On optimal cross-over designs when carry-over effects are proportional to direct effects

R. A. Bailey,
School of Mathematical Sciences,
Queen Mary, University of London,
Mile End Road, London, E1 4NS, U.K.
r.a.bailey@qmul.ac.uk

J. Kunert,
Fachbereich Statistik,
Universität Dortmund,
D 44221 Dortmund, Germany
kunert@statistik.uni-dortmund.de

May 11, 2005

Summary

There are a number of different models for cross-over designs which take account of carry-over effects. Since it seems plausible that a treatment with a large direct effect should generally have a larger carry-over effect, Kempton, Ferris and David (2001) considered a model where the carry-over effects are proportional to the direct effects. The advantage of this model lies in the fact that there are fewer parameters to be estimated. Its problem lies in the non-linearity of the estimates. Kempton et al. (2001) considered the least squares estimate. They point out that this estimate is asymptotically equivalent to the estimate in a linear model which assumes the true parameters to be known.

For this estimate they determine optimal designs numerically for some cases. The present paper generalizes some of their results. Our results are derived with the help of a generalization of the methods used in Kunert and Martin (2000).

Some key words: Carry-over effects; Cross-over designs; Universal optimality.

1 Introduction

In cross-over designs the experimental subjects are exposed to a series of treatments, one after the other. One important example of cross-over designs is the case of a sensory trial. Here, products are described with the human senses. Each assessor tastes and evaluates a series of products, such as the bitterness of several brands of beer. A problem with this kind of experiment is the liability to have carry-over effects. This may be a lingering taste of a product that influences the perception of the next product, or it may be a tendency to give a lower rating to the next product after a very intense product. For instance, it is well known that assessors have a tendency to give a lower bitterness rating to a product if it is evaluated directly after a very bitter one.

There are a number of models for cross-over designs which take account of carry-over effects. It seems plausible from what was said above that a treatment with a large direct effect should generally have a larger carry-over effect. Kempton, Ferris and David (2001) considered a model where the carry-over effects are proportional to the direct effects. The advantage of this model lies in the fact that there are fewer parameters to be estimated. The problem lies in the non-linearity of the estimates. Kempton et al. (2001) considered the least squares estimate. They point out that this estimate is asymptotically equivalent to the estimate in a linear model which assumes the true parameters to be known. For this estimate they numerically determine optimal designs for some cases. The present paper generalizes some of their results.

We consider cross-over designs for t treatments in p periods and n subjects. We will restrict attention to the case that $p \leq t$. We assume that there is no carry-over effect in the first period, while there are carry-over effects in later periods which are proportional to the direct effect of the treatment in the previous period. A design $d \in \Omega_{t,n,p}$ is a mapping of $\{1, \ldots, n\} \times \{1, \ldots, p\}$ to $\{1, \ldots, t\}$, which determines the treatment assigned to subject i in period j. Let T_d and F_d be the plots-by-treatments incidence matrices for direct and carry-over effects in design d respectively, and $U = I_n \otimes u_p$ be the plots-by-subjects incidence matrix and $P = u_n \otimes I_p$ be the plots-by-periods incidence matrix. Here I_s is the identity matrix of size s, while u_s is the s-vector of ones.

Let y_{ij} be the measurement on subject i in period j, and

$$y = [y_{11}, y_{12}, \ldots, y_{1p}, y_{21}, \ldots, y_{np}]^\top.$$

Like Kempton et al. (2001), we assume the model

$$y = T_d\tau + F_d\tau\lambda + P\alpha + U\beta + e, \tag{1}$$

where e is a vector of independent identically distributed errors with zero mean and variance σ^2, and we analyse the data by least squares. For this analysis, we assume that we are interested in the estimation of τ and not in the estimation of the unknown proportionality factor λ, which we restrict to lie between -1 and 1.

If, in model (1), we add a constant κ to every element of τ, then we increase every response in the first period by κ and every other response by $\kappa(1+\lambda)$. Since model (1) includes period effects, we may suppose that the elements in τ sum to zero.

As was observed by Kempton et al. (2001), the least squares solution $(\hat{\tau}, \hat{\lambda})$ is asymptotically equivalent to the least squares solution of the linear model

$$\tilde{y} = (T_d + \lambda_0 F_d)\tau + (F_d\tau_0)\lambda + P\alpha + U\beta + e, \tag{2}$$

where τ_0 is the unknown true value of τ, λ_0 is the unknown true value of λ, and $\tilde{y} = y + \lambda_0 F_d \tau_0$.

Define

$$\mathcal{C}_d = (T_d + \lambda_0 F_d)^\top \omega^\perp([F_d\tau_0, P, U])(T_d + \lambda_0 F_d),$$

where for a matrix M we define $\omega(M) = M(M^\top M)^- M^\top$ and $\omega^\perp(M) = I - \omega(M)$. Since $\lambda_0 F_d \tau_0$ is a constant vector, assumed known, and the elements in τ sum to zero, the covariance matrix $\mathrm{cov}(\hat{\tau})$ of the estimate $\hat{\tau}$ in model (2) is $\sigma^2 \mathcal{C}_d^+$, where M^+ is the Moore–Penrose generalized inverse of the matrix M. Hence, the covariance of the estimates depends on the unknown true values τ_0 and λ_0.

A class of designs which have excellent optimality properties in cross-over experiments under various aspects are the totally balanced designs (Kunert and Stufken, 2002). These are defined as follows.

Definition 1 A design $d \in \Omega_{t,n,p}$ is called *totally balanced* if

(i) d is a generalized Youden design,

(ii) d is a balanced block design in the carry-over effects,

(iii) d is balanced for carry-over effects, and

(iv) the number of subjects where both treatments i and j appear $[p/t]+1$ times and treatment j does not appear in the last period is the same for every pair $i \neq j$.

2 An upper bound for the $\bar{A}$-criterion

Put $B_t = I_t - t^{-1}u_t u_t^\top$. Then

$$\mathcal{C}_d = (B_t T_d^\top + \lambda_0 B_t F_d^\top)\omega^\perp([B_t F_d \tau_0, P, U])(T_d B_t + \lambda_0 F_d B_t).$$

This is because $T_d u_t$ and $F_d u_t$ are in the column span of P: see, for example, Kunert (1983).

It follows from Kunert (1983) that

$$\mathcal{C}_d \leq (B_t T_d^\top + \lambda_0 B_t F_d^\top)\omega^\perp([B_t F_d \tau_0, U])(T_d B_t + \lambda_0 F_d B_t) = \tilde{\mathcal{C}}_d, \tag{3}$$

say. Equality holds in (3) if and only if

$$(B_t T_d^\top + \lambda_0 B_t F_d^\top)\omega^\perp([B_t F_d \tau_0, U])P = 0. \tag{4}$$

We define

$$\mathcal{C}_{d11} = T_d^\top \omega^\perp(U) T_d,\ \mathcal{C}_{d12} = T_d^T \omega^\perp(U) F_d B_t,\ \mathcal{C}_{d22} = B_t F_d^\top \omega^\perp(U) F_d B_t.$$

Note that these three matrices were called U, $W^\top$ and V by Kempton et al. (2001). Then the bound in (3) can be written as

$$\tilde{\mathcal{C}}_d = \mathcal{E}_d - (\mathcal{C}_{d12}\tau_0 + \lambda_0 \mathcal{C}_{d22}\tau_0)(\tau_0^\top \mathcal{C}_{d22}\tau_0)^{-1}(\tau_0^\top \mathcal{C}_{d12}^\top + \lambda_0 \tau_0^\top \mathcal{C}_{d22}),$$

where $\mathcal{E}_d = \mathcal{C}_{d11} + \lambda_0 \mathcal{C}_{d12} + \lambda_0 \mathcal{C}_{d12}^\top + \lambda_0^2 \mathcal{C}_{d22}$; while (4) can be written as

$$\begin{aligned} & B_t T_d^\top \omega^\perp(U) P + \lambda_0 B_t F_d^\top \omega^\perp(U) P \\ - \quad & (\mathcal{C}_{d12}\tau_0 + \lambda_0 \mathcal{C}_{d22}\tau_0)(\tau_0^\top \mathcal{C}_{d22}\tau_0)^{-1}\tau_0^\top B_t F_d^\top \omega^\perp(U) P = 0. \end{aligned}$$

Now, $\omega^\perp(U)P = P - p^{-1}u_{np}u_p^\top$, so, if the design d is such that in each period all treatments appear equally often, then $T_d^\top \omega^\perp(U) P = F_d^\top \omega^\perp(U) P = 0$. This implies that equation (4) is true and we have equality in (3). A problem is that the information matrix depends on the unknown parameters λ_0 and τ_0. Like Kempton et al. (2001), we decided to take two different approaches to deal with this problem. With respect to the parameter τ_0, we considered the average performance over a distribution of τ_0. To do this, we assume that the distribution of τ_0 is permutation invariant. Note that this assumption is valid if we randomize the treatment labels. For the parameter λ_0, there is no canonical distribution. We may, however, assume that λ_0 is small in absolute size. Again, this is justifiable by practical considerations. The experimenter will try to carry out the experiment in such a way that the carry-over effects are small. This can (and has to) be achieved by non-statistical measures like washout periods.

We are interested in determination of an $\bar{A}$-optimal design. That is, we want to determine a design that minimizes the average A-criterion, where the average is taken over the distribution of the unknown parameter τ_0. We do this in several steps.

Step 1 Consider a fixed but arbitrary τ_0 and an arbitrary design $d \in \Omega_{t,n,p}$. We assume that $\theta_{1,d} \geq \theta_{2,d} \geq \cdots \geq \theta_{t-1,d}$ are the nonnegative eigenvalues of $\mathcal{C}_d$. The local A-criterion given by Kempton et al. (2001) is $2\phi_A(\mathcal{C}_d, \tau_0)\sigma^2/(t-1)$, where

$$\phi_A(\mathcal{C}_d, \tau_0) = \sum_{i=1}^{t-1} \frac{1}{\theta_{i,d}}.$$

Because τ_0 is orthogonal to u_t, we can find an orthonormal basis $\{x_1, \ldots, x_t\}$ of $\mathbb{R}^t$ such that x_{t-1} and x_t are scalar multiples of τ_0 and u_t respectively. Then it is well known that the A-criterion satisfies

$$\phi_A(\mathcal{C}_d, \tau_0) \geq \sum_{i=1}^{t-1} \frac{1}{x_i^\top \mathcal{C}_d x_i},$$

because $[x_1^T \mathcal{C}_d x_1, \ldots, x_{t-1}^\top \mathcal{C}_d x_{t-1}]$ is majorized by $[\theta_1, \ldots, \theta_{t-1}]$, see Fan (1949, Theorem 1).

Since $\mathcal{C}_d \leq \tilde{\mathcal{C}}_d$, it follows that

$$\phi_A(\mathcal{C}_d, \tau_0) \geq \sum_{i=1}^{t-1} \frac{1}{x_i^\top \tilde{\mathcal{C}}_d x_i}.$$

The convexity of $1/x$ further implies that

$$\phi_A(\mathcal{C}_d, \tau_0) \geq \frac{(t-2)^2}{\sum_{i=1}^{t-2} x_i^\top \tilde{\mathcal{C}}_d x_i} + \frac{1}{x_{t-1}^\top \tilde{\mathcal{C}}_d x_{t-1}}.$$

Some straightforward algebra shows that

$$x_{t-1}^\top \tilde{\mathcal{C}}_d x_{t-1} = \frac{1}{\tau_0^\top \tau_0} \left(\tau_0^\top \mathcal{C}_{d11} \tau_0 - \frac{(\tau_0^\top \mathcal{C}_{d12} \tau_0)^2}{\tau_0^\top \mathcal{C}_{d22} \tau_0} \right) = \ell_{\tau_0,d},$$

say. Now,

$$\sum_{i=1}^{t-2} x_i^\top \tilde{\mathcal{C}}_d x_i \leq \sum_{i=1}^{t-2} x_i^\top \mathcal{E}_d x_i = \operatorname{tr} \mathcal{E}_d - x_{t-1}^\top \mathcal{E}_d x_{t-1},$$

because $\tilde{\mathcal{C}}_d \leq \mathcal{E}_d$, trace is invariant to change of basis and $\mathcal{E}_d x_t = 0$. Hence, defining

$$s_{\tau_0,d} = \operatorname{tr} \mathcal{E}_d - \frac{1}{\tau_0^\top \tau_0} \tau_0^\top \mathcal{E}_d \tau_0,$$

we get that

$$\phi(\mathcal{C}_d, \tau_0) \geq \frac{1}{\ell_{\tau_0,d}} + \frac{(t-2)^2}{s_{\tau_0,d}}.$$

Now assume the design d is such that all $\mathcal{C}_{dij}$, $1 \leq i \leq j \leq 2$, are completely symmetric, that is $\mathcal{C}_{dij} = c_{dij}(t-1)^{-1} B_t$, where $c_{dij} = \operatorname{tr} \mathcal{C}_{dij}$. Then the vectors $x_1, \ldots, x_t$ are eigenvectors of $\tilde{\mathcal{C}}_d$, and, for $1 \leq i \leq t-2$ we have

$$x_i^\top \tilde{\mathcal{C}}_d x_i = x_i^\top \mathcal{E}_d x_i = \frac{s_{\tau_0,d}}{t-2}.$$

The first equality is due to the fact that for those i

$$x_i^\top \mathcal{C}_{dij} \tau_0 = \frac{c_{dij}}{t-1} x_i^T B_t \tau_0 = 0.$$

The second equality is due to the fact that $x_i^\top B_t x_i$ is the same for all $1 \leq i \leq t-1$.

If the design additionally satisfies equation (4), we therefore have

$$\phi(\mathcal{C}_d, \tau_0) = \frac{1}{\ell_{\tau_0,d}} + \frac{(t-2)^2}{s_{\tau_0,d}}.$$

Note that for this design the numbers $\ell_{\tau_0,d}$ and $s_{\tau_0,d}$ do not depend on τ_0.

Step 2 We assume that the distribution of τ_0 is permutation invariant. Let $\mathcal{S}$ denote the set of all $t \times t$ permutation matrices and assume that for each $\pi \in \mathcal{S}$ we have determined $s_{\tau_\pi,d}$ and $\ell_{\tau_\pi,d}$ as in Step 1, where $\tau_\pi = \pi\tau_0$. Then, due to the to the convexity of $1/x$, we have

$$\frac{1}{t!} \sum_{\pi \in \mathcal{S}} \left(\frac{1}{\ell_{\tau_\pi,d}} + \frac{(t-2)^2}{s_{\tau_\pi,d}} \right) \geq \frac{t!}{\sum_{\pi \in \mathcal{S}} \ell_{\tau_\pi,d}} + \frac{t!\,(t-2)^2}{\sum_{\pi \in \mathcal{S}} s_{\tau_\pi,d}}.$$

If $\ell_{\tau_\pi,d}$ and $s_{\tau_\pi,d}$ are the same for all π, then we have equality.

To continue, we need to determine $\sum_{\pi \in \mathcal{S}} \ell_{\tau_\pi,d}$ and $\sum_{\pi \in \mathcal{S}} s_{\tau_\pi,d}$. For $\sum_{\pi \in \mathcal{S}} s_{\tau_\pi,d}$, this is straightforward. We have

$$\begin{aligned}
\frac{1}{t!} \sum_{\pi \in \mathcal{S}} s_{\tau_\pi,d} &= \operatorname{tr} \mathcal{E}_d - \frac{1}{t!} \sum_{\pi \in \mathcal{S}} \left(\frac{1}{\tau_0^\top \pi^\top \pi \tau_0} \tau_0^\top \pi^\top \mathcal{E}_d \pi \tau_0 \right) \\
&= \operatorname{tr} \mathcal{E}_d - \frac{1}{\tau_0^\top \tau_0} \frac{1}{t!} \tau_0^\top \left(\sum_{\pi \in \mathcal{S}} \pi^\top \mathcal{E}_d \pi \right) \tau_0
\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\mathcal{E}_d - \frac{1}{\tau_0^\top \tau_0}\tau_0^\top \left(\frac{\operatorname{tr}\mathcal{E}_d}{t-1}B_t\right)\tau_0 \\
&= \frac{t-2}{t-1}\operatorname{tr}\mathcal{E}_d = \frac{t-2}{t-1}\left(c_{d11} - 2\lambda_0 c_{d12} + \lambda_0^2 c_{d22}\right).
\end{aligned}
$$

Note that this does not depend on τ_0.

The same argument shows that

$$
\frac{1}{t!}\sum_{\pi\in\mathcal{S}} \frac{1}{\tau_0^\top \pi^\top \pi \tau_0}\tau_0^\top \pi^\top \mathcal{C}_{d11}\pi\tau_0 = \frac{\operatorname{tr}\mathcal{C}_{d11}}{t-1} = \frac{c_{d11}}{t-1}.
$$

Put $a_\pi = \tau_0^\top \pi^\top \mathcal{C}_{d12}\pi\tau_0$ and $b_\pi = \tau_0^\top \pi^\top \mathcal{C}_{d22}\pi\tau_0$. Since the a_π and b_π are positive, it follows that

$$
\begin{aligned}
\sum_{\pi\in\mathcal{S}} \frac{1}{\tau_0^\top \pi^\top \pi\tau_0}\frac{a_\pi^2}{b_\pi} &= \frac{1}{\tau_0^\top\tau_0}\sum_{\pi\in\mathcal{S}}\left(\frac{a_\pi}{\sqrt{b_\pi}}\right)^2 \geq \frac{1}{\tau_0^\top\tau_0}\frac{1}{t!}\left(\sum_{\pi\in\mathcal{S}}\left(\frac{a_\pi}{\sqrt{b_\pi}}\right)\right)^2 \\
&\geq \frac{1}{\tau_0^\top\tau_0}\frac{1}{t!}\left(\frac{\sum a_\pi}{\sum\sqrt{b_\pi}}\right)^2 \geq \frac{1}{\tau_0^\top\tau_0}\frac{(\sum a_\pi)^2}{\sum b_\pi} \\
&= \frac{1}{\tau_0^\top\tau_0}\left(\frac{t!\,\tau_0^\top\tau_0 c_{d12}}{t-1}\right)^2 \frac{(t-1)}{t!\,\tau_0^\top\tau_0 c_{d22}} = \frac{t!}{t-1}\frac{c_{d12}^2}{c_{d22}}.
\end{aligned}
$$

Hence

$$
\frac{1}{t!}\sum_{\pi\in\mathcal{S}}\ell_{\tau_\pi,d} \leq \frac{1}{t-1}\left(c_{d11} - \frac{(c_{d12})^2}{c_{d22}}\right).
$$

Again, the right hand side does not depend on τ_0.

In all, we have shown that for every τ_0 we have

$$
\frac{1}{t!}\sum_{\pi\in\mathcal{S}}\phi_A(\mathcal{C}_d,\pi\tau_0) \geq \frac{(t-1)c_{d22}}{c_{d11}c_{d22}-c_{d12}^2} + \frac{(t-1)(t-2)}{c_{d11}+2\lambda_0 c_{d12}+\lambda_0^2 c_{d22}},
$$

with equality holding if all the $\mathcal{C}_{dij}$ are completely symmetric.

We therefore have shown our first result.

Proposition 1 *For any design $d\in\Omega_{t,n,p}$ define*

$$
\bar{\phi_A}(\mathcal{C}_d) = \int \phi_A(\mathcal{C}_d,\tau_0)\mathrm{d}P(\tau_0),
$$

where the distribution P of τ_0 is permutation invariant. Then

$$
\bar{\phi_A}(\mathcal{C}_d) \geq \frac{(t-1)c_{d22}}{c_{d11}c_{d22}-c_{d12}^2} + \frac{(t-1)(t-2)}{c_{d11}+2\lambda_0 c_{d12}+\lambda_0^2 c_{d22}} = \bar{\phi}_A^*(\mathcal{C}_d),
$$

say. Equality holds if in each period all treatments appear equally often and all $\mathcal{C}_{dij}$, for $1\leq i\leq j\leq 2$, are completely symmetric.

Note that $\bar{\phi}_A(\mathcal{C}_d) = (t-1)\bar{A}/(2\sigma^2))$, where $\bar{A}$ is the $\bar{A}$-criterion defined by Kempton et al. (2001). It is a local criterion, depending on the unknown true λ_0.

As a corollary, we get a slight generalization of Theorem 1 of Kempton et al. (2001).

Corollary *Assume that a design $d^* \in \Omega_{t,n,p}$, where $p \leq t$, is totally balanced, that λ_0 is arbitrary and that we consider the $\bar{\phi}_A(\mathcal{C}_d)$ criterion. Then d^* is optimal over all designs for which each treatment appears at most once for each subject.*

Proof The design d^* fulfils the conditions in Proposition 1 which guarantee that

$$\bar{\phi}_A(\mathcal{C}_{d^*}) = \frac{(t-1)c_{d^*22}}{c_{d^*11}c_{d^*22} - c_{d^*12}^2} + \frac{(t-1)(t-2)}{c_{d^*11} + 2\lambda_0 c_{d^*12} + \lambda_0^2 c_{d^*22}}.$$

Since for all competing designs d we have $c_{dij} = c_{d^*ij}$, for $1 \leq i \leq j \leq 2$, the proof is complete. $\square$

Kempton et al. (2001) defined the IA-criterion as the average A-criterion over the joint distribution of both τ_0 and λ_0. If λ_0 and τ_0 are independent and the distribution of τ_0 is permutation invariant, then the corollary implies that a totally balanced design is IA-optimal over all designs for which each treatment appears at most once for each subject, whatever distribution of λ_0 we might assume.

The foregoing optimality results use the average of the A-criterion over permutations of τ_0 (and possibly over a distribution of λ_0). The following example shows that we cannot strengthen this either to A-optimality for all τ_0 or to maximality, in the Loewner order, of the average of the $\mathcal{C}_d$ matrix itself, which would be needed for an analogue of universal optimality. More precisely, let $\bar{\mathcal{C}}_d$ be the average of $\mathcal{C}_d$ over permutations of τ_0.

Example Here are two designs for four treatments in three periods and 24 subjects. Since Designs 1–5 are named in the paper by Kempton et al. (2001), the new designs are called Design 6 and Design 7. Design 6 consists of two copies of

$$\begin{matrix} A & B & C & D & B & C & A & D & C & A & B & D \\ B & A & D & C & C & B & D & A & A & C & D & B \\ C & D & A & B & A & D & B & C & B & D & C & A \end{matrix}$$

while Design 7 consists of three copies of

$$\begin{matrix} A & B & C & D & A & B & C & D \\ B & A & D & C & B & A & D & C \\ C & D & A & B & D & C & B & A \end{matrix}$$

Periods are shown as rows and subjects as columns. Both designs have all treatments equally often in each period, so $\mathcal{C}_6 = \tilde{\mathcal{C}_6}$ and $\mathcal{C}_7 = \tilde{\mathcal{C}_7}$. Moreover, Design 6 is totally balanced.

In Design 6, $\mathcal{C}_{611} = 16B_4$, $\mathcal{C}_{612} = -(16/3)B_4$ and $\mathcal{C}_{622} = (28/3)B_4$. Hence

$$\begin{aligned}\mathcal{C}_6 &= \frac{1}{3}(48 - 32\lambda_0 + 28\lambda_0^2)B_4 - \frac{4(7\lambda_0 - 4)^2}{21}\frac{\tau_0\tau_0^\top}{\tau_0^\top\tau_0} \\ &= \frac{4}{3}(12 - 8\lambda_0 + 7\lambda_0^2)\left(B_4 - \frac{\tau_0\tau_0^\top}{\tau_0^\top\tau_0}\right) + \frac{272}{21}\frac{\tau_0\tau_0^\top}{\tau_0^\top\tau_0}.\end{aligned}$$

The average of $\tau_0\tau_0^\top/\tau_0^\top\tau_0$ over all permutations of τ_0 is $(1/3)B_4$, so

$$\begin{aligned}\bar{\mathcal{C}_6} &= \frac{4}{3}\left(12 - 8\lambda_0 + 7\lambda_0^2 - \frac{(49\lambda_0^2 - 56\lambda_0 + 16)}{21}\right)B_4 \\ &= \frac{4}{63}(98\lambda_0^2 - 112\lambda_0 + 236)B_4.\end{aligned}$$

Put $\theta_1 = (1, 1, -1, -1)^\top$, $\theta_2 = (1, -1, 1, -1)^\top$ and $\theta_3 = (1, -1, -1, 1)^\top$. The permutations of θ_1 consist of $\pm\theta_1$, $\pm\theta_2$ and $\pm\theta_3$ equally often. We shall take τ_0 to be one of the θ_i. For i = 1, 2, 3, put $S_i = \theta_i\theta_i^\top/\theta_i^\top\theta_i$. Then the S_i are mutually orthogonal idempotents of rank 1 whose sum is B_4. Further, put

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Then $S_1 = 2^{-1}G - 4^{-1}u_4u_4^\top$ and $S_2 + S_3 = I - 2^{-1}G = B_4 - S_1$.

In Design 7, $\mathcal{C}_{711} = 16B_4$, $\mathcal{C}_{712} = -6B_4 + GB_4 = -4S_1 - 6S_2 - 6S_3$ and $\mathcal{C}_{722} = 12B_4 - 4GB_4 = 4S_1 + 12S_2 + 12S_3$. Therefore

$$\begin{aligned}&\frac{(\mathcal{C}_{712} + \lambda_0\mathcal{C}_{722})\theta_i\theta_i^\top(\mathcal{C}_{712}^\top + \lambda_0\mathcal{C}_{722})}{\theta_i^\top\mathcal{C}_{722}\theta_i} \\ &= \frac{[4(\lambda_0 - 1)S_1 + 6(2\lambda_0 - 1)(S_2 + S_3)]\theta_i\theta_i^\top[4(\lambda_0 - 1)S_1 + 6(2\lambda_0 - 1)(S_2 + S_3)]}{\theta_i^\top(4S_1 + 12(S_2 + S_3))\theta_i},\end{aligned}$$

which is equal to $4(\lambda_0 - 1)^2S_1$ if $i = 1$ and to $3(2\lambda_0 - 1)^2S_i$ if $i = 2$ or 3. Hence, if $\tau_0 = \theta_1$ then

$$\begin{aligned}\mathcal{C}_7 &= (16 - 8\lambda_0 + 4\lambda_0^2)S_1 + (16 - 12\lambda_0 + 12\lambda_0^2)(S_2 + S_3) - 4(\lambda_0 - 1)^2S_1 \\ &= 12S_1 + 4(4 - 3\lambda_0 + 3\lambda_0^2)(S_2 + S_3);\end{aligned}$$

while if $\tau_0 = \theta_i$ for $i = 2$ or 3 then

$$\begin{aligned}
\mathcal{C}_7 &= (16 - 8\lambda_0 + 4\lambda_0^2)S_1 + (16 - 12\lambda_0 + 12\lambda_0^2)(S_2 + S_3) - 3(2\lambda_0 - 1)^2 S_i \\
&= 4(4 - 2\lambda_0 + \lambda_0^2)S_1 + 13S_i + 4(4 - 3\lambda_0 + 3\lambda_0^2)S_{5-i}.
\end{aligned}$$

Therefore,

$$\bar{\mathcal{C}}_7 = \frac{4}{3}(2\lambda_0^2 - 4\lambda_0 + 11)S_1 + (8\lambda_0^2 - 8\lambda_0 + 15)(S_2 + S_3).$$

First we examine the performance of Designs 6 and 7 without permuting the entries in τ_0. The eigenvalues of $\mathcal{C}_6$ are $4(12 - 8\lambda_0 + 7\lambda_0^2)/3$, twice, and $272/21$, no matter what τ_0 is: in particular they are 16, 16 and 12.95 when $\lambda_0 = 0$. If $\tau_0 = \theta_1$ then the eigenvalues of $\mathcal{C}_7$ are $4(4 - 3\lambda_0 + 3\lambda_0^2)$, twice, and 12, which reduce to 16, 16 and 12 if $\lambda_0 = 0$; while if $\tau_0 = \theta_2$ then the eigenvalues of $\mathcal{C}_7$ are $4(4 - 3\lambda_0 + 3\lambda_0^2)$, $4(4 - 2\lambda_0 + \lambda_0^2)$ and 13, which reduce to 16, 16 and 13 if $\lambda_0 = 0$. Thus if $\lambda_0 = 0$ then Design 6 is A-better than Design 7 if $\tau_0 = \theta_1$ but A-worse than Design 7 if $\tau_0 = \theta_2$. Thus total balance does not guarantee A-optimality for all values of τ_0 and λ_0 even over the binary designs.

Secondly, we compare the average performance of Designs 6 and 7 over permutations of τ_0. We obtain $\bar{\mathcal{C}}_6 - \bar{\mathcal{C}}_7 = \mu_1 S_1 + \mu_2(S_2 + S_3)$, where $\mu_1 = 4(56\lambda_0^2 - 28\lambda_0 + 5)/63$ and $\mu_2 = -(112\lambda_0^2 - 56\lambda_0 + 1)/63$. Now, μ_1 is always positive, while μ_2 is positive if and only if $0.019 < \lambda_0 < 0.481$. For other values of λ_0, neither of $\bar{\mathcal{C}}_6$ and $\bar{\mathcal{C}}_7$ is greater than the other in the Loewner order. However, Design 6 is $\bar{A}$-better than Design 7 for all λ_0.

For further optimality results, we need a further generalization of Kushner's (1997) method: see also Kunert and Martin (2000).

3 Optimality of totally balanced designs

Step 3 For a given design each unit receives a sequence of treatments. Two sequences are called equivalent if one can be transformed to the other by relabelling of treatments. For given p and t there is a number K, say, of possible equivalence classes of sequences. For an arbitrary sequence from a given class k, we can define matrices T_k and F_k, which are the $p \times t$ design matrices for direct and carry-over effects for this sequence. We then define

$$\begin{aligned}
c_{11}(k) &= \operatorname{tr}\left(B_t T_k^\top B_p T_k B_t\right), \\
c_{12}(k) &= \operatorname{tr}\left(B_t T_k^\top B_p F_k B_t\right), \\
c_{22}(k) &= \operatorname{tr}\left(B_t F_k^\top B_p F_k B_t\right).
\end{aligned}$$

Then the $c_{ij}(k)$ do not depend on the special choice of the sequence, only on the equivalence class k. We denote by $\pi_{d,k}$ the proportion of subjects to which design d assigns a sequence from equivalence class k. Then the c_{dij} can be written as linear combinations of $c_{ij}(k)$. More precisely,

$$c_{dij} = n\left(\sum_{k=1}^{K} \pi_{d,k} c_{ij}(k)\right),$$

for $1 \le i \le j \le 2$. Hence, one possibility to find an $\bar{\mathrm{A}}$-optimal design would be to determine a lower bound for $\bar{\phi}_A^*(\mathcal{C}_d)$, or, equivalently, an upper bound for

$$\begin{aligned} q_d^* \quad = \quad & -\frac{\sum \pi_{d,k} c_{22}(k)}{\sum \pi_{d,k} c_{11}(k) \sum \pi_{d,k} c_{22}(k) - \left(\sum \pi_{d,k} c_{12}(k)\right)^2} \\ & -\frac{t-2}{\sum \pi_{d,k}\left(c_{11}(k) + 2\lambda_0 c_{12}(k) + \lambda_0^2 c_{22}(k)\right)}. \end{aligned}$$

For real x, y, z and v, put $g_k(v) = c_{11}(k) + 2vc_{12}(k) + v^2 c_{22}(k)$,

$$h_k(x, y, z) = (t-2) g_k(\lambda_0) x^2 + (t-2)2x + g_k(z) y^2 + 2y,$$

and $q_d(x, y, z) = \sum_{k=1}^{K} \pi_{d,k} h_k(x, y, z)$. Then q_d^* is the minimum of $q_d(x, y, z)$ over x, y and z. We therefore are looking for a design d such that

1. in every period all treatments appear equally often,
2. all $\mathcal{C}_{dij}$ are completely symmetric
3. the proportions $\pi_{d,k}$ of units with sequences from the equivalence class $k \in \{1, \ldots, K\}$ are such that

   $$\min_x \min_y \min_z \sum_{k=1}^{K} \pi_{d,k} h_k(x, y, z)$$

   is as large as possible.

We consider a totally balanced design d^* and restrict attention to the case $p \le t$. Then d^* consists entirely of sequences which are equivalent to $[p, p-1, \ldots, 2, 1]$. We denote the class of all these sequences by $k = 1$. We may assume that $p \ge 3$; for if $p = 2$ then all sequences not confounded with subjects are equivalent to sequence 1. Hence $t - 2$ is positive and $h_k(x, y, z)$ does depend on λ_0.

It is easily seen that, for a general sequence class k,

$$\begin{aligned}
c_{11}(k) &= p - \frac{1}{p}\sum_{i=1}^{t} n_{ik}^2, \\
c_{12}(k) &= \sum_{i=1}^{t} m_{ik} - \frac{1}{p}\sum_{i=1}^{t} n_{ik}\tilde{n}_{ik}, \\
c_{22}(k) &= (p-1)\frac{(pt-1)}{pt} - \frac{1}{p}\sum_{i=1}^{t} \tilde{n}_{ik}^2,
\end{aligned}$$

where n_{ik} is the number of appearances of treatment i in the representative sequence for class k, $\tilde{n}_{ik}$ is the number of appearances of treatment i in the first $p-1$ periods of the representative sequence, while m_{ik} is the number of appearances of treatment i preceded by itself. For the sequence class 1, we have

$$\begin{aligned}
c_{11}(1) &= p-1, \\
c_{12}(1) &= -\frac{p-1}{p}, \\
c_{22}(1) &= \frac{(p-1)}{p}\frac{(pt-t-1)}{t},
\end{aligned}$$

and the minimum of $h_1(x, y, z)$ is attained for

$$\begin{aligned}
z &= -\frac{c_{12}(1)}{c_{22}(1)} = z^*, \text{ say,} \\
y &= -\frac{1}{g_1(z^*)} = y^*, \text{ say,} \\
x &= -\frac{1}{g_1(\lambda_0)} = x^*, \text{ say.}
\end{aligned}$$

Note that $z^* > 0$ and does not depend on λ_0. Neither does y^* depend on λ_0. Also, $g_1(v) > 0$ for all real v, and z^* minimizes g_1.

For a given sequence class k, there is one treatment, say treatment 1, that appears in the last period of the representative sequence. Then $\tilde{n}_{ik} = n_{ik}$ for all $2 \le i \le t$, while $\tilde{n}_{1k} = n_{1k} - 1$. This implies that

$$\sum_{i=1}^{t} n_{ik}\tilde{n}_{ik} = \sum_{i=2}^{t} n_{ik}n_{ik} + n_{1k}(n_{1k}-1) = \sum_{i=1}^{t} n_{ik}^2 - n_{1k} = \sum_{i=1}^{t} n_{ik}^2 - \tilde{n}_{1k} - 1$$

and

$$\sum_{i=1}^{t} \tilde{n}_{ik}^2 = \sum_{i=2}^{t} n_{ik}^2 + (n_{1k}-1)^2 = \sum_{i=1}^{t} n_{ik}^2 - 2n_{1k} + 1 = \sum_{i=1}^{t} n_{ik}^2 - 2\tilde{n}_{1k} - 1.$$

We therefore have

$$\begin{aligned}
c_{11}(k) &= c_{11}(1) + \frac{1}{p}(p - \sum n_{ik}^2), \\
c_{12}(k) &= c_{12}(1) + \sum m_{ik} + \frac{1}{p}(p - \sum n_{ik}^2 + \tilde{n}_{1k}), \\
c_{22}(k) &= c_{22}(1) + \frac{1}{p}(p - \sum n_{ik}^2 + 2\tilde{n}_{1k}).
\end{aligned}$$

From this we get that

$$g_k(v) = g_1(v) - \frac{a}{p} + 2v\left(m - \frac{a}{p} + \frac{\tilde{n}_{1k}}{p}\right) + v^2\left(-\frac{a}{p} + 2\frac{\tilde{n}_{1k}}{p}\right),$$

where $a = \sum n_{ik}^2 - p$ and $m = \sum m_{ik}$. Note that a and m are both non-negative integers.

If, for a given design, we can show that all sequences in the design come from classes k such that $h_k(x^*, y^*, z^*) \leq h_1(x^*, y^*, z^*)$, then this design cannot perform better than the totally balanced design d^*. Hence, when trying to find a design that performs better than d^*, we try to choose $\sum n_{ik}^2$, $\sum m_{ik}$, and $\tilde{n}_{1k}$ in such a way that $h_k(x^*, y^*, z^*)$ is as large as possible.

Let s be the number of treatments occurring in the sequence k. Then $m \leq p-s$, with equality if and only if every treatment has all its occurrences consecutive. Moreover, if $n_{ik} \geq 1$ then $n_{ik}^2 \geq 3n_{ik}-2$, so $a+p \geq 3\sum n_{ik} - 2s = 3p-2s$, so $a \geq 2(p-s)$, with equality if and only if no treatment occurs more than twice in the sequence. Hence $a \geq 2m$: compare this with Proposition 4.2 of Kunert (1984). Similarly, $a + p \geq \sum_{i\neq 1} n_{ik} + 3n_{1k} - 2 = p + 2(n_{1k} - 1) = p + 2\tilde{n}_{1k}$, so $a \geq 2\tilde{n}_{1k}$, with equality if and only if $n_{1k} \leq 2$ and $n_{ik} \leq 1$ for $i > 1$.

In all, we get that $a \geq 2\max\{\tilde{n}_{1k}, m\}$. We have $a = 2m = 2\tilde{n}_{1k}$ if and only if either $a = 0$, which happens only for sequences in class 1, or the sequence is equivalent to $[p-1, p-2, \ldots, 3, 2, 1, 1]$.

Put $F(v) = p(g_k(v) - g_1(v)) = -a(1+v)^2 + 2mpv + 2\tilde{n}_{1k}v(1+v)$ for real numbers v. Then

$$h_k(x^*, y^*, z^*) - h_1(x^*, y^*, z^*) = \frac{(t-2)(x^*)^2}{p}F(\lambda_0) + \frac{(y^*)^2}{p}F(z^*).$$

This equation is linear in $\tilde{n}_{1k}$ and in m. For given $a > 0$, therefore, the largest $h_k(x^*, y^*, z^*)$ will be bounded above by one of the following four cases: $m = \tilde{n}_{1k} = 0$; $m = 0$ and $\tilde{n}_{1k} = a/2$; $m = a/2$ and $\tilde{n}_{1k} = 0$; $m = \tilde{n}_{1k} = a/2$.

If $m = 0 = \tilde{n}_{1k}$ then $F(v) = -a(1+v)^2 \leq 0$ for all v so $h_k(x^*, y^*, z^*) \leq h_1(x^*, y^*, z^*)$.

If $m = 0$ and $\tilde{n}_{1k} = a/2$ then $F(v) = -a(1+v) \le 0$ for $v \ge -1$. Since $-1 \le \lambda_0 \le 1$ and $z^* > 0$, it follows that $h_k(x^*, y^*, z^*) \le h_1(x^*, y^*, z^*)$.

If $m = a/2$ and $\tilde{n}_{1k} = 0$ then $F(v) = a(pv - (1+v)^2)$ and so $F'(v) = a(p-2(1+v))$. Thus $F'(v) \ge 0$ for $v \le 1/2$. Moreover, $F(1/(p-2)) = -a(p-2)^{-2} < 0$, so $F(\lambda_0) < 0$ if $\lambda_0 \le 1/(p-2)$. Also, $z^* = t/(pt-t-1) < 1/(p-2)$ so $F(z^*) < 0$. Hence $h_k(x^*, y^*, z^*) \le h_1(x^*, y^*, z^*)$ for all λ_0 in $[-1, 1/(p-2)]$.

Finally, if $m = \tilde{n}_{1k} = a/2$ then $F(v) = a(-1 + (p-1)v)$ so

$$
\begin{aligned}
&h_k(x^*, y^*, z^*) - h_1(x^*, y^*, z^*) \\
&\quad = \frac{a}{p}\left[(t-2)(-1+(p-1)\lambda_0)(x^*)^2 + (-1+(p-1)z^*)(y^*)^2\right] \\
&\quad = \frac{a}{p}[(t-2)f(\lambda_0) + f(z^*)],
\end{aligned} \tag{5}
$$

where $f(v) = [-1+(p-1)v]/[g_1(v)]^2$. Then

$$
f'(v) = \frac{(p-1)g_1(v) - 2[-1+(p-1)v]g_1'(v)}{[g_1(v)]^3}.
$$

The denominator is always positive. The numerator is a quadratic polynomial in v, which is negative when v is large in absolute value and positive on an interval containing z^*, because $g_1(z^*) > 0$ and $g_1'(z^*) = 0$. It follows that $f(v)$ is decreasing for all v less than some $v_0 < z^*$ and increasing for all v between v_0 and z^*. Therefore, if $(t-2)f(-1) < -f(z^*)$ and $(t-2)f(\lambda^*) < -f(z^*)$ for some λ^* in $(-1, z^*)$ then the expression in (5) is negative for all λ_0 in $[-1, \lambda^*]$. Note that

$$
f(z^*) = \frac{1}{pt-t-1}\frac{1}{[g_1(z^*)]^2} = \frac{1}{pt-t-1}\frac{1}{[1-z^*/p]^2}\frac{1}{(p-1)^2}. \tag{6}
$$

First we examine $f(-1)$. Now, g_1 increases away from z^* and $-(p-1)z^* < -1$, so

$$
g_1(-1) \le g_1(-(p-1)z^*) = (p-1)(1 - \frac{z^*}{p} + pz^*).
$$

Since $t \ge p$, we have

$$
z^* \le \frac{1}{p-1-1/p} = \frac{p}{p^2-p-1}, \tag{7}
$$

so

$$
\frac{pz^*}{1-z^*/p} \le \frac{p^2}{p^2-p-2} \le \frac{9}{4}
$$

because $p \geq 3$. Hence $g_1(-1)/g_1(z^*) \leq 13/4$, and so

$$\begin{aligned}
(t-2)f(-1) &= \frac{-p(t-2)}{[g_1(-1)^2]} \leq -\frac{4^2}{13^2}\frac{p(t-2)}{[g_1(z^*)]^2} \\
&= -\frac{4^2}{13^2}p(t-2)(pt-t-1)f(z^*) \\
&\leq -\frac{4^2}{13^2}p(p-2)(p^2-p-1)f(z^*), \qquad \text{because } t \geq p, \\
&\leq \frac{-16 \times 15}{169}f(z^*), \qquad \text{because } p \geq 3, \\
&< -f(z^*).
\end{aligned}$$

Secondly, we define

$$\lambda^* = \frac{1}{(t-2)(p-1)}\left[t-2-\frac{1}{pt-t-1}\frac{1}{(1-z^*/p)^2}\right]. \tag{8}$$

Note that λ^* is slightly smaller than $1/(p-1)$. We have $g_1(0) = p-1$ and g_1 increases away from z^*, so if $0 \leq v \leq z^*$ then $g_1(v) \leq p-1$. But $(p-1)^{-1} < z^*$, so if $0 \leq v < (p-1)^{-1}$ then $f(v) \leq [-1+(p-1)v]/(p-1)^2$. It follows that

$$\begin{aligned}
(t-2)f(\lambda^*) &\leq (t-2)[-1+(p-1)\lambda^*]/(p-1)^2 \\
&= \frac{-1}{pt-t-1}\frac{1}{[1-z^*/p]^2}\Big/(p-1)^2
\end{aligned}$$

Equation (6) then shows that $(t-2)f(\lambda^*) \leq -f(z^*)$.

Now, inequality (7) shows that $1-z^*/p \geq (p^2-p-2)/(p^2-p-1)$; moreover $pt-t-1 \geq p^2-p-1$. Hence,

$$\lambda^* \geq \frac{1}{(t-2)(t-1)}[t-2-\frac{p^2-p-1}{(p^2-p-2)^2}]$$

which is positive since $t \geq p \geq 3$.

We have therefore proven the following two propositions.

Proposition 2 *For all $-1 \leq \lambda_0 \leq 1$, a totally balanced design $d^* \in \Omega_{t,n,p}$ is $\bar{A}$-optimal over all designs $d \in \Omega_{t,n,p}$ for which no treatment is directly preceded by itself.*

Proposition 3 *If $-1 \leq \lambda_0 \leq \lambda^*$, where $\lambda^* > 0$ is defined by equation (8), then a totally balanced design $d^* \in \Omega_{t,n,p}$ is $\bar{A}$-optimal over all possible designs $d \in \Omega_{t,n,p}$.*

Note that the situation where λ_0 is small in absolute size is an important case. Generally, experimenters will try to run the experiment in such a way that carry-over effects can be avoided as much as possible, see also Jones, Kunert and Wynn (1992).

If λ_0 gets large and positive, then designs with positive m will get better than d^*, as was shown by Kempton et al. (2001).

Because g_1 is rather flat on $[1, z^*]$, Equation (8) gives a good approximation to the upper bound for values of λ_0 for which the totally balanced design is optimal, but the actual upper bound is slightly higher. For example, when $p = t = 4$ then $\lambda^* = 0.315$ but numerical investigation of Equation (5) shows that $(t-2)f(0.318) \approx -f(z^*)$. Hence, for all $\lambda_0 \leq 0.318$, the totally balanced design "Design 1" of Kempton et al. (2001) is optimal over $\Omega_{4,12,4}$.

References

Fan, K. (1949). On a theorem of Weyl concerning eigenvalues of linear transformations I. *Proc. Nat. Acad. Sci. U.S.A.* **35**, 652–655.

Jones, B., Kunert, J. & Wynn, H.P. (1992). Information matrices for mixed effects models with applications to the optimality of repeated measurements designs. *Journ. Statist. Plann. Inf.* **33**, 261–274.

Kempton, R. A., Ferris, S. J. & David, O. (2001). Optimal change-over designs when carry-over effects are proportional to direct effects of treatments. *Biometrika* **88**, 391–399.

Kunert, J. (1983). Optimal design and refinement of the linear model with applications to repeated measurements designs. *Ann. Statist.* **11**, 247–257.

Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs. *Ann. Statist.* **12**, 1006–1017.

Kunert, J. & Martin, R. J. (2000). On the determination of optimal designs for an interference model. *Ann. Statist.* **28**, 1728–1742.

Kunert, J. & Stufken, J. (2002). Optimal crossover designs in a model with self and mixed carryover effects. *Journal of the American Statistical Association* **97**, 898–906.

Kushner, H. B. (1997). Optimal repeated measurements designs: the linear optimality equations. *Ann. Statist.* **25**, 2328–2344.