Circular designs balanced for neighbours at distances one and two

R. A. Bailey
University of St Andrews / QMUL (emerita)
rab@mcs.st-and.ac.uk, r.a.bailey@qmul.ac.uk

Queen Mary, University of London / Southampton
joint Statistics seminar, 2 May 2013

Ongoing joint work with Tank Aldred (University of Otago, New Zealand), Brendan McKay (ANU, Australia) and Ian Wanless (Monash University, Australia)

Abstract

We consider experiments where the experimental units are arranged in a circle or in a single line in space or time.

If neighbouring treatments may affect the response on an experimental unit, then we need a model which includes the effects of direct treatments, left neighbours and right neighbours. It is desirable that each ordered pair of treatments occurs just once as neighbours and just once with a single unit in between. A circular design with this property is equivalent to a special type of quasigroup.

In one variant of this, self-neighbours are forbidden. In a further variant, it is assumed that the left-neighbour effect is the same as the right-neighbour effect, so all that is needed is that each unordered pair of treatments occurs just once as neighbours and just once with a single unit in between.

I shall report progress on finding methods of constructing the three types of design.

An experiment on sunflowers

Sunflowers are traditionally very tall plants.
When new, short-stalked varieties were introduced, agricultural research stations wanted to do experiments to compare the new varieties with the old.

Problem: If we grow each variety in a separate field, then any perceived differences may be caused by differences in fertility between the fields.

Problem: If we grow the varieties mixed up in the same field, with several plots per variety, then each tall variety may shade the variety growing on the plot to its immediate North.

Solution: Use a neighbour-balanced design in which each ordered pair (i,j) of different varieties occurs the same number of times as (South, North) neighbours.

Two designs for four varieties of sunflower

A	*B*	*C*	*D*
D	*A*	*B*	*C*
B	*C*	*D*	*A*
C	*D*	*A*	*B*
C	*D*	*A*	*B*

The Southern row consists of treated border plots on which no response is measured.

Each variety has each variety (including itself) just once as a Southern neighbour.

A	*B*	*C*	*D*
D	*A*	*B*	*C*
B	*C*	*D*	*A*
C	*D*	*A*	*B*
X	*X*	*X*	*X*

The Southern row is simply whatever is at the edge of the field.

Each variety has other each variety, and the field edge, just once as a Southern neighbour.

An experiment on control of aphids

Entomologists wanted to compare several sprays to deter aphids from the crop without killing them.
The sprays should be applied to a square array of rectangular plots in a single field, using a Latin square (each spray occurs on one plot per row and one plot per column).

Problem: If one spray is effective, it may actually increase the number of aphids on neighbouring plots.
The aphids are as likely to spread East as West, so direction in one dimension is not an issue, but the North–South effect may be different from the East–West one, because the plots are not square.

Solution: Use a quasi-complete Latin square, in which each unordered pair $\{i,j\}$ of sprays occurs the same number of times as neighbours within rows and the same number of times as neighbours within columns.

Five sprays on aphids

P	*X*	*D*	*G*	*M*
X	*G*	*P*	*M*	*D*
D	*P*	*M*	*X*	*G*
G	*M*	*X*	*D*	*P*
M	*D*	*G*	*P*	*X*

Each pair of different treatments occurs twice as row neighbours and twice as column neighbours.

Unequal replication (X denotes 'control')

X	P	D	M	G	X
M	X	P	G	X	D
D	G	M	P	X	X
G	D	X	X	M	P
X	M	X	D	P	G
P	X	G	X	D	M

The experiment at Rothamsted on control of aphids

Some issues in neighbour designs

- Border plots or not?
- Self-neighbours or not?
- One or two dimensions?
- Neighbour effects in one or both directions in each dimension?
- Left-neighbour effects the same as right-neighbour effects?
- Equal replication or not?
- Do we want to estimate direct effects (tasting coffee) or neighbour effects (the neighbour design reduces bias, and over-fitting reduces error degrees of freedom) or total effects (what happens when the farmer uses only the short variety)?

An experiment in marine biology

Richard Cormack (St Andrews) posed me this question in 1993.

A marine biologist wanted to compare 5 genotypes of bryozoan by suspending them in sea water around the circumference of a cylindrical tank. Each genotype was replicated 5 times, so that altogether 25 items were suspended in the tank.

The marine biologist required that

(i) each ordered pair of items should occur just once as ordered neighbours around the circumference of the tank;

(ii) each ordered pair of items should occur just once with a single item in between them, in order.

A circular design for 5 treatments with neighbour balance at distances one and two

1 1 3 4 3 0 0 1 0 2 2 0 3 3 1 2 1 4 0 4 4 2 3 2 4

The lazy way to write the design

(1 1 3 4 3 0 0 1 0 2 2 0 3 3 1 2 1 4 0 4 4 2 3 2 4)

1 1 3 4 3 0 0 1 0 2 2 0 3 3 1 2 1 4 0 4 4 2 3 2 4

Statistical model

Denote by $\tau(i)$ the treatment on plot i.

Denote by Y_i the response on plot i.

$$Y_i = \lambda_{\tau(i-1)} + \delta_{\tau(i)} + \rho_{\tau(i+1)} + \varepsilon_i$$

where the ε_i are independent random variables with mean 0 and common variance σ^2.

The direct treatment effects δ,
the left neighbour effects λ
and the right neighbour effects ρ
can be estimated orthogonally of each other
in a experiment of this size
if and only if each pair (λ_j, δ_k) occurs equally often
and each pair (δ_j, ρ_k) occurs equally often
and each pair (λ_j, ρ_k) occurs equally often;
in other words, the design has neighbour balance at distances one and two.

Those conditions again

Among the triples of the form

$$(\tau(i-1), \tau(i), \tau(i+1)),$$

each ordered pair of treatments occurs once in positions 1 and 2, once in positions 1 and 3, and once in positions 2 and 3.

Among the triples of the form

$$(\text{row}, \text{column}, \text{symbol}),$$

each ordered pair of symbols occurs once in positions 1 and 2, once in positions 1 and 3, and once in positions 2 and 3.

These are conditions for a Latin square
whose rows and columns have the same labels as the symbols
—a quasigroup.

Building the design from a quasigroup (Latin square)

The quasigroup operation $\circ$ is defined by

$a \circ b =$ symbol in row a and column b of the Latin square.

In the circular design, each triple should have the form

$$(a, b, a \circ b).$$

We can start with any ordered pair (x, y) and successively build the circular design from the quasigroup as

$$x \quad y \quad x \circ y \quad y \circ (x \circ y) \quad (x \circ y) \circ (y \circ (x \circ y)) \quad \cdots$$

Latin square to circle

$\circ$	A	B	C	D
A	B	A	D	C
B	C	D	A	B
C	D	C	B	A
D	A	B	C	D

$(\; A \quad A \quad B \quad A \quad C \quad D \quad A \quad A$ oops!

This quasigroup gives a design with four separate circles, not one.

$(\; A \quad A \quad B \quad A \quad C \quad D \;)$

$(\; A \quad D \quad C \quad C \quad B \quad C \;)$

$(\; B \quad B \quad D \;)$

$(\; D \;)$

Eulerian quasigroups

Let's call a quasigroup Eulerian if it gives a single large circle: that is, a sequence with maximal period.

	0	1	2	3	4
0	1	0	2	3	4
1	2	3	1	4	0
2	3	4	0	2	1
3	0	2	4	1	3
4	4	1	3	0	2

(1 1 3 4 3 0 0 1 0 2 2 0 3 3 1 2 1 4 0 4 4 2 3 2 4)

Do Eulerian quasigroups of order n exist?

If $n \leq 4$, a manual check shows that there are none.

For $n = 5$, we have shown an example.

For every other value of n that we have tried,
we have found an Eulerian quasigroup by computer search;
and we can prove that existence for coprime n and m implies existence for mn;
BUT we have been unable to prove that they always exist.

Exercise

Show that, if $Q = \mathbb{Z}_{p^s}$ or $Q = \mathrm{GF}(p^s)$, then no binary operation of the form

$$x \circ y = ax + by + c$$

makes Q into an Eulerian quasigroup.

Variant I: no self-neighbours

Sometimes it is undesirable to have the same treatment on neighbouring plots.

We need a circular design with $n(n-1)$ plots in which each

- each ordered pair of distinct treatments occurs just once as ordered neighbours.
- each left-neighbour treatment occurs just once with all but one of the right-neighbour treatments.

The incidence of
direct treatments with left-neighbour treatments is a symmetric balanced incomplete-block design (BIBD, aka 2-design);
direct treatments with right-neighbour treatments is a symmetric BIBD;
left-neighbour treatments with right-neighbour treatments is a symmetric BIBD.

Preece (1976) showed that, for overall balance, the missing pairs at distance two must also be the self-pairs.

Idempotent Eulerian circular sequences

We need a circular design with $n(n-1)$ plots in which each

- each ordered pair of distinct treatments occurs just once as ordered neighbours.
- each left-neighbour treatment occurs just once with every right-neighbour treatment except itself.

The results of Druilhet (1999) show that such designs are optimal for the estimation of direct effects and neighbour effects, in the sense of minimizing average variance of these estimators.

A quasigroup is idempotent if $x \circ x = x$ for all x.

Our circular design is equivalent to an idempotent quasigroup in which the $n(n-1)$ off-diagonal cells give a single circle.

Construction when $n = 6$ (in general, when n is even)

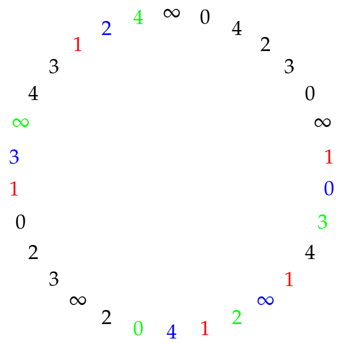
The treatments are the integers modulo 5, together with ∞.

sequence	$[4,3,1,2]$	all different, non-zero
neighbour sums	$[2,4,3]$	all different, non-zero, non-1
sum of ends	1	must be 1
cumulative sums	$[0,4,2,3,0]$	

(∞ 0 4 2 3 0 ∞ 1 0 3 4 1 ∞ 2 1 4 0 2 ∞ 3 2 0 1 3 ∞ 4 3 1 2 4)

Neighbours of ∞ at distances one and two are OK, by cyclic construction.
Differences at distance one come from the original sequence;
most differences at distance two are the neighbour sums.
$1 -$ last cumulative sum $= 1 - 0 = 1 =$ missing neighbour-sum
so differences at distance two either side of ∞ give this.

A circular design for 6 treatments with no self-neighbours at distance one or two



(∞ 0 4 2 3 0 ∞ 1 0 3 4 1 ∞ 2 1 4 0 2 ∞ 3 2 0 1 3 ∞ 4 3 1 2 4)

Construction when $n = 7$ (in general, when n is odd)

The treatments are the integers modulo 6, together with ∞.

sequence	$[4,1,2,5,3]$	all different, non-zero
neighbour sums	$[5,3,1,2]$	all different, non-zero, non-4
sum of ends	1	must be 1
cumulative sums	$[0,4,5,1,0,3]$	

(∞ 0 4 5 1 0 3 ∞ 1 5 0 2 1 4 ∞ 2 0 1 3 2 5
... ∞ 3 1 2 4 3 0 ∞ 4 2 3 5 4 1 ∞ 5 3 4 0 5 2)

Neighbours of ∞ at distances one and two are OK, by cyclic construction.
Differences at distance one come from the original sequence;
most differences at distance two are the neighbour sums.
$1 -$ last cumulative sum $= 1 - 3 = 4 =$ missing neighbour-sum
so differences at distance two either side of ∞ give this.

Solution for variant I

Theorem
Given an initial sequence of the non-zero integers modulo $n-1$ satisfying those conditions, that construction always produces an idempotent Eulerian circular sequence.

Theorem
Such an initial sequence can be constructed whenever $n \geq 6$.

Variant II: undirectional neighbour effects

Suppose that the effect of the neighbouring treatment is the same whether it is from the left or the right.

$$Y_i = \lambda_{\tau(i-1)} + \delta_{\tau(i)} + \lambda_{\tau(i+1)} + \varepsilon_i,$$

where the ε_i are independent random variables with mean 0 and common variance σ^2.

Can we arrange that every treatment have every treatment as neighbour just once, on one side or the other?

A self-pair gives a self-neighbour on both sides, so we must ban self-pairs. So we need a circle of $n(n-1)/2$ plots.

Each plot has two neighbours, so each treatment has an even number of neighbours, so $n-1$ must be even.

Any triple (a, b, a) gives b as a neighbour of a on both sides, so there can be no such triples.

Model and variance

$$Y_i = \lambda_{\tau(i-1)} + \delta_{\tau(i)} + \lambda_{\tau(i+1)} + \varepsilon_i,$$

In vector form

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\delta} + \mathbf{X}_2\boldsymbol{\lambda} + \boldsymbol{\varepsilon}.$$

The neighbour conditions in the design imply that each treatment occurs r times, where $r = (n-1)/2$.
Also $\mathbf{X}_1^\top\mathbf{X}_1 = r\mathbf{I}$, $\mathbf{X}_1^\top\mathbf{X}_2 = \mathbf{J} - \mathbf{I}$ and $\mathbf{X}_2^\top\mathbf{X}_2 = 2r\mathbf{I} + (\mathbf{J} - \mathbf{I})$,
where $\mathbf{I}$ is the $n \times n$ identity matrix
and $\mathbf{J}$ is the $n \times n$ all-1 matrix.

Some calculations show that the variance of the estimator of the difference between two direct effects is

$$\frac{2(2r-1)}{(r-1)(2r+1)}\sigma^2$$

while that for the difference between two neighbour effects is

$$\frac{2r}{(r-1)(2r+1)}\sigma^2.$$

Construction when $n = 9$

The treatments are the integers modulo 9.

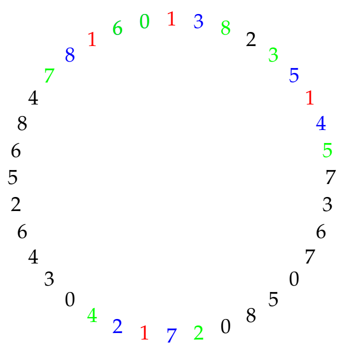
circular sequence	$(1, 2, 5, 3)$	$\pm$ entries are all different
circular neighbour sums	$(3, 7, 8, 4)$	$\pm$ entries are all different
cumulative sums	$[1, 3, 8, 2]$	last one is coprime to 9

(1 3 8 2 3 5 1 4 5 7 3 6 7 0 5 8 0 2 7 1 2 4 0 3 4 6 2 5 6 8 4 7 8 1 6 0)

We keep adding 2 to the original sequence of length 4.
Because 2 is coprime to 9, every pair in the original sequence gets all its shifts modulo 9.

Differences at distance one come from the original sequence; difference at distance two are the neighbour sums.

A circular design for 9 treatments with undirectional neighbour balance at distances one and two



(1 3 8 2 3 5 1 4 5 7 3 6 7 0 5 8 0 2 7 1 2 4 0 3 4 6 2 5 6 8 4 7 8 1 6 0)

Solution for variant II

Theorem
Given an initial circular sequence of $(n-1)/2$ of the integers modulo n satisfying those conditions, that construction always produces a circular sequence balanced for undirected neighbours at distances one and two.

Theorem
Such an initial sequence can be constructed whenever n is odd and $n \geq 9$. There is also such a circular sequence when $n = 7$.

Back to the original question

A quasigroup of order n with operation $\circ$ is Eulerian if the sequence

$$x \quad y \quad x \circ y \quad y \circ (x \circ y) \quad (x \circ y) \circ (y \circ (x \circ y)) \quad \cdots$$

does not repeat before n^2 steps.

Conjecture

If $n \geq 5$ then there exists an Eulerian quasigroup of order n.

Coprime sizes

Theorem

If Q_1 and Q_2 are Eulerian quasigroups of orders n and m, where n and m are coprime, then $Q_1 \otimes Q_2$ is an Eulerian quasigroup of order nm.

Proof.

In the sequence

$$(a, x) \quad (b, y) \quad (a \Box b, x \circ y) \quad (b \Box (a \Box b), y \circ (x \circ y)) \quad \cdots$$

the first coordinates repeat every n^2 steps, but not earlier, and the second coordinates repeat every m^2 steps, but not earlier. □

Strategy

So all we have to do is to find an Eulerian quasigroup for all of the following orders:

- q where q is an odd prime power and $q \geq 5$
- $3q$ where q is an odd prime power
- $2q$ where q is an odd prime power
- $4q$ where q is an odd prime power
- powers of 2 bigger than 4.

Homework

If $Q = \mathbb{Z}_{p^s}$ or $Q = \mathrm{GF}(p^s)$, then no binary operation of the form

$$x \circ y = ax + by + c$$

makes Q into an Eulerian quasigroup.

Technique

If q is odd, try taking $Q = \mathbb{Z}_q$ and putting

$$x \circ y = \pi(x + y)$$

where π is a relatively simple permutation.

For example, when $q = 7$ put $\pi = (0\,1\,2)(3\,4)$ so that

$$4 \circ 5 = \pi(4 + 5) = \pi(2) = 0.$$

Obstacle

Theorem

If n is even then no Eulerian quasigroup can be obtained from a group of order n by permutions of rows, columns or symbols.

. . . for all practical purposes

Theorem

If $n \geq 5$ and there is no Eulerian quasigroup of order n then n is divisible by a prime power exceeding 500.

References I

- Aldred, R. E. L., Bailey, R. A., McKay, B. D. and Wanless, I.M.:
  Circular designs balanced for neighbours at distances one and two.
  Preprint, 2010.
- Azaïs, J.-M., Bailey, R. A. and Monod, H.:
  A catalogue of efficient neighbour-designs with border plots.
  Biometrics, **49** (1993), 1252–1261.
- Bailey, R. A.:
  Quasi-complete Latin squares: construction and randomization.
  Journal of the Royal Statistical Society, Series B **46** (1984), 323–334.

References II

- Bayer, M. M. and Todd, C. D.:
  Effect of polypide regression and other parameters on colony growth in the cheilostomate Electra pilosa (L.).
  In *Bryozoans in Space and Time* (eds. D. P. Gordon, A. M. Smith and J. A. Grant-Mackie), pp. 29–38.
  Wellington, NZ: National Institute of Water and Atmospheric Research (1996).
- Druilhet, P.:
  Optimality of neighbour-balanced designs.
  Journal of Statistical Planning and Inference, **81** (1999), 141–152.
- Preece, D. A.:
  Non-orthogonal Graeco-Latin designs.
  In *Combinatorial Mathematics IV* (eds. L. R. A. Casse and W. D. Wallis), Lecture Notes in Mathematics, **560**, pp. 7–26.
  Berlin: Springer (1976).