Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 4059–4070
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Locally D-optimal designs for multistage models and heteroscedastic polynomial regression models

Zhide Fang[a,*], Douglas P. Wiens[b], Zheyang Wu[a]

[a] *Department of Mathematics, University of New Orleans, New Orleans LA 70148, USA*
[b] *Department of Mathematical and Statistical Sciences, University of Alberta, USA*

Received 31 December 2003; accepted 19 November 2004
Available online 23 May 2005

Abstract

We consider the construction of locally D-optimal designs for a nonlinear, multistage model in which one observes a binary response variable with expected value $P(x;\boldsymbol{\theta}) = H(\theta_0 + \theta_1 x + \cdots + \theta_k x^k)$. Here H is any twice differentiable distribution function. Our results apply as well to heteroscedastic polynomial regression models, under mild conditions on the efficiency function.

MSC: Primary 62K05; 62F35; Secondary 62J05

Keywords: Bioassay; D-optimality; Dose response; Optimal design

1. Introduction

Consider the design problem for a dose-response model, in which one observes, with error, a binary response variable $Y(x)$ with expected value

$$P(x;\boldsymbol{\theta}) = H(\theta_0 + \theta_1 x + \cdots + \theta_k x^k), \tag{1}$$

corresponding to an input variable (dose level) x lying in an interval $[0, b]$, $0 < b < \infty$. Here H is any twice differentiable distribution function. The parameters θ_j are unknown, and are

* Corresponding author. Tel.: +1 504 2806108; fax: +1 504 2805516.
E-mail addresses: zfang@math.uno.edu (Z. Fang), doug.wiens@ualberta.ca (D.P. Wiens), ywu2@uno.edu (Z. Wu).

0378-3758/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.11.015

to be estimated. In oncological studies, (1) with $H(z) = 1 - \exp(-z)$ for $z = \sum_{i=0}^{k} \theta_i x^i \geqslant 0$ is called a k-stage model, and $P(x; \boldsymbol{\theta})$ gives the lifetime probability of developing cancer after exposure at level x. The k-stage model can be adopted to assess the carcinogenic risks resulting from exposure to environmental chemicals. The model assumes that the mechanism of carcinogenesis can be interpreted as a series of k somatic mutations at the cellular level. After going through this series of mutational stages, a cell becomes malignant and proceeds to develop into a tumour. See, for example, Armitage and Doll (1954), Portier and Hoel (1983), Portier and Elder (1990) and references therein for detailed discussion of the biological background of the model.

Suppose that n subjects are tested at n, not necessarily distinct, dose levels $x_1, \ldots, x_n$. The design, denoted by ξ, is defined to be the measure placing mass n^{-1} at each x_j. Set $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_k)'$ and $\mathbf{f}(x) = (1, x, \ldots, x^k)'$. The vector $\boldsymbol{\theta}$ is typically estimated by maximum likelihood. From the Bernoulli likelihood

$$Y(x_i) \sim \text{bin}(1, P(x_i; \boldsymbol{\theta}) = H(\boldsymbol{\theta}'\mathbf{f}(x_i))),$$

one obtains the information matrix of the design as

$$\begin{aligned} I(\boldsymbol{\theta}; \xi) &= \sum_{i=1}^{n} \mathbf{f}(x_i)\lambda(x_i; \boldsymbol{\theta})\mathbf{f}'(x_i) \\ &= \int_0^b \mathbf{f}(x)\lambda(x; \boldsymbol{\theta})\mathbf{f}'(x)\xi(\mathrm{d}x), \end{aligned} \tag{2}$$

where, with $h = H'$,

$$\lambda(x; \boldsymbol{\theta}) = \left\{ \frac{h^2(z)}{H(z)(1 - H(z))} \right\}_{|z=\boldsymbol{\theta}'\mathbf{f}(x)}. \tag{3}$$

In the theory of optimal designs, the criterion for optimality is usually defined through some scalar function of the information matrix. Then the optimal design is the one which maximizes (or minimizes) this function among a class of candidate designs. For the multi-stage model (1), the information matrix depends on the unknown parameter vector $\boldsymbol{\theta}$. Thus the design problem becomes more difficult than that for linear regression models. One way to handle this difficulty is to guess the true value, or obtain a good initial estimate, of $\boldsymbol{\theta}$, and to then construct the *locally optimal* design. Locally optimal designs are introduced and discussed by Chernoff (1953), Fedorov (1972), Silvey (1980), Ford et al. (1989) and others.

In this paper, we seek the locally D-optimal designs for model (1), i.e., designs maximizing $\log |I(\boldsymbol{\theta}; \xi)|$ for a prespecified choice of $\boldsymbol{\theta}$. Although $\boldsymbol{\theta}$ is of course never known in practice, locally optimal designs are of interest, especially when good initial parameter estimates are available or when sequential designs can be carried out in batches—the estimates of parameters based on the previous batches can serve as the initial estimates in order to construct the locally optimal design for the next batch. See Ford et al. (1992) and Sitter and Fainaru (1997) for a detailed discussion.

Our model is different from those considered in Sitter and Fainaru (1997). They construct, by geometric methods, locally optimal designs for bioassay models $P(x; \boldsymbol{\theta}) = H(\theta_1(x - \theta_2))$, where H is a distribution function with a symmetric density. Logistic and probit models are