# 6. Optimal Designs

## 6.1. Continuous Designs

Recall our definition of an *exact* design:

$$d = \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \\ r_1, \ldots, r_n \end{array} \right\}, \qquad \text{for } \boldsymbol{x}_i \in \chi \subset \mathbb{R}^m$$

e.g. $\chi = [-1, 1]^m$ . Here, $\boldsymbol{x}_i$ is a *support point* of the design; $\chi$ is the design space; $r_j$ is integer with $0 < r_j \leq N$;

$$\sum_{j=1}^{n} r_j = N;$$

and $n$ is the number of support points, i.e. the number of distinct points.

**Example 13**: $2^2$ factorial design

$$d = \left\{ \begin{array}{cccc} (-1,-1) & (-1,+1) & (+1,-1) & (+1,+1) \\ r_1 & r_2 & r_3 & r_4 \end{array} \right\},$$

- four support (distinct) points;
- $r_j = 1 \quad j = 1, \ldots, 4$: unreplicated $2^2$ factorial design

  $r_j = 2 \quad j = 1, \ldots, 4$: two replicates

  $r_i \neq r_j$ for at least one $i, j$: unbalanced design.

We can *normalise* the design by defining

$$r_j^* = r_j / N\,,$$

with $0 < r_j^* < 1$; $\quad \sum_j r_j^* = 1$. If we relax the assumption that $N r_j^*$ must be an integer, we can define an *approximate* or *continuous* design:

$$\xi = \left\{ \begin{array}{c} \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \\ w_1, \ldots, w_n \end{array} \right\}$$

with $0 < w_j \leq 1$ and $\sum_j w_j = 1$. Here, $w_j$ can be *any* number in $(0, 1]$ subject to $\sum_j w_j = 1$ with no restriction that $Nw_j$ is an integer. Hence, a continuous design is independent of $N$. In practice, $Nw_j$ needs to be rounded to obtain an integer.

**Example 13 cont.**: $2^2$ factorial design

$$\xi = \left\{ \begin{array}{cccc} (-1,-1) & (-1,+1) & (+1,-1) & (+1,+1) \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}$$

- we place $w_j = \frac{1}{4}$ of our experimental resource at each support point;
- $w_j$ need not be equal, but often are for "good" designs.

**(Normalised) Information matrix**: for a linear model $\boldsymbol{Y} = \tilde{X}\boldsymbol{\beta} + \varepsilon$, let $X$ be a $n \times p$ model matrix including the $n$ unique rows of $\tilde{X}$. The $j$-th row of $X$ is given by

$$\mathbf{f}(\mathbf{x}_j)^T,$$

for the $j$th support point, $\mathbf{x}_j$. Thus

$$\tilde{X} = \begin{bmatrix} \mathbf{f}(\mathbf{x}_1)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_1)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_n)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_n)^T \end{bmatrix}$$

**Example 14**: single factor quadratic regression

Consider $Y(x) = \mathbf{f}(x)^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, for $x \in [-1, 1]$, where

$$\mathbf{f}(x)^{\mathrm{T}} = \begin{pmatrix} 1 & x & x^2 \end{pmatrix}$$

If we take the design

$$\xi = \left\{ \begin{array}{cc} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{array} \right\},$$

then

$$X = \begin{bmatrix} \mathbf{f}(-1) \\ \mathbf{f}(1) \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

If $N = 4$, then two runs are observed for each support point and

$$\tilde{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Define

$$M(\xi) = X^{\mathrm{T}} W X$$

to be the (normalised) information matrix, where

$$W = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix}.$$

Hence $M(\xi)$ is a $p \times p$ matrix and can be written as

$$M(\xi) = \sum_{j=1}^{n} w_j \mathbf{f}(\mathbf{x}_j) \mathbf{f}(\mathbf{x}_j)^T.$$

This is true for all information matrices, as they are additive - add the information for different support points, weighted by the proportion of runs assigned to that point.

Now,

(i)

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= (\tilde{X}^{\mathrm{T}}\tilde{X})^{-1}\sigma^2 \\
&= (NX^{\mathrm{T}}WX)^{-1}\sigma^2 \\
&= \frac{1}{N}(X^{\mathrm{T}}WX)^{-1}\sigma^2 \\
&= \frac{1}{N}M(\xi)^{-1}\sigma^2
\end{aligned}$$

(ii)

$$\begin{aligned}
\mathrm{Var}(\hat{Y}(\mathbf{x})) &= \mathrm{Var}(\mathbf{f}(\mathbf{x})^{\mathrm{T}}\hat{\boldsymbol{\beta}}) = \mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{f}(\mathbf{x}) \\
&= \frac{1}{N}\mathbf{f}(\mathbf{x})^{\mathrm{T}}M(\xi)^{-1}\mathbf{f}(\mathbf{x})\sigma^2.
\end{aligned}$$

**Example 13 cont**: Consider $Y(x) = \mathbf{f}(x)^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon$, for $x \in [-1, 1]$, with $\mathbf{f}(x)^{\mathrm{T}} = \begin{pmatrix} 1 & x & x^2 \end{pmatrix}$. If we take the design

$$\xi = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & +\frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\},$$

then

$$X = [] , \qquad W = [] \ ,$$

and

$$M(\xi) = X^{\mathrm{T}}WX = \tag{6.1}$$

with

$$M(\xi)^{-1} = \begin{bmatrix} 2.56 & 0 & -2.81 \\ 0 & 1.8 & 0 \\ -2.81 & 0 & 5.06 \end{bmatrix}.$$

This means that

$$\begin{aligned}
\mathrm{Var}(\hat{\boldsymbol{\beta}}) &= \frac{1}{N}M(\xi)^{-1}\sigma^2 = \frac{1}{N}(X^{\mathrm{T}}WX)^{-1}\sigma^2 \\
&= \frac{1}{N}\begin{bmatrix} 2.56 & 0 & -2.81 \\ 0 & 1.8 & 0 \\ -2.81 & 0 & 5.06 \end{bmatrix}\sigma^2
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}(\hat{Y}(x)) &= \frac{1}{N}\mathbf{f}(x)^{\mathrm{T}}M(\xi)^{-1}\mathbf{f}(x)\sigma^2 \\
&= \frac{1}{N}\begin{pmatrix} 1 & x & x^2 \end{pmatrix} M(\xi)^{-1}\begin{pmatrix} 1 \\ x \\ x^2 \end{pmatrix}\sigma^2 \\
&= \frac{1}{N}(2.56 - 3.82x^2 + 5.06x^4)\sigma^2
\end{aligned}$$

**Standardised variance** is defined as:

$$\begin{aligned}
\nu(\boldsymbol{x},\xi) &= \frac{N\mathrm{Var}(\hat{Y}(\boldsymbol{x}))}{\sigma^2} \\
&= \mathbf{f}(\boldsymbol{x})^{\mathrm{T}}M(\xi)^{-1}\mathbf{f}(\boldsymbol{x})
\end{aligned}$$

Two results on the standardised variance

**Result 1**

$$\sum_{j=1}^{n} w_j\nu(\boldsymbol{x}_j,\xi) = p$$

**Proof**

$$\begin{aligned}
\sum_j w_j \nu(\boldsymbol{x}_j, \xi) &= \sum_j w_j \mathbf{f}(\boldsymbol{x})^{\mathrm{T}} M^{-1}(\xi) \mathbf{f}(\boldsymbol{x}) \\
&= \sum_j w_j \mathrm{tr}\left\{ M^{-1}(\xi) \mathbf{f}(\boldsymbol{x}) \mathbf{f}(\boldsymbol{x})^{\mathrm{T}} \right\} \\
&= \mathrm{tr}\left\{ M^{-1}(\xi) \sum_j w_j \mathbf{f}(\boldsymbol{x}) \mathbf{f}(\boldsymbol{x})^{\mathrm{T}} \right\} \\
&= \mathrm{tr}\left( M^{-1}(\xi) M(\xi) \right) \\
&= \mathrm{tr}(I_p) = p
\end{aligned}$$

$[\boldsymbol{a}^{\mathrm{T}} A \boldsymbol{a} = \mathrm{tr}(\boldsymbol{a}^{\mathrm{T}} A \boldsymbol{a}) = \mathrm{tr}(A \boldsymbol{a} \boldsymbol{a}^{\mathrm{T}})]$

**Result 2**

$$\max_{\boldsymbol{x} \in \mathcal{X}} \nu(\boldsymbol{x}, \xi) \geq p$$

**Proof**

$$\begin{aligned}
p &= \sum_j w_j \nu(\boldsymbol{x}_j, \xi) \\
&\leq \sum_j w_j \max_{\boldsymbol{x} \in \mathcal{X}} \nu(\boldsymbol{x}, \xi) \\
&= \max_{\boldsymbol{x} \in \mathcal{X}} \nu(\boldsymbol{x}, \xi) \sum_j w_j \\
&= \max_{\boldsymbol{x} \in \mathcal{X}} \nu(\boldsymbol{x}, \xi)
\end{aligned}$$

$$\Rightarrow \max_{\boldsymbol{x} \in \mathcal{X}} \nu(\boldsymbol{x}, \xi) \geq p$$

## 6.2. Optimality Criteria

In order to choose a design $\xi$ to use in an experiment, we can define various optimality criteria

- mathematical functions that encapsulate the performance of a design for a particular objective;
- choosing $\xi$ entails choosing $\boldsymbol{x}_j$ and $w_j$ $(j = 1, \ldots, n)$ **and** the value of $n$ (how many support points).

Most popular optimality criteria are based on functions of the information matrix $M(\xi)$.

1. $D$-optimality

$$\begin{aligned}\Psi_D(\xi) &= \log\left[|M(\xi)|^{1/p}\right] \\ &= \frac{1}{p}\log|M(\xi)|.\end{aligned}$$

[$|A| = \det(A)$]

A design $\xi^*$ is $D$-optimal if

$$\Psi_D(\xi^*) = \max_{\xi}\log\left[|M(\xi)|^{1/p}\right] .$$

This is equivalent to minimising the log determinant of the covariance matrix of $\hat{\boldsymbol{\beta}}$ since

$$\Psi_D(\xi^*) = \min_{\xi}\left(-\log\left[|M(\xi)|^{1/p}\right]\right) = \min_{\xi}\log\left[|M(\xi)^{-1}|^{1/p}\right],$$

[recall $\log|A| = \log 1/|A^{-1}| = -\log|A^{-1}|$].

Noticing that

$$\text{var}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{N}M(\xi)^{-1},$$

thus

$$\Psi_D(\xi^*) = \min_{\xi}\log\left[|(N/\sigma^2)\text{var}(\hat{\boldsymbol{\beta}})|^{1/p}\right].$$

A $D$-optimal design minimises the volume of confidence ellipsoid for $\hat{\boldsymbol{\beta}}$, i.e. provides the best accuracy of parameter estimators.

Note that $\xi^*$ can be found by maximising $\log|M(\xi)|$.

2. $G$-optimality

$$\max_{\boldsymbol{x}\in\chi}\nu(\boldsymbol{x},\xi)$$

A design $\xi^*$ is $G$-optimal if

$$\Psi_G(\xi^*) = \min_{\xi} \max_{\boldsymbol{x} \in \chi} \nu(\boldsymbol{x}, \xi)\,.$$

This criterion minimises the maximum prediction variance, and gives the best "worst-case" for prediction accuracy.

There are lots of other "alphabetic" optimality criteria, e.g.

A-optimality: $\Psi_A(\xi^*) = \min_\xi \mathrm{tr}\,\{M(\xi)^{-1}\}$;

V-optimality: $\Psi_V(\xi^*) = \min \int_\chi \nu(\boldsymbol{x}, \xi) g(\boldsymbol{x}) d\boldsymbol{x}$,

where $g(\boldsymbol{x})$ is a pdf across $\chi$.

**Example 14 cont.**: quadratic regression

$$\xi = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}.$$

$$M(\xi) = \begin{bmatrix} 1 & 0 & \frac{5}{9} \\ 0 & \frac{5}{9} & 0 \\ \frac{5}{9} & 0 & \frac{41}{81} \end{bmatrix};$$

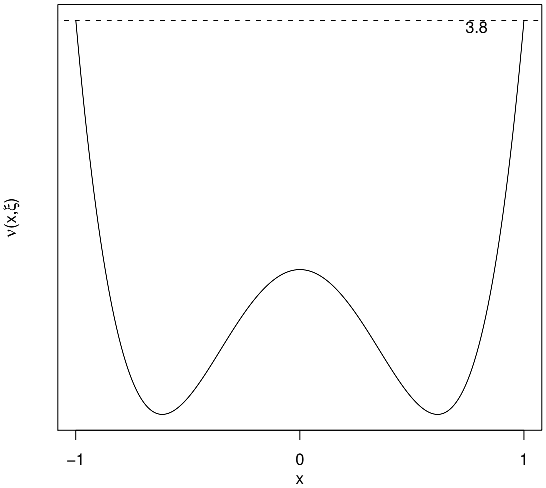
$$M(\xi)^{-1} = \begin{bmatrix} 2.56 & 0 & -2.81 \\ 0 & 1.8 & 0 \\ -2.81 & 0 & 5.06 \end{bmatrix};$$

$$\begin{aligned} |M(\xi)| &= \left(\frac{5}{9}\right)\left(\frac{41}{81}\right) - \frac{5}{9}\left(\frac{5}{9}\right)^2 \\ &= 0.1097\,; \end{aligned}$$

$$\Psi_D(\xi) = \frac{1}{p}\log(0.1097) = -0.7365\,;$$

$$\nu(x, \xi) = 2.56 - 3.82x^2 + 5.06x^4\,;$$

We can plot $\nu(x, \xi)$ as a function of $x$, and mark the maximum:




$$\max \nu(x, \xi) = 3.8 \quad (\text{at } x = -1, +1)\,.$$

**Design efficiency**: we can assess a design $\xi$ by comparing to the optimal design $\xi^*$.

$D$**-efficiency**:

$$D\text{-eff}(\xi) = \left\{ \frac{|M(\xi)|}{|M(\xi^*)|} \right\}^{1/p}$$

- $0 \le D\text{-eff}(\xi) \le 1$;
- $D\text{-eff}(\xi) = 1 \Rightarrow \xi$ is $D$-optimal;
- $D\text{-eff}(\xi) = 0 \Rightarrow \xi$ cannot estimate the model and $|M(\xi)| = 0$.

$G$**-efficiency**: We know that $\max_{\boldsymbol{x} \in \mathcal{X}} \nu(\boldsymbol{x}, \xi) \ge p$, and so we have a lower bound on $\Psi_G(\xi)$. Hence $G$-efficiency is defined as

$$G\text{-eff}(\xi) = p/\Psi_G(\xi)\,.$$

**Example 14 cont.**: quadratic regression with

$$\xi = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\}$$

$$G\text{-eff} = \frac{3}{3.8} = 0.79$$

To calculate the $D$-eff, we first need the $D$-optimal design for $x \in [-1, +1]$. In general, this is done using numerical optimisation on the computer (e.g. using SAS). This is beyond the scope of this course.

For this example, the $D$-optimal design has 3 support points and for $x \in [-1, 1]$ is given by

$$\xi^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\} .$$

Now

$$\begin{aligned} M(\xi^*) &= \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{3} \end{bmatrix} ; \end{aligned}$$

The determinant of $M(\xi^*)$ is given by

$$|M(\xi^*)| = 0.1481,$$

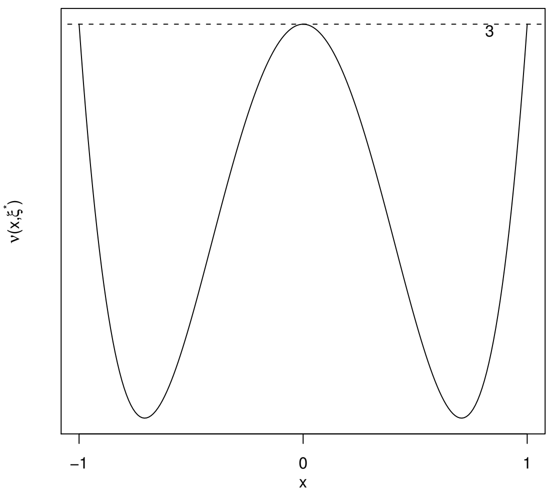
so that

$$\Psi_D(\xi^*) = \log |M(\xi^*)|^{1/3} = -0.6365 ;$$

$$\begin{aligned} D\text{-eff}(\xi) &= \left\{ \frac{|M(\xi)|}{|M(\xi^*)|} \right\}^{1/p} \\ &= \left( \frac{0.1097}{0.1481} \right)^{1/3} \\ &= 0.905 ; \end{aligned}$$

$$
\begin{aligned}
\nu(\boldsymbol{x}, \xi^*) &= \begin{bmatrix} 1 & x & x^2 \end{bmatrix} \begin{bmatrix} 3 & 0 & -3 \\ 0 & 1.5 & 0 \\ -3 & 0 & 4.5 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= \begin{bmatrix} 3 - 3x^2 & 1.5x & -3 + 4.5x^2 \end{bmatrix} \begin{bmatrix} 1 \\ x \\ x^2 \end{bmatrix} \\
&= 3 - 3x^2 + 1.5x^2 - 3x^2 + 4.5x^4 \\
&= 3 - 4.5x^2 + 4.5x^4 \, ;
\end{aligned}
$$

We can again plot $\nu(\boldsymbol{x}, \xi^*)$, and mark the maximum.




$\max_{x \in \chi} \nu(x, \xi^*) = 3$ at $x = -1, +1, 0$, the support points of design.

Note $G$-eff$(\xi^*) = p/\Psi_G(\xi^*) = 3/3 = 1$, implying that the $D$-optimal design is also $G$-optimal. This is the basic result of the *General Equivalence Theorem.*

### 6.3. General Equivalence Theorem

The following three conditions on a continuous design $\xi^*$ are equivalent

1. $\Psi_D(\xi^*) = \max_\xi \Psi_D(\xi)$

- that is, $\xi^*$ is $D$-optimal

2. $\max_{\boldsymbol{x}\in\mathcal{X}} \nu(\boldsymbol{x},\xi^*) = \min_\xi \max_\mathcal{X} \nu(\boldsymbol{x},\xi)$

   - that is, $\xi^*$ is $G$-optimal

3. $\nu(\boldsymbol{x},\xi^*) \leq p$, with equality for $\boldsymbol{x}$ belonging to the support points of $\xi^*$.

Hence, we can use the standardised variance $\nu(\boldsymbol{x},\xi)$ to establish if a design is $D$-optimal, as any non-optimal design will have $\max_{\boldsymbol{x}\in\mathcal{X}} \nu(\boldsymbol{x},\xi) > p$.

**Example 14 cont.** quadratic regression with

$$\xi^* = \left\{ \begin{array}{ccc} -1 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array} \right\} .$$

$$\nu(\boldsymbol{x},\xi^*) = 3 - 4.5x^2 + 4.5x^4$$

$\max \nu(\boldsymbol{x},\xi^*) = 3$ at $x = -1, 0, 1$.

This maximum occurs at the support points of $\xi^*$; for $0 < |x| < 1$, $x^2 > x^4$ and hence $\nu(\boldsymbol{x},\xi^*) < 3$, thus implying that $\xi^*$ is a $D$-optimal and $G$-optimal design.

We also considered

$$\xi_1 = \left\{ \begin{array}{cccc} -1 & -\frac{1}{3} & \frac{1}{3} & 1 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{array} \right\} ,$$

and showed it had $D$-eff$< 1$. Also,

$$\nu(\boldsymbol{x},\xi_1) = 2.56 - 3.82x^2 + 5.06x^4$$

$$\max \nu(\boldsymbol{x},\xi_1) = 3.8 > 3\,,$$

and hence this design is not $D$-optimal (for $|x| < 1$, $x^2 \geq x^4$ and hence $\nu(\boldsymbol{x},\xi_1) \leq 3.8$).