Journal of Statistical Planning and Inference 154 (2014) 12–25

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Optimal designs for nonlinear regression models with respect to non-informative priors

Ina Burghaus [a], Holger Dette [b,*]

[a] Ruhr-Universität Bochum, Abteilung für Medizinische Informatik, Biometrie und Epidemiologie, 44780 Bochum, Germany
[b] Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany

ARTICLE INFO

Article history:
Received 23 October 2013
Received in revised form
19 May 2014
Accepted 21 May 2014
Available online 20 June 2014

Keywords:
Optimal design
Bayesian optimality criteria
Non-informative prior
Jeffreys prior
Reference prior
Polynomial regression
Canonical moments
Heteroscedasticity

ABSTRACT

In nonlinear regression models the Fisher information depends on the parameters of the model. Consequently, optimal designs maximizing some functional of the information matrix cannot be implemented directly but require some preliminary knowledge about the unknown parameters. Bayesian optimality criteria provide an attractive solution to this problem. These criteria depend sensitively on a reasonable specification of a prior distribution for the model parameters which might not be available in all applications. In this paper we investigate Bayesian optimality criteria with non-informative prior distributions. In particular, we study the Jeffreys and the Berger–Bernardo prior for which the corresponding optimality criteria are not necessarily concave. Several examples are investigated where optimal designs with respect to these criteria are calculated and compared to Bayesian optimal designs based on a uniform and a functional uniform prior.

1. Introduction

Nonlinear regression models provide an important tool to describe the relation between a response and a predictor and have many applications in engineering, physics, biology, economics and medicine, among others (see Ratkowsky, 1983). It is well known that a good design can improve the accuracy of the statistical analysis substantially and numerous authors have worked on the problem of constructing optimal designs for nonlinear regression models. An intrinsic difficulty of these optimization problems consists in the fact that the Fisher information, say $I(x,\boldsymbol{\theta})$, at an experimental condition x depends on the unknown parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ of the model. A common approach in the literature is to assume some prior knowledge of the unknown parameter, which can be used for the construction of optimal designs. Chernoff (1953) proposed the concept of local optimality where a fixed value of the unknown parameter is specified, and a design is determined by maximizing a functional of the information matrix for this specified parameter.

Since this pioneering work numerous authors have constructed locally optimal designs for various regression models (see He et al., 1996; Khuri et al., 2006; Fang and Hedayat, 2008; Yang and Stufken, 2009; Yang, 2010; Dette and Melas, 2011, among many others). On the other hand, the concept of local optimality has been criticized by several authors, because it depends sensitively on a precise specification of the unknown parameters and can lead to inefficient designs if these parameters are misspecified (see for example Dette et al., 2013, Example 2.1). As a robust alternative Pronzato and Walter

* Corresponding author.
E-mail addresses: ina.burghaus@rub.de (I. Burghaus), holger.dette@rub.de (H. Dette).

http://dx.doi.org/10.1016/j.jspi.2014.05.009
0378-3758/

(1985) and Chaloner and Larntz (1989) proposed Bayesian optimal designs which maximize an expectation of the information criterion with respect to a prior distribution for the unknown parameters (see also Chaloner and Verdinelli, 1995 for a review). Bayesian optimal designs for various prior distributions have been discussed by numerous authors (see Haines, 1995; Dette and Neugebauer, 1997; Han and Chaloner, 2003 or Braess and Dette, 2007 among others). However, there exist many applications where the specification of a prior distribution is difficult and several authors advocate the use of a uniform prior as a pragmatic approach if no preliminary knowledge about the unknown parameter is available. In a recent paper it was pointed out by Bornkamp (2012) that for several models the use of a uniform prior as a non-informative prior does not yield reasonable designs. This author proposed the concept of a functional uniform prior in order to construct Bayesian optimality criteria with respect to non-informative prior distributions.

In this paper we consider two alternative criteria for the construction of Bayesian optimal designs with respect to non-informative prior distributions. Roughly speaking, the criteria maximize the predicted Kullback–Leibler distance between the prior and the posterior distribution for the unknown parameter of the model with respect to the choice of the experimental design, where – in contrast to the classical approach to Bayesian optimality – the prior distribution depends also on the design of experiment. This dependence is a natural consequence of the Bayesian point of view in constructing non-informative priors (see for example Polson, 1992; Ibrahim and Laud, 1991). If the prior is chosen in order to maximize a distance between the prior and the posterior distribution in a regression model it usually depends on the predictors, i.e. the experimental design. The criteria based on the Jeffreys prior or the Berger-Bernardo prior are introduced in Section 2, which also gives an introduction into the field of optimal experimental design. Here it is demonstrated that Bayesian optimal design problems corresponding to non-informative priors are in general not concave. Necessary conditions for the optimality of a given design are also derived. In Section 4 we discuss specific mathematical models to obtain a better understanding of the criteria based on the Jeffreys prior or the Berger-Bernardo prior. In particular we use the theory of canonical moments which is introduced in Section 3 (see also Dette and Studden, 1997) in order to determine saturated Bayesian optimal designs with respect to non-informative priors for polynomial regression models with a heteroscedastic error structure. Finally, in Section 5 we consider two frequently used nonlinear regression models and compare the optimal designs with respect to the criteria based on the Jeffreys prior or the Berger-Bernardo prior with optimal designs with respect to "classical" Bayesian optimality criteria based on a uniform and a functional uniform distribution. Most of the designs derived in this paper are saturated designs and do not allow to check the goodness-of-fit of the postulated model. An interesting topic for future research consists in the development of efficient designs which are robust with respect to the model assumptions.

2. Optimal design and non-informative priors

An approximate design is defined as a probability measure ξ on the design space $\mathcal{X}$ with finite support (see Kiefer, 1974). If the design ξ has masses ξ_i at the points x_i $(i=1,\dots,m)$ and N observations can be made by the experimenter, this means that the quantities $\xi_i N$ are rounded to integers, say N_i, satisfying $\sum_{i=1}^m N_i = N$, and the experimenter takes N_i observations at each location x_i $(i=1,\dots,m)$. The corresponding design with masses N_i/N at the points x_i $(i=1,\dots,m)$ will be denoted as exact design ξ_N. Assume that ξ_N is an exact design with masses N_i/N at points x_i $(i=1,\dots,m)$ and that N_i independent observations $Y_{i1},\dots,Y_{iN_i}$ are taken at each x_i with density

$$p(y_{ij}|\boldsymbol{\theta},x_i);\quad j=1,\dots,N_i,\quad i=1,\dots,m; \tag{2.1}$$

such that

$$\lim_{N\to\infty}\frac{N_i}{N}=\xi_i>0,\quad i=1,\dots,m, \tag{2.2}$$

where $\boldsymbol{\theta}\in\Theta$ is a k-dimensional parameter. For the design ξ_N we define by

$$p(\boldsymbol{y}|\boldsymbol{\theta},\xi_N)=\prod_{i=1}^m\prod_{j=1}^{N_i}p(y_{ij}|\boldsymbol{\theta},x_i)$$

the joint density of the N-dimensional vector $\boldsymbol{Y}=(Y_{11},\dots,Y_{mN_m})^T$. In the following we assume that the prior distribution for the parameter $\boldsymbol{\theta}$ may depend on the design (such as the Jeffreys prior) and consider the problem of maximizing the expected Kullback–Leibler distance between the prior and the posterior distribution with respect to the choice of the design ξ_N, that is

$$U(\xi_N)=\int\log\left(\frac{p(\boldsymbol{\theta}|\boldsymbol{y},\xi_N)}{p(\boldsymbol{\theta}|\xi_N)}\right)\cdot p(\boldsymbol{\theta},\boldsymbol{y}|\xi_N)\,d\boldsymbol{\theta}\,d\boldsymbol{y}. \tag{2.3}$$

Here $p(\boldsymbol{\theta}|\xi_N)$ denotes the density of the prior distribution of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\boldsymbol{y},\xi_N)$ is the density of the posterior distribution of $\boldsymbol{\theta}$ given $\boldsymbol{y}$ and $p(\boldsymbol{\theta},\boldsymbol{y}|\xi_N)$ is the density of the joint distribution of $(\boldsymbol{\theta},\boldsymbol{Y})$. Note that all distributions may depend on the design ξ_N.

Under regularity assumptions it can be shown by similar arguments as in Chaloner and Verdinelli (1995) that the expected Kullback–Leibler distance can be approximated by

$$U(\xi_N)\approx-\frac{k}{2}\log\left(\frac{2\pi}{N}\right)-\frac{k}{2}+\frac{1}{2}\int\log(|M(\xi_N,\boldsymbol{\theta})|)p(\boldsymbol{\theta}|\xi_N)\;d\boldsymbol{\theta}-\int\log p(\boldsymbol{\theta}|\xi_N)\;p(\boldsymbol{\theta}|\xi_N)\,d\boldsymbol{\theta}, \tag{2.4}$$