

# A Note on Bayesian D-optimality

Matt Simpson

September 1, 2016

## 1 Introduction

This is a quick note outlining the class of problems we need to look for in order to use PSO for finding Bayesian D-optimal designs, or nearly optimal designs. Strictly speaking, D-optimality is more narrow than necessary.

## 2 General Framework

Suppose we wish to design an experiment with  $n$  experimental units in order to determine the effect of some covariates on some response  $y_i, i = 1, 2, \dots, n$ . Let  $\mathbf{y} = \mathbf{y}_{1:n} = (y_1, y_2, \dots, y_n)'$ . The covariates are split into two types — fixed covariates and adjustable covariates. Fixed covariates are immutable features of the experimental units, while adjustable covariates are features which are chosen by the experimenter. For example, any treatment applied to the experimental unit is an adjustable covariate. Let  $\mathbf{X}^f$  denote the  $n \times p$  matrix of fixed covariates,  $\mathbf{X}^a$  denote the  $n \times q$  matrix of adjustable covariates, and  $\mathbf{X} = [\mathbf{X}^f, \mathbf{X}^a]$  the full  $n \times (p+q)$  matrix of covariates. Finally let  $\boldsymbol{\theta}$  denote a vector of model parameters. Then the generic model is

$$\mathbf{y}|\mathbf{X}, \boldsymbol{\theta} \sim [\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}], \quad \boldsymbol{\theta}|\mathbf{X} \sim [\boldsymbol{\theta}]$$

where  $[\cdot|\cdot]$  denotes the probability density or mass function of the enclosed random variables. Often  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi)$ ,  $\mathbf{y}|\mathbf{X}, \boldsymbol{\theta} \sim [\mathbf{y}|\boldsymbol{\mu}, \phi]$  with  $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$  and  $g(\cdot)$  is a known link function.

The design problem is to choose  $\mathbf{X}^a \in \mathcal{X}$  in order to maximize some design criterion. The space  $\mathcal{X}$  contains all constraints on the elements of  $\mathbf{X}^a$ ,

for example if  $x_{i2} = x_{i1}^2$  in a quadratic regression model. What is being chosen here is not which covariates to include, but rather the values of the covariates. For example in a simple linear regression model with  $\mathbf{X}^f = (1, 1, \dots, 1)'$  corresponding to the intercept and  $\mathbf{X}^a = (x_1, x_2, \dots, x_n)$  corresponding to the slope, the design problem is to choose the values of  $x_1, x_2, \dots, x_n$  for a fixed  $n$ , typically where  $x_i$  lives in a constrained space.

A standard Bayesian design criterion is the expected shannon information gain:

$$E_{\mathbf{y}, \boldsymbol{\theta}} \left\{ \log \frac{[\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}]}{[\boldsymbol{\theta}]} \right\} = \iint \log \frac{[\boldsymbol{\theta}|\mathbf{y}, \mathbf{X}]}{[\boldsymbol{\theta}]} [\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}] [\boldsymbol{\theta}] d\boldsymbol{\theta} d\mathbf{y}.$$

Maximizing this in  $\mathbf{X}^a$  is equivalent to maximizing

$$U(\mathbf{X}^a) = \iint \log \frac{[\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}]}{[\mathbf{y}|\mathbf{X}]} [\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}] [\boldsymbol{\theta}] d\boldsymbol{\theta} d\mathbf{y}.$$

When  $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \sigma^2 \mathbf{S}_0)$  and  $\mathbf{y}|\mathbf{X}, \boldsymbol{\theta} \sim N(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n)$  where  $\mathbf{I}_n$  is the  $n \times n$  identity matrix and  $\sigma^2$  is known, then maximizing  $U(\mathbf{X}^a)$  is equivalent to maximizing the Bayesian D-optimality criterion:

$$D(\mathbf{X}^a) = |\mathbf{X}'\mathbf{X} + \mathbf{S}_0^{-1}|.$$

However in general, maximizing  $U$  and  $D$  are not equivalent, and when they are not, typically  $U(\mathbf{X}^a)$  must be approximated via monte carlo simulation as well as an approximation for the model's marginal likelihood,  $[\mathbf{y}|\mathbf{X}]$ .

Whether the model is  $U$  or  $D$ , finding the optimal design is a difficult optimization problem, and near-optimal designs are often desirable. As such, this is a perfect problem for heuristic optimization algorithms such as particle swarm optimization. The problem for us, then, is to find an example that fits into this framework and ideally uses federal data.