Optimal experimental design, an introduction

Jesus.LopezFidalgo@uclm.es
University of Castilla-La Mancha
Department of Mathematics
Institute of Applied Mathematics to Science and Engineering
Books (just a few)

Concepts and notation

Process:

- Select a model,
  \[ E(y \mid x) = f^T(x)\theta, \quad \text{Var}(y \mid x) = \sigma^2 \] (linear model).
- \( f \) continuous with components linearly independent.
- Select an experimental condition \( x \) on a **design space** \( \chi \) (typically a compact on a Euclidian space).
- Examples:
  - \( y = \alpha_0 + \alpha_1 x + \epsilon, \ f(x) = (1, x)^T, \ x \in \chi = [a, b]. \)
  - \( y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \epsilon, \ f(x) = x = (x_1, x_2, x_3)^T, \)
    \( x \in \chi = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]. \)
  - \( y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \epsilon, \ f(x) = (1, x, x^2)^T, \ x \in \chi = [a, b]. \)
  - ANOVA (1-way, 3 levels):
    \( y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \ f(x_1, x_2) = (1, x_1, x_2)^T, \)
    \( x = (x_1, x_2) \in \chi = \{(0, 1), (1, 0), (1, 1)\}. \)
Perform the experiment and observe the response ($r.v. \ y$).

Repeat the process for the experimental conditions $x_1, \ldots, x_n$ obtaining responses $y_1, \ldots, y_n$.

Estimate the parameters (MLE) $\theta_1, \ldots, \theta_m, \sigma^2$ and make inferences.
EXPERIMENTAL DESIGN

- **Aim:** Choosing $x_1, \ldots, x_n$ (*exact design of size* $n$).
- Some may be repeated (probability measure):
  $\xi_n(x) = \frac{n_x}{n}$, where $n_x =$ times $x$ appears in the design.
- Covariance matrix:
  $\Sigma\hat{\theta} = \sigma^2 (X^T X)^{-1} = \sigma^2 n^{-1} M^{-1}(\xi),$
  $M(\xi) = \sum_{x \in \chi} f(x) f^T(x) \xi_n(x)$ (*information matrix*):
  - Symmetric.
  - Non-definite positive.
  - Singular if the number of different points in the design $k < m$. 
Extended concept of experimental design

- **Approximate design (Kiefer):** probability measure $\zeta$ on the Borel field generated by the open sets of $\chi$.
- **Information matrix:** $M(\zeta) = \int_{\chi} f(x)f^T(x)\zeta(dx)$.
- Convex set of approximate designs: $\Xi$.
- If $\zeta$ is discrete and $\xi$ the probability function associated:
  \[
  M(\xi) = \sum_{x \in \chi} f(x)f^T(x)\xi(x)
  \]
- If $\zeta$ is absolutely continuous there exists a density function (pdf), $\xi$, associated:
  \[
  M(\xi) = \int_{\chi} f(x)f^T(x)\xi(x)dx
  \]
- Compact and convex set of information matrices: $\mathcal{M}$
Caratheodory’s theorem

- **Caratheodory’s theorem**: For any information matrix there is always a design with at most $\frac{1}{2}m(m+1)+1$ different points.

- Finite support,

$$\xi = \left\{ x_1, x_2, \ldots, x_k, p_1, p_2, \ldots, p_k \right\}$$

- Restricting $\Xi$ to finite designs $\mathcal{M}$ is still the same (Caratheodory, definition of integral and continuity).

- In practice, perform $n_i \approx n\xi(x_i)$ experiments at $x_i$, $\sum_i n_i = n$.

- From now on we will use $\xi$ or $\zeta$ as convenience.
Example

\[ y = \alpha_0 + \alpha_1 x + \varepsilon, \quad x \in \chi = [0, 1], \quad f(x) = (1, x)^T. \]

\[ \xi = \begin{cases} 0 & 0.5 & 1 \\ 0.2 & 0.4 & 0.4 \end{cases}. \]

\[
M(\xi) = \sum_{i=1}^{3} \begin{pmatrix} 1 & x_i & x_i^2 \\ x_i & x_i^2 \end{pmatrix} p_i \\
= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} 0.2 + \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.25 \end{pmatrix} 0.4 + \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} 0.4 \\
= \begin{pmatrix} 1 & 0.6 \\ 0.6 & 0.5 \end{pmatrix}
\]
Example

\[ \xi = \begin{pmatrix} 0 & 0.5 & 1 \\ 0.2 & 0.4 & 0.4 \end{pmatrix} . \]

If just \( n = 12 \) experiments may be performed, possible rounding off:

about \( 0.2 \times 12 = 2.4 \) \((2, 3, 3)\) experiments at 0,
about \( 0.4 \times 12 = 4.8 \) \((5, 4, 5)\) experiments at 0.5,
about \( 0.4 \times 12 = 4.8 \) \((5, 5, 4)\) experiments at 1.

**Remark:** It is not as simple as the usual rounding off.
CRITERION FUNCTION

- $\Phi : \Xi \rightarrow [0, +\infty)$ or $\Phi : \mathcal{M} \rightarrow [0, +\infty)$ to be minimized (maximized).
- The first is more general.
- The second
  - has to be considered as a function of $M^{-1}$,
  - has to be non-increasing (better estimates of the parameters):
    $$M \leq N \Rightarrow \Phi(M) \geq \Phi(N),$$
- may be global or partial (interest on part of the parameters),
- may be positive homogeneous: $\Phi(\delta M) = \frac{1}{\delta} \Phi(M), \quad \delta > 0$,
- $\mathcal{M}$ is within the cone of the non-negative definite matrices in the Euclidean space of the symmetric matrices. Thus, the usual concept of differentiability applies here.
- A **Φ–optimal design** $\zeta^*$ minimizes $\Phi$.
- Convex: $\Phi[(1 - \epsilon)\zeta + \epsilon\zeta'] \leq (1 - \epsilon)\Phi(\zeta) + \epsilon\Phi(\zeta')$. 
Equivalence theorem for differentiable criteria

▶ Sensitive function:
\[
\psi(x, \xi) = \partial \Phi[M(\xi), f(x)f^T(x)] \\
= f^T(x)\nabla \Phi[M(\xi)]f(x) - \text{tr} M(\xi)\nabla \Phi[M(\xi)].
\]

▶ Equivalence theorem:
\( \xi^* \) is \( \Phi \)-optimal if and only if \( \psi(x, \xi^*) \geq 0, \ x \in \chi \).
Equality for \( x \in S_{\xi^*} \).
Comments on the equivalence theorem

- Just for approximate designs.
- $\mathcal{M}$ is convex and it is the same for finite or general designs.
- From now on we assume finite designs and use the symbol $\xi$ (pdf) instead of $\zeta$ (measure).
- Still valid for a restricted search in a convex subset.
Checking condition

\[ \psi(z, \xi^*) = 0, \ z \in S_{\xi}^* \]

\[ \left( \frac{\partial \psi(x, \xi^*)}{\partial x} \right)_{x=z} = 0, \ z \in S_{\xi}^* \cap \text{Int}(\chi) \]

E.g., for a two parameter model assume \( \xi = \begin{pmatrix} x_1 & x_2 \\ 1 - p & p \end{pmatrix} \),

then there are two equations:

\[ \psi(x_1, \xi) = 0, \ \psi(x_2, \xi) = 0 \]

and 0, 1 or 2 of the equations:

\[ \left( \frac{\partial \psi(x, \xi)}{\partial x} \right)_{x=x_1} = 0, \ \left( \frac{\partial \psi(x, \xi)}{\partial x} \right)_{x=x_2} = 0 \]

to be solved.
EFFICIENCY (goodness of a design)

\[ \text{eff}_\Phi(\xi) = \frac{\Phi[M(\xi^*)]}{\Phi[M(\xi)]}. \]

If \( \Phi \) homogenous, e.g. for 60% efficiency and \( n \) observations:

\[ 0.6 = \frac{\Phi[M(\xi^*)]}{\Phi[M(\xi)]} = \frac{\Phi(\sigma^{-2} n \Sigma^{-1} \hat{\theta})}{\Phi(\sigma^{-2} n \Sigma^{-1} \hat{\theta})} = \frac{\sigma^2 n^{-1} \Phi(\Sigma^{-1})}{\sigma^2 n^{-1} \Phi(\Sigma^{-1})} = \frac{\Phi(\Sigma^{-1})}{\Phi(\Sigma^{-1})}. \]

For \( n \) observations with \( \xi \) and \( n^* \) with \( \xi^* \) the efficiency is 1 if:

\[ 1 = \frac{\Phi(\Sigma^{-1})}{\Phi(\Sigma^{-1})} = \frac{\sigma^2 n^{-1} \Phi[M(\xi^*)]}{\sigma^2 n^{-1} \Phi[M(\xi)]} = \frac{\sigma^2 n^{-1}}{\sigma^2 n^{-1}} 0.6, \]

\( n^* = 0.6 n \) (60%) observations would be enough with \( \xi^* \).
A $\Phi$–optimal design has $m$ points at least: Any design with less than $m$ points gives a singular matrix.

For strictly convex criteria there is always a $\Phi$–optimal design with no more than $\frac{m(m+1)}{2}$ points:

- The optimal must be in the boundary (dimension $\frac{m(m+1)}{2} - 1$).
- Caratheodory in the boundary gives a limit of $\frac{m(m+1)}{2}$ points.

If $\sigma^2(x)$ is not constant, the whole theory is applicable for

$$\tilde{f}(x) = f(x)/\sigma(x).$$

If the observations are correlated ($\Sigma_y$), just exact designs.

Information matrix: $X^T\Sigma_y^{-1}X$. 
Determinant of the inverse:

\[
\Phi_D[M(\xi)] = \begin{cases} 
\log \det M^{-1}(\xi) = -\log \det M(\xi) & \text{if } \det M(\xi) \neq 0, \\
\infty & \text{if } \det M(\xi) = 0.
\end{cases}
\]

Homogeneous definition:

\[
\Phi_D[M(\xi)] = \begin{cases} 
\det M^{-1/m}(\xi) & \text{if } \det M(\xi) \neq 0, \\
\infty & \text{if } \det M(\xi) = 0.
\end{cases}
\]
Properties

- $\Phi_D$ continuous on $\mathcal{M}$.
- Convex on $\mathcal{M}$ and strictly convex on $\mathcal{M}_+$ (nonsingular information matrices).
- Differentiable on $\mathcal{M}_+$:
  \[ \nabla(-\log \det M) = -M^{-1} \]

- Minimum volume of the confidence ellipsoid:
  \[ (\hat{\Theta} - \Theta)^T (X^T X)^{-1} (\hat{\Theta} - \Theta) \leq mF_{m, n-m, \gamma} S_R^2 \equiv c^2. \]

Volume = $c^m V_m [\det M^{-1}(\xi)]^{1/2}$, where $V_m$ is the volume of the $m$–dimensional sphere.
G-optimality

- Kirstine Smith (1918).
- Kiefer & Wolfovitz gave the name.

\[
\Phi_G[M(\xi)] = \begin{cases} 
\max_{x \in \chi} f^T(x) M^{-1}(\xi)f(x) & \text{if } \det M(\xi) \neq 0, \\
\infty & \text{if } \det M(\xi) = 0.
\end{cases}
\]

- \(\Phi_G\) continuous and convex on \(M\).
- \(\Phi_G[M(\xi)] \propto \max_{x \in \chi} \text{var}_{\xi} \hat{y}(x)\).
A–optimality

- Trace of the inverse:

\[
\Phi_A[M(\xi)] = \begin{cases} 
\text{tr}M^{-1}(\xi) & \text{if } \det M(\xi) \neq 0 \\
\infty & \text{if } \det M(\xi) = 0, 
\end{cases}
\propto \sum_{i=1}^{m} \text{var}_{\xi} \hat{\theta}_i.
\]

The second equality is a consequence of

\[
\text{var}_{\xi} \hat{\theta}_i \propto e_i^t M^{-1}(\xi)e_i
\]

if the parameters are estimable.

- \(\Phi_A\) continuous on \(\mathcal{M}\).
- Convex on \(\mathcal{M}\) and strictly convex on \(\mathcal{M}_+\).
- Differentiable on \(\mathcal{M}_+\):

\[
\nabla (\text{tr}M^{-1}(\xi)) = -M^{-2}(\xi)
\]
Atwood (1976): 

\[ \Phi_L[M(\xi)] = \begin{cases} 
\text{tr} WM^{-1}(\xi) & \text{if } \det M(\xi) \neq 0, \\
\infty & \text{if } \det M(\xi) = 0, 
\end{cases} \]

where \( W \) is a positive definite matrix of dimension \( m \).

\( \Phi_L \) continuous on \( M \).

Convex on \( M \) and strictly convex on \( M_+ \).

Differentiable on \( M_+ \):

\[ \nabla [\text{tr} WM^{-1}(\xi)] = -M^{-1}(\xi)WM^{-1}(\xi). \]
c–optimality

- Variance of the estimator of $c^T \theta$, linear combination of the parameters.
- $\Phi_c[M(\xi)] = c^T M^- (\xi) c$. 
Elfving procedure in practice

- Plot the curve $x(t) = f_1(t)$, $y(t) = f_2(t)$, $t \in \chi$ and its symmetric through origin.
- Plot the convex hull of both curves.
- Plot the line through the origin defined by vector $c$.
- Optimal design defined by the boundary point.
Elfving plot (One–point design)

\[
\begin{align*}
\{ t^* \} \\
\{ 1 \}
\end{align*}
\]

Elfving plot (Two–point design)

\[
\begin{bmatrix}
a & t^* \\
1 - p^* & p^*
\end{bmatrix}
\]
Elfving procedure for three parameters

\[ y = \theta_0 + \theta_1 t + \theta_2 t^2 + \varepsilon, \quad t \in [0, 1] \]

One, two or three design points.
ξ* is D–optimal iff it is G–optimal:

\[
\max_{x \in \chi} d(x, \xi^*) = m,
\]

where \( d(x, \xi) = f^T(x)M^{-1}(\xi)f(x). \)

Efficiency bound: \( \text{eff}_D[M(\xi)] \geq 2 - \frac{\max_x d(x, \xi)}{m} \)
Example: Checking condition for a quadratic model

\[
E[y] = \theta^T f(x) = \theta_1 + \theta_2 x + \theta_3 x^2, \sigma^2(x) = 1,
\]
\[
x \in \chi = [-1, 2], f(x) = (1, x, x^2)^T.
\]

\[
M(\xi) = \sum_{x \in [-1, 2]} f(x)f^T(x)\xi(x)
\]
\[
= \sum_{x \in [-1, 2]} \xi(x) \begin{pmatrix}
1 & x & x^2 \\
x & x^2 & x^3 \\
x^2 & x^3 & x^4 
\end{pmatrix}
\]

Assume the D-optimal design is of the type:

\[
\xi = \left\{ \begin{array}{c}
-1 \\
1/3 \\
1/3 \\
1/3 
\end{array} \right\}
\]
Maximize the determinant:

$$\det M(\xi) = \frac{1}{3} (-z^2 + z + 2)^2.$$ 

The maximum is reached at $z^* = 0.5$:

$$\xi^* = \begin{bmatrix} -1 & 0.5 & 2 \\ 1/3 & 1/3 & 1/3 \end{bmatrix},$$

Checking whether $\xi^*$ is the D-optimal:

$$d(x, \xi^*) = f^T(x)M^{-1}(\xi^*)f(x) = 0.89 \left( x^2 - 3.43x + 3.3 \right) \left( x^2 + 1.43x + 0.87 \right).$$
Example: Elfving procedure

\[ E[y] = \theta^T f(x) = \theta_1 x + \theta_2 x^2, \sigma^2(x) = 1, \]
\[ x \in \chi = [0, 1], f(x) = (1, x, x^2)^T. \]

Plot the curve \( x(t) = t, y(t) = t^2, t \in [0, 1] \):
Symmetric through the origin
Tangential point, $z$:

\[
\frac{x(z) - x_0}{x'(z)} = \frac{y(z) - y_0}{y'(z)}
\]

\[
\frac{z + 1}{1} = \frac{z^2 + 1}{2z}, \quad z = \sqrt{2} - 1
\]
Line through the origin defined by vector $c = (0.5, 0.3)$

\[(\lambda 0.5)^2 = \lambda 0.3, \ t = 3/5\]

\[\xi_c^* = \begin{bmatrix} 3/5 \\ 1 \end{bmatrix}\]
Line through the origin defined by vector \( \mathbf{c} = (0.5, 1) \)

Cutting point:

\[
\begin{align*}
\left\{ \begin{array}{l}
y = 2x, \\
\frac{x-1}{z+1} = \frac{y-1}{z^2+1},
\end{array} \right.
\end{align*}
\]

\[
x = \frac{1}{4} \left( 2 - \sqrt{2} \right) = 0.15, \quad y = 0.30.
\]
Weights

Convex combination:

\[(0.15, 0.30) = (1 - p)(-z, -z^2) + p(1, 1), \quad p = \frac{1}{4} \left( 3 - \sqrt{2} \right) = 0.30\]

\[\xi^*_c = \left\{ \begin{array}{cc} \sqrt{2} - 1 & 1 \\ \frac{1}{4} (1 + \sqrt{2}) & \frac{1}{4} (3 - \sqrt{2}) \end{array} \right\} .\]
THANK YOU