Biostat 279 HW 2

Elvis Cui

Han Cui

Dept. of Biostat

UCLA

Biostat 279 Homework 2 due 5/17/2021

1 Suppose you have a 2-parameter linear model Ey = a + bx defined on [-1,1] and the variance of y is inversely proportional to a known function $\lambda(x)$, apart from an unimportant positive constant. Assuming that all observations are independent, find D-optimal designs (in terms of c) for estimating the two parameters a and b if

(a) $\lambda(x) = c-x^2$, $c>1$;

(b) $\lambda(x) = x+c$, $c>1$.

2 Let Tr A denote the trace of the square matrix A. Show that the function Tr M^{-1} is a convex function of M over the space of all information matrices generated by approximate designs on a compact interval X, i.e. show that if μ_1 and μ_2 are two approximate designs defined on X and $0 \le a \le 1$, then

$$\mathrm{Tr}\,(aM(\mu_1) + (1-a)M(\mu_2))^{-1} \le a\mathrm{Tr}\,M(\mu_1)^{-1} + (1-a)\mathrm{Tr}\,M(\mu_2)^{-1}.$$

Use the equivalence theorem and show that given a model, the A and D optimal designs are the same if the information matrices of the two optimal designs are proportional to the identity matrix.

3 Verify the 2x2 normalized information matrix of the two-parameter logistic model with the probability of response $p_i = 1/(\exp(-b(x_i-a))$ at dose x_i, i= 1, 2, …, n, shown in the lecture notes. Confirm that the locally D-optimal design is supported equally at the 17.6 and 82.4 percentiles of the logistic curve by displaying the sensitivity plot for your choices of 3 pairs of nominal values. Is the locally D-optimal design still the same if the model was parameterized by $p_i = 1/(\exp(-bx_i-a)$ instead? Justify.

4 I have posted several reading materials on the entire class website to date, and you will continue to see more today. Browse through them and you may pick your project based on one of the papers. You can expand on ideas in the paper or propose interesting design problems motivated by one or more papers. Alternatively, you can also propose an optimal design project. The contents on the project can be theoretical or based on developing a software program using PSO or other metaheuristic or traditional algorithms for finding new optimal designs.

Q1:

1 Suppose you have a 2-parameter linear model Ey = a + bx defined on [-1,1] and the variance of y is inversely proportional to a known function λ(x), apart from an unimportant positive constant. Assuming that all observations are independent, find D-optimal designs (in terms of c) for estimating the two parameters a and b if

(a) $\lambda(x) = c-x^2$, c>1;

(b) $\lambda(x) = x+c$, c>1.

Sol.
$$M(\xi) = \sum_{i=1}^{n} \lambda(x_i)\begin{pmatrix}1\\x_i\end{pmatrix}(1 \;\; x_i)\,w_i$$
$$= \sum_{i=1}^{n}\lambda(x_i)\begin{pmatrix}1 & x_i\\ x_i & x_i^2\end{pmatrix} w_i$$

where $\xi = \begin{pmatrix} x_1 & \cdots & x_n \\ w_1 & \cdots & w_n\end{pmatrix}$

(a): $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $\epsilon_i \sim N(0, \frac{1}{c-x_i^2})$

Let $\xi = \begin{pmatrix} -a & a \\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$, then

$$m(\xi) = (c-a^2)\left[\frac{1}{2}\begin{pmatrix}1 & -a\\ -a & a^2\end{pmatrix} + \frac{1}{2}\begin{pmatrix}1 & -a\\ -a & a^2\end{pmatrix}\right]$$
$$= (c-a^2)\begin{bmatrix}1 & 0\\ 0 & a^2\end{bmatrix}$$
$$\Rightarrow m^{-1}(\xi) = \frac{1}{c-a^2}\begin{bmatrix}1 & \\ & \frac{1}{a^2}\end{bmatrix}$$

Sensitivity function:

$$d(x,\xi) = \lambda(x) f(x)^T m^{-1}(\xi) f(x) - 2$$
$$= \frac{c-x^2}{c-a^2}(1 \;\; x)\begin{pmatrix}1 & \\ & \frac{1}{a^2}\end{pmatrix}\begin{pmatrix}1\\x\end{pmatrix} - 2$$
$$= \frac{1}{c-a^2}(c-x^2)\left(1 + \frac{x^2}{a^2}\right) - 2$$
$$= -\frac{x^4}{(c-a^2)a^2} + \frac{1}{a^2}x^2 + \frac{c}{(c-a^2)} - 2$$

So $d(x,\xi)$ attains maximum when
$x^2 = \frac{c-a^2}{2}$

If plug-in $x = \pm a$, then

$d(\pm a, \xi) = 0$!

Simplification

Thus, if $\frac{c-a^2}{2} \geq 1 \; \forall a$, i.e.,

$c \geq 3$, then $a = \pm 1$,

$$\xi_D = \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}.$$

if $c < 3$, then $\frac{c-a^2}{2} = a^2$

$\Rightarrow a = \pm\sqrt{\frac{c}{3}}$, which results in

$$\xi_D = \begin{pmatrix} -\sqrt{\frac{c}{3}} & +\sqrt{\frac{c}{3}} \\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}.$$

(b): $\lambda(x) = x + c$.

Let $\xi = \begin{pmatrix} -1 & +1 \\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$,

$$m(\xi) = \frac{-1+c}{2}\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix} + \frac{1+c}{2}\begin{pmatrix}1 & 1\\ 1 & 1\end{pmatrix}$$
$$= \begin{pmatrix} c & 1 \\ 1 & c\end{pmatrix} \Rightarrow$$
$$m^{-1}(\xi) = \frac{1}{c^2-1}\begin{pmatrix} c & -1 \\ -1 & c\end{pmatrix}$$

$$d(x,\xi) = \lambda(x) f(x)^T M^{-1}(\xi) f(x) - 2$$

$$= \frac{(x+c)}{c^2-1} \times (1 \quad x) \begin{pmatrix} c & -1 \\ -1 & c \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} - 2$$

$$= \frac{x+c}{c^2-1} (cx^2 - 2x + c) - 2$$

$$= \frac{1}{c^2-1} \left[cx^3 + (-2+c^2)x^2 - cx + c^2 \right] - 2$$

Let $g(x) = cx^3 + (c^2-2)x^2 - cx + c^2$

$\Rightarrow g'(x) = 3cx^2 + 2(c^2-2)x - c$

At $x = 1$: $g(1) = 2(c^2-1)$

$g'(1) = 2(c-1)(c+2) > 0$

$\Rightarrow d(1,\xi) = 0$ &

$d(1^-,\xi) < 0$.

At $x=-1$: $g(-1) = 2(c^2-1)$

$g'(-1) = -2(c+1)(c-2)$

if $c > 2$, $g'(-1) < 0$ &

$d(-1,\xi) = 0$ &

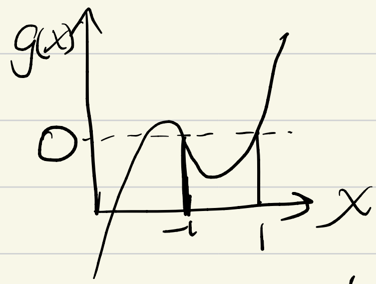
$d(-1^+,\xi) < 0$

Thus,

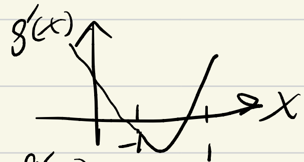
$\xi_D = \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$ is D-opt

if $c \geq 2$.

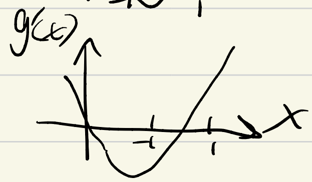
The reason is, $g(x)$ looks like this: (if $c > 2$)



Since $g'(x)$ is like this



or



$\square$

Next, if $c < 2$:

Let $\xi = \begin{pmatrix} a & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

$$M(\xi) =$$

$$\frac{(a+c)}{2} \begin{pmatrix} 1 & a \\ a & a^2 \end{pmatrix} + \frac{1+c}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} c + \frac{a+1}{2} & \frac{a^2+ac+c+1}{2} \\ \frac{a^2+ac+c+1}{2} & \frac{a^3+a^2c+c+1}{2} \end{pmatrix}$$

→ Next page

By equi. thm, the checking condition is indeed (for D-opt design)

$$\lambda(x) f(x)^T M^{-1}(\xi) f(x) - 2 \le 0 \quad (*)$$

$\forall x \in [-1,1]$ & when $x=1$ & a, equality is attained.

After some tedious manipulation on Mathematica, I got

$$LHS(*) = \frac{2(a-x)(x-1)\left[\overbrace{1+a^2-2c^2-x-a(2+c+x)-c(1+2x)}^{(\triangle)}\right]}{(a-1)^2(1+c)(a+c)}$$

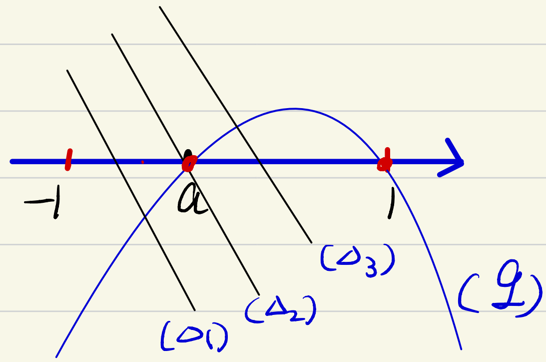
which is 0 when $x=a, x=1$.

Further Hand movement manually simplification outputs

$$(\triangle) = -(1+a+2c)x-(a-1)^2-c(a+1+2c)$$

Next, I will show that $(*)$ attains maxima iff $(\triangle)$ is 0 when $x=a$.

Pf: The quadratic part & linear part of $(*)$ should be ploted as:



- For $(\triangle_1)$: Since $(*)$ is continuous, so $\exists\ a^* \in [-1, a)$ such that $(g)<0$, $(\triangle_1)<0 \Rightarrow (g)\cdot(\triangle_1)>0$, thus, $(*)>0$, a contradiction to equivalence theorem.
- For $(\triangle_3)$: $\exists\ a^* (a, 1]$ s.t. $(g)>0$, $(\triangle_3)>0$ & thus $(*)>0$. A contradiction.
- Thus, $(\triangle_2)$ is the only possible choice for the linear part $(\triangle)$ & we have

$$(\triangle)=0 \text{ at point } x=a. \quad \square$$

This gives

$$a = \frac{1-2c^2-c}{3(c+1)} \text{ for } c<2.$$

after some calculation.

Q2:

2 Let Tr A denote the trace of the square matrix A. Show that the function Tr M^{-1} is a convex function of M over the space of all information matrices generated by approximate designs on a compact interval X, i.e. show that if μ_1 and μ_2 are two approximate designs defined on X and $0 \le a \le 1$, then

$$\mathrm{Tr}\,(aM(\mu_1) + (1-a)M(\mu_2))^{-1} \le a\mathrm{Tr}\,M(\mu_1)^{-1} + (1-a)\mathrm{Tr}\,M(\mu_2)^{-1}.$$

Use the equivalence theorem and show that given a model, the A and D optimal designs are the same if the information matrices of the two optimal designs are proportional to the identity matrix.

Sol. (1) $\mathrm{Tr}\,M^{-1}$ is a cvx fct. of M over the sp. of all info. matrices on a compact X.

Pf: ETS $\forall \mu_1, \mu_2$ approx. designs, $\forall a \in [0,1]$,

$$\mathrm{Tr}(aM(\mu_1)+(1-a)M(\mu_2))^{-1} \le a\mathrm{Tr}[M(\mu_1)^{-1}] + (1-a)\mathrm{Tr}[M(\mu_2)^{-1}]$$

By 250A & Simultaneous diag. thm, $\exists\, \mathcal{C}$, $|\mathcal{C}| \neq 0$ s.t.

$$\mathcal{U}^T M(\mu_1)\mathcal{U} = I,\quad \mathcal{U}^T M(\mu_2)\mathcal{U} = \Lambda,$$

diag. of Λ are eigvals of $M(\mu_1)^{-\frac{1}{2}} M(\mu_2) M(\mu_1)^{-\frac{1}{2}}$

And $\mathcal{U} = M(\mu_1)^{-\frac{1}{2}} P$, $P^TP = PP^T = I$

Thus, $\begin{cases} M(\mu_1) = \mathcal{U}^{-T}\mathcal{U}^{-1} \\ M(\mu_2) = \mathcal{U}^{-T}\Lambda\,\mathcal{U}^{-1} \end{cases}$.

$$\Rightarrow \mathrm{Tr}(aM(\mu_1)+(1-a)M(\mu_2))^{-1}$$

$$= \mathrm{Tr}(a\mathcal{U}^{-T}\mathcal{U}^{-1} + (1-a)\mathcal{U}^{-T}\Lambda\mathcal{U}^{-1})^{-1}$$

$$= \mathrm{Tr}[\mathcal{U}^{-T}(aI+(1-a)\Lambda)\mathcal{U}^{-1}]^{-1}$$

$$= \mathrm{Tr}[\mathcal{U}(aI+(1-a)\Lambda)^{-1}\mathcal{U}^T]$$

$$\overset{(*)}{\le} \mathrm{Tr}[\mathcal{U}(aI+(1-a)\Lambda^{-1})\mathcal{U}^T]$$

$$= a\,\mathrm{Tr}\,\mathcal{U}\mathcal{U}^T + (1-a)\mathrm{Tr}(\mathcal{U}\Lambda^{-1}\mathcal{U}^T)$$

where $(*)$ comes from:

$$\frac{1}{a+(1-a)\lambda} \le \frac{1}{a} + \frac{1}{(1-a)\lambda}$$

$\forall \lambda > 0$. $a \in [0,1]$.

(Harmonic mean $\le$ arithmatic mean)

(2) Show A, D-optimal are equivalent if $m(\xi) \propto I$.

Pf: if $m(\xi) = cI$ for $c > 0$.

Then the sensitivity fct for

D-opt: $f(x)^T m^{-1}(\xi) f(x) - p$

$$= \frac{1}{c} f(x)^T f(x) - p$$

A-opt: $f(x)^T m^{-2}(\xi) f(x) - \mathrm{Tr}(m^{-1}(\xi))$

$$= \frac{1}{c^2} f(x)^T f(x) - \frac{1}{c} p$$

$$= \frac{1}{c}\left[\frac{1}{c} f(x)^T f(x) - p\right]$$

So up to a const., D-opt and A-opt designs are equivalent.

Q3:

3 Verify the 2x2 normalized information matrix of the two-parameter logistic model with the probability of response $p_i = 1/(\exp(-b(x_i-a))$ at dose x_i, i= 1, 2, …, n, shown in the lecture notes. Confirm that the locally D-optimal design is supported equally at the 17.6 and 82.4 percentiles of the logistic curve by displaying the sensitivity plot for your choices of 3 pairs of nominal values. Is the locally D-optimal design still the same if the model was parameterized by $p_i = 1/(\exp(-bx_i-a)$ instead? Justify.

Sol. ① $y_i = \log\frac{P_i}{1-P_i} = +b(x_i-a)$, let

$l_i = \log\left(e^{Y_i y_i - \log(1+e^{y_i})}\right)$, $\theta = (a,b)$, then

$$-\frac{\partial^2 l_i}{\partial\theta\partial\theta^T} = \frac{-\partial}{\partial y_i}\left(\frac{\partial l_i}{\partial y_i}\frac{\partial y_i}{\partial\theta}\right)\frac{\partial y_i}{\partial\theta^T}$$

$$= \frac{e^{y_i}}{(1+e^{y_i})^2} \times \begin{pmatrix}-b\\ (x_i-a)\end{pmatrix}\left(-b \quad (x_i-a)\right)$$

$$= P_i(1-P_i)\begin{pmatrix} b^2 & -b(x_i-a)\\ -b(x_i-a) & x_i^2\end{pmatrix}$$

Then the info matrix associated with

$$\xi = \begin{pmatrix} x_1 & \cdots & x_n\\ w_1 & & w_n\end{pmatrix} \text{ is}$$

$$-E\sum_{i=1}^{n}\frac{\partial^2 l_i}{\partial\theta\partial\theta^T} =$$

$$\begin{bmatrix} \sum_{i=1}^{n} P_i(1-P_i)b^2 & +b\sum_{i=1}^{n} P_i(1-P_i)(x_i-a)\\ +b\sum_{i=1}^{n} P_i(1-P_i)(x_i-a) & \sum_{i=1}^{n} x_i^2 P_i(1-P_i)\end{bmatrix}$$

② Let $\xi = \begin{pmatrix}\gamma & 2a-\gamma\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$

and note that

$$P_i(1-P_i) = \frac{e^{-b(x_i-a)}}{(1+e^{-b(x_i-a)})^2}$$

Then, Let $t = -b(\gamma-a)$, we have

$$m(\xi) = \begin{bmatrix} b^2 & 0\\ 0 & (\gamma-a)^2\end{bmatrix}\frac{e^t}{(1+e^t)^2}$$

using the fact

$$P_1(1-P_1) = P_2(1-P_2) = \frac{e^t}{(1+e^t)^2}$$

Then

$$\det|m(\xi)| = \frac{t^2 e^{2t}}{(1+e^t)^4} \equiv e^{g(t)}$$

$\Rightarrow$

$$g(t) = 2\log t + 2t - 4\log(1+e^t)$$

$$g'(t) = \frac{2}{t} + 2 - \frac{4e^t}{1+e^t} = 0$$

produces

$t_1 = 1.543405,\ t_2 = -1.543405$

Finally,

$\gamma_1 = -\frac{t_1}{b} + a$

$\gamma_2 = -\frac{t_2}{b} + a$ $(2a-\gamma)$

plug-in them into $\log\frac{P_i}{1-P_i} = -b(x_i-a)$

$$\Rightarrow \begin{cases} P_1 = 0.82396\\ P_2 = 0.17604\end{cases}$$

To check

$$\xi_D = \begin{pmatrix} \gamma & 2a-\gamma \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

for $\gamma = -\frac{t}{b} + a$,

$t = 1.543405$

is indeed D-opt $\forall\, a, b$,

use equivalence thm:

$$\nabla P_x(\theta)\, M^{-1}(\xi_D)\, \nabla P_x(\theta) - 2$$

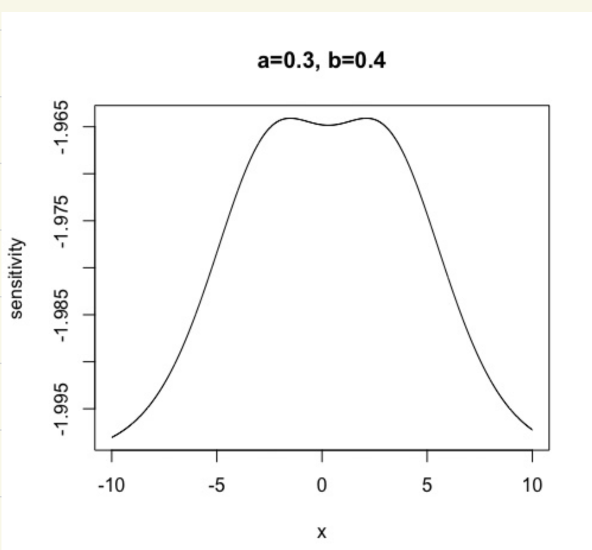
with nominal values

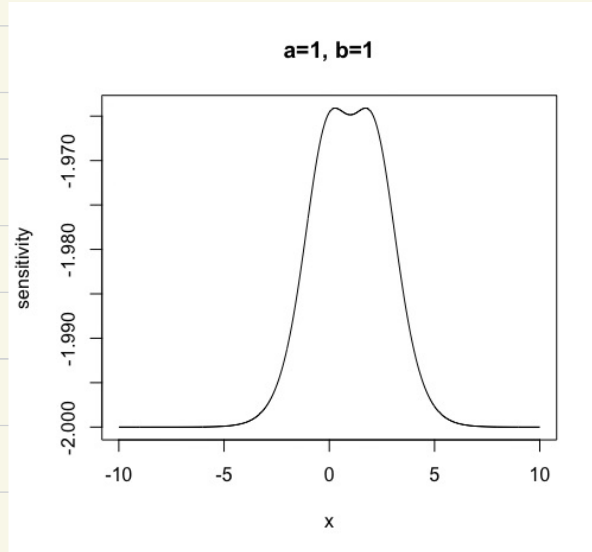
$\theta_1 = (0.3, 0.4)$

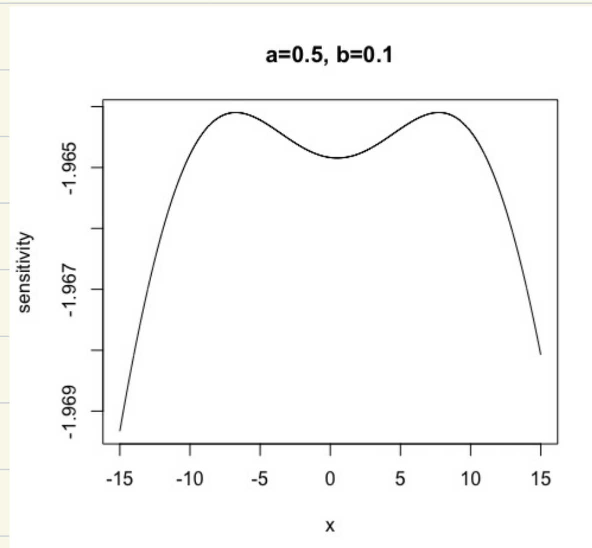
$\theta_2 = (1, 1)$

$\theta_3 = (0.5, 1)$

Plots are shown below.







③ what if

$y_i = \log\frac{P_i}{1-P_i} = a + bx_i$?

Sol.

$$-\frac{\partial^2 l_i}{\partial\theta\partial\theta^T} = P_i(1-P_i)\begin{pmatrix}1\\x_i\end{pmatrix}(1 \quad x_i)$$

$$= \begin{pmatrix} P_i(1-P_i) & x_i P_i(1-P_i) \\ x_i P_i(1-P_i) & x_i^2 P_i(1-P_i) \end{pmatrix}$$

The info. mat $M(\xi)$ is

$$-\sum_{i=1}^{n} \frac{\partial^2 l_i}{\partial\theta\partial\theta^T} w_i$$

Note $P_i(1-P_i) = \frac{e^{a+bx_i}}{(1+e^{a+bx_i})^2}$

Let $\xi = \begin{pmatrix} \gamma & -\frac{2a}{b} - \gamma \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$

and $t = a + b\gamma$

Then we have

$$M(\xi) = \frac{e^t}{(1+e^t)^2}\begin{pmatrix} 1 & -\frac{a}{b} \\ -\frac{a}{b} & (\gamma+\frac{a}{b})^2 + \frac{a^2}{b^2} \end{pmatrix}$$

$$|M(\xi)| = \frac{e^{2t}}{(1+e^t)^4}(\gamma + \frac{a}{b})^2 \equiv e^{g(t)}$$

$\Rightarrow$

$$g(t) = 2t - 4\log(1+e^t) + 2\log\left(\frac{t}{b}\right)$$

$$g'(t) = 2 - \frac{4e^t}{1+e^t} + \frac{2}{t} = 0$$

produces

$t_1 = +1.5434$;

$t_2 = -1.5434$.

$\Rightarrow \gamma_1 = \frac{t_1 - a}{b}$, $\gamma_2 = \frac{t_2 - a}{b}$

which results in

$$\begin{cases} P_1 = 0.82396 \\ P_2 = 0.17604 \end{cases}$$

This is the same locally D-opt design as we had before.