Journal of Statistical Planning and Inference 144 (2014) 81–91

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Elemental information matrices and optimal experimental design for generalized regression models

Anthony C. Atkinson [a], Valerii V. Fedorov [b,*], Agnes M. Herzberg [c], Rongmei Zhang [d]

[a] Department of Statistics, London School of Economics, London WC2A 2AE, UK
[b] Quintiles, 5927 South Miami Boulevard, Morrisville, NC 27560, USA
[c] Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada K7L 3N6
[d] Department of Biostatistics, Perelman School of Medicine, University of Pennsylvania 423 Guardian Drive, Philadelphia, PA 19104, USA

ARTICLE INFO

Available online 3 October 2012

Keywords:
Adaptive design
Convex optimal design
Elemental information matrix
Equivalence theorem

ABSTRACT

The construction of optimal experimental designs for regression models requires knowledge of the information matrix of a single observation. The latter can be found if the elemental information matrix corresponding to the distribution of the response is known. We present tables of elemental information matrices for distributions that are often used in statistical work. The tables contain matrices for one- and two-parameter distributions. Additionally we describe multivariate normal and multinomial cases. The parameters of response distributions can themselves be parameterized to provide dependence on explanatory variables, thus leading to regression formulations for wide classes of models. We present essential results from optimal experimental design and illustrate our approach with a few examples including bivariate binary responses and gamma regression.

1. Introduction

We are concerned with optimal experimental design for a rather general class of regression models. The observations need not be normally distributed, but might, for example, follow a beta or binomial distribution, or, indeed, any distribution for which the Fisher information matrix exists, together with regular maximum-likelihood estimators. By regression we intend that the parameters of these distributions are themselves functions of further parameters and of explanatory variables, at least some of which can be chosen and controlled in the experiment. Frequently a link function will be required, for example in the parameterization of the variance of a normal distribution or the probability in a Bernoulli model. We introduce the term "elemental information matrix" to emphasize that the information matrix is associated with a specific distribution in its standard form and is not related to the further parameterization that comes from the inclusion of explanatory variables.

Our main objective (see Sections 2 and 4) is to provide a collection of elemental information matrices for specific distributions which are crucial for the solution of optimal design problems. In this way, we bring together results that are repetitively scattered throughout the statistical literature. These matrices are essential for constructing information matrices for "single" observations. The latter may in many cases consist of a series of observations on a single subject. A typical example is when, together with a controlled variable, for example the dose of a drug, the model includes the times at which this dose is administered. For applications to experimental design we require the possibility of independent

* Corresponding author. Tel.: +1 919 9981855.
E-mail addresses: Valerii.Fedorov@quintiles.com (V.V. Fedorov), rongmei@mail.upenn.edu (R. Zhang).

0378-3758/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.09.012

replications of these series. Combining the concept of a single observation (supported by tables of information matrices) with a collection of sensitivity functions for various popular optimality criteria (see Section 3), the reader can address numerous tasks in the optimal design of experiments.

We begin in Section 2 with an introduction to the generalized regression model, continuing in Section 3 with some basic ideas about optimal experimental design. The core of our paper is Section 4 where we provide tables of elemental information matrices for one and two parameter univariate distributions. The multivariate normal distribution, including a parameterized variance, and multinomial distributions receive special attention in Sections 4.2 and 4.3. Examples 1 and 2 in Section 5 include the equivalence theorem for D-optimality for normal observations with a specific parameterized variance. In particular, Example 3 includes the family of designs for univariate generalized linear models, whereas, in Example 4 there are two binary responses. Example 5 is gamma regression when both the parameters are functions of explanatory variables.

2. Generalized regression

2.1. Main model

We assume that the observed response Y is distributed as

$$Y \sim p(y|\eta), \tag{1}$$

where Y and y are k-dimensional vectors, with η of dimension l. The parameters η depend on controls $x \in \mathcal{X}$, where $\mathcal{X}$ is a design set (region). Usually $\mathcal{X} \in \mathbb{R}^s$, but in general $\mathcal{X}$ could be a compact set of more complicated structure, for instance, part of a functional space (see Fedorov and Hackl, 1997, Section 5.7). In what follows we assume that $\mathcal{X} \in \mathbb{R}^s$ unless otherwise stated. In the standard setting for regression models, it is assumed that the expected values of responses Y are parameterized, as in Pázman (1986), Pukelsheim (1993), Fedorov and Hackl (1997) or Atkinson et al. (2007); see Example 1. The most popular examples are "normal" regression, binary regression and Poisson regression. Actually, all regression models with responses belonging to a one-parameter family can be treated in a similar way. For multiparameter distributions, η is a vector, all components of which may again depend on controls and other unknown parameters. When the components of η are of the same type (e.g. expectations in a multivariate normal model) the generalization is straightforward. However, in the case of other multi-parameter distributions (such as the gamma or Weibull), the forms of the various dependencies are less obvious.

As far as we know, in the experimental design literature only in the "normal" case have different types of components of η been parameterized, namely, the expectation(s) of the response(s) and its variance (variance–covariance matrix); see, for instance, Atkinson and Cook (1995), Dette and Wong (1999) or Fedorov and Leonov (2004). A rare exception is the beta regression considered by Wu et al. (2005) where again the mean and variance are separately parameterized.

2.2. Likelihood estimators and Fisher information matrices

Before proceeding with the design problem for generalized regression, we recall a few commonly known facts and introduce the required notation. Let

$$\eta = \eta(x,\theta), \tag{2}$$

where $\eta^T(x,\theta) = \{\eta_1(x,\theta), \ldots, \eta_k(x,\theta)\}$ are given functions of controls x and unknown regression parameters θ. Further, let n independent observations $\{Y_i\}_1^n$ be made at $\{x_i\}_1^n$. Eqs. (1) and (2) constitute a generalized regression model.

The maximum-likelihood estimator (MLE) $\hat{\theta}_n$ of θ can be defined as

$$\begin{aligned}\hat{\theta}_n &= \arg\max_{\theta\in\Theta} \prod_{i=1}^{n} p(y_i|\eta(x_i,\theta)) \\ &= \arg\max_{\theta\in\Theta} L(\{y_i\}_1^n,\{x_i\}_1^n,\theta),\end{aligned} \tag{3}$$

where Θ is compact and the true value θ_t of θ is an internal point of Θ. Under rather mild conditions on $p(y|\eta)$, $\eta(x,\theta)$ and on the sequence $\{x_i\}_1^n$ (see, for example, Lehmann and Casella, 1998, Chapter 2), the MLE $\hat{\theta}_n$ is strongly consistent and its normalized asymptotic variance–covariance matrix is

$$nD(\theta_t,\{x_i\}_1^n) = M^{-1}(\theta_t,\{x_i\}_1^n), \tag{4}$$

where

$$M(\theta,\{x_i\}_1^n) = \sum_{i=1}^{n} \mu(x_i,\theta), \tag{5}$$

and

$$\mu(x,\theta) = \mathrm{Var}\left[\frac{\partial}{\partial\theta}\ln p\{y|\eta(x,\theta)\}\right]. \tag{6}$$