Proceedings of the 6th International Conference on Statistics: Theory and Applications (ICSTA'24)
Barcelona, Spain – August 19 - 21, 2024
Paper No. 155
DOI: 10.11159/icsta24.155

Study Over Some Approaches in Formulating Optimal Designs

Akram Mahmoudi[1], Saumen Mandal[2]
[1]Jönköping international business school, Jönköping University
Jönköping, Sweden
akram.mahmoudi@ju.se; saumen.mandal@umanitoba.ca
[2]Department of Statistics, University of Manitoba
Winnipeg, Canada

Extended Abstract

Due to the extensive utilization of optimal designs our interest lies in some methods used for achieving optimum designs. Consequently, we consider two optimization algorithms to construct approximate optimal designs. One of these methods is a gradient-based algorithm and the other one is a gradient-free algorithm. Accordingly, we investigate the effectiveness of the multiplicative algorithm (MA) and simulated annealing (SA) method. Our analysis involves simulating these optimization algorithms in both quadratic and cubic models. Through diverse scenarios, we examine the performance of the estimators, identifying the strengths and weaknesses of each method. Ultimately, we present a comprehensive comparison between these methods.

In this study, our primary focus is centred around the widely used information-based optimality criteria. Specifically, we explore the well-known D-optimality criterion introduced by [1] and [2], which aims to enhance the precision of estimating regression coefficients. In this study, our main objective is to evaluate the performance of two optimization algorithms under two distinct models. Through a comprehensive simulation study, we thoroughly examine the performance of studied optimal design methods and compare their accuracy, convergence time, and complexity under various scenarios. The focus is primarily on two optimization algorithms: the gradient-based multiplicative algorithm and the gradient-free simulated annealing method.

We consider the problem of selecting an experimental design to provide information on a model $y \sim p(y|\boldsymbol{x}, \boldsymbol{\theta}, \sigma)$ where y is the response variable; $p(.)$ is a probability model; $\boldsymbol{x}$ is a vector of design variables, $\boldsymbol{x} \in X$, X is the design space; $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)'$ is a vector of unknown parameters and σ is a nuisance parameter, fixed but unknown. In linear models it is further assumed that $y(\boldsymbol{x})$ has an expected value of $E(y|\boldsymbol{x}) = \boldsymbol{v}'\boldsymbol{\theta}$ where $\boldsymbol{v} \in V$, and $V = \{\boldsymbol{v} \in R^k : \boldsymbol{v} = \boldsymbol{\eta}(\boldsymbol{x}),\ \boldsymbol{x} \in X\}$ $with$ $\boldsymbol{\eta}(\boldsymbol{x}) = (\eta_1(\boldsymbol{x}), \eta_2(\boldsymbol{x}), \ldots, \eta_k(\boldsymbol{x}))'$.

Assuming the set V=$\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_J\}$, then p is characterized by a set of weights p_1, p_2, ..., p_J with each weight p_j assigned to vector $\boldsymbol{v}_j$ and satisfying $p_j \geq 0$ for all $j = 1, 2, \ldots, J$. The objective is to optimally select these ps (weights). Given that $\widehat{\boldsymbol{\theta}}$ is the least square estimator of $\boldsymbol{\theta}$, the covariance of $\widehat{\boldsymbol{\theta}}$ is proportional to $M^{-1}(p)$, where $M(p)$ represents the information matrix. The goal is to maximize $g(p) = (1/k)\log(\det(M(p)))$ over the set p defined as $\{p = (p_1, p_2, ..., p_J) : p_j \geq 0, \sum_{j=1}^{J} p_j = 1\}$.

The multiplicative algorithm, extensively discussed by [3], [4] and [5], is an iterative approach for generating optimal designs. The process starts with an initial design and updates it iteratively. The algorithm is defined as following,

$$p_j^{(r+1)} = \frac{p_j^{(r)} f(d_j^{(r)}, \delta)}{\sum_{j=1}^{J} p_j^{(r)} f(d_j^{(r)}, \delta)}, \tag{1}$$

where $d_j^{(r)}$ is the partial derivative of g with respect to p_j evaluated at $p = p^{(r)}$, and $f(.)$ is a positive, strictly increasing function that may depend on a parameter δ. For the D-optimal criterion, $d_j > 0$.

Another optimization algorithm discussed in this paper is Simulated Annealing, which is a gradient-free algorithm developed by [6]. The SA method offers the advantage of escaping from local optimums, allowing for more effective optimization. A comprehensive description of simulated annealing and its practical implementation can be found in the seminal work by [7], which highlights its characteristics and performance in various applications.

The Simulated annealing has the following steps:

- we start with $\boldsymbol{p}^{(0)}$. Stage $i = 0, 1, 2, \ldots$ has m_i iterations; first we set $i = 0$,
- Given iteration $\boldsymbol{p}^{(r)}$, we generate $\boldsymbol{p}^{(r+1)}$ as follows:

1- sample a candidate $\boldsymbol{p}^*$ from a proposal distribution $p(.|\boldsymbol{p}^{(r)})$,

2- compute $h(\boldsymbol{p}^{(r)}, \boldsymbol{p}^*) = \exp(\frac{g(\boldsymbol{p}^*) - g(\boldsymbol{p}^{(r)})}{\tau_i})$, where τ_i is a temperature and gets any values greater than zero,

3- define next iteration $\boldsymbol{p}^{(r+1)}$ according to,

$$\boldsymbol{p}^{(r+1)} = \begin{cases} \boldsymbol{p}^*, & \text{with probability } min\{h(\boldsymbol{p}^{(r)}), \boldsymbol{p}^*), 1\}, \\ \boldsymbol{p}^{(r)}, & otherwise \end{cases}$$

4- set $r <- r + 1$ and repeat steps 1-3, m_i times,

5- update $\tau_i = \alpha\tau_{i-1}$ and $m_i = \beta(m_{i-1})$, where function α should slowly decrease the temperature, while function β should be increasing; set $i <- i + 1$; go to step 1.

In the SA method, gradual decrease in temperature ensures that only improvements are accepted. This strategic approach allows the algorithm to thoroughly explore the solution space instead of prematurely converging to a local optimum. As a result, the algorithm increases its chances of discovering the global optimum, even in the presence of multiple local optima.

The performance of the MA and SA methods are investigated through a simulation study considering quadratic and cubic models. In the quadratic model, both methods of the MA and SA find the first, middle and last design points as the optimal ones, accordingly both methods have the same accuracy in this model. But in the cubic model, the MA finds the optimal points more accurately than the SA.

The general result of the paper in terms of comparing the two optimization algorithms of MA and SA under the quadratic and cubic models are summarized in the following:

In terms of the performance time of the algorithms, MA is much faster than the SA. Further, the MA is not very sensitive to its components (function f, δ and number of simulation), so it is easy to adjust the values of its components. Conversely, the SA is very sensitive to its components (τ, m, α, β and number of simulation), and an improper value for them, makes the SA to diverge. Therefore, it is important to choose the components of the SA properly.

Regarding the convergence of the methods, sometimes SA fails to converge, but MA has a good convergence behaviour. Finally, regarding the need of gradient in the algorithms, since MA is gradient based algorithm, it is necessary to obtain gradient in MA. However, in SA, we don't need to obtain gradient analytically; therefore, the SA is very useful when it is extremely difficult or maybe impossible to acquire gradients.

References

[1] Fedorov VV. "Theory of optimal experiments". Elsevier; 2013 Apr 20.

[2] Rady EA, Abd El-Monsef MM, Seyam MM. "Relationships among several optimality criteria". Interstat. 2009 Jun;15(6):1-1.

[3] Torsney B. "Contribution to discussion of 'Maximum Likelihood estimation via the EM algorithm'by Dempster et al". J. Roy. Statist. Soc. B. 1977;39:26-7.

[4] Torsney B, Martín-Martín R. "Multiplicative algorithms for computing optimum designs". Journal of Statistical Planning and Inference. 2009 Dec 1;139(12):3947-61.

[5] Fu L, Ma F, Yu Z, Zhu Z. "Multiplication algorithms for approximate optimal distributions with cost constraints". Mathematics. 2023 Apr 21;11(8):1963.

[6] Kirkpatrick S, Gelatt Jr C. D, and Vecchi M. P. "Optimization by simulated annealing". Science. 1983; 220(4598):671–680.

[7] Bertsimas D, Tsitsiklis J. "Simulated annealing. Statistical science". 1993 Feb;8(1):10-5.