Optimal design for the bounded log-linear regression model

HaiYing Wang, Andrey Pepelyshev and Nancy Flournoy

Abstract Wang and Flournoy (2012) developed estimation procedures for the bounded log-linear regression model, an alternative to the four parameter logistic model which has a bounded response with non-homogeneous variance. In the present paper, we theoretically obtain that an optimal design that minimizes an information based criterion consists at most five design points including the two boundary points of the design space. The D-optimal design does not depend on the two parameters representing the boundaries of the response but it does depend on the variance of the error. Furthermore, if the error variance is known and bigger than a constant, we prove that the D-optimal design is the two-point design supported at boundary points with equal weights. Numerical examples are provided.

1 The statement of the problem

Consider the bounded log-linear regression model defined by

$$\log\left(\frac{B-Y}{Y-A}\right) = a+bx+\varepsilon, \text{ or equivalently, } Y = B - \frac{B-A}{1+e^{-(a+bx+\varepsilon)}}, \tag{1}$$

HaiYing Wang
Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65203 USA, e-mail: hwzq7@mail.missouri.edu

Andrey Pepelyshev
Institute of Statistics, RWTH Aachen University, Wullnerstr 3, 52056, Germany, e-mail: pepelyshev@stochastik.rwth-aachen.de

Nancy Flournoy
Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65203 USA, e-mail: flournoyn@missouri.edu

where $\varepsilon \sim N(0,\sigma^2)$, a, b, σ, A and B are unknown parameters, x is a non-random covariate, $x \in \mathbb{X}$ and Y is the response. Note that model (1) is closely related to the four parameter logistic (4PL) model. The inferential procedure for model (1) as well as its advantages over the 4PL model can be found in [3].

For estimating $\theta = (a,b,\sigma,A,B)$ by the local maximum likelihood method, the Fisher information matrix based on a single observation at x is

$$I(\theta,x) = \begin{bmatrix} I_{11} & 0 & I_{13} \\ 0 & I_{22} & I_{23} \\ I_{13}^T & I_{23}^T & I_{33} \end{bmatrix},$$

where $I_{11} = \frac{1}{\sigma^2}\begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix}$, $I_{13} = -\frac{1}{\sigma^2(B-A)}\begin{bmatrix} 1+\delta e^{c} & 1+\delta e^{-c} \\ (1+\delta e^{c})x & (1+\delta e^{-c})x \end{bmatrix}$, $I_{22} = \frac{2}{\sigma^2}$,
$I_{23} = \frac{2\delta}{\sigma(B-A)}\left[-e^{c},\ e^{-c}\right]$, $I_{33} = \begin{bmatrix} \frac{\delta^4 e^{2c}}{(B-A)^2} + \frac{1+2\delta e^{c}+\delta^4 e^{2c}}{\sigma^2(B-A)^2} & -\frac{1}{(B-A)^2} + \frac{2+\delta e^{c}+\delta e^{-c}}{\sigma^2(B-A)^2} \\ -\frac{1}{(B-A)^2} + \frac{2+\delta e^{c}+\delta e^{-c}}{\sigma^2(B-A)^2} & \frac{\delta^4 e^{2c}}{(B-A)^2} + \frac{1+2\delta e^{-c}+\delta^4 e^{-2c}}{\sigma^2(B-A)^2} \end{bmatrix}$,
$c = a+bx$ and $\delta = e^{\sigma^2/2}$; see [3] for derivation of the matrix $I(\theta,x)$.

Denote an approximate design by $\xi = \{x_i, w_i\}_1^K$, where $w_i > 0$ is the design weight at the point x_i and $\sum_{i=1}^K w_i = 1$. Under the design ξ, the average information matrix for θ is

$$M_\xi(\theta) = \sum_{i=1}^K w_i I(\theta, x_i). \tag{2}$$

We consider optimality criteria that minimize a statistically meaningful concave functional of this information matrix. In the rest of the paper we assume that the design space is defined such that $c = a+bx \in [l,u]$.

2 Main results

2.1 The case of unknown σ

In the following theorem we obtain an upper bound for the number of support points of optimal designs that improves on the classical upper bound based on Carathéorody's theorem [1, 2].

Theorem 1. *An optimal design that minimizes an information based criterion for model* (1) *is supported at no more than* 5 *design points. In addition, the optimal design is always supported at boundary points.*

Proof. Reformulating the design problem in terms of c rather than x, we rewrite the design as $\xi = \{c_i, w_i\}_1^K$. By matrix manipulation, the matrix $I(\theta,x)$ has the form $I(\theta,c) = P_\theta C_c P_\theta^T$, where

$$C_c = \begin{bmatrix} 1 & c & 0 & e^c & e^{-c} \\ c & c^2 & 0 & ce^c & ce^{-c} \\ 0 & 0 & \frac{1}{2\sigma^2} & e^c & -e^{-c} \\ e^c & ce^c & e^c & (\sigma^2+1)\delta^2 e^{2c} & \frac{1-\sigma^2}{\delta^2} \\ e^{-c} & ce^{-c} & -e^{-c} & \frac{1-\sigma^2}{\delta^2} & (\sigma^2+1)\delta^2 e^{-2c} \end{bmatrix}$$

and

$$P_\theta = \frac{1}{\sigma} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{-a}{b} & \frac{1}{b} & 0 & 0 & 0 \\ 0 & 0 & 2\sigma & 0 & 0 \\ \frac{-1}{B-A} & 0 & 0 & \frac{-\delta}{B-A} & 0 \\ \frac{-1}{B-A} & 0 & 0 & 0 & \frac{-\delta}{B-A} \end{bmatrix}.$$

Using notations from [5], we define $\Psi_1(c) = e^{-2c}$, $\Psi_2(c) = e^{-c}$, $\Psi_3(c) = ce^{-c}$, $\Psi_4(c) = c$, $\Psi_5(c) = e^c$, $\Psi_6(c) = c^2$, $\Psi_7(c) = ce^c$ and $\Psi_8(c) = e^{2c}$. As described in [5], we find $f_{1,1} = -2e^{-2c}$, $f_{2,2} = e^c/2$, $f_{3,3} = 1$, $f_{4,4} = -2e^c$, $f_{5,5} = 6e^c$, $f_{6,6} = -e^{-c}/3$, $f_{7,7} = -3e^c$, $f_{8,8} = 24e^c$ and $F = \prod_{i=1}^{8} f_{i,i} = 288e^{2c} > 0$. Therefore, by Theorem 2 in [5], any optimal design based on the Fisher information matrix is supported at no more than $8/2+1 = 5$ points including two boundary points. □

Note that C_c is independent of the boundary parameters A and B, and P_θ does not involve c. Thus, the D-optimal design does not depend on the boundary parameters A and B.

Let us numerically study the sharpness of the derived upper bound. We focus on D-optimality for all the numerical studies. Since the D-optimal design does not depend on A and B, without loss of generality we define $A = 0$ and $B = 10$. We assume that the design interval is $\mathbb{X} = [-2,2]$. Suppose that $a = 0$ and $b = 1$. Straightforward calculus gives the D-optimal designs: $\xi^*_{2p} = \{(-2,0.5),(2,0.5)\}$ for $\sigma = 1$; $\xi^*_{3p} = \{(-2,0.41),(0,0.18),(2,0.41)\}$ for $\sigma = 0.4$; $\xi^*_{4p} = \{(-2,0.26),(-0.94,0.24),(0.94, 0.24),(2,0.26)\}$ for $\sigma = 0.1$. In Figure 1 the sensitivity function $d(x,\xi^*,\theta) = \mathrm{tr}\left\{I(\theta,x)M_{\xi^*}(\theta)^{-1}\right\}$ is depicted for these three designs. Note that the D-optimality of the computed designs is confirmed by the equivalence theorem. We have not found cases when the D-optimal design is supported at 5 points.

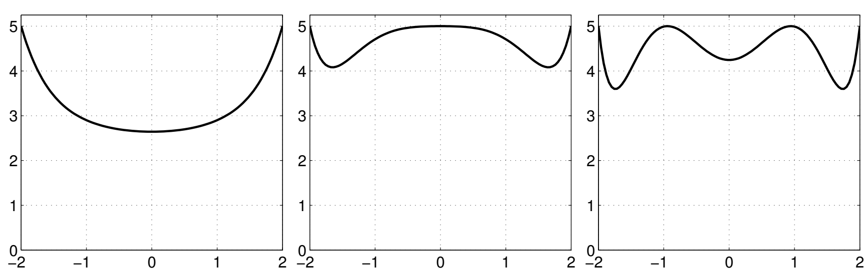

Fig. 1 The sensitivity function $d(x,\xi^*,\theta)$ for model (1) with unknown σ in three cases. Left: $\sigma = 1$. Middle: $\sigma = 0.4$. Right: $\sigma = 0.1$.

Now we evaluate the asymptotic efficiency of designs ξ^*_{2p}, ξ^*_{3p}, ξ^*_{4p}, the uniform design $\xi_{\text{unif}} = \{(-2,0.2),(-1,0.2),(0,0.2),(1,0.2),(2,0.2)\}$ and the D-optimal designs under different values of σ. Figure 2 displays the asymptotic efficiency of ξ^*_{2p}, ξ^*_{3p}, ξ^*_{4p} and ξ_{unif} relative to the D-optimal design. It is seen that ξ^*_{2p} is optimal when σ is large whereas ξ^*_{3p} and ξ^*_{4p} are each optimal only at one value of σ. Note that the D-efficiency of ξ_{unif} is about 0.9 for small σ and 0.7 for large σ.

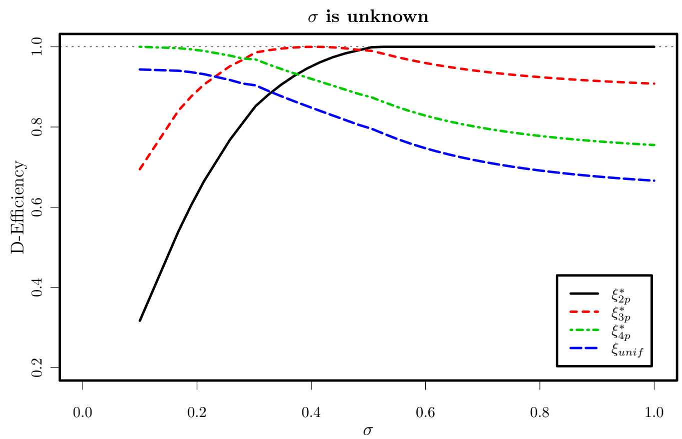

Fig. 2 Values of $\{\det M_\xi(\theta)/\det M_{\xi^D}(\theta)\}^{1/5}$ the D-efficiency of design ξ for the model (1) with unknown σ, different σ and cases when ξ is ξ^*_{2p}, ξ^*_{3p}, ξ^*_{4p} and ξ_{unif}, $A = 0$, $B = 10$.

For finite sample sizes, we compare the mean square error (MSE) for each parameter estimate under different designs by simulation. When sample sizes are small, the MLE may not exist, see [3]. For these cases, the smallest and largest observations are used as estimators of A and B. We consider the two point design, the uniform design and a design in which design points are randomly taken from a continuous uniform distribution on $\mathbb{X} = [-2,2]$. The last we call the random design. Table 1 presents the relative MSE (RMSE) of each parameter estimate calculated from 1000 repetitions of the simulation. The two point design outperforms the other two designs for most scenarios. It does not perform well for estimating σ when $\sigma = 0.5$. The two point design is not optimal in this scenario.

2.2 *The case of known σ*

When σ is known, there are four unknown parameters and the matrix C_c in the Fisher information matrix reduces to $C_c = \begin{bmatrix} C_{c11} & C_{c12} \\ C_{c21} & \Sigma_c \end{bmatrix}$, where $C_{c11} = \begin{bmatrix} 1 & c \\ c & c^2 \end{bmatrix}$, $C_{c21} = C^T_{c12} = (\mathbf{Z}_c, c\mathbf{Z}_c)$, $\mathbf{Z}_c = (e^c, e^{-c})^T$ and $\Sigma_c = \begin{bmatrix} (\sigma^2+1)\delta^2 e^{2c} & (1-\sigma^2)/\delta^2 \\ (1-\sigma^2)/\delta^2 & (\sigma^2+1)\delta^2 e^{-2c} \end{bmatrix}$. Using

Table 1 Values of $MSE(e_k^T\hat{\theta}|\xi_{2p}^*)/MSE(e_k^T\hat{\theta}|\xi)\times 100\%$, the relative performance of estimating individual parameters for the model (1) with unknown σ in cases when ξ is the uniform design and the random design of size n, $k=1,\ldots,5$. "NS" is the number of cases that consistent solution to the likelihood equation cannot be found for the given design and "NST" is the number of cases of no consistent solution for the two point design among 1000 repetitions of simulation

Design		a	b	σ	A	B	NS	NST
$\sigma=1.0$								
Uniform	$n=20$	98.5	44.6	105.2	39.8	40.8	252	263
	$n=40$	98.3	50.6	113.5	39.2	44.7	11	16
	$n=80$	96.3	53.7	117.9	45.7	43.2	0	0
Random	$n=20$	76.5	29.9	86.0	21.7	19.3	293	263
	$n=40$	74.5	31.9	89.3	22.2	21.4	10	16
	$n=80$	89.0	30.7	100.6	22.6	20.7	0	0
$\sigma=0.5$								
Uniform	$n=20$	139.3	88.4	270.6	74.6	77.0	66	175
	$n=40$	131.4	108.6	283.4	87.2	92.8	4	27
	$n=80$	120.1	114.2	249.8	90.6	101.8	0	0
Random	$n=20$	79.0	47.6	174.2	33.5	32.0	108	175
	$n=40$	83.5	55.4	181.8	40.5	43.2	16	27
	$n=80$	94.5	57.8	182.4	44.4	43.0	0	0

arguments from the proof of Theorem 1 we can obtain that an optimal design that minimizes an information based criterion for this model is also supported at no more than 5 points.

In the following theorem we explicitly derive the D-optimal design in some cases.

Theorem 2. *For model (1) with known σ, there exists a constant $\zeta<9$ such that if $(\sigma^2+1)e^{\sigma^2}>\zeta$ the D-optimal design ξ^* is the two-point design supported at boundary points with equal weights.*

Proof. From the extended general equivalence theorem in [4], it suffices to show that $\sup_{c\in[l,u]} d(c,\xi^*,\theta)=4$. Note that

$$\mathrm{tr}\left\{I_c(\theta)M_{\xi^*}(\theta)^{-1}\right\} = \mathrm{tr}\left\{P_\theta C_c P_\theta^T\left(P_\theta A P_\theta^T\right)^{-1}\right\} = \mathrm{tr}\left(A^{-1}C_c\right),$$

where $A=(C_l+C_u)/2$, and C_l and C_u has the same form as C_c with c replaced by l and u, respectively. Thus, we need to prove that $\sup_{c\in[l,u]}\mathrm{tr}(A^{-1}C_c)=4$. By tedious calculation, we have $\mathrm{tr}\{(C_l-C_u)(C_l+C_u)^{-1}\}=0$, which implies that $\mathrm{tr}(A^{-1}C_l)=\mathrm{tr}(A^{-1}C_u)=4$.

Now we will prove that $\mathrm{tr}(A^{-1}C_c)$ reaches its maximum at the boundary points l and u. By direct calculation, we obtain

$$
\begin{aligned}
\operatorname{tr}(A^{-1}C_c) =& 2\frac{(c-l)^2+(c-u)^2}{(l-u)^2}+\operatorname{tr}\left(D^{-1}\Gamma_c\right)\\
&+\frac{\operatorname{tr}\left(D^{-1}\{(u-c)(\mathbf{Z}_l-\mathbf{Z}_c)-(c-l)(\mathbf{Z}_u-\mathbf{Z}_c)\}^{\otimes 2}\right)}{(l-u)^2}\\
\le& 2+\operatorname{tr}(D^{-1}\Gamma_c)+\operatorname{tr}(D^{-1}\kappa),
\end{aligned}
$$

where $\kappa=\{2(u-c)^2(\mathbf{Z}_l-\mathbf{Z}_c)^{\otimes 2}+2(c-l)^2(\mathbf{Z}_u-\mathbf{Z}_c)^{\otimes 2}\}/(l-u)^2$, $\Gamma_c=\Sigma_c-\mathbf{Z}_c\mathbf{Z}_c^T$, $D=(\Gamma_l+\Gamma_u)/2$, Γ_l and Γ_u has the same form as Γ_c with c replaced by l and u, respectively, and $M^{\otimes 2}=MM^T$ for any matrix M.

Note that $\operatorname{tr}(D^{-1}\kappa)+\operatorname{tr}(D^{-1}\Gamma_c)=2$ at the boundary l or u. Consequently, we need to show that $\operatorname{tr}(D^{-1}\kappa)+\operatorname{tr}(D^{-1}\Gamma_c)$ or equivalently $\operatorname{tr}(D^*\kappa)+\operatorname{tr}(D^*\Gamma_c)$ achieves its maximum at boundary points l or u, where

$$
D^*=\frac{1}{2}\begin{bmatrix}\phi\left(e^{-2l}+e^{-2u}\right) & 2\psi\\ 2\psi & \phi\left(e^{2l}+e^{2u}\right)\end{bmatrix}
$$

is the co-factor matrix of D, and $\phi=(\sigma^2+1)\delta^2-1$ and $\psi=\{1+\delta^2-\sigma^2\}/\delta^2>0$.

By direct calculation, we have

$$
\begin{aligned}
\operatorname{tr}(D^*\kappa) &\le 2\phi\frac{(u-c)^2}{(l-u)^2}(e^{2l}+e^{2u})(e^{-c}-e^{-l})^2+2\phi\frac{(c-l)^2}{(l-u)^2}(e^{-2l}+e^{-2u})(e^{c}-e^{u})^2\\
&=\Delta_1+\Delta_2
\end{aligned}
$$

and

$$
\operatorname{tr}(D^*\Gamma_c)=\frac{\phi^2}{2}\{(e^{-2l}+e^{-2u})e^{2c}+(e^{2l}+e^{2u})e^{-2c}\}-2\psi^2.
$$

Note that if $\Delta_i<\operatorname{tr}(D^*\Gamma_l)/2-\operatorname{tr}(D^*\Gamma_c)/2$ for $c\in(l,u)$, $i=1,2$, then it follows that $\operatorname{tr}(D^*\kappa)+\operatorname{tr}(D^*\Gamma_c)$ achieves its maximum at $c=l$. This is true because Δ_i vanishes at the two boundary points, $i=1,2$.

By direct calculation, we obtain

$$
\frac{1}{2}\{\operatorname{tr}(D^*\Gamma_l)-\operatorname{tr}(D^*\Gamma_c)\}=\frac{\phi^2}{4}\left(e^{2l}+e^{2u}\right)\left\{1-e^{2(x-u)}\right\}\left(e^{-2l}-e^{-2x}\right). \tag{3}
$$

Thus, assuming $\phi\ge 8$, to prove that $\Delta_1<\operatorname{tr}(D^*\Gamma_l)/2-\operatorname{tr}(D^*\Gamma_c)/2$, we need

$$
\frac{(u-c)^2}{(l-u)^2}\le\left\{1-e^{2(c-u)}\right\}\frac{1+e^{l-c}}{1-e^{l-c}}. \tag{4}
$$

Note that the inequality (4) is true if $\left\{1-e^{2(c-u)}\right\}(1+e^{l-c})/(1-e^{l-c})>1$. Otherwise, we have $2e^{l-c}<e^{2(c-u)}+e^{l+c-2u}<2e^{2(c-u)}$, which implies $(l+2u)/3<c$. Thus, if $u-l<1$, then by the mean value theorem we have

$$
\left\{1-e^{2(c-u)}\right\}\frac{e^{c-l}+1}{e^{c-l}-1}>2(u-c)e^{\frac{2(l-u)}{3}}\frac{2}{e(c-l)}>\frac{4(u-c)}{e^2(c-l)}>\frac{1}{3}\frac{(u-c)}{(u-l)}>\frac{(u-c)^2}{(l-u)^2}.
$$

If $u-l \geq 1$, we obtain $\{1-e^{2(c-u)}\}(e^{c-l}+1)/(e^{c-l}-1) > \{1-e^{2(c-u)}\}$. Let $h(c) = \{1-e^{2(c-u)}\} - (u-c)/\{3(l-u)\}$. Then $h''(c) = -4e^{2(c-u)} < 0$ and, therefore, $h(x)$ is convex. Note that $h(u) = 0$ and $h\{(l+2u)/3\} = 8/9 - e^{2(l-u)/3} \geq 8/9 - e^{-2/3} > 0$ and, thus, $h(c) > 0$ for $c \in [(l+2u)/3, u]$. $\square$

Let us now investigate the dependence of the D-optimal design on σ. As previously, we suppose that $A = 0$, $B = 10$, $\mathbb{X} = [-2,2]$, $a = 0$ and $b = 1$. Then we obtain the D-optimal designs: $\xi_{2p}^* = \{(-2, 0.5), (2, 0.5)\}$ for $\sigma = 1$; $\zeta_{3p}^* = \{(-2, 0.425), (0, 0.15), (2, 0.425)\}$ for $\sigma = 0.2$; $\zeta_{4p}^* = \{(-2, 0.33), (-0.86, 0.17), (0.86, 0.17), (2, 0.33)\}$ for $\sigma = 0.1$. In Figure 3 we depict the sensitivity function $d(x, \xi^*, \theta)$ for these three designs. We can observe that $d(x, \xi^*, \theta) \leq 4$ for all $x \in \mathbb{X}$ that proves the D-optimality. Note that the D-optimal design is the two-point design if $\sigma > 0.31$ (for $a = 0$ and $b = 1$). The relative asymptotic efficiency graphs given

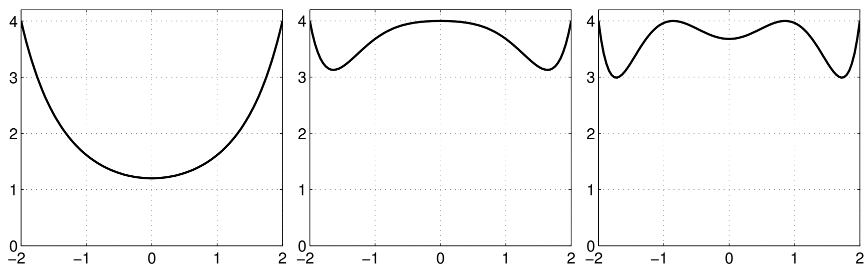

Fig. 3 The sensitivity function $d(x, \xi^*, \theta)$ for model (1) with known σ in three cases. Left: $\sigma = 1$. Middle: $\sigma = 0.2$. Right: $\sigma = 0.1$.

in Figure 4 for the case of known σ are similar to the case when σ is unknown. However, the design ξ_{2p}^* is D-optimal for a larger range of σ.

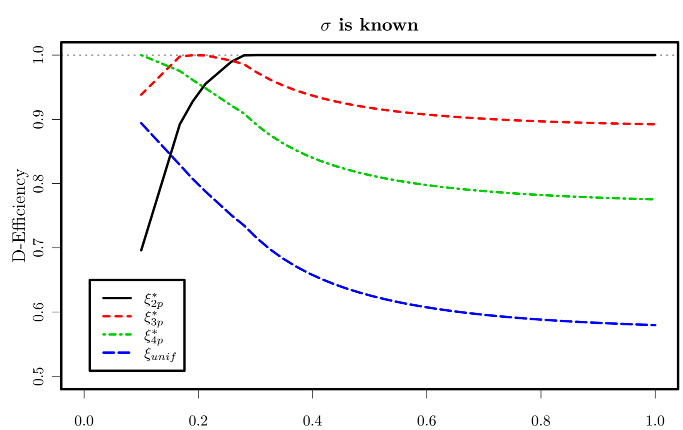

Fig. 4 Values of $\{\det M_\xi(\theta)/\det M_{\xi^D}(\theta)\}^{1/4}$ the D-efficiency of design ξ for the model (1) with known σ, different σ and cases when ξ is ξ_{2p}^*, ζ_{3p}^*, ζ_{4p}^* and ξ_{unif}, $A = 0$, $B = 10$.

Table 2 reports finite sample comparisons. The two point design dominates the other designs when $\sigma = 1.0$. When $\sigma = 0.5$, the uniform design provides better accuracy for estimating only the parameter a and the two point design is preferable for estimating the other parameters.

Table 2 Values of $MSE(e_k^T\hat{\theta}|\xi_{2p}^*)/MSE(e_k^T\hat{\theta}|\xi) \times 100\%$, the relative performance of estimating individual parameters for the model (1) with known σ in cases when ξ is the uniform design and the random design of size n, $k = 1, \ldots, 4$. "NS" is the number of cases that consistent solution to the likelihood equation cannot be found for the given design and "NST" is the number of cases of no consistent solution for the two point design among 1000 repetitions of simulation

Design		a	b	A	B	NS	NST
$\sigma = 1.0$							
Uniform	$n = 20$	96.5	40.6	35.3	35.3	0	0
	$n = 40$	98.6	45.2	33.7	37.3	0	0
	$n = 80$	96.3	48.5	40.7	38.6	0	0
Random	$n = 20$	74.2	28.5	19.6	17.6	4	0
	$n = 40$	75.4	31.2	20.8	19.7	0	0
	$n = 80$	90.3	30.9	21.9	21.1	0	0
$\sigma = 0.5$							
Uniform	$n = 20$	120.9	42.4	47.4	44.0	20	13
	$n = 40$	129.0	49.0	46.2	49.1	0	1
	$n = 80$	118.4	53.6	50.8	53.0	0	0
Random	$n = 20$	75.0	32.0	21.9	19.3	23	13
	$n = 40$	81.6	33.5	24.5	23.4	0	1
	$n = 80$	94.6	34.9	27.4	26.4	0	0

References

[1] C. Carathéodory. Über den Variabilitätsbereich der Fourierschen Konstanten von positiven harmonischen Funktionen. *Rendiconti del Circolo Matematico di Palermo*, 32:193–217, 1911.

[2] Ernst Steinitz. Bedingt konvergente Reihen und konvexe Systeme. *J. Reine Angew. Math.*, 143:128–175, 1913.

[3] HaiYing Wang and Nancy Flournoy. Inference on a new sigmoid regression model with unknown support and unbounded likelihood function, 2012. NI12001-DAE: http://www.newton.ac.uk/preprints/NI12001.pdf.

[4] Lynda V. White. An extension of the general equivalence theorem to nonlinear models. *Biometrika*, 60:345–348, 1973.

[5] Min Yang. On the de la Garza phenomenon. *Annals of Statistics*, 38:2499–2524, 2010.