Introduction to Computational Algebra for Design of Experiments

A. Di Bucchianico

Eindhoven University of Technology
Department of Mathematics

Slides partially based on slides from H.P. Wynn and E. Riccomagno

Dortmund, January 12, 2005

Contents of this talk

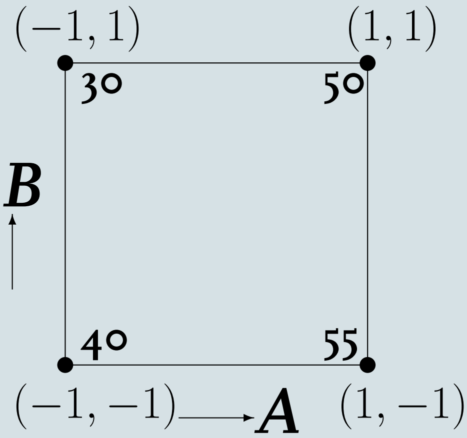
- terminology 2-level factorial designs
- functional description of Least Squares
- design and ideals
- estimable sets
- replications
- mixture designs

Recommended reading:

Pistone and Wynn, Biometrika 83 (1996), 653-666.

Pistone, Riccomagno and Wynn, *Algebraic Statistics. Computational Commutative Algebra in Statistics*, Chapman & Hall, 2001.

Full factorial 2^2 design



$$\widehat{A} = \tfrac{1}{2}\left((50-30)+(55-40)\right) \text{ etc.}$$

setting	I	A	B	AB
$(-1,-1) = (1)$	$+$	$-$	$-$	$+$
$(1,-1) = a$	$+$	$+$	$-$	$-$
$(-1,1) = b$	$+$	$-$	$+$	$-$
$(1,1) = ab$	$+$	$+$	$+$	$+$

2^{3-1} factorial design

setting	I	A	B	C	AB	AC	BC	ABC
a	+	+	−	−	+	−	+	+
b	+	−	+	−	−	+	−	+
c	+	−	−	+	−	−	−	+
abc	+	+	+	+	+	+	+	+

Aliasing relations (Abelian group theory: Fisher, Ann. Eugenics 11 (1942))

$$
\begin{aligned}
I &= ABC \\
A &= BC \\
B &= AC \\
C &= AB
\end{aligned}
$$

Limitations group theoretic approach

- interpretation aliasing relations + computational rules
- only works for "regular" fractions
  - Plackett-Burman designs (main effect designs)
  - unsuccessful runs of experimental design (cf. Holliday et al., Comp. Stat. 14 (1999))
  - ...
- how about more than 2 levels (including mixed number of levels)?

full factorial 2^3

$$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_{12} x_1 x_2 + \theta_{13} x_1 x_3 + \theta_{23} x_2 x_3 + \theta_{123} x_1 x_2 x_3$$

nonregular fraction $2^3 \setminus \{(0,0,0),(1,1,0)\}$

$$y = ??$$

Pistone-Wynn approach in a nutshell

- designs are finite subsets of $\mathbb{R}^d$
- view finite subsets of $\mathbb{R}^d$ as solutions of polynomial equations
- write models as polynomial functions on the design points
- equivalence class of models (identifiability) through polynomials that vanish on the design points (*ideals in ring of polynomials*)
- compute minimal representations of ideals (Gröbner bases)
- build estimable models from Gröbner bases

Estimability and identifiability are translated to operations with polynomials and zeros of polynomials (algebraic geometry).

Least squares estimation can also be put in polynomial context (Cohen et al., mODa 6 proceedings (2001), 37–44).

Polynomial approach to Least Squares

A design $\mathcal{D}$ is a finite set of points $\mathbb{R}^d$.

$\mathcal{L}(\mathcal{D})$ is the vector space of all functions $\mathcal{D} \mapsto \mathbb{R}$. Any element of $\mathcal{L}(\mathcal{D})$ can be represented as a polynomial (interpolation!).

If $f, g \in \mathcal{L}(\mathcal{D})$, then an inner product is defined by

$$\langle f \, , \, g \rangle_{\mathcal{D}} := \sum_{a \in \mathcal{D}} f(a) g(a)$$

A norm is defined on $\mathcal{L}(\mathcal{D})$ by

$$\| f \|_{\mathcal{D}} = \sqrt{\langle f \, , \, f \rangle_{\mathcal{D}}}$$

Least Squares and Inner Product

$$Y(x) = f(x, \theta) + \varepsilon(x)$$

Observations $Y_1, \ldots, Y_n$ where Y_i taken at design point a_i.
Polynomial g interpolates data: $g(a_i) = y_i$, $i = 2, \ldots, n$

$$\begin{aligned}\widehat{\theta} &= \min_{\theta \in \Theta} \sum_{i=1}^{n} \mid Y_i - f(a_i, \theta) \mid^2 \\ &= \min_{\theta \in \Theta} \langle g - f(.,\theta)\,,\ g - f(.,\theta)\rangle_{\mathcal{D}} \\ &= \min_{\theta \in \Theta} \|g - f(.,\theta)\|_{\mathcal{D}}^2\end{aligned}$$

Hence, $\widehat{\theta}$ is related to orthogonal projection in $(\mathcal{L}(\mathcal{D}), \langle .\,,\ .\rangle_{\mathcal{D}})$.

$$Y = X^t\theta + \varepsilon \Rightarrow \widehat{\theta} = (X^tX)^{-1}X^tY$$

Is there an analogue in terms of the inner product $\langle .\,,\ .\rangle_{\mathcal{D}}$?

Least Squares and Orthonormal Expansions

Restrict to linear models:

$$\begin{aligned} Y(x) &= \sum_{\alpha \in \mathcal{M}} \theta_\alpha \, p_\alpha(x) + \varepsilon(x) \\ Y &= X^t \theta + \varepsilon \end{aligned}$$

where the design matrix X is given by $\begin{pmatrix} 1 & p_{\alpha_1}(a_1) & p_{\alpha_2}(a_1) & \ldots \\ 1 & p_{\alpha_1}(a_2) & p_{\alpha_2}(a_2) & \ldots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$

$\sum_{\alpha \in \mathcal{M}} \widehat{\theta}_\alpha \, p_\alpha$ is the orthogonal projection of g (interpolator of data) onto span$\{p_\alpha \mid \alpha \in \mathcal{M}\}$.

If $\{p_\alpha \mid \alpha \in \mathcal{M}\}$ is an orthogonal subset of $(\mathcal{L}(\mathcal{D}), \langle . \, , \, . \rangle_{\mathcal{D}})$, then

$$\widehat{\theta}_\alpha = \frac{\langle g \, , \, p_\alpha \rangle_{\mathcal{D}}}{\langle p_\alpha \, , \, p_\alpha \rangle_{\mathcal{D}}}.$$

Confounding and identifiability

A linear model is identifiable by a design $\mathcal{D}$ if the functions p_α ($\alpha \in \mathcal{M}$) are linearly independent elements of $\mathcal{L}(\mathcal{D})$.

$$Y = X\theta + \varepsilon$$

If the design matrix X is not of full rank, then θ is not identifiable since different values of θ yield the same value of $X\theta$. This actually means that the model coincides for different parameter values when restricted to the design points. In other words, the functions on $\mathcal{D}$ that take as values the components of the columns of X are linearly dependent.

Further development of this idea in Galetto et al., J. Stat. Plann. Inf. 117 (2003), 345–363.

2^{3-1} factorial design revisited

$I = ABC$ really means $1 = x_A x_B x_C$ *when restricted to* $\mathcal{D}$.

$x = \pm 1 \Rightarrow x = 1/x$ yields the other aliasing relations like $A = BC$ etc.

setting	I	A	B	C	AB	AC	BC	ABC
$a = (1, -1, -1)$	$+$	$+$	$-$	$-$	$+$	$-$	$+$	$+$
$b = (-1, 1, -1)$	$+$	$-$	$+$	$-$	$-$	$+$	$-$	$+$
$c = (-1, -1, 1)$	$+$	$-$	$-$	$+$	$-$	$-$	$-$	$+$
$abc = (1, 1, 1,)$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$

Assume identifiable model $y = \theta_I + \theta_A x_A + \theta_B x_B + \theta_C x_C$.

$\langle g \, , \, x_A \rangle_{\mathcal{D}} = Y(a) x_A|_{(1,-1,-1)} + Y(b) x_A|_{(-1,1,-1)} + Y(c) x_A|_{(1,1,-1)} + Y(abc) x_A \,|_{(1,1,1)}$

This is the reason behind coding levels as -1 and $+1$.

Further applications of polynomial approach to Least Squares

The polynomial approach also yields :

- formulas of covariance of estimators $\widehat{\theta}_\alpha$
- simple proof of Gauss-Markov theorem
- unbiasedness of estimator for the variance
- symbolic orthonormalisation through Gram-Schmidt procedure on polynomials
- interpretation of sums of squares in terms of norms of monomials

Contrasts and orthogonality is treated in a polynomial way in Fontana et al., J. Stat. Plann. Inf. 87 (2000), 149–172.

Sums of squares

$$\widehat{\theta}_I = \frac{\langle g\,,\,1\rangle_{\mathcal{D}}}{\|1\|_{\mathcal{D}}} = \frac{\sum_{i=1}^n Y_i}{n} = \overline{Y}$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n \left(\widehat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^n \left(Y_i - \widehat{Y}_i\right)^2 .$$

g interpolates data: $g(a_i) = Y_i$

$$\begin{aligned} g - P_1 g &= P_{\mathcal{M}}(g - P_1 g) + P_{\mathcal{M}^\perp}(g - P_1 g) \\ &= (P_{\mathcal{M}} g - P_1 g) + (I - P_{\mathcal{M}})(g - P_1 g) \\ &= (P_{\mathcal{M}} g - P_1 g) + (g - P_{\mathcal{M}} g) \end{aligned}$$

$\widehat{Y}_i$ corresponds to $P_{\mathcal{M}} g$ evaluated at the ith design point.

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \|g - P_1 g\|_{\mathcal{D}}^2 \text{. etc.}$$

Back to Wynn-Pistone approach

Direct problem: given design $\mathcal{D}$, how to find identifiable models $\sum_{\alpha \in \mathcal{M}} \theta_\alpha x^\alpha$.

$x^\alpha = x_1^{\alpha_1} \ldots x_d^{\alpha_d}$ where $\alpha_i \in \mathbb{Z}_{\geq 0}$

Full computational answer in terms of Gröbner bases (details follow). Designs are interpreted in algebraic geometric fashion.

Inverse problem: given model $\sum_{\alpha \in \mathcal{M}} \theta_\alpha x^\alpha$, find all designs $\mathcal{D}$ for which this model is identifiable.

Work in progress by Robbiano and Caboara (latest paper ISSAC 2001 conference).

Designs as variety

Full factorial 2^2 design with 0 and 1 coding: $\{(0,0),(0,1),(1,0),(1,1)\}$.
These points are the solutions of the polynomial equations

$$\begin{cases} x_1^2 - x_1 &= 0 \\ x_2^2 - x_2 &= 0 \end{cases}$$

An (affine) variety is the solution set of a system of polynomial equations.
For arbitrary polynomials s_1, s_2, we have

$$s_1(x_1^2 - x_1) + s_2(x_2^2 - x_2) = 0 \text{ on } \mathcal{D}$$

Polynomials that vanish on $\mathcal{D}$ are polynomial ideal $I(\mathcal{D})$:

1. for all $f, g \in I(\mathcal{D})$ then $f + g \in I(\mathcal{D})$,
2. for all $f \in I(\mathcal{D})$ and h polynomial, then $hf \in I(\mathcal{D})$.

Polynomial ideals and confounding

If f, g are polynomials such that $f = g$ on $\mathcal{D}$, then $f \sim g$.

A polynomial ideal defines an equivalence relation on the set of polynomials:

$$f \sim g \Longleftrightarrow f - g \in I(\mathcal{D})$$

$$e.g., \quad f = x_2(x_1^2 - x_1) \quad \text{and} \quad g = x_1(x_2^2 - x_2)$$

Identifiable models are elements of quotient space $\mathcal{L}(\mathcal{D})/I$.

$$\langle f_1, \ldots, f_v \rangle = \left\{ \sum_{i=1}^{v} f_i s_i : s_i \in k[x_1, \ldots, x_d] \right\}$$

This is the ideal generated by $(f_1 \ldots, f_v)$.

- polynomial ideals are finitely generated (Hilbert basis theorem)
- Gröbner bases are bases (in algebraic sense!) with suitable properties

Term ordering

Term ordering τ is a relation on $\mathbb{Z}_+^d$ (monomials) such that for all α, β, γ $(x^\alpha, x^\beta, x^\gamma)$

1. $\alpha \succ_\tau \alpha$ (reflexive)
2. if $\alpha \succ \beta$ and $\beta \succ \gamma$ then $\alpha \succ \gamma$ (transitive)
3. $\alpha \succ \beta$ and $\beta \succ \alpha$ then $\alpha = \beta$ (antisymmetric)
4. either $\alpha \succ \beta$ or $\beta \succ \alpha$ or $\alpha = \beta$ (total ordering)
5. every subset of $\mathbb{Z}_+^d$ has a smallest element (well-ordering)

5'. $1 \prec x^\alpha$ for all $\alpha \neq (0, \ldots, 0)$

5". there is not an infinite decreasing sequence

6. if $\alpha \succ \beta$ and $\gamma \in \mathbb{Z}_+^d$ then $\alpha + \gamma \succ \beta + \gamma$ (compatible with simplification)

Examples of term orderings

(a) The *lexicographic order* : $a \succ_{\text{lex}} b$ (or $x^a \succ_{\text{lex}} x^b$) if the first nonzero entry from the left in $a - b$ is positive.

$x_1x_2^2x_3 \succ_{\text{lex}} x_1x_2x_3^2$; $x_1^2 \succ_{\text{lex}} x_1x_2x_3^2$

(b) The *lexicographic total degree order* or *graded lex order*: $a \succ_{\text{grlex}} b$ (or $x^a \succ_{\text{grlex}} x^b$) if $\deg x^a \succ \deg x^b$ or $\deg a = \deg b$ and $a \succ_{\text{lex}} b$. In words: graded lex order orders by total degree first and breaks ties using lex order.

$x_1x_2^2x_3 \succ_{\text{grlex}} x_1x_2x_3^2; \quad x_1x_2x_3^2 \succ_{\text{grlex}} x_1^2$

(c) The *graded reverse lexicographic order*: $a \succ_{\text{lex}} b$ (or $x^a \succ_{\text{lex}} x^b$) if and only if $\deg x^a \succ \deg x^b$ or $\deg x^a = \deg x^b$ and the right-most nonzero entry in $a - b$ is negative.

$x_1x_2x_3^2 \succ_{\text{degrevlex}} x_1x_2^2x_3; \quad x_1x_2x_3^2 \succ_{\text{degrevlex}} x_1^2$

For connection with regression strategies, see Giglio et al. , J. Appl. Stat. 7 (2000), 923–938.

Leading terms

The leading term of a monomial w.r.t. a certain term ordering is the largest element w.r.t. that term ordering.

$f = x_1^2 - x_1$
$\mathrm{LT}(f) = x_1^2$ for all τ because
$x_1^2 \succ x_1 \iff x_1 \succ 1$

$g = x_1x_2^2x_3 + x_1x_2x_3^2 + x_1^2$
$\mathrm{LT}_{\texttt{grlex}(x_2 \succ x_1 \succ x_3)}(g) = x_1x_2^2x_3$
$\mathrm{LT}_{\texttt{lex}(x_1 \succ x_2 \succ x_3)}(g) = x_1^2$
$\mathrm{LT}_{\texttt{degrevlex}(x_1 \succ x_2 \succ x_3)}(g) = x_1x_2x_3^2$

Gröbner bases

A finite subset G of a polynomial ideal I is a *Gröbner basis* of I with respect to a term-ordering τ if and only if

$$\langle \mathrm{LT}_\tau(f) : f \in I \rangle = \langle \mathrm{LT}_\tau(g) : g \in G \rangle$$

Gröbner bases are not unique:
both $\{x_2^2 - x_2x_1, x_1^2\}$ and $\{x_2^2 - x_1x_2 + x_1^2, x_1^2\}$ are Gröbner bases of the same ideal with respect to `degrevlex`($x_2 \succ x_1$).

Some properties of Gröbner bases

- Every nonzero ideal has Gröbner bases, and a Gröbner basis is a basis in the algebraic sense.
- Let G be a Gröbner basis. For $f \in \mathbb{Q}[x]$ there exist a unique r such that

$$f = \sum_{g \in G} s_g g + r$$

  such that $\mathrm{LT}(r)$ is not divisible by $\mathrm{LT}(g)$ for all $g \in G$.
- Gröbner bases are computable (Buchberger algorithm); available in software (Maple, Mathematica, CoCoA, Singular, ...).
- Convenient representation of the quotient space.

The quotient space

Let D be a design

$$\mathbb{Q}[x]/\text{Ideal}(D) = \{[f] : f \in \mathbb{Q}[x] \text{ and } [f] = \{g : f - g \in \text{Ideal}(D)\}\ \}$$

- Vector space
- Finite
- Given τ,and Gröbner basis G

$$\begin{aligned} \text{Est} &= \{x^\alpha : x^\alpha \text{ is not divisible by } \text{LT}(g),\ \ g \in G\} \\ &= \{x^\alpha : \alpha \in \mathcal{M}\} \end{aligned}$$

The 2^{3-1} fractional factorial design

A	B	C			
1	1	1	ABC	$=$	I
1	−1	−1	BC	$=$	A
−1	1	−1	AC	$=$	B
−1	−1	1	AB	$=$	C

$$\begin{cases} a^2-1, & b^2-1, & c^2-1 \\ bc-a, & ac-b, & ab-c \end{cases}$$

For `degrevlex`

$$\mathrm{LT} = \{a^2, \quad b^2, \quad c^2, \quad ab, \quad ac, \quad bc\}$$

and

$$\mathrm{Est} = \{1, \quad a, \quad b, \quad c\}$$

Any function over 2^{3-1} is represented as linear combinations of those four monomials.

Replicates

Replicates cause problems because interpolation is no longer possible.
Work-around: introduce dummy variable as counting label
Consider the standard 2^{3-1} design with generator $I = ABC$ and 4 additional centre points. *I.e.*, the design points $\mathcal{D}$ are

$$\{(1,-1,-1,1),(-1,1,-1,1),(-1,-1,1,1),(1,1,1,1)\,, \\ (0,0,0,1),(0,0,0,2),(0,0,0,3),(0,0,0,4)\}\,.$$

Using the `degrevlex` with elimination of the counting variable t, we obtain

$$Est(\mathcal{D}) = \left\{1, x_1, x_2, x_3, x_3^2, t, t^2, t^3\right\}.$$

An orthonormal basis for the linear span of the terms not involving t is given by

$$\left\{\frac{1}{\sqrt{8}}, \frac{x_1}{2}, \frac{x_2}{2}, \frac{x_3}{2}, \frac{x_3^2 - \frac{1}{2}}{\sqrt{2}}\right\}.$$

Mixture designs

In chemical experiments concentrations add up to 100%. Thus we have the extra constraint $x_1 + \ldots + x_d = 1$.

Example:

x_1	x_2	x_3
1	0	0
0	1	0
0	0	1
$\frac{1}{2}$	$\frac{1}{2}$	0
0	$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	0	$\frac{1}{2}$
$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

For `degrevlex` we obtain Est $= \{1, x_1, x_2, x_1^2, x_1x_2, x_2^2, x_3^2\}$.
Note that x_3 is missing because $x_3 = 1 - x_1 - x_2$.

More information: Giglio et al., in: Optimum design 2000, 33–44.