Statistics and Applications
Vol. 11, Nos. 1 & 2, 2013 (New Series), pp. 1-13

Some finer aspects of the de la Garza phenomenon: a study of exact designs in linear and quadratic regression models

Bikas K. Sinha[1], Nripes K. Mandal[2], Manisha Pal[3]
[1] *Retired Faculty, Indian Statistical Institute, Kolkata, India*
[2] *Calcutta University, Kolkata, India*
[3] *Calcutta University, Kolkata, India*

Abstract

The well-known de la Garza Phenomenon [de la Garza, 1954] relates to the information matrix of the parameters in a standard Gauss-Markov linear model involving a single covariate in polynomial regression. It works well in the framework of approximate or continuous designs. For discrete or exact designs, one has to be careful in extracting its full spirit. We propose to discuss some features of this highly fascinating area of research.

Key words Linear models, regression designs, continuous designs, exact designs, information matrix, equivalence, domination.

1 Introduction

In the context of optimality studies in regression designs, the importance of de la Garza phenomenon has been emphasized repeatedly in various forms and settings. The most general known application is in the context of polynomial regression models involving a single covariate. Vide Pukelsheim (1993, 2006). Less known are most recent studies in non-linear parameter settings. Vide Yang (2010). In between, Liski et al (2002), Pukelsheim (1993, 2006), Khuri et al (2006) and Yang and Stufken (2009) have dealt with this phenomenon and discussed its importance in the search for optimal designs. However, all these studies [as also the original article by de la Garza in 1954, at least in spirit] relate to what is known as approximate or, continuous design set-up. In this article our focus is exclusively on exact or discrete design set-up.

Consider the problem of selecting an experimental design to furnish information on the parameters in a linear model $y = f'(x)\beta + e$, where y is the observable response variable, β is a $p \times 1$ vector of unknown parameters; the errors are assumed to be independent and identically distributed with mean 0 and constant variance σ^2, and f is a $p \times 1$ vector-valued known continuous function of the $q \times 1$ vector of design variables x that is constrained to lie in a compact subset $\chi(x)$ of R^q. An experimental design can be described as a probability mass function ξ that places total mass on a finite collection of s points in the design region $\chi(x)$. In exact [or, small sample i.e., discrete] design, $n(x_i)$ is required to be an integer for $i = 1, ..., s$, where n is the total sample size and $n(x_i)$ is the number of observations attributed to the design point $x_i; i = 1, 2, ..., s; \sum_i n(x_i) = N$. In a continuous [or, approximate] design set-up, this integer restriction is not imposed and $n(x_i)/N$ is replaced by the mass function ξ_i, for each $i = 1, 2, ..., s$. Virtually all optimality criteria characterize the worth of a design through a concave functional $\phi(M)$ that depends only on the $p \times p$ information matrix $M(\xi) = \sum_i \xi_i f(x_i) f'(x_i)$. In an exact design set-up defined by $[(x_i, n_i); i = 1, 2, ..., s; \sum_i n_i = N]$, for a given N, we may write M as $M = \sum_i n_i f(x_i) f'(x_i)$, for brevity, with the obvious interpretation that it refers to the total information content of the design based on N observations.

In Section 2, we present de la Garza phenomenon, as is known in the literature with reference to continuous designs. We focus only on linear regression and quadratic regression models. We also discuss related issues as are relevant for optimality studies. Next, in Section 3, we take up the models again and examine the validity of de la Garza phenomenon in the context of exact designs for given N. In Section 4, we examine Loewner Order Domination property of information matrices for discrete or exact designs, and that too, for linear and quadratic regression only. Finally, in Section 5, we make some concluding observations.

2 de la Garza Phenomenon

Consider a polynomial regression model of degree d in a single non-stochastic quantitative regressor x assuming values in a finite non-degenerate interval which is taken, without any loss of generality, to be the interval $[-1, 1]$. That is,

$$\Xi = [-1, 1], -1 \leq x \leq 1 \tag{1}$$

where at the value $x = x_i$, we have a set of uncorrelated responses $[y_{ij}, j = 1, 2, , n_i]$, each with the model expectation

$$E(y_{ij}) = f'(x_i)\beta \tag{2}$$

where β represents the $(d+1)\times 1$ vector of unknown parameters and $f'(x) = (1, x, x^2, ..., x^d)$ for each x in Ξ. The associated errors in the model are assumed to be independently and identically distributed with mean 0 and constant variance σ^2.

For estimating the model parameters, the optimum experimental design is obtained so as to maximize some function of the information matrix. Typically, the optimal designs are studied under approximate theory. In this context, de la Garza (1954) established that for every s-point approximate design, $s > d + 1$, given by

$$D_s = [(x_i, \xi_i), i = 1, 2, ..., s; \sum_i \xi_i = 1] \tag{3}$$

there exists a $(d + 1)$-point approximate design D^*_{d+1}

$$D^*_{d+1} = [(x^*_i, p^*_i), i = 1, 2, .., d + 1; \sum_i p^*_i = 1]. \tag{4}$$

which is based precisely on $d+1$ distinct support points with $x_{min} \leq x^*_i \leq x_{max}$, such that $M(D_s) = M(D^*_{d+1})$, where $M(D_s) = \sum_i \xi_i f(x_i) f'(x_i)$ and $M(D^*_{d+1}) = \sum_i p^*_i f(x^*_i) f'(x^*_i)$ represent the moment/information matrices for D_s and D^*_{d+1} respectively. Thus de la Garza Phenomenon applies to dth degree polynomial regression model in terms of Information Equivalence of any $s[> d + 1]$-point supported continuous design with that of a suitably chosen exactly $(d + 1)$-point supported continuous design! Of course, the nature of support points x^* in D^*_{d+1} and the choice of the revised weight vector $\mathbf{p}^* = (p^*_1, p^*_2, ..., p^*_{d+1})$

are both governed by the description of the original design D_s we start with. Moreover, x^*-values are covered by the range of the original x-values.

Remark 2.1 In the context of a very general form of the model, an important result due to Caratheodory (Caratheodory's theorem) provides useful guidance.[cf. Silvey (1980).] According to Caratheodory's theorem, any moment matrix based on an arbitrary continuous design with k distinct support points can be realized by an alternative design with at the most $p(p+1)/2+1$ distinct support points where p is the number of parameters of the model, whenever $k > p(p+1)/2+1$. However, it may be noted that such designs may not be easy to actually construct. Anyway, for the polynomial model in one independent variable, de la Garza provides much better result since it achieves the Information Equivalence with fewer number of distinct design points. Afterwards, several workers contributed in this area and extended it in different directions (see e.g., Yang (2010) for all related references).

As it transpires, this equivalence of information matrices relates to continuous design set-up. It is not clear, how far it holds in the exact design case, especially when the sample size is small. For large N, a continuous design can well be approximated by an exact design. In this paper, as mentioned earlier, we are concerned with the study of the true nature of this equivalence in polynomial regression set-up and in the exact sense for small values of N and provide a lower bound to the value of N for the information equivalence to hold. This should serve as a guiding rule in practice when we deal with small overall sample sizes N.

3 Exact Design : Analysis of Information Equivalence

3.1 Case of Linear Regression

Consider a linear regression model:

$$y = \alpha + \beta x + e \tag{5}$$

with usual assumption on error. The problem is to estimate the unknown fixed parameters α and β by a suitably chosen exact or discrete design. To examine the full spirit of the de la Garza phenomenon in the discrete case, let us start with a 3-point design $D(1) : [-1, 0, 1]$ so that $s = 3 > 2$. According to de la Garza Phenomenon, under continuous design theory, whatever the weights $0 < w_{-1}, w_0, w_{+1} < 1; w_{-1} + w_0 + w_{+1} = 1$ assigned to the three points $-1, 0, +1$, we can find one 2-point design, say $[(a, p); (b, 1-p)]$ such that $-1 \leq a < b \leq 1, 0 < p < 1$ and we have 'Information Equivalence' between the two designs. Let us consider its discrete version below. Given a point-symmetric 3-point design with a given total number of observations N and its decomposition into n_-, n_0, n_+ - being assigned to $-1, 0, 1$ respectively, can we now find a 2-point design i.e., $[(a, n_a); (b, n_b)]$ satisfying

$$(i) - 1 \leq a < b \leq 1, (ii) n_a + n_b = N$$

where n_a and n_b both assume integer values, and attaining Information Equivalence? The solution to this problem depends on the nature of the original design and also on the value of N. In the above, the design points are already specified [viz., $-1, 0, 1$] and so it depends on the total number N of design points and also its decomposition into three parts, viz., n_-, n_0, n_+ we start with. We set

$$D_1 : [(-1, n_-); (0, n_0); (1; n_+); n_- + n_0 + n_+ = N], \tag{6}$$

$$D_2 : [(a, n_a); (b, n_b); -1 \leq a < b \leq 1; n_a + n_b = N]. \tag{7}$$

This suggests, for information equivalence,

$$an_a + bn_b = n_+ - n_- \tag{8a}$$

$$a^2 n_a + b^2 n_b = n_+ + n_-. \tag{8b}$$

Set

$$u = n_a, v = n_b, T_1 = n_+ - n_-, T_2 = n_+ + n_-. \tag{9}$$

From (8a) and (8b), in terms of (9), we obtain

$$a = [T_1/(u+v)] + / - \sqrt{v[(u+v)T_2 - T_1^2]/u(u+v)^2}, \tag{10a}$$

$$b = [T_1/(u+v)] + / - \sqrt{u[(u+v)T_2 - T_1^2]/v(u+v)^2}. \tag{10b}$$

It can be readily verified that $(u+v)T_2 > T_1^2$ so that the above solutions to a and b are real. Let us choose

$$a = [T_1/(u+v)] + \sqrt{v[(u+v)T_2 - T_1^2]/u(u+v)^2} \tag{11a}$$

$$b = [T_1/(u+v)] - \sqrt{u[(u+v)T_2 - T_1^2]/v(u+v)^2}. \tag{11b}$$

so that $b < a$. Note that T_1 and T_2 are both known. We will now sort out values of u and v subject to $u + v = N$ so as to satisfy the requirement that

$$[-1 \leq b < a \leq 1]. \tag{12}$$

It is to be noted that a and b given by (11a) and (11b) depend on u and v only through u/v or v/u. It can be shown, after a little algebra, that the requirement (12) is realized if and only if

$$\begin{aligned}&[P_0(1-P_0) + 4(P_+(P_-)]/[2(P_-) + P_0]^2 \leq u/v \\ &\leq [2(P_+) + P_0]^2/[P_0(1-P_0) + 4(P_+)(P_-)]\end{aligned} \tag{13}$$

holds, where

$$n_-/N = P_-, n_0/N = P_0, n_+/N = P_+. \tag{14}$$

Let us write

$$L = [P_0(1-P_0) + 4(P_+)(P_-)]/[P_0(1-P_0) + 4(P_+)(P_-) + [2(P_-) + P_0]^2], \tag{15a}$$

$$U = [2(P_+) + P_0]^2/[P_0(1 - P_0) + 4(P_+)(P_-) + [2(P_+) + P_0]^2]. \tag{15b}$$

Then (13) can be equivalently expressed as

$$NL \le n_a \le NU. \tag{16}$$

Thus for a given set of frequencies n_-, n_0, n_+ with $n_- + n_0 + n_+ = N$, it is possible to find a two-point exact design achieving the same information matrix as that of the original three-point design, provided an integer n_a satisfying (16) exists. This shows that, for a given N, it is not always possible to have the information equivalence by a two-point design unless (16) includes at least one integer. A sufficient condition for this to happen is, of course, that the length of the interval viz. $N(U - L) \ge 1$. Even otherwise, a choice of n_a could be ensured.

Below we examine different specific situations involving the original 3-point design.

(1) $P_0 = P_+ = P_- = 1/3$. This design is point and mass symmetric. Here we find $L = 2/5$ and $U = 3/5$. So, for N = 3, $NL = 6/5$ and $NU = 9/5$, which do not include any integer. So 3-point design with point and mass symmetry cannot be replaced by a 2-point design whenever $N = 3$.

However, for $N = 6$, we have $NL = 12/5, NU = 18/5$ and these include the integer 3. So there is a solution with support points at $-(2/3), (2/3)$, each with 3 observations, as was mentioned before. For $N = 9$, we have $NL = 18/5$ and $NU = 27/5$. These include 2 integers viz. 4 and 5. So we have two solutions : $[(-5/30, 4); (4/30, 5)]$ and $[(-4/30, 5); (5/30, 4)]$.

(2) $P_0 = 2/7, P_+ = 4/7, P_- = 1/7$ i.e., the initial exact design has size which is a multiple of 7, say $N = 7k$. This design is point-symmetric but mass-asymmetric and explicitly it is : $[(-1, k); (0, 2k); (+1, 4k)]$ where k is an integer. From (15a) - (15b), we observe that both L and U are independent of k and are given by :$L = 13/21, U = 50/63$. Let us consider different values of k and identify corresponding solutions for the 2-point designs:

$$(a) \quad k = 1 : N = 7; NL = 13/3 < NU = 50/9$$

so that we have one solution viz.

$$n_a = 5, a = 3/7 + \sqrt{(1040)}/70;$$

$$n_b = 2, b = 3/7 - 5\sqrt{(1040)}/140$$

and the corresponding design is

$$[(3/7 - 5\sqrt{(1040)}/140, 2); (3/7 + \sqrt{(1040)}/70, 5)].$$

$$(b) \quad k = 2 : N = 14; NL = 26/3 < NU = 100/9.$$

Here we have 3 solutions given by :

$$n_a = 9, a = 3/7 + \sqrt{(520)}/42; n_b = 5, b = 3/7 - 3\sqrt{(2080)}/140;$$

$$n_a = 10, a = 3/7 + \sqrt{(1040)}/70; n_b = 4, b = 3/7 - \sqrt{(260)}/14;$$
$$n_a = 11, a = 3/7 + \sqrt{(3432)}/154; n_b = 3, b = 3/7 - \sqrt{(3432)}/42.$$

The corresponding designs are given by

$$[(3/7 - 3\sqrt{(2080)}/140, 5); (3/7 + \sqrt{(520)}/42, 9)];$$
$$[(3/7 - \sqrt{(260)}/14, 4); (3/7 + \sqrt{(1040)}/70, 10)];$$

$$[(3/7 - \sqrt{(3432)}/42, 3); (3/7 + \sqrt{(3432)}/154, 11)].$$

(3) $P_0 = 3/5, P_+ = P_- = 1/5$. The corresponding design is again point and mass-symmetric and the initial exact design has size multiple of 5, say $N = 5k$. Explicitly it is : $[(-1, k); (0, 3k); (1, k)]$ where k is an integer. From (15a) and (15b) we have $L = 2/7, U = 5/7$ which are independent of k. Then for $k = 1$, we get $N = 5, 10/7 \leq n_a \leq 25/7$ so that we have the following solutions:

$$(i)\ a = 3/\sqrt{(15)}, b = -2/\sqrt{(15)}; (n_a, n_b) = (2, 3)$$
$$(ii)\ a = 2/\sqrt{(15)}, b = -3/\sqrt{(15)}; (n_a, n_b) = (3, 2).$$

Again, for $k = 2, N = 10, 20/7 \leq n_a \leq 50/7$ so that the admissible values of n_a are $n_a = 3, 4, 5, 6, 7$. Thus we have the following solutions:

$$(i)\ a = 6/\sqrt{(210)}, b = -14/\sqrt{(210)}; (n_a, n_b) = (7, 3);$$
$$(ii)\ a = 14/\sqrt{(210)}, b = -6/\sqrt{(210)}; (n_a, n_b) = (3, 7);$$
$$(iii)\ a = 4/\sqrt{(60)}, b = -6/\sqrt{(60)}; (n_a, n_b) = (6, 4);$$
$$(iv)\ a = 6/\sqrt{(60)}, b = -4/\sqrt{(60)}; (n_a, n_b) = (4, 6);$$
$$(v)\ a = 2/\sqrt{(10)}, b = -2/\sqrt{(10)}; (n_a, n_b) = (5, 5).$$

We have a few more examples in the point symmetric case which can be proved in a similar way:

$$(i)\ D[(-1, 1); (0, 2); (1, 1)] = D[(-1/(2), 2); ((1/(2), 2)]$$
$$(ii)\ D[(-1, 4); (0, 2); (1, 2)] = D[(-1/4 - \sqrt{(165)}/20; 5); (-1/4 + \sqrt{(165)}/12, 3].$$

However, for $[(-1, 2); (0, 1); (1, 1)]$, it is not possible to find a two-point exact design satisfying the sufficient condition $N(U - L) \geq 1$.

Remark 3.1.. Since the condition $N(U - L) \geq 1$ is only sufficient, there is scope for having an exact design which attains the information equivalence but does not satisfy the condition. We may expand further on this matter. Writing $P_- = x, P_0 = y, P_+ = z$ so that $x + y + z = 1$, we see that

$$U = \frac{(y + 2z)^2}{y + 4z}; L = \frac{(y + 4z) - (y + 2z)^2}{y + 4x}.$$

The expression for $U - L$ can be simplified to

$$U - L = \frac{y[2y(1-y) + y + 8xz]}{(y + 4z)(y + 4x)}.$$

In the particular case of $x = z = (1-y)/2, U - L$ simplifies to $y/(2-y)$. Therefore, $N(U - L) < 1$ iff $y < 2/(N+1)$, i.e., $Ny < 2N/(N+1)$. This forces $Ny = 1$ and hence N to be necessarily an odd integer, say $N = 2k+1$. In that case, we deduce $NU = k + \frac{3k+1}{4k+1}$ and $NL = k + \frac{k}{4k+1}$. Hence, there is no integer in this range whenever $N(U - L) < 1$, in case $P_- = P_+$.

We now pass on to point asymmetric designs with three distinct points and examine possibility of information equivalence.

Let us consider an example of a 3-point asymmetric design with $N = 3$: $[(-1, 1); (a, 1); (1, 1)]$ with $a \neq 0$ and without loss of generality, we take $a > 0$. Let us examine whether it is possible to have information equivalence of this design by a two-point design. Consider an arbitrary two-point design $[(u, 2); (v, 1)]$ with $-1 \leq v < u \leq 1$. Equating the elements of the information matrices of the two designs we get :

$$a = 2u + v, 2 + a^2 = 2u^2 + v^2.$$

This yields : $u = a/3 - (2/3)\sqrt{(a^2+3)}$, or, $u = a/3 + (2/3)\sqrt{(a^2+3)}$ and for $0 < a < 1$, it turns out that $a/3 - (2/3)\sqrt{(a^2+3)} < -1$ and $1 < a/3 + (2/3)\sqrt{(a^2+3)}$. Hence, in this case, we fail to attain information equivalence by a two-point discrete design ! For $N = 6$, naturally, equal allocation of 2 at each of the 3 points: $[-1, a, 1](a \neq 0)$ will yield the same negative result when we opt for $[(u, 4); (v, 2)]$. It follows that $[(u, 5); (v, 1)]$ also fails to yield any affirmative result. We next try out equal allocation two-point design $[(u, 3); (v, 3)]$. For identical information matrices of the two designs we require $2a = 3(u + v)$ and $4 + 2a^2 = 3(u^2 + v^2)$ which yield $a/3 - (1/3)\sqrt{(6 + 2a^2)} = u < v = a/3 + (1/3)\sqrt{(6 + 2a^2)}$.

Note that for $a = 0$, we have the point symmetric design and then the solution is : $-\sqrt{(2/3)} = u < v = \sqrt{(2/3)}$ which has already been discussed. If $a > 0$, the condition $-1 < u < 1$ leads to $0 < a < 2\sqrt{3} - 3$.

It thus transpires that not all values of N are amenable to supporting the equivalence theorem of the information matrix. We need a minimum value and only then it works!

3.2 Case of Quadratic Regression

Consider a symmetric n-point design D_n : $[-1 \leq x_1 < x_2 < ... < x_n \leq 1$, each x-value used only once so that $N = n$. Can we replace D_n by a suitably defined and Information-Equivalent 3-point design, say D_3^* : $[(x, f); (y, n-2f), (z, f)]$, where $-1 \leq x < y < z \leq 1$? Towards the solution, we have the following :

Theorem 3.1: *In a quadratic regression model with the usual model assumptions, for every exact symmetric regression design* D_{2k} *having support points* $[-1 \leq -a_k < -a_{k-1} < , < -a_2 < -a_1 < 0 < a_1 < a_2 < \; < a_{k-1} < a_k \leq 1]$*, each occurring* *once, with $k \geq 2$, and $N = 2k$, there exists an exact symmetric 3-point regression design $D_3^* = [(-x*, f); (0, N-2f); (x*, f)]$, satisfying $I(\theta; D_{2k}) = I(\theta; D_3^*)$, where $x* = \sqrt{\frac{\sum_i a_i^4}{\sum_i a_i^2}}, 0 < x* < a_k$ and $f = \frac{[\sum_i a_i^2]^2}{\sum_i a_i^4}$ provided $[\sum_i a_i^2]^2$ is an integer multiple of $\sum_i a_i^4$.*

Proof: Consider a 3-point exact mass-symmetric regression design $D_3 = [(x, f); (y, N-2f); (z, f)]$ where $-1 \leq x < y|z \leq 1$. Equating the first and the third moments of the designs D_{2k} and D_3, we derive

$$f(x+z) + (N-2f)y = 0, f(x^3+z^3) + (N-2f)y^3 = 0.$$

After some algebraic manipulation, this leads to

$$(N-2f)y[y^2 \frac{(N-2f)^2}{f^2} - 1 - 3xz] = 0.$$

Now, to have a 3-point design, $(N-2f)$ must be greater than zero, i.e. we must have $f < k$. Hence, either $y = 0$ or $3xz = 1 + y^2 \frac{(N-2f)^2}{f^2}$. However, the latter implies $(x-z)^2 = -N(N-2f)y^2/f^2 < 0$, which is impossible. Hence, for the above to hold, we must have that $y = 0$. For $y = 0$, it is easy to check that $x = -z$. Now, equating the second and fourth moments of the two designs, we derive

$$z = x* = \sqrt{\frac{\sum_i a_i^4}{\sum_i a_i^2}}; f = \frac{[\sum_i a_i^2]^2}{\sum_i a_i^4}.$$

Clearly, $0 < x* < a_k$. Further, for f to be an integer we must have that $[\sum_i a_i^2]^2$ is an integer multiple of $\sum_i a_i^4$. The design, so arrived at, is now identified as D_3^* as described in the statement of the Theorem 3.1.

A generalization to the case where the design D_{2k} is point symmetric, but not necessarily mass symmetric, results in the following theorem.

Theorem 3.2: *In a quadratic regression model with the usual model assumptions, for every discrete, point-and mass-symmetric regression design*

$$D_{2k} = [-1 \leq a_k < -a_{k-1} < ... < -a_2 < -a_1 < 0;$$

$$0 < a_1 < a_2 < ... < a_{k-1} < a_k;$$

$$n_k, n_{k-1}, ..., n_2, n_1, n_1, n_2, ..., n_{k-1}, n_k],$$

with $k \geq 2$, there exists an exact point and mass symmetric 3-point regression design

$$D_3^* = [(-x*, f*); (0, N-2f*); (x*, f*)],$$

satisfying $I(\theta; D_{2k}) = I(\theta; D_3^*)$, *where* $x* = \sqrt{\frac{\sum_i n_i a_i^4}{\sum_i n_i a_i^2}}, 0 < x* < a_k$, *provided* $f* = \frac{(\sum_i n_i a_i^2)^2}{\sum_i n_i a_i^4}$ *is an integer.*

Proof: Consider a 3-point exact regression design $D_3 = [(x, f_x); (y, f_y); (z, f_z)]$ satisfying $-1 \le x < y < z \le 1; f_x + f_y + f_z = N = \sum n_i$.

In order that the designs D_3 and D_{2k} be information equivalent we must have

$$f_x + f_y + f_z = N$$

$$xf_x + yf_y + zf_z = 0$$

$$x^2 f_x + y^2 f_y + z^2 f_z = 2\sum_i n_i a_i^2$$

$$x^3 f_x + y^3 f_y + z^3 f_z = 0$$

$$x^4 f_x + y^4 f_y + z^4 f_z = 2\sum_i n_i a_i^4.$$

Putting $y = 0$ and solving the above equations, we get $x = -z, f_x = f_z$ and

$$z = x* = \sqrt{\frac{\sum_i n_i a_i^4}{\sum_i n_i a_i^2}}; f_z = f* = \frac{(\sum_i n_i a_i^2)^2}{\sum_i n_i a_i^4}.$$

Clearly, $0 < x* < a_k$. Hence the theorem.

Remark 3.2: For a $(2k+1)$-point discrete, symmetric design of the form

$$D(2k+1) = [-1 \le -a_k < -a_{k-1} < ... < -a_2 < -a_1 < 0;$$

$$0 < a_1 < a_2 < ... < a_{k-1} < a_k;$$

$$n_k, n_{k-1}, ..., n_2, n_1, n_0, n_1, n_2, ..., n_{k-1}, n_k]$$

an information equivalent 3-point discrete, symmetric regression design will be same as that obtained in Theorem 3.2, except that the mass at zero will increase by n_0.

Examples.

(i) $D_6 : [(-1,1); (-2/3,1); (-1/3,1); (1/3,1); (2/3,1); (1,1)]$ is equivalent to the 3-point symmetric design $D_3^* : [(\sqrt{(7)}/3, 2); (0, 2); (\sqrt{(7)}/3; 2)]$,

(ii) $D_4 : [(-1,1); (-1/2,2); (1/2,2); (1,1)]$ is equivalent to the 3-point symmetric design given by $D_3^* : [(-\sqrt{3}/2, 2); (0, 2); (\sqrt{(3)}/2; 2)]$.

4 Exact Design: Loewner Order Dominance

In Section 3, we studied the existence of information-equivalent exact designs in the light of the de la Garza phenomenon. It is noted that, though de la Garza phenomenon holds in the continuous design set-up, it is not generally true in the discrete or exact design set-up. A similar question also arises involving domination of one information matrix by another in Loewner Order sense. Thus, in a linear regression set-up, we may ask : Given an exact design with three or more support points inside the closed interval $[-1, 1]$, is it possible to have a 2-point exact design whose information matrix dominates that of the given design in the Loewner Order Domination [LOD] sense? A similar question may also be asked in the context of quadratic regression model. Note that LOD refers to the difference of the two matrices being an nnd matrix. Once LOD holds, the dominating design outperforms the dominated design in the best sense and in respect of any sensible criterion of bestness. However, it is generally hard to achieve, if not impossible. We examine the domination aspect in this section.

4.1 Linear Regression

Let us first consider some specific cases involving the 3-point designs considered in Section 3.

(i) $D_3 : [(-1, n_-); (0, n_0); (1, n_+)]$, where n_0 is even. Let us define a 2-point design $D_2^* = [(-1, N_-); (1, N_+)$, where $N_+ = n_+ + n_0/2, N_- = n_- + n_0/2$. It is easy to see that D_2^* dominates D_3 in the Loewner Order sense. The domination holds even if we replace the support points $+1, -1$ in D_3 by $a, -a$, where $a < 1$. However, when n_0 is odd, it is difficult to find a 2-point design dominating D_3.

(ii) In the point and mass symmetric case with $[(-1, 2); (0, 2); (1, 2)]$, the information-equivalent 2-point symmetric design replicates each of the points $-\sqrt{2}/3$, $+\sqrt{2}/3$ thrice. We can have Loewner Order Domination by stretching these two points further to $-1, +1$, each replicated thrice. This also holds true for other point and mass symmetric cases. Now, let us consider any arbitrary n-point point- and mass-symmetric design D_n with total frequency N (even). The first two moments of the design are $\mu_1^{'} = 0; \mu_2^{'} > 0$. In order that a two point symmetric design $D_2^* = [(-b, N/2); (b, N/2]$ dominates D_n in the Loewner Order sense, we require that the second order moment of D_2^* is greater than $\mu_2^{'}$, which gives $\mu_2^{'} < b^2$. Also, we must have $b^2 < 1$. This result can be formalized in the following theorem:

Theorem 4.1: *Given any arbitrary n-point point- and mass-symmetric design D_n defined in the closed interval $[-1, 1]$, with total frequency N (even) observations and second order moment $\mu_2^{'}$, there exists a 2-point symmetric design $D_2^* = [(-b, N/2); (b, N/2)]$ that dominates D_n in the Loewner Order sense whenever $\mu_2^{'} < b^2 < 1$. However for N odd, it is not possible to find a discrete 2-point symmetric design having Loewner Order domination over D_n.*

Examples

(i) Consider $D_3 = [(-1,1);(0,2);(1,1)]$. Here $N = 4$ and $\mu_2' = 1/2$. Then, in order that the design $D_2^* = [(-b,2);(b,2)]$ dominates D_3 in the Loewner Order sense we must have $1/2 < b^2 \leq 1$.

(ii) Consider $D_4 = [(-0.75,1);(-0.25,2);(0.25,2);(0.75,1)]$. Here $N = 6$ and $\mu_2' = 11/48$. Then, the design $D_2^* = [(-b,3);(b,3)]$ dominates D_4 in the Loewner Order sense whenever $11/48 < b^2 \leq 1$.

4.2 Quadratic Regression

In the case of quadratic regression, we wish to examine whether there exists a 3-point design that dominates a given n-point design ($n > 3$) in the Loewner Order Domination sense. Let us consider an n-point point- and mass-symmetric design D_n, with total frequency N and second and fourth order moments given by μ_2' and μ_4', respectively. Let us further consider a 3-point symmetric design $D_3^* = [(-b,f);(0,N-2f):(b,f)]$ with $b < 1$. The second and fourth order moments of D_3^* are $\mu_2^* = 2b^2f/N, \mu_4^* = 2b^4f/N$. In order that D_3^* dominates D_n in the Loewner Order sense, we must have [Vide Pukelsheim (1993, 2006)] $\mu_2^* = \mu_2$ and $\mu_4^* > \mu_4$, which reduces to

$$(i) f = N\mu_2'/2b^2,$$

which should be an integer,

$$(ii) \mu_4'/\mu_2' < b^2 < 1.$$

Further, we must have $f < N/2$, which is satisfied for $b^2 > \mu_4'/\mu_2'$. From (i) and (ii), we have that $N\mu_2'/2 \leq f < N(\mu_2')^2/2\mu_4')$ which is possible only when the stated interval contains at least one integer which gives f, and, then $b = \sqrt{\frac{N\mu_2'}{2f}}$.

The above may be summarized as follows:

Theorem 4.2: *Given an arbitrary n-point point- and mass-symmetric design D_n, $n > 3$, with total frequency N and second and fourth moments given by μ_2' and μ_4' respectively, a sufficient condition for a 3-point symmetric design $D_3^* = [(-b,f);(0,N-2f):(b,f)]$ to dominate D_n in the Loewner Order sense is that the interval $[N\mu_2'/2, N(\mu_2')^2/2\mu_4'))$ contains at least one integer.*

As an example, let us consider a 4-point symmetric design
$D_4 = [(-a,1);(-c,1);(c,1);(a,1)]$ and a 3-point design
$D_3^* = [(-b,1);(0,2);(b,1)]$ to dominate D_4. Here we have $f = 1$. Hence, for D_3^* to dominate D_4 in Loewner Order sense, we must have that the interval $[(a^2+c^2),(a^2+c^2)^2/(a^4+c^4))$ contains the integer 1, which is true provided $a^2+c^2 < 1$. Thus, if $a = 0.4$ and $c = 0.3$, so that $a^2+c^2 = 0.25$, the design $[(-0.5,1);(0,2);(0.51)]$ dominates D_4.

Now suppose $D_4 = [(-a,1),(-c,2);(c,2);(a,1)], 0 < c < a \leq 1$ and $D_3^* = [(-b,f);(0,6-2f);(b,f)]$. Then, by the theorem, D_3^* dominates D_4 if the interval $[(a^2+2c^2),(a^2+2c^2)^2/(a^4+2c^4))$ contains an integer.

For $a = 1, c = 0.5$, the stated interval is $[1.5, 2)$, which does not contain an integer. Hence, there does not exist any 3-point design that dominates the design $[(-1,1);(-0.5,2);(0.5,1);(1,1)$ in the Loewner Order sense. However, in this case, by taking $f = 2$ and $b = \sqrt{(3/4)}$, we get an information-equivalent 3-point design.

If $a = 0.5, c = 0.25$, then the stated interval is $[3/8, 2)$, which contains the integer 1. Hence, $f = 1$ and $b = \sqrt{(3/8)}$. Thus, the 3-point symmetric design $[(-\sqrt{(3/8)}, 1); (0, 4); (\sqrt{(3/8)}, 1)]$ possesses Loewner Order domination over the 4-point symmetric design $[(-0.5,1);(-0.25,2);0.25,2);(0.5,1)]$. For $a = 1/25, c = 1/32$, the stated interval is $[0.0036, 2.8260)$. This time we have two choices : $f = 1, b = 0.0596$ and $f = 2, b = 0.0421$. Thus, both the designs $[(-0.0596,1);(0,4);(0.0596,1)]$ and $[(-0.0421,2),(0,2);(0.0421;2)]$ dominate the design $[(-1/25,1);(-1/32,2);(1/32,2);(1/25,1)]$ in Loewner Order dominantion sense.

5 Concluding Observations

Some of the relevant references on characterization of optimal continuous designs for inference on parameters of a very popular model viz., logistic linear regression model [LLRM] include Abdelbasit and Plackett (1983), Minkin (1987), Khan and Yazdi (1988), Wu (1988), Ford et. al. (1992), Sitter and Wu (1993), Hedayat et.al. (1997), Mathew and Sinha (2001), and Stufken and Yang (2009). In particular, Mathew and Sinha (2001) provided a unified approach for the derivation of D- and A-optimal designs for the model parameters and functions thereof. However, for such non-linear models, exact design theory is not applicable. With reference to exact design framework, the so-called optimal design for estimation of 'ED50' in LLRM turns out to provide biased and inconsistent results. Vide Guiard and Pichlmeier (1999). Again, with reference to logistic quadratic regression model [LQRM], a study by Fornius and Nyquist (2010) hinted on a 1-point optimal design. Follow-up studies by Nandy and Nandy (2013a, 2011, 2013b) suggested that the 1-point design, in the exact sense, produces biased and inconsistent estimator of the parameter under question. The latter authors suggested a 3-point design to produce asymptotically unbiased and consistent estimator of the model parameter with high efficiency. The study of exact designs in such non-linear model settings seems to be difficult. At this stage, we need further research in the context of polynomial regression models, involving exact designs, to check the validity of de la Garza Phenomenon.

References

Abdelbasit, K. M. and Plackett, R. L. (1983). Experimental design for binary data. *J. Amer. Statist. Assoc.*, **78**, 90-98.

de la Garza, A. (1954). Spacing of information in polynomial regression. *Ann. Math. Statist.* **25**, 123-130.

Ford, I., Torsney, B. and Wu, C.F.J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Royal Statist.*

Soc. Series B. **54**, 569-583.

Fornius, E. F. and Nyquist, H. (2009). Using the Canonical Design Space to Obtain c-optimal Designs for the Quadratic Logistic Model. *Comm. Statist. Th. and Meth.*, **39**, 144-157.

Guiard, V. and Pichlmeier,U. (1999). Critical assessment of the c-optimality design criteria for estimating the median effective dose in quantal dose-response curves. *Biometrical J.* **41**, 799-816.

Hedayat, A. S., Yan, B. and Pezzuto, J. M. (1997). Modeling and identifying optimum designs for fitting dose-response curves based on raw optical density data. *J. Amer. Statist. Assoc.* **92**, 1132-1140.

Khan, M. K. and Yazdi, A. A. (1988). On D-optimal designs for binary data. *J. Statist. Plann. Inf.* **18**, 83-91.

Khuri, A. I., Mukherjee, B., Sinha, B. K. and Ghosh, M. (2006). Design issues for generalized linear models: A review. *Statist. Sci.* **21**, 376-399.

Liski, E., Mandal, N. K., Shah, K. R. and Sinha, Bikas K.(2002). *Topics in optimal design.* Springer-Verlag Lecture Notes in Statistics Series. **163**.

Mathew, T. and Sinha, B. K. (2001). Optimal designs for binary data under logistic regression. *J. Statist. Plann. Inf.* **93**, 295-307.

Minkin, S. (1987). Optimal design for binary data. *J. Amer. Statist. Assoc.* **82**, 1098-1103.

Nandy, K. and Nandy, R. (2013a). A study of the performance of two-stage optimal designs in a logistic dose-response model. *Private communication.*

Nandy, K, and Nandy, R. (2011, 2013b). Some aspects of optimal designs for quadratic dose-response models. Presented at Statistical Meeting in Halifax, Canada. *Private Communication.*

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York; Also (2006). SIAM, Philadelphia, PA.

Silvey, S. D. (1980). *Optimal Design.* Chapman and Hall, London.

Sitter, R. R. and Wu, C.F.J. (1993). Optimal designs for binary response experiments: D- and A- criteria. *Scand. J. Statist.* **20**, 329-342.

Wu, C.F.J. (1988). Optimal design for percentile estimation of a quantal response curve. In: Dodge, Y., Federov, V. and Wynn, H.P. (Eds.), *Optimal Design and Analysis of Experiments.* Elsevier, Amsterdam, 213-223.

Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for non-linear models with two parameters. *Ann. Statist.* **37**, 518-541.

Yang, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38**, 2499-2524.