Optimal Paired Choice Block Designs

Rakhi Singh[1], *Ashish Das*[2] *and Feng-Shun Chai*[3]

[1]*IITB-Monash Research Academy, Mumbai, India*

[2]*Indian Institute of Technology Bombay, Mumbai, India*

[3]*Academia Sinica, Taipei, Taiwan*

Supplementary Material

Design for Example 3.1. $k = 3, v_1 = 2, v_2 = 3, v_3 = 4, b = 1, N = s = 72.$

(000,111)	(000,222)	(000,333)	(100,211)	(100,322)	(100,033)	(020,131)	(020,313)	(120,302)
(001,112)	(001,223)	(001,330)	(101,212)	(101,323)	(101,030)	(021,132)	(021,310)	(121,303)
(002,113)	(002,220)	(002,331)	(102,213)	(102,320)	(102,031)	(022,133)	(022,311)	(122,300)
(003,110)	(003,221)	(003,332)	(103,210)	(103,321)	(103,032)	(023,130)	(023,312)	(123,301)
(010,121)	(010,232)	(010,303)	(110,221)	(110,332)	(110,003)	(020,202)	(120,231)	(120,013)
(011,122)	(011,233)	(011,300)	(111,222)	(111,333)	(111,000)	(021,203)	(121,232)	(121,010)
(012,123)	(012,230)	(012,301)	(112,223)	(112,330)	(112,001)	(022,200)	(122,233)	(122,011)
(013,120)	(013,231)	(013,302)	(113,220)	(113,331)	(113,002)	(023,201)	(123,230)	(123,012)

Lemma 1. *A necessary and sufficient condition for* $\tilde{C}_M = C_M$ *to hold is that for each block and each attribute, the frequency distribution of the levels of the attribute are same for the two options.*

Proof. Let $P_{Mj} = ((P_j)'_1 \cdots (P_j)'_t \cdots (P_j)'_b)'$ where $(P_j)_t$ represents P_{Mj} for the tth block. Then the condition $W'P_M = 0$ is equivalent to the condition $1'(P_1)_t = 1'(P_2)_t$, $t = 1, \ldots, b$. Let $(P_j)_t = ((P_j)_t^1 \cdots (P_j)_t^w \cdots (P_j)_t^k)$ where $(P_j)_t^w$ is of order $s \times (v_w - 1)$ and represents $(P_j)_t$ for the wth attribute. Therefore for $t = 1, \ldots, b$, if $1'(P_1)_t = 1'(P_2)_t$, then $1'(P_1)_t^w = 1'(P_2)_t^w$ for every w and t. Now, since the ith column of $(P_j)_t^w$ provides frequency of level i and level v_w in the wth attribute of the jth option in the tth block, therefore, $1'(P_1)_t^w = 1'(P_2)_t^w$ implies that the frequency of each of the levels of attribute w is same in the two options among the s choice pairs in block t.

The converse follows by noting that if for each block and each attribute, the frequency distribution of the levels of the attribute are same for the two options, then $1'(P_1)_t = 1'(P_2)_t$ for every t. □

Proof of Theorem 1. The proof follows as a special case of Lemma 1. □

Proof of Theorem 2. Under the linear paired comparison model, a design d optimally estimates the main effects if $C_M = diag(C_{(1)}, \ldots, C_{(k)})$ (see Großmann and Schwabe (2015)) where $C_{(i)} = z_i(I_{v_i-1} + J_{v_i-1})$ with $z_i = 2N/(v_i - 1)$, $i = 1, \ldots, k$. This implies that C_M normalized by number of pairs would attain an optimal structure if $C_{(i)} = z_i(I_{v_i-1} + J_{v_i-1})$ with $z_i = 2/(v_i - 1)$, $i = 1, \ldots, k$.

Since the $OA + G$ method of construction entails adding generators to the orthogonal array of strength $t, (t \geq 2)$, the off-diagonal elements of $P_M'P_M$ corresponding to two different attributes is zero since under each level of the first attribute, all the levels of the second attribute occur equally often. Also, since in an orthogonal array, under each column (attribute) the levels are equally replicated, to establish that each $C_{(i)}$ attains an optimal structure of the form $z_i(I_{v_i-1} + J_{v_i-1})$, it is enough to show that normalized $P_M'P_M$ corresponding to a paired choice design with one attribute, say at v levels, attains the structure $z(I_{v-1} + J_{v-1})$, where $z = 2/(v - 1)$.

Without loss of generality, we consider only v choice pairs for a typical attribute since under each column, the n rows of the orthogonal array involves v symbols each replicated n/v times. While using the generator g_j, let P_1^0, P_2^j be the $v \times (v-1)$ effects-coded matrix for the main effects for the first and second options, respectively, corresponding to any one attribute at v levels. When $h > 1$, note that P_M is the collection of different matrices generated out of the corresponding $\{P_1^0, P_2^j\}, j = 1 \ldots, h$ of choice pairs. For notational simplicity, we denote P_1^0 by P_0 and P_2^j by P_j, $j = 1, \ldots, v-1$. Also, note that $1'P_j = 0$ and $\sum_{j=0}^{v-1} P_j = 0$.

Consider the information matrix $P_M'P_M$ normalized for v even. $v(v-1)P_M'P_M = \sum_{j=1}^{v-1}(P_0 - P_j)'(P_0 - P_j) = \sum_{j=1}^{v-1}(P_0'P_0 + P_j'P_j - P_0'P_j - P_j'P_0) = \sum_{j=1}^{v-1}\{2(I_{v-1} + J_{v-1})\} - P_0'(\sum_{j=1}^{v-1} P_j) - (\sum_{j=1}^{v-1} P_j')P_0 = \{2(v-1)(I_{v-1}+J_{v-1})\} - P_0'(-P_0) - (-P_0')P_0 = 2\{(v-1)(I_{v-1}+J_{v-1})\} + 2P_0'P_0 = 2v(I_{v-1} + J_{v-1})$. Thus, for v even, $h = v-1$ generators of the type $g_j = 1, \ldots, v-1$ leads to the optimal structure of normalized $P_M'P_M$.

For v odd, we note that, if say, mth row of P_0 corresponds to the level i, then the mth row of P_{v-j} corresponds to the level $i - j \pmod v$. Similarly, if say, lth row of P_j corresponds to the level i, then the lth row of P_0 corresponds to the level $i - j \pmod v$. This makes the lth row of P_j and P_0 same as the mth row of P_0 and P_{v-j} for every two rows $l \neq m = 1, \ldots, v$. Therefore, for v odd, $P_j'P_0 = P_0'P_{v-j}$. Now, $v(v-1)/2P_M'P_M = \sum_{j=1}^{(v-1)/2}(P_0 - P_j)'(P_0 - P_j) = \sum_{j=1}^{(v-1)/2}(P_0'P_0 + P_j'P_j - P_0'P_j - P_j'P_0) = \sum_{j=1}^{(v-1)/2}\{2(I_{v-1} + J_{v-1})\} - \sum_{j=1}^{(v-1)/2}(P_0'P_j + P_j'P_0) = (v-1)(I_{v-1}+J_{v-1}) - \sum_{j=1}^{(v-1)/2}(P_0'P_j + P_0'P_{v-j}) = (v-1)(I_{v-1}+J_{v-1}) - P_0'\sum_{j=1}^{(v-1)/2}(P_j + P_{v-j}) = (v-1)(I_{v-1}+J_{v-1}) - P_0'\sum_{j=1}^{v-1} P_j = (v-1)(I_{v-1}+J_{v-1}) - P_0'(-P_0) = v(I_{v-1}+J_{v-1})$. Thus, for v odd, $h = (v-1)/2$ generators of the type $g_j = 1, \ldots, (v-1)/2$ leads to the optimal structure of normalized $P_M'P_M$. $\square$

Proof of Theorem 3. For a given $OA(n_1, k+1, v_1 \times \cdots \times v_k \times \delta, 2)$, corresponding to the k attributes at levels $v_i, i = 1, \ldots, k$, let d_1 be the design constructed through $OA + G$ method using $h = lcm(v_1, \ldots, v_k)$ generators. Then d_1 with parameters k, $v_1, \ldots, v_k$, $b = 1$, $s = hn_1$ is an optimal paired choice design. From d_1, the choice pairs obtained through each of the h generators constitute a block of size n_1. This is true since n_1 rows of a block form the orthogonal array in the first option and, with labels re-coded through the generator, in the second option and hence the conditions in Theorem 1 are satisfied.

Finally, we use the δ symbols of the $(k+1)$th column of the orthogonal array for further blocking giving a paired choice block design d_2 with parameters k, $v_1, \ldots, v_k$, $b = h\delta$, $s = n_1/\delta$. This is true since for every attribute in each of the blocks so formed, each of the v_i levels occurs equally often under ith attribute and hence by Theorem 1, d_2 is optimal in $\mathcal{D}_{k,b,s}$. □

Proofs for Theorem 4 and Theorem 5 require a result from Dey (2009) that is given below.

Lemma 2 (Dey (2009)). *Consider $v(v-1)/2$ combinations involving v levels taken two at a time. Then, for v odd, the combinations can be grouped into $(v-1)/2$ replicates each comprising v combinations. The groups are $\{(i, v-2-i), (i+1, v-1-i), \ldots, (i+v-1, v-2-(i-(v-1)))\}$ and the levels are reduced modulo v; $i = 0, \ldots, (v-3)/2$.*

Proof of Theorem 4. Theorem 3 of Graßhoff et al. (2004) states that from $m(\geq k)$ rows of a Hadamard matrix H_m of order m, an optimal paired choice design d_3 with parameters $k, v, b = 1, s = mv(v-1)/2$ is constructed using the $v(v-1)/2$ combinations of v levels taken two at a time. From every row of $\{H_m, -H_m\}$, $v(v-1)/2$ choice pairs are obtained by replacing '1' in the row by the first column of the combinations and '-1' in the row by the second column of the combinations. If v is odd, then $(v-1)/2$ is an integer and the $v(v-1)/2$ combinations can be arranged in rows such that each of the two columns have every level appearing equally often. Such an arrangement is always possible and follows from systems of distinct representatives. Therefore, corresponding to each of the rows of $\{H_m, -H_m\}$, using $v(v-1)/2$ choice pairs as a block, a paired choice block design with parameters $k, v, b = m, s = v(v-1)/2$ is obtained which, following Theorem 1 is optimal. Now for v odd, from Dey (2009), $v(v-1)/2$ combinations involving v levels taken two at a time can be grouped into $(v-1)/2$ replicates each comprising v combinations. Therefore, the blocks generated by each row of H_m can be further broken into $(v-1)/2$ blocks each of size v, which gives us d_4. □

Proof of Theorem 5. Construction 3.2 of Demirkale, Donovan, and Street (2013) uses an $OA(n_2, k+1, v^k \times v_{k+1}, 2)$ with $v_{k+1} = n_2/v$ and forms v_{k+1} parallel sets each having v rows. Then, an optimal paired choice design with parameters $k, v, b = 1, s = v_{k+1}\binom{v}{2}$ is constructed using the $v(v-1)/2$ combinations of v numbers $\{1, \ldots, v\}$ taken two at a time. Let $\{i, j\}$ be a typical row. Then, for each such row of size two, corresponding rows i and j from each of the v_{k+1} parallel sets are chosen to form the choice pairs of the optimal paired choice design d_6. Again as earlier, for v odd, the $v(v-1)/2$ combinations can be arranged in rows such that each of the two columns have every number appearing equally often. Considering the $v(v-1)/2$ choice pairs, obtained from a parallel set, as a block, we get the paired choice block design with parameters $k, v, b = v_{k+1}, s = v(v-1)/2$ which is optimal in $\mathcal{D}_{k,b,s}$. Further proof follows on the same lines as the proof of Theorem 4 by treating the pairs generated by each parallel set as blocks. □

Proof of Theorem 6. Theorem 4 of Graßhoff et al. (2004) uses an $OA(n_3, k+1, m_1 \times \cdots \times m_k \times \delta, 2)$ with $m_i = v_i(v_i-1)/2$ for some odd v_i to construct an optimal paired choice design d_7 with parameters $k, v_i, \ldots, v_k, b = 1, s = n_3$. This method involves a one-one mapping between m_i levels of orthogonal array to the $v_i(v_i-1)/2$ combinations on v_i symbols. For a combination $\{i, j\}$ corresponding to a symbol of an orthogonal array, the first option in a pair is obtained by replacing i in place of that symbol and the second option has j in the corresponding position. Then, similar to construction of Theorem 3, using the δ (≥ 1) symbols of the $(k+1)$th column of the orthogonal array for blocking gives us an optimal paired choice block design d_8 with parameters $k, v_i, \ldots, v_k, b = \delta, s = n_3/\delta$. Note that this method is applicable only for odd v_i since for even v_i, it is not possible to arrange $v_i(v_i - 1)/2$ combinations in a position-balanced manner. □

Proof of Theorem 7. From Theorem 1, for each of the h generators, a paired choice design using the $OA+G$ method of construction is optimal under the broader main effects block model if $P_M' P_I = 0$.

For a given generator, to show that $P_M' P_I = 0$, it suffices to show that the inner product of the columns of P_M corresponding to the mth main effect and the columns of P_I corresponding to the two-factor interaction effect of ith and jth attribute is zero. Using an $OA(n_1, k, v_1 \times \cdots \times v_k, 3)$ in the $OA+G$ method of construction, we establish the result through the following two cases.

Case (i) $m = i$: In an orthogonal array of strength 2, each of the $v_i v_j$ combinations occur equally often $n_1/(v_i v_j)$ times as rows. Therefore, since the paired choice design is based on the orthogonal array, for showing that $P_M' P_I = 0$, it suffices to show that $P_M' P_I = 0$ for one of the $n_1/(v_i v_j)$ sets of $v_i v_j$ rows of the type $(i, j); i = 0, \ldots, v_i - 1; j = 0, \ldots, v_j - 1$. For such $v_i v_j$ rows, note that P_{My}, $(y = 1, 2)$, corresponding to the jth attribute, can be partitioned into v_i sets $P_{My(j)}$ each of v_j distinct rows. Then, $1' P_{My(j)} = 0$. Let P_{Iy} corresponding to the ith attribute fixed at level i_l $(i_l = 0, \ldots, v_i - 1)$ and the jth attribute taking v_j distinct levels be represented by $P_{Iy(i_l j)}$. Then, the columns of $P_{Iy(i_l j)}$ are multiples of either $P_{My(j)}$ or 0_v. Therefore, $1' P_{Iy(i_l j)} = 0$ for $y = 1, 2$.

Let P_M corresponding to the ith attribute at level i_l be represented by X_{i_l}. Then, $X_{i_l} = 1 x_{i_l}'$ where x_{i_l}' is a row vector of size $v_i - 1$. Therefore, $P_M' P_I = \sum_{i_l=0}^{v_i-1} X_{i_l}' (P_{I1(i_l j)} - P_{I2(i_l j)}) = \sum_{i_l=0}^{v_i-1} x_{i_l} (1' P_{I1(i_l j)} - 1' P_{I2(i_l j)}) = 0$.

Case (ii) $m \neq i$: In an orthogonal array of strength 3, each of the $v_i v_j v_m$ combinations occur equally often $n_1/(v_m v_i v_j)$ times as rows. Therefore, as in Case (i), for showing that $P_M' P_I = 0$, it suffices to show that $P_M' P_I = 0$ for one of the $n_1/(v_m v_i v_j)$ sets of $v_m v_i v_j$ rows of the type $(m, i, j); m = 0, \ldots, v_m - 1; i = 0, \ldots, v_i - 1; j = 0, \ldots, v_j - 1$.

For such $v_m v_i v_j$ rows, note that P_{Iy}, $(y = 1, 2)$, corresponding to the ith and jth attribute, can be partitioned into v_m sets $P_{Iy(ij)}$ each of $v_i v_j$ distinct rows. Therefore, $1' P_{Iy(ij)} = 0$ for $y = 1, 2$, since from Case (i), $1' P_{Iy(i_l j)} = 0$ for the ith attribute at level i_l.

Finally, since for the mth attribute at level m_l $(m_l = 0, \ldots, v_m - 1)$, the $v_i v_j$ combinations under attributes i and j occur equally often, therefore $P_M' P_I = \sum_{m_l=0}^{v_m-1} X_{m_l}' (P_{I1(ij)} - P_{I2(ij)}) = \sum_{m_l=0}^{v_m-1} x_{m_l} (1' P_{I1(ij)} - 1' P_{I2(ij)}) = 0$. □

Proof of Theorem 10. From Lemma 1, $W' P_M = 0$ if and only if for each attribute under the choice pairs having foldover in the second option of a choice pair, the level l $(l = 0, 1)$ appears equally often in both the options in every block and thus, the frequency of the pair $(1, 0)$ is same as the frequency of the pair $(0, 1)$ under every attribute in each block.

Let $P_I = (Y_1' \cdots Y_t' \cdots Y_b')'$ where Y_t is the $s \times k(k-1)/2$ matrix corresponding to the tth block. With $(P_{Ij})_t$ representing P_{Ij} for the tth block, $Y_t = (P_{I1})_t - (P_{I2})_t$. Then, the condition $W' P_I = 0$ is equivalent to the condition $1'(P_{I1})_t = 1'(P_{I2})_t$ for every $t = 1, \ldots, b$. Consider $(P_{Ij})_t = ((P_{Ij})_t^{12} \cdots (P_{Ij})_t^{lm} \cdots (P_{Ij})_t^{(k-1)k})$ where $(P_{Ij})_t^{lm}$ is of order $s \times 1$ and represents $(P_{Ij})_t$ for the two-factor interaction between the lth and the mth attribute. Therefore, the necessary and sufficient condition for $1'(P_{I1})_t = 1'(P_{I2})_t$ is that $1'(P_{I1})_t^{lm} = 1'(P_{I2})_t^{lm}$ for every l and m.

In the tth block, for the choice pairs where either both the attributes have a foldover in the second option or both do not have a foldover in the second option, the corresponding rows in $(P_{I2})_t^{lm}$ are same as the corresponding rows in $(P_{I1})_t^{lm}$.

However, for the pairs in which one attribute has a foldover in the second option and another does not have foldover in the second option, the corresponding rows in $(P_{I2})_t^{lm}$ are negative of the corresponding rows in $(P_{I1})_t^{lm}$. In such a case, $1'(P_{I1})_t^{lm} = 1'(P_{I2})_t^{lm}$ if and only if $1'(P_{I1})_t^{lm} = -1'(P_{I2})_t^{lm} = 0$. Now, $1'(P_{I1})_t^{lm} = 0$ if and only if the frequency of the pairs from the set $\{(01,00),(01,11),$ $(10,00)$,$(10,11)\}$ is same as the frequency of the pairs from the set $\{(00,01)$, $(00,10)$, $(11,01)$, $(11,10)\}$ under the lth and the mth attribute. □

Proof of Theorem 11. In steps (iii)-(iv), corresponding to an element f of F, make the first set of $2^{\alpha-1}2^{k-\alpha-2} = 2^{k-3}$ blocks having choice pairs $(ab, a'b)$, $(ab', a'b')$, $(a'c, ac)$, $(a'c', ac')$. Similarly, following the steps (iii)-(iv), we make an additional set of 2^{k-3} blocks having choice pairs $(ac, a'c)$, $(ac', a'c')$, $(a'b, ab)$, $(a'b', ab')$. Note that each of the constructed blocks satisfy conditions (i) and (ii) of Theorem 10. This gives rise to a total of 2^{k-2} sets of blocks each of size 4. The way we have constructed the choice pairs in steps (iii)-(iv), it follows that the collection of first option in the 2^k choice pairs forms a complete factorial having 2^k combinations. Furthermore, the additional set of 2^{k-3} blocks, in the construction, is identical to the first set of 2^{k-3} blocks. Accordingly, we retain only the first set of 2^{k-3} blocks. This gives rise to a total of 2^{k-1} choice pairs divided into 2^{k-3} blocks each of size 4. Therefore, step (v) gives an optimal paired choice block design d_2^I with parameters $k, v = 2, s = 4, b$ where $b = 2^{k-3}\binom{k}{q}$ for k odd and $b = 2^{k-3}\binom{k+1}{q+1}$ for k even.

□

References

Demirkale, F., D. Donovan, and D. J. Street (2013). Constructing D-optimal symmetric stated preference discrete choice experiments. *J. Statist. Plann. Inference 143*(8), 1380–1391.

Dey, A. (2009). Orthogonally blocked three-level second order designs. *J. Statist. Plann. Inference 139*(10), 3698–3705.

Graßhoff, U., H. Großmann, H. Holling, and R. Schwabe (2004). Optimal designs for main effects in linear paired comparison models. *J. Statist. Plann. Inference 126*(1), 361–376.

Großmann, H. and R. Schwabe (2015). Design for discrete choice experiments. In A. Dean, M. Morris, J. Stufken, and D. Bingham (Eds.), *Handbook of Design and Analysis of Experiments*, pp. 791–835. Boca Raton, FL: Chapman and Hall.

Rakhi Singh, IITB-Monash Research Academy, Mumbai, India

E-mail: agrakhi@gmail.com

Ashish Das, Indian Institute of Technology Bombay, Mumbai, India

E-mail: ashish@math.iitb.ac.in

Feng-Shun Chai, Academia Sinica, Taipei, Taiwan

E-mail: fschai@stat.sinica.edu.tw