Statistical Methods - Part 2

Philip Dahlqvist-Sjöberg

29^{th} of May 2020

1

a)
When conducting an experiment, one aims to minimize the costs of sampling and other circumstances, as well as secure the internal validity (Berger & Wong 2009, p. 1). By planning and optimizing the experimental design, one can assure that the attributes of the variables are not merely unwanted variation, but represents the true treatment effects, i.e., estimates of parameters.

For the first example, we consider,

$$\eta(x, \boldsymbol{\beta}) = \beta_0 + \beta_1 x_i^2$$
$$x = \mathcal{X} : [0, 6].$$

From this,

$$\boldsymbol{J} = \frac{d\eta}{d\boldsymbol{\beta}} = f'(x) = \begin{bmatrix} \frac{\partial \eta}{\partial \beta_0} \\ \frac{\partial \eta}{\partial \beta_1} \end{bmatrix} = \begin{bmatrix} 1 \\ x^2 \end{bmatrix}, \tag{1}$$

where $\boldsymbol{J}$ is the $1 \times p$ *Jacobian* matrix, where p is the number of parameters. O'Brien & Funk (2003) provides a clear introduction how to get the standardized matrix for the least squares estimators. One multiplies the Jacobian (1), with itself,

$$\boldsymbol{Z} = \boldsymbol{J}\boldsymbol{J}^T = \begin{bmatrix} 1 \\ x_i^2 \end{bmatrix} \begin{bmatrix} 1 & x_i^2 \end{bmatrix}^T = \begin{bmatrix} 1 & x_i^2 \\ x_i^2 & x_i^4 \end{bmatrix}.$$

Further, they illustrates the standardized (Fisher) information matrix, where one uses the least squares estimator matrix, and the given design,

$$\xi_1 = \begin{Bmatrix} 0 & 3 & 6 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}, \tag{2}$$

to find the information matrix,

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{n} w_i \boldsymbol{Z}_i \tag{3}$$

$$\boldsymbol{M}(\xi_1, \boldsymbol{\beta}) = \begin{bmatrix} 1 & 15 \\ 15 & 459 \end{bmatrix}$$

b)
O'Brien & Funk (2003) continues to derive the standardized predictor variance of the design as,

$$d(x, \xi, \boldsymbol{\beta}) = \boldsymbol{J}^T \boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}) \boldsymbol{J}.$$

The standardized predictor variance of the design in (2), is plotted over the given interval in Figure 1. This illustrates the value for different points of the predictor variance. For a good design, points with

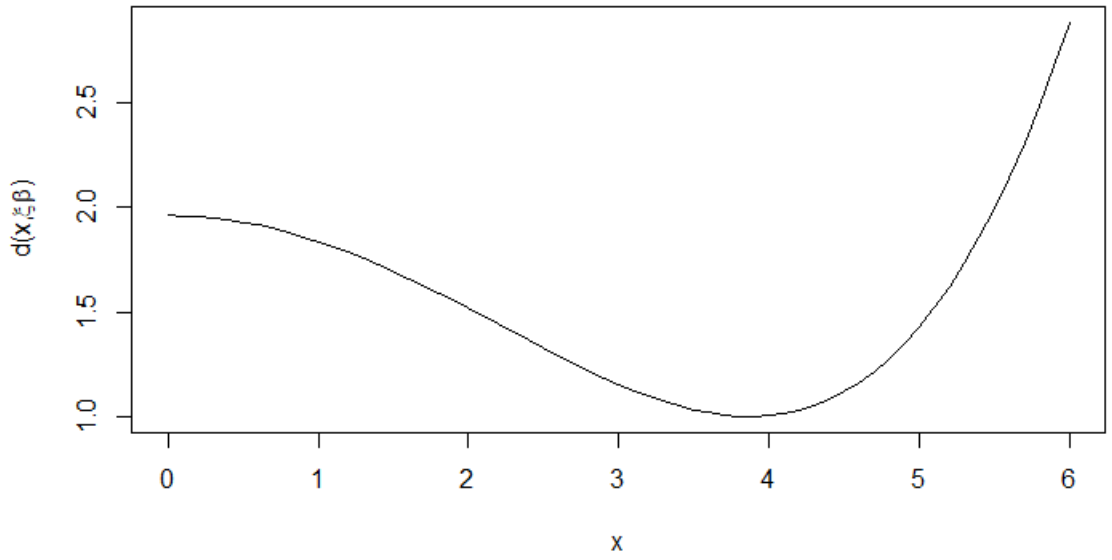

Figure 1: Standardized predictor variance of the design ξ_1, over $x = \mathcal{X} : [0, 6]$. The three points selected in the design are $0, 3$ and 6, which for a optimal design, would all be maximum in this graph. Hence, the conclusion of an unoptimal design.

high variance will capture the needed information to estimate parameters.

c)
O'Brien & Funk (2003) uses a similar graph to argue that an optimal design, will for all given points in the design on the x-axis be local or global maximum points in the given interval. Clearly in Figure 1, only $x = 6$ is a maximum point. Hence, this design is not an optimal design.

To find a optimal design, one can use different criteria methods. D-optimality criterion (Berger & Wong 2009, p. 40) is one of them, which minimizes the *volume* of the parameter confidence ellipse to find the most accurate estimators. The criterion looks like,

$$D - Criterion = \det(\mathrm{Cov}(\hat{\boldsymbol{\beta}})) = \ln(\det(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}))) = -\ln(\det(\boldsymbol{M}(\xi, \boldsymbol{\beta}))). \tag{4}$$

In R, there is a package called ICADO (Masoudi et al. 2019) with an inbuilt function called *locally*. This function has as default the D-criterion, and was iterated 100 times over the given interval $x = \mathcal{X} : [0, 6]$, initial parameter values of $\beta_0 = 1$ and $\beta_1 = 0$, and number of points, k, optimal at three. This result indicated a reduction of k to two was enough, which yielded the optimal design,

$$\xi_1^* = \begin{Bmatrix} 0 & 6 \\ \frac{1}{2} & \frac{1}{2} \end{Bmatrix}.$$

In Figure 2, the same graph as in Figure 1 is shown, but for the optimal design ξ_1^*. Clearly in the graph, these two new points are both local maximum of variance in the function, over the given interval $x = \mathcal{X} : [0, 6]$. Hence, these two points are an optimal design.

d)
Relative D-efficiency (RE_d) is a ratio between two different designs, in order to calculate the difference in sampling efficiency, usually for an unoptimal design relative to an optimal design (Berger & Wong

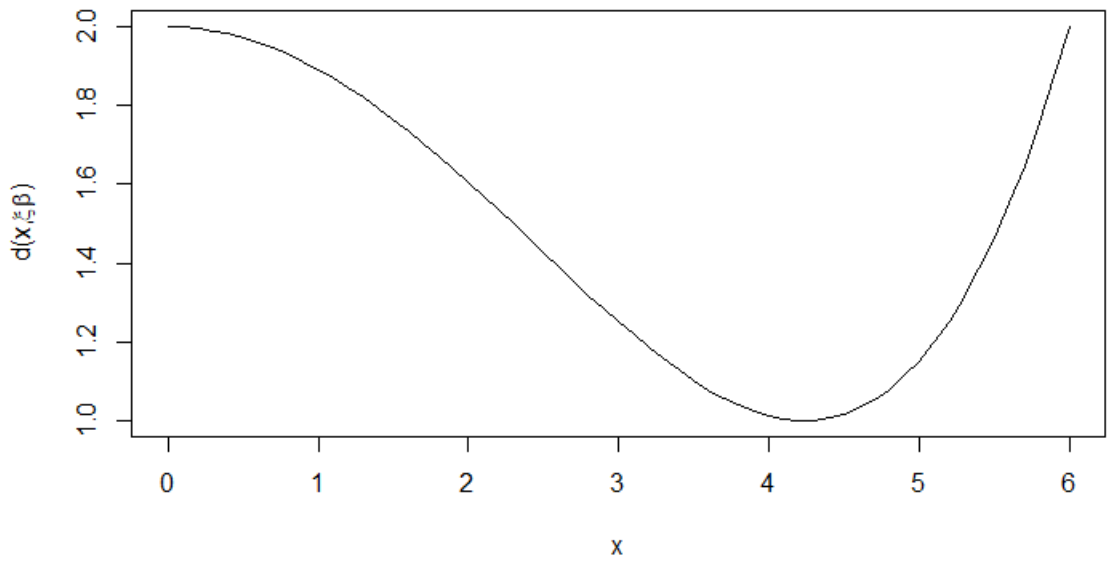

Figure 2: Standardized predictor variance of the design ξ_1^*, over $x = \mathcal{X} : [0, 6]$. The two points selected in the optimal design are 0 and 6, which are both maximum in this graph. Hence, the conclusion of an optimal design.

2009, p. 36-37). The ratio follows the criterion function, and with D-criterion, the ratio looks like,

$$RE_d = \frac{var(\hat{\boldsymbol{\beta}}^*)}{var(\hat{\boldsymbol{\beta}})} = \left(\frac{\det(\boldsymbol{M}^{-1}(\xi_1^*, \boldsymbol{\beta}))}{\det(\boldsymbol{M}^{-1}(\xi_1, \boldsymbol{\beta}))} \right)^{\frac{1}{p}} = 0.8498, \tag{5}$$

where p is the number of parameters in the model, which is two in this example. Further, to find percentage difference the calculation looks like,

$$(RE_d^{-1} - 1)100\% = 17.7\%. \tag{6}$$

Hence, the unoptimal design requires 17.7% more observations to achieve the same efficiency as the optimal design, i.e., the optimal design can reduce sampling marginal costs by 17.7% and have as much validity as the unoptimal design. However, this calculation assumes ceteris paribus when changing the sample size.

e)
A-optimal criterion is another criterion to find an optimal design (Berger & Wong 2009, p. 41). The algorithm itself works similar as in the D-optimal design, however the criterion function minimizes,

$$A - criterion = Trace[Cov(\hat{\boldsymbol{\beta}})] = \sum_{i=1}^{p} \frac{1}{\lambda_{\boldsymbol{M_i}}},$$

where p is the number of parameters, and $\lambda_{\boldsymbol{M_i}}$ is the i^{th} parameter eigenvalues of the standardized information matrix. For the given example, there are two parameters. The algorithms was iterated 100 times over the given interval $x = \mathcal{X} : [0, 6]$, initial design points was randomly generated, optimized individually then simultaneous, and number of points, k, optimal at two.

Looking in Table 1, the result are shown for both the hand-written algorithm and the build in, *locally*, function. Both results are identical at points $x = \{0, 6\}$ with almost all weight on 0, as it's optimal design. Both D-criterion and A-criterion attained the same optimal design points, however, their weights differ. See R-code for algorithm.

Table 1: A-optimal design 1e). Result from using the hand-written algorithm from lecture 2, and the same result from the *locally* built in function in R. Clearly, both algorithms conclude the same two values as the optimal design.

ξ^*	x	w	$x_{locally}$	$w_{locally}$
2	0	0.973	0	0.973
1	6	0.027	6	0.027

2

a)
We have a data set containing number of killed beetles, exposed with different dosing of gaseous carbon disulphide. Firstly, the data set it extended so each beetle is it's own observation. Then using the standard *glm* function in R, a logistic regression is run on the model,

$$p(x) = \frac{\exp \beta_0 + \beta_1 x + \beta_2 x^2}{1 + \exp \beta_0 + \beta_1 x + \beta_2 x^2}.$$

The estimated parameters are,

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 431.1 \\ -520.6 \\ 156.4 \end{bmatrix}.$$

b)
Firstly the $\eta(x, \boldsymbol{\beta})$ function is calculated from the log of odds,

$$\eta(x, \boldsymbol{\beta}) = \ln \left(\frac{p(x)}{1 - p(x)} \right).$$

The D-optimal design is calculated similarly to the design in Section 1, but for the logistic model, the standardized information matrix is dependent on the parameter values. The new matrix function looks like,

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{n} w_i p(x_i)(1 - p(x_i)) \boldsymbol{Z}_i, \tag{7}$$

where in this case,

$$\boldsymbol{Z} = \begin{bmatrix} 1 & x & x^2 \\ x & x^2 & x^3 \\ x^2 & x^3 & x^4 \end{bmatrix}.$$

The algorithms was then iterated 100 times over the interval $x = \mathcal{X} : [1.5, 2]$, initial parameter values of $\beta_0 = 431.1$, $\beta_1 = -520.6$ and $\beta_2 = 156.4$, initial design points was randomly generated, optimized individually then simultaneous, and number of points, k, optimal at four.

Note that the interval was chosen, taking the lowest and highest values of the recorded dose in the data set, and giving it some marginal. This could probably be selected more accurate with more specific knowledge of beetles. However, this interval will be used for all tests on this data.

Looking in Table 2, the results from both the hand-written algorithm and the *locally* function is displayed. Both algorithm find similar points for an optimal design. See R-code for algorithm.

c)
The sensitivity of the design from 2b) can be analyzed by changing the parameter input and using relative efficiency, equation (5), and calculating the number of extra subjects needed for the same

Table 2: D-optimal design 2b). Result from using the hand-written algorithm from lecture 2, and the same result from the *locally* built in function in R. Both algorithms conclude similar four values as the optimal design.

ξ^*	x	w	$x_{locally}$	$w_{locally}$
4	1.51731	0.3114205	1.52655	0.295
1	1.590779	0.1885823	1.57464	0.204
3	1.737716	0.1885764	1.75384	0.205
2	1.811185	0.3114207	1.80195	0.296

efficiency, equation (6). Hence, keeping the design points from 2b) constant, and only changing the $\boldsymbol{\beta}$ parameter values when calculating the standardized information matrix.

By using the parameter values from the *glm* model in the numerator, and altered values in the denominator, the change in efficiency can show how sensitive the design is.

Here are results, from taking all three parameters, minus one,

RE_d	% Extra subjects
0.0363	2655%

The altered parameters achieve a very low relative efficiency to the *glm* parameters. Hence, to attain same efficiency, over 26 times more subjects are needed. This shows how such a small change of the parameters, can affect the efficiency of the design. Hence, if the parameters are wrongly estimated, which is very likely, the design choice may be very wrong for the true parameters.

d)
To solve the problem described above, two strategies to deal with parameter dependency issue in logistic model is Bayesian design, and sequential design. Bayesian design uses prior distribution to find parameters and the optimal design is attained by averaging over the prior distribution (Berger & Wong 2009, p. 244-245).

Firstly, a prior distribution is selected, either discrete or continuous, which then constitutes the D-criterion as,

$$\mathrm{E}_\beta[\det(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta}))].$$

Problems with the Bayesian design, is that it rarely comes in closed form. To find the design, computer algorithms are almost always necessary. Much like with inference problems, the Bayesian approach trust on a good prior distribution for the parameters. This is however, chosen arbitrarily by the researcher, whom might not be an expert in the subject.

Sequential design is discussed in Fackle Fornius (2008) PhD thesis, which is a stepwise approach to handle the problem. Firstly, create an optimal design with it's points and respective weights. Then estimate the parameters in the model. This first design and estimated parameters are then used to generate the second design. This is sequentially done, to a curtain stopping criteria is fulfilled. The idea is that the previous estimation is assumed true.

This method can use both a parametric and non-parametric approach, as described in Fackle Fornius (2008) article, but wont be discussed further here. This method is effective, however, it can sometimes be unpractical to use.

e)
Analogous of the D-optimal design in b), the E-optimal design can be attain by only changing the optimal criterion. The E-criterion is,

$$E - criterion = \lambda_{max}(\boldsymbol{M}^{-1}(\xi, \boldsymbol{\beta})).$$

Table 3: E-optimal design 2e). Result from using the hand-written algorithm from lecture 2, and the same result from the *locally* built in function in R. Clearly both algorithms conclude similar three values as the optimal design.

ξ^*	x	w	$x_{locally}$	$w_{locally}$
2	1.5	0.2614	1.5368	0.333
1	1.6657	0.4846	1.7472	0.333
3	1.8343	0.2540	1.8057	0.333

This criterion is to be minimized in the search for the optimal design, as the other criterion's described. However, due to matrix singularity problems when taking inverse of the standardized information matrix, the maximum criterion has been used. To use the maximum criterion, while searching the minimum, the criterion has been taken as negative,

$$E - criterion = -\lambda_{min}(\boldsymbol{M}(\xi, \boldsymbol{\beta})).$$

The algorithms was then iterated 100 times over the interval $x = \mathcal{X} : [1.5, 2]$, initial parameter values of $\beta_0 = 431.1$, $\beta_1 = -520.6$ and $\beta_2 = 156.4$, initial design points was randomly generated, optimized individually then simultaneous, and number of points, k, optimal at three.

Looking in Table 3, the results from both the hand-written algorithm and the *locally* function is displayed. Both algorithm find similar points for an optimal design. See R-code for algorithm.

3

a)
The three parameter E_{max} model is given by,

$$y_i = E_0 + \frac{E_{max} x_i}{ED_{50} + x_i} + \epsilon_i.$$

Using the D-criterion, equation (4), and the *locally* function over the given interval $x = \mathcal{X} : [0, 100]$, and initial parameter vector given as $\boldsymbol{\theta}^T = (E_0, E_{max}, ED_{50})^T = (22, 11.2, 70)^T$, for 100 iterations and number of points, k, optimal at three. The result from the *locally* optimal design is,

$$\xi^*_{E_{max}} = \begin{Bmatrix} 0 & 29.17 & 100 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{Bmatrix}.$$

As expected, the two endpoints are part of the optimal design, all with equal weight.

b 1)
According to Berger & Wong (2009), deriving the 3PL mathematically is very complex, however, here is an attempt at simplifying the calculations.

The model's parameter vector is $\boldsymbol{\beta} = \{a, b, c\}$, and the function is,

$$p(\theta) = c + \frac{1 - c}{1 + e^{-a(\theta - b)}}.$$

Firstly, we use $p(\theta)$ and finding the log of odds function,

$$\eta(\theta, \boldsymbol{\beta}) = \ln\left(\frac{p(\theta)}{1 - p(\theta)}\right) = \ln\left(\frac{c + \frac{1-c}{1+e^{-a(\theta-b)}}}{1 - \left(c + \frac{1-c}{1+e^{-a(\theta-b)}}\right)}\right).$$

Analogous that of Section 2, we calculate the standardized information matrix by first finding the *Jacobian* matrix,

$$\boldsymbol{J} = \frac{d\eta}{d\boldsymbol{\beta}} = \begin{bmatrix} \frac{\partial \eta}{\partial a} \\ \frac{\partial \eta}{\partial b} \\ \frac{\partial \eta}{\partial c} \end{bmatrix},$$

Table 4: D-optimal design 3b2). Result from using the hand-written algorithm from lecture 2, and the same result from the *locally* built in function in R. Clearly, both algorithms conclude similar three values as the optimal design.

ξ^*	x	w	$x_{locally}$	$w_{locally}$
3	-3	0.333	-3	0.333
1	-0.2696	0.333	-0.1002	0.333
2	3	0.333	3	0.333

then calculating the standardized information matrix similarly as in equation (7)

$$\boldsymbol{Z} = \boldsymbol{J}\boldsymbol{J}^T = \begin{bmatrix} -\left(\frac{-\theta+b}{ce^{-a(\theta-b)}+1}\right) \\ -\left(\frac{a}{ce^{-a(\theta-b)}+1}\right) \\ \frac{1}{e^{a(\theta-b)}+c} + \frac{1}{1-c} \end{bmatrix} \times \left[-\left(\frac{-\theta+b}{ce^{-a(\theta-b)}+1}\right), \quad -\left(\frac{a}{ce^{-a(\theta-b)}+1}\right), \quad \frac{1}{e^{a(\theta-b)}+c} + \frac{1}{1-c} \right],$$

$$\boldsymbol{M}(\xi, \boldsymbol{\beta}) = \sum_{i=1}^{n} w_i p_i(\theta)(1 - p_i(\theta))\boldsymbol{Z}_i.$$

b 2)
Finding the D-optimal design for the 3PL model is straight forward when the information matrix has been established. Using the same D-criterion as equation (4), over the given interval $x = \mathcal{X} : [-3, 3]$ and the parameter values of $a = 0.5, b = 1$ and $c = 0.05$ while initial design points was randomly generated, optimized individually then simultaneous, gave three expected results for k optimal at three.

In Table 4, the results from both the hand-written algorithm and the *locally* functions results are shown. Clearly, both method attain almost identical results, with the expected two values at the bounds of the design region.

References

Berger, M. & Wong, W. (2009), 'An introduction to optimal designs for social and biomedical research'.

Fackle Fornius, E. (2008), Optimal Design of Experiments for the Quadratic Logistic Model, PhD thesis, Stockholm University, Faculty of Social Sciences, Department of Statistics.

Masoudi, E., Holling, H. & Wong, W. K. (2019), *An R Package for Finding Optimal Designs for Nonlinear Models Using Imperialist Competitive Algorithm.* R package version 0.9.9.
URL: *https://github.com/ehsan66/ICAOD*

O'Brien, T. & Funk, G. (2003), 'A gentle introduction to optimal design for regression models', *The American Statistician* **57**, 265–267.