetsky pr. 28, St. Petersburg, 198504, Russia

[c] School of Mathematics, Cardiff University, Cardiff, CF24 4AG, UK

ARTICLE INFO

Article history:
Received 11 July 2012
Received in revised form 1 February 2013
Accepted 1 February 2013
Available online 8 February 2013

Keywords:
Optimal design
Correlated observations
Universally optimal design
Multiplicative algorithms

ABSTRACT

The problem of determining optimal designs for least squares estimation is considered in the common linear regression model with correlated observations. The approach is based on the determination of 'nearly' universally optimal designs, even in the case where the universally optimal design does not exist. For this purpose, a new optimality criterion which reflects the distance between a given design and an ideal universally optimal design is introduced. A necessary condition for the optimality of a given design is established. Numerical methods for constructing these designs are proposed and applied for the determination of optimal designs in a number of specific instances. The results indicate that the new 'nearly' universally optimal designs have good efficiencies with respect to common optimality criteria.

1. Introduction

Consider the common linear regression model

$$y(x) = \theta_1 f_1(x) + \cdots + \theta_m f_m(x) + \varepsilon(x), \tag{1}$$

where $f_1(x), \ldots, f_m(x)$ are linearly independent continuous functions, $\varepsilon(x)$ denotes a random error process or field, $\theta_1, \ldots, \theta_m$ are unknown parameters, and x is the explanatory variable, which varies in a compact design space $\mathcal{X} \subset \mathbb{R}^d$. We assume that N observations, say $y_1, \ldots, y_N$, can be taken at experimental conditions $x_1, \ldots, x_N$ to estimate the parameters in the linear regression model (1). Suppose that $\varepsilon(x)$ is a stochastic process with

$$\mathrm{E}[\varepsilon(x)] = 0, \qquad \mathrm{E}[\varepsilon(x)\varepsilon(x')] = K(x, x'), \quad x, x' \in \mathcal{X}. \tag{2}$$

Throughout this paper, we call the function $K(x, x')$ a covariance kernel. An important case appears when the error process is stationary and the covariance kernel is of the form $K(x, x') = \sigma^2 \rho(\|x - x'\|)$. If $\rho(0) = 1$, the function $\rho(\cdot)$ is called the correlation function, and, if $\rho(t) \to \infty$ as $t \to 0$, the function $\rho(\cdot)$ is a singular covariance function. Regression models with correlated errors are often used in practice, for example, in analysis of spatial models (Fedorov, 1996; Müller, 2007), computer experiments (Bates et al., 1996), and nonlinear models of chemical processes (Dette et al., 2010; Ucinski and Atkinson, 2004).

If N observations, say $Y = (y_1, \ldots, y_N)^T$, are available at experimental conditions $x_1, \ldots, x_N$, and the covariance kernel is known, then the vector of parameters can be estimated by the weighted least squares method, that is, by

* Corresponding author.
E-mail addresses: holger.dette@rub.de (H. Dette), andrey@ap7236.spb.edu (A. Pepelyshev), ZhigljavskyAA@cf.ac.uk (A. Zhigljavsky).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2013.02.002

$\hat{\theta} = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} Y$, where $\mathbf{X} = (f_i(x_j))_{j=1,\dots,N}^{i=1,\dots,m}$ is an $N \times m$ matrix and $\boldsymbol{\Sigma} = (K(x_i, x_j))_{i,j=1,\dots,N}$ is an $N \times N$ matrix. We assume that points $x_1, \dots, x_N$ are such that matrices $\boldsymbol{\Sigma}$ and $\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}$ are invertible. Note that the estimator $\hat{\theta}$ is the best unbiased linear estimator (BLUE) of θ, and its variance–covariance matrix is given by

$$\mathrm{Var}(\hat{\theta}) = (\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}.$$

If the correlation structure of the process is not known, one usually uses the ordinary least squares estimator $\tilde{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y$, which has the covariance matrix

$$\mathrm{Var}(\tilde{\theta}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \tag{3}$$

An exact experimental design $\{x_1, \dots, x_N\}$ is a collection of N points from the design space $\mathcal{X}$, which defines the time points or experimental conditions where observations are taken. Optimal designs for weighted or ordinary least squares estimation minimize a functional of the covariance matrix of the weighted or ordinary least squares estimator, respectively, and numerous optimality criteria have been proposed in the literature to discriminate between competing designs (see Pukelsheim, 2006).

Exact optimal designs for specific linear models with correlated observations have been investigated in Dette et al. (2008b), Kiseľák and Stehlík (2008), and Harman and Štulajter (2010). Because even in simple models exact optimal designs are difficult to find, most authors use asymptotic arguments to determine efficient designs for the estimation of the model parameters (see Sacks and Ylvisaker, 1966, 1968; Bickel and Herzberg, 1979, or Zhigljavsky et al., 2010).

Sacks and Ylvisaker (1966, 1968) and Näther (1985, Chapter 4) assumed that the design points $\{x_1, \dots, x_N\}$ are generated by the quantiles of a distribution function; that is, $x_i = a((i-1)/(N-1))$, $i = 1, \dots, N$, where the function $a : [0, 1] \to \mathcal{X}$ is the inverse of a distribution function. Let ξ_N denote a normalized design supported at N points $\{x_1, \dots, x_N\}$ with the weight $1/N$ assigned to each point. Then the covariance matrix of the least squares estimator $\tilde{\theta}$ given in (3) can be represented as

$$\mathrm{Var}(\tilde{\theta}) = D(\xi_N) = M^{-1}(\xi_N) B(\xi_N, \xi_N) M^{-1}(\xi_N), \tag{4}$$

where the matrices $M(\xi)$ and $B(\xi, \nu)$ are defined by

$$M(\xi) = \int f(u) f^T(u) \xi(\mathrm{d}u), \tag{5}$$

$$B(\xi, \nu) = \iint K(u, v) f(u) f^T(v) \xi(\mathrm{d}u) \nu(\mathrm{d}v), \tag{6}$$

respectively (the integration is always taken over the set $\mathcal{X}$), and $f(u) = \big(f_1(u), \dots, f_m(u)\big)^T$ denotes the vector of regression functions. We call any probability measure ξ on $\mathcal{X}$ an approximate design or simply a design; however, its interpretation in practice is different from the one given in Kiefer (1974), where, in the case of a discrete design $\xi = \{x_1, \dots, x_n; w_1, \dots, w_n\}$, the weight w_i means the relative proportion of observations performed at the point x_i. In the case of correlated errors, only one realization of a stochastic process is usually observed, implying that no replication of design points is needed, and in practice design points are computed as quantiles of the cumulative distribution function defined by ξ; this rule applies independently of whether ξ is a discrete or continuous probability measure (or a mixture of the two). If some points in the collection of quantiles replicate (this can happen if N is small and the optimal design is discrete), we can replace the points which replicate by other points that are near to them. The definitions of the matrices $M(\xi)$ and $B(\xi, \xi)$ can be extended to an arbitrary design ξ, provided that the corresponding integrals exist. The matrix

$$D(\xi) = M^{-1}(\xi) B(\xi, \xi) M^{-1}(\xi) \tag{7}$$

is called the covariance matrix for the design ξ, and can be defined for any probability measure ξ supported on the design space $\mathcal{X}$ such that the matrices $B(\xi, \xi)$ and $M^{-1}(\xi)$ are well defined. This set will be denoted by Ξ. We assume that the design set $\mathcal{X}$ has enough points so that the set Ξ is non-empty; that is, there exists at least one design $\xi \in \Xi$.

Optimal designs for regression models with dependent data have been investigated mainly for the location scale model. The difficulties in a general development of the optimal design theory for correlated observations can be explained by the different structure of the covariance of the least squares estimator in model (1), which is of the form $M^{-1} B M^{-1}$. As a consequence, the corresponding design problems are in general not convex (except for the location scale model where $m = 1$ and $f_1(u) \equiv 1$). Recently, Dette et al. (2011) derived universally optimal designs for regression models of arbitrary dimension if the corresponding regression functions are eigenfunctions of an integral operator defined by the covariance kernel of the error process. For example, the design with arcsine density is universally optimal for the polynomial model with logarithmic covariance kernel. On the other hand, there are many situations where this assumption is not satisfied, and in these cases there may not exist a universally optimal design.

The present paper is devoted to the numerical construction of 'nearly' universally optimal designs for regression models in such situations. This means that we consider model (1) with $m > 1$ parameters in the case where a universally optimal design does not exist. In Section 2, we introduce a new optimality criterion which reflects the distance between a given design and an ideal universally optimal design. A necessary condition for the optimality of a given design is established in