Using Particle Swarm Optimization to identify D-optimal designs for binary response generalized linear models with mixed factors

Joshua Lukemire

Emory University

Design Problem

Our goal is to design experiments to obtain parameter estimates for the k factor model taking binary response

$$logit(\mu) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i. \tag{1}$$

D-optimal designs aim to minimize the area of the confidence ellipsoid for the parameter estimates, which corresponds to maximizing the determinant of the Fisher information matrix. The Fisher information matrix for an approximate design ξ is given by

$$\mathbf{I}_{\xi} = \sum_{i=1}^{m} p_i \Psi(X^T\beta) X_i X_i^T. \tag{2}$$

where for the logit link $\Psi(X^T\beta) = \frac{1}{2+e^{X^T\beta}+e^{-X^T\beta}} = \frac{e^{X^T\beta}}{(1+e^{X^T\beta})^2}$, p_i is the proportion of observations assigned to experimental setting i, and X_i is the i^{th} row of the design matrix, $i = 1, \ldots, m$.

Identifying optimal designs for generalized linear models is difficult because the information matrix depends on the model parameters. An optimal design at one set of parameter values may perform quite poorly at another set. To deal with this problem we use the local optimality approach and take an assumed set of values for the model parameters to construct designs.

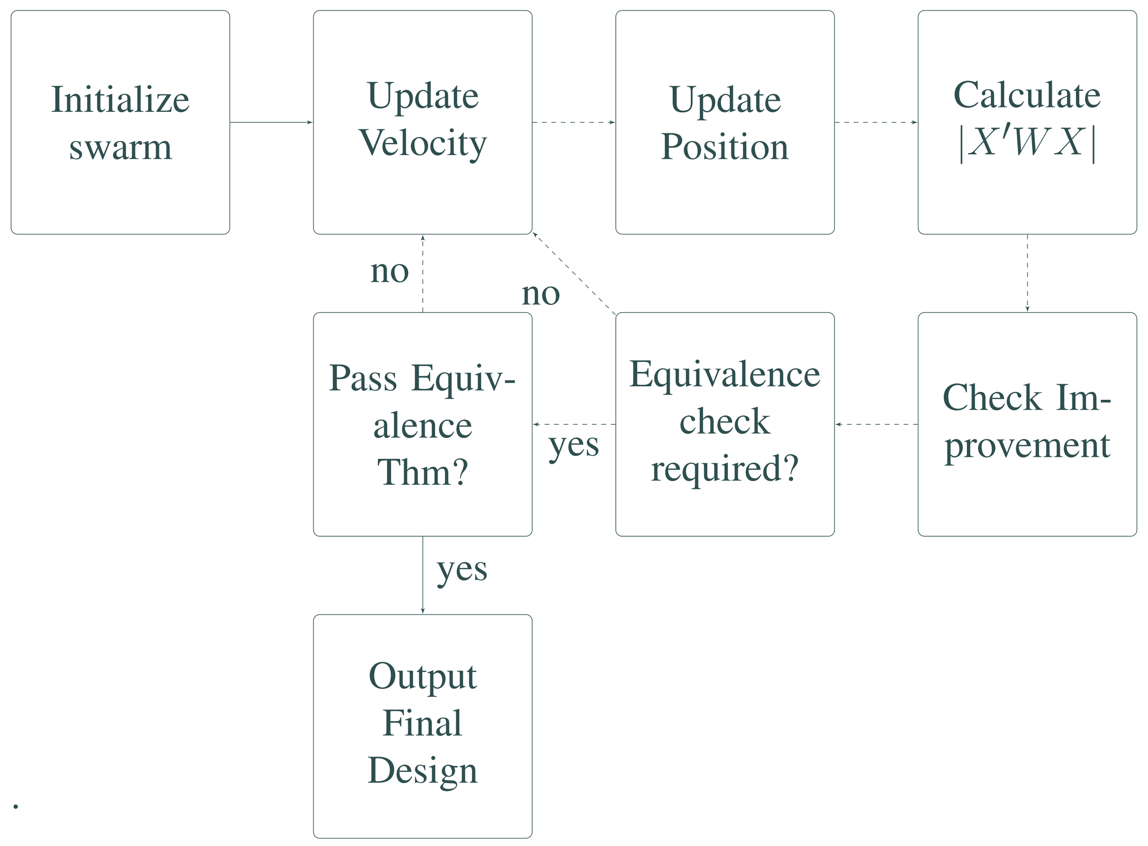
Theoretical results are not yet available for experiments with mixed factors. Instead the continuous factors are often treated as discrete, which can result in a large loss of efficiency. Numerical methods such as the popular Fedorov-Wynn type algorithms generally require a set of candidate points or an explicit objective function. As the number of factors increases an explicit objective function often cannot be found and the set of candidate points quickly becomes prohibitively large. We propose the use of Particle Swarm Optimization (PSO) to handle generating D-optimal designs without either of these requirements.

Particle Swarm Optimization

- Particle swarm optimization is a metaheuristic optimization algorithm based on animal behavior first introduced by Kennedy and Eberhart in 1995. While popular in other areas such as engineering, PSO is not yet commonly used to identify optimal designs.
- Goal: replicate the behavior of a swarm of birds as they search for food. Each member of the swarm, known as a particle, represents a candidate solution to the problem, and the food represents the value obtained at this solution.
- Each member of the swam has an idea of where the best solution is based on its previous positions. Each member also has knowledge of the best position any particle has achieved, known as the global best position. At each iteration each particle is pulled in the direction of these two positions.
- Algorithm implementation is written in C++ and called from R

.

PSO for Optimal Design of Experiments



.

Results

Structure of 2^2 Designs

We construct D-optimal designs for the model

$$logit(\mu) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \tag{3}$$

where $x_1 \in \{-1, 1\}$ and $x_2 \in [-1, 1]$. Our goal is to identify regions where minimally supported designs can be constructed as well as to identify regions in which designs can be constructed on the factor boundaries.

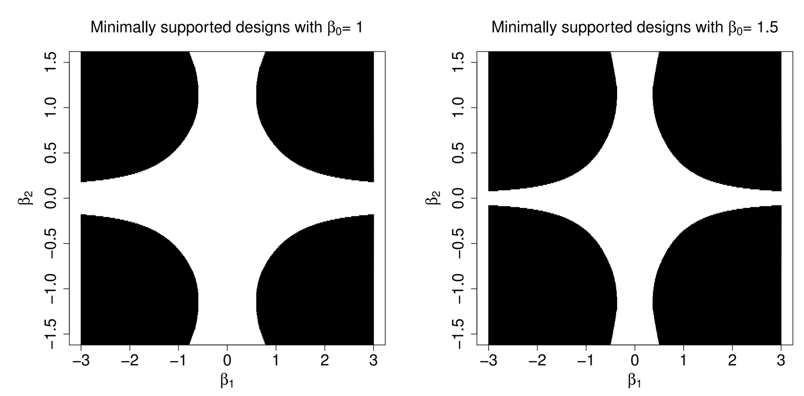
Simulation 1 Setup

- PSO settings: 25 particles, 100 maximum iterations, 500 maximum resets, convergence tolerance 0.0001, and minimum lower efficiency bound 99.9%
- Nominal values of $\beta_0 \in \{1, 1.5, 2\}$, and a grid of resolution 0.01 over $\beta_1 \in [-1.5, 1.5]$, and $\beta_2 \in [-3, 3]$
- Total of 180,901 optimal designs at each β_0

Simulation 1 Results

The black curvilinear areas in the first 3 panels of Figure 1 show parameter values for β_1 and β_2 for which minimally supported designs could be constructed with $\beta_0 = \{1, 1.5, 2\}$

The final panel in Figure 1 provides the region in which non-minimally supported designs can be constructed using only points on the boundary. The lines correspond to the boundaries within which 4 point designs could be constructed with $x_2 \in \{-1, 1\}$.



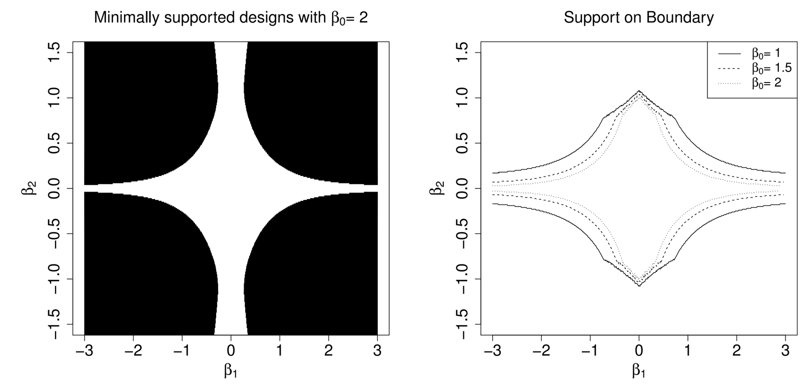


Figure 1: Minimally supported designs with $\beta_0 = 1, 1.5, 2$ and locations in which designs could be constructed using points only on the boundary.

Simulation 2: Loss of Efficiency

While the logit link is commonly used in practice, in some cases it may not be correct. Simulation 2 investigates the performance of logit-based designs when true link is not logit.

Simulation 2 Setup

- Nominal value $\beta_0 = 1$, search grid of resolution 0.01 over $\beta_1 \in [-1.5, 1.5]$, $\beta_2 \in [-3, 3]$.
- For each combination of parameter values use PSO to identify D-optimal design using logit, probit, log-log, and complementary log-log links.
- Compare the relative efficiency of the logit-based design to the design identified using the correct link function.

Simulation 2 Results

	True Link		
Quantile	Probit	Log-log	C-log-log
0.99	1.0000	1.0000	1.0000
0.95	1.0000	1.0000	0.9900
0.90	1.0000	1.0000	0.9488
0.80	0.9900	0.9737	0.8692
0.70	0.9670	0.9106	0.7925

Table 1: Relative efficiencies of designs constructed assuming the logit link when the true link was probit, log-log, and complementary log-log.

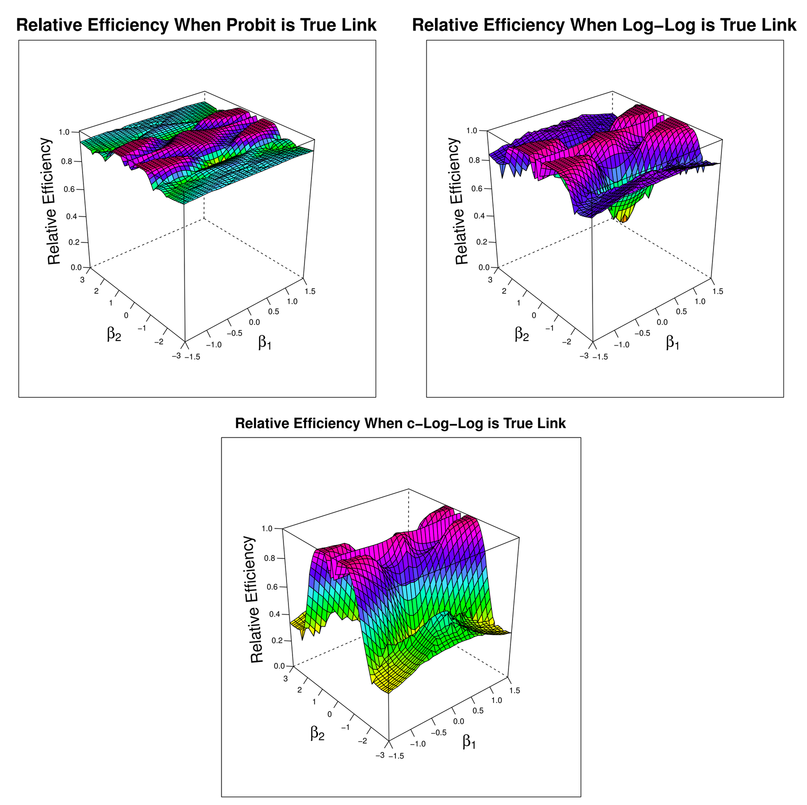


Figure 2: Relative efficiencies of the logit design to the design obtained using correct link function.

Irregular Design Regions

- Hypothetical plastic molding experiment similar to the one in (Anderson and Whitcomb, 2004).
- 2 continuous factors, temperature and pressure constrained to $5600 \leq 10 \times Temperature + Pressure \leq 5800$

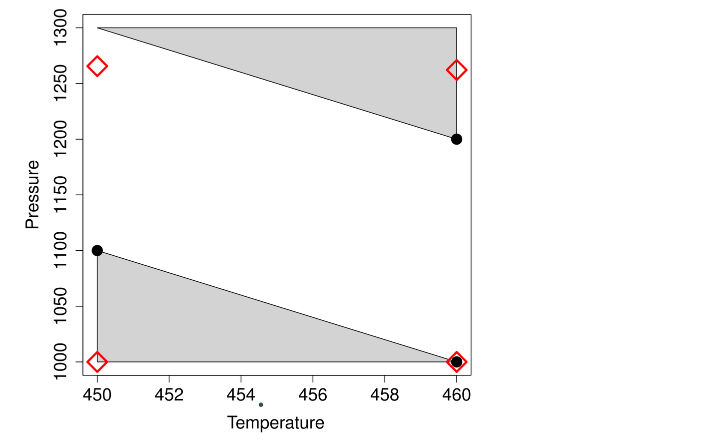


Figure 3: Plot of identified designs over the design space. The red diamonds correspond to the design that ignores the constraint, and the black dots correspond to the design that conforms to the constraint.

Temp	Pressure	p_i
450	1100.00	0.334
460	1200.00	0.335
460	1000.00	0.331

Temp	Pressure	p_i
450	1000.00	0.3220
450	1265.56	0.1899
460	1000.00	0.3214
460	1262.13	0.1667

Table 2: D-optimal designs obtained with (left) and without (right) conforming to the constraint.

Conclusions

- Particle swarm optimization is a powerful tool for identifying D-optimal designs for mixed factor experiments taking a binary response.
- For 2 factor experiments, minimally supported designs are often available. For those designs which are not minimally supported, it is often not optimal to place all experimental units at the factor boundaries (as is often done when the continuous factors are treated as discrete).
- The designs obtained by PSO are generally highly robust against misspecification of the link function.

References

[1] Mark Anderson and Patrick Whitcomb. *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments*. Productivity Press, 2004.

[2] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, volume 4, pages 1942–1948, Nov/Dec 1995.

Acknowledgements

Abhyuday Mandal [a] and Weng Kee Wong [b]

[a] Department of Statistics, University of Georgia
[b] Department of Biostatistics, University of California at Los Angeles