Designing experiments for semi-parametric B-spline models

David Woods
Department of Mathematics, University of Southampton
Southampton, SO17 1BJ, UK
D.C.Woods@maths.soton.ac.uk

Sue Lewis
Department of Mathematics, University of Southampton
Southampton, SO17 1BJ, UK
S.M.Lewis@maths.soton.ac.uk

1. Introduction

This work was motivated by the design of industrial factorial experiments, where previous knowledge indicated that a specific variable influenced a continuous response Y in a non-smooth way across a particular region, whilst the remaining variables had smooth effects. The use of B-spline basis functions (Eubank, 1999) enables the formulation of identifiable, parsimonious models which are extensions of additive semi-parametric models (Green and Silverman, 1995, p. 64). Efficient designs can then be found via algorithmic search using, for example, V- or D-optimality. Criteria for design choice have been developed that take account of uncertainty in the knot locations. This methodology can also be used when no prior information is available on the location of potential non-smooth behaviour, and when the effects of several variables are modelled using basis functions (Lewis, Woods and DeWynne, 2001).

2. Methodology

Non-additive behaviour between the $n-1$ parametric variables $x_2, \ldots, x_n$ and the complete set of B-spline basis functions, $B_i(x_1)$, $i = 1, \ldots, m$ (for specified number of knots, degree and smoothness at each knot), can be expressed in identifiable models of the form

$$E(Y) = \sum_{i_1=1}^{m} \alpha_{i_1} B_{i_1}(x_1) + \sum_{i_1=1}^{m} \sum_{i_2=2}^{n} \beta_{i_1 i_2} B_{i_1}(x_1) x_{i_2} + \sum_{j=2}^{d} \sum_{i_1=2}^{n} \sum_{i_2 \geq i_1}^{n} \cdots \sum_{i_j \geq \ldots \geq i_1}^{n} \gamma_{i_1 \ldots i_j} x_{i_1} x_{i_2} \ldots x_{i_j} \,,$$

where the α, β, γ are constants and $2 \leq d \leq n-1$. Additive iid Normal errors are assumed. A modified Fedorov exchange algorithm is applied to the $B_i(x_1)$ $(i = 1, \ldots, m)$ and the x_j $(j = 2, \ldots, n)$. Three optimality criteria are incorporated: local optimality, and a Bayesian approach and minimax argument that find designs robust to uncertainty in the knot locations.

3. Example

Consider $n = 3$, a maximally smooth cubic basis with one knot in [0.3, 0.5] for x_1, and the following simple model with $d = 2$, where each variable is scaled to lie between 0 and 1:

$$E(Y) = \sum_{i_1=1}^{5} \alpha_{i_1} B_{i_1}(x_1) + \sum_{i_1=1}^{5} \sum_{i_2=2}^{3} \beta_{i_1 i_2} B_{i_1}(x_1) x_{i_2} + \sum_{i_1=2}^{3} \sum_{i_2 \geq i_1}^{3} \gamma_{i_1 i_2} x_{i_1} x_{i_2} .$$

Figure 1 compares 3 designs in 28 runs found by using local D-optimality (knot at 0.4), minimization of the expected determinant of the inverse information matrix assuming a uniform knot distribution on 0.3, 0.4, 0.5 (a Bayesian approach), and minimization of the maximum determinant for knot positions 0.3, 0.4 and 0.5. It indicates that the first method is inferior for the upper half of the anticipated range for the knot location.

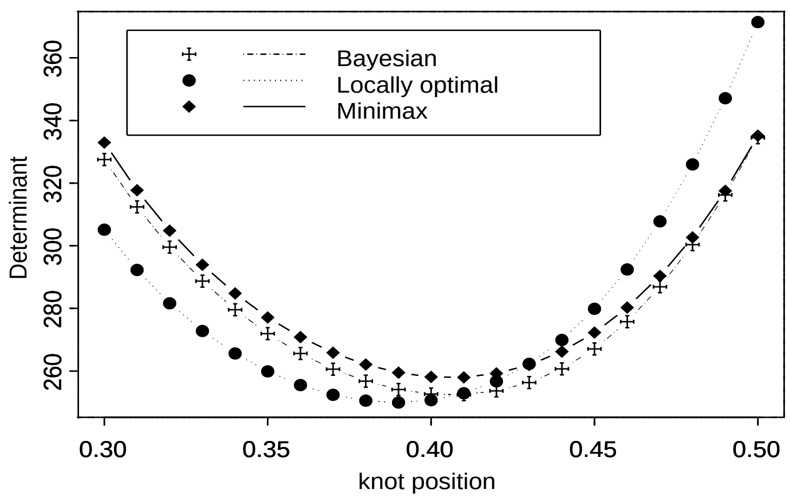

Figure 1: Determinant of the inverse information matrix for three designs.

REFERENCES

Eubank, E.L. (1999). *Nonparametric Regression and Spline Smoothing. Marcel Dekker.*

Lewis, S.M., Woods, D.C. and DeWynne, J.N. (2001). Design of experiments for semiparametric models based on B-splines. *Tech. rep. 356, Mathematics, Southampton Univ.*

Green, P.J. and Silverman, B.W. (1995). *Nonparametric Regression and Generalized Linear Models. Chapman and Hall.*

RESUME

On décrit une méthode pour le design factoriel dans lesquelles on utilise un modèle semiparamètrique impliquant des fonctions B-splines. On applique un algorithme de recherche qui incorpore un choix de critères prenant en compte l'incertitude sur la position des noeuds.