OPTIMAL DESIGNS FOR DIALLEL CROSS EXPERIMENTS

Aloke Dey
Indian Statistical Institute, 7 SJS Sansanwal Marg, New Delhi – 110 016

1. Introduction

The diallel cross is a type of mating design used in plant breeding to study the genetic properties of a set of inbred lines. The purpose of such an experiment is to compare the lines with respect to their *general combining abilities (g.c.a.)*. The general combining ability of an inbred line is the average of performance of the hybrids that this line produces with other lines and, can be looked upon as analogous to a main effect in the context of a factorial experiment. Often, apart from inferring on general combining abilities, an experimenter is also interested in *specific combining abilities*. The specific combining ability refers to a pair of inbred lines involved in a cross and is the deviation of a particular cross from the average of the general combining abilities of the two lines involved in the cross. Thus, the specific combining ability is analogous to a two-factor interaction in a factorial experiment. For more on the genetic interpretation of these parameters, see Griffing (1956) and Hinkelmann (1975).

A common type of diallel cross experiment involves $p(p-1)/2$ crosses of the type $(i \times j)$, $i < j$, $i, j = 1, 2, \cdots, p$ where p is the number of inbred lines under consideration. A diallel cross of the above type has also been called in the literature a complete diallel cross. Traditionally, such experiments have been conducted using a completely randomized design or, a randomized (complete) block design. With the increase in the number of lines p, the number of crosses in a complete diallel cross increases rapidly. In such a situation, assuming that heterogeneity (in the experimental material) exists in one direction, adoption of a randomized complete block design with crosses as treatments would result in a large error variance, even with a moderate number of lines.

In order to control the error, one would therefore look for an appropriate incomplete block design in preference to a complete block design. One possibility is to use available incomplete block designs, for instance, a balanced incomplete block (BIB) design, for the experiment, identifying crosses with treatments. This approach has been advocated, for instance, by Das and Giri (1986, pp. 441-442) and Ceranka and Mejza (1988). Another approach advocated is to start with an incomplete block design for the usual treatment-block structure, treat the treatments as lines and make all possible pairwise crosses among the lines within a block; see e.g., Ghosh and Divecha (1997) and Sharma (1998). Both these approaches, however, are inappropriate if one is interested in the optimality of the design. The following examples illustrate this fact.

Example 1.1: Let $p = 7$ so that there are 21 crosses in a complete diallel experiment. Consider the following two design d_1 and d_2 both involving $b = 7$ blocks, each of size $k = 3$. For convenience, the cross $(i \times j)$ is represented as (i, j)

d_1			d_2		
$(1,2)$	$(1,4)$	$(2,4)$	$(1,7)$	$(2,6)$	$(3,5)$
$(2,3)$	$(2,5)$	$(3,5)$	$(1,2)$	$(3,7)$	$(4,6)$
$(3,4)$	$(3,6)$	$(4,6)$	$(2,3)$	$(1,4)$	$(5,7)$
$(4,5)$	$(4,7)$	$(5,7)$	$(3,4)$	$(2,5)$	$(1,6)$
$(5,6)$	$(1,5)$	$(1,6)$	$(4,5)$	$(3,6)$	$(2,7)$
$(6,7)$	$(2,6)$	$(2,7)$	$(5,6)$	$(4,7)$	$(1,3)$
$(1,7)$	$(1,3)$	$(3,7)$	$(6,7)$	$(1,5)$	$(2,7)$

The rows under both d_1 and d_2 are the blocks. The design d_1 is obtained by first choosing a BIB design with parameters $v = 7 = b, r = 3 = k, \lambda = 1$ and then making all possible pairwise crosses within the treatments (lines) within each block. The design d_2 is obtained in an altogether different manner. It turns out that the average variance of the best linear unbiased estimator of all elementary comparisons among g.c.a. effects under design d_1 is $\frac{6}{7}\sigma^2$ while the same under d_2 is $\frac{3}{7}\sigma^2$, σ^2 being the per observation variance. Thus, judged by the average variance criterion, d_2 is twice as efficient as d_1, even though d_1 is based on an optimal block design viz. a BIB design.

Example 1.2: Now, let $p = 12$. Following Ghosh and Divecha (1997), one can start with a group divisible design with parameters $v = 12, b = 9, r = 3, \ k = 4, \ \lambda_1 = 0,$ $\lambda_2 = 1, \ \ m = 4, \ \ n = 3$ and generate a diallel cross design by the same technique as in Example 1.1. Call this design d_3. A design d_4, comparable to d_3 is shown below:

$(1,2)$	$(5,6)$	$(9,10)$	$(3,4)$	$(7,8)$	$(11,12)$
$(1,3)$	$(5,7)$	$(9,11)$	$(2,4)$	$(6,8)$	$(10,12)$
$(1,4)$	$(5,8)$	$(9,12)$	$(2,3)$	$(6,7)$	$(10,11)$
$(1,6)$	$(5,10)$	$(9,2)$	$(3,8)$	$(7,12)$	$(4,11)$
$(1,7)$	$(5,11)$	$(3,9)$	$(2,8)$	$(6,12)$	$(4,10)$
$(1,5)$	$(5,12)$	$(4,9)$	$(2,7)$	$(6,11)$	$(3,10)$
$(1,10)$	$(2,5)$	$(6,9)$	$(3,12)$	$(4,7)$	$(8,11)$
$(1,11)$	$(3,5)$	$(7,9)$	$(2,12)$	$(4,6)$	$(8,10)$
$(1,12)$	$(4,5)$	$(8,9)$	$(2,11)$	$(3,6)$	$(7,10)$

The average variance of the best linear unbiased estimator (BLUE) of all elementary comparisons among *g.c.a.* effects under d_3 is $0.414\ \sigma 2$ whereas, that under d_4 is $0.253\ \sigma 2$.

Such examples can obviously be multiplied to demonstrate that techniques different from the ones described above are needed to obtain efficient (block) designs for diallel cross experiments.

2. Nested Designs and Optimality

We now describe some optimal block designs for complete diallel cross experiments under the model with no specific combining ability effect. Let *d* be a block design for a complete diallel cross experiment involving *p* inbred lines, *b* blocks each of size $k\ (\geq 2)$. This means that there are *k* crosses in each of the blocks of *d*. Further, let r_{di} denote the number of times the i^{th} cross appears in *d*, $i = 2,\cdots,p(p-1)/2$, and similarly, let w_{dj} denote the number of times the j^{th} line occurs in crosses in the whole design *d*, $j = 1,2,\cdots,p$. It is then easy to see that

$$\sum_{i=1}^{p(p-1)/2} r_{di} = bk \qquad \text{and} \qquad \sum_{j=1}^{p} w_{dj} = 2bk.$$

We also let $n = bk$ denote the number of observations generated by *d*. For the data obtained from the design *d*, we postulate the model

$$\mathbf{Y} = \mu \mathbf{1}_n + \mathbf{\Delta}_1 \mathbf{g} + \mathbf{\Delta}_2 \boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\mathbf{Y}$ is the $n \times 1$ vector of observed responses, μ is a general mean effect, $\mathbf{g}$ and $\boldsymbol{\beta}$ are vectors of *p* general combining ability effects and *b* block effects respectively, Δ_1, Δ_2 are the corresponding design matrices, *i.e.*, the $(s, t)^{th}$ element of Δ_1 is *1* if the s^{th} observation pertains to the t^{th} line, and is zero, otherwise; Δ_2 is defined similarly. $\boldsymbol{\varepsilon}$ is the vector of random error components, these components being distributed with mean zero and constant variance σ^2. In (2.1), the elements of $\boldsymbol{\varepsilon}$ take care of specific combining ability effects as well as unassignable variation. Under the model (2.1), it can be shown that the coefficient matrix of the reduced normal equations for estimating linear functions of general combining ability effects using a design *d* is

$$\mathbf{C}_d = \mathbf{G}_d - \mathbf{N}_d \mathbf{N}_d' / k \tag{2.2}$$

where $\mathbf{G}_d = (g_{dii'})$, $\mathbf{N}_d = ((n_{dij}))$, $g_{dii} = w_{di}$, and for $i \neq i'$, $g_{dii'}$ is the number of times the cross $(i \times i')$ appears in *d*; n_{dij} is the number of times the line *i* occurs in block *j* of *d*.

The matrix $\mathbf{C}_d$ is often referred as the information matrix of *d*. A design *d* is called connected if and only if all elementary comparisons among general combining ability effects are estimable using *d*. A necessary and sufficient condition for a design *d* to be connected is that $\text{Rank}(\mathbf{C}_d) = p-1$. We denote by $\mathbf{D}$ *(p, b, k)* the class of all such connected block designs *{d}* with *p* lines, *b* blocks each of size *k*. We then have

Theorem 2.1: For any design $d \in$ **D** *(p, b, k)*,

$$tr(\mathbf{C}_d) \leq k^{-1}b\{2k(k - 1 - 2x) + px(x + 1)\},$$

where $x = [2k/p]$ $[.]$ is the greatest integer function and for a square matrix **A**, *tr(***A***)* stands for the trace. Equality holds if and only if $n_{dij} = x$ *or* $x + 1$ for all $i = 1 2, \cdots, p$, $j = 1, 2, \cdots, b$.

Note that if $2k < p$ then $x = 0$ and in that case we have

$$tr\,(\mathbf{C}_d) \leq 2b(k - 1),\ d \in \mathbf{D}(p,\ b,\ k). \qquad (2.3)$$

The above development helps one to find *universally optimal* designs for diallel cross experiments. A universally optimal design is in particular also *minimizes the average variance of the best linear unbiased estimators of all elementary contrasts among the general combining ability effects*. Making an appeal to a result of Kiefer (1975) and to Theorem 2.1, we have the following result:

Theorem 2.2: Let $d^* \in$ **D** *(p, b, k)* be a block design for diallel crosses, and suppose $\mathbf{C}_{d^*}$ satisfies

(i) $tr(\mathbf{C}_{d^*}) = k^{-1}b\{2k(k - 1 - 2x) + px(x + 1)\}$, and

(ii) $\mathbf{C}_{d^*}$ is completely symmetric.

Then d^* is universally optimal in **D** *(p, b, k)* and in particular minimizes the average variance of the best linear unbiased estimators of all elementary contrasts among the general combining ability effects.

There is an interesting connection between nested balanced incomplete block (NBIB) design of Preece (1967) and optimal designs for diallel crosses. For completeness, we recall the definition of a nested balanced incomplete block design.

Definition 1. A nested balanced incomplete block design with parameters $(v, b_1, k_1, r^*, b_2, k_2\ \lambda_1, \lambda_2, m)$ is a design for v treatments, each replicated r^* times with two systems of blocks such that

a) The second system is nested within the first, with each block from the first system, called henceforth as '*block*' containing exactly m blocks from the second system, called hereafter as *'sub-blocks'*:

b) Ignoring the second system leaves a balanced incomplete block design with parameters $v, b_1, k_1, r^*, \lambda_1;$

c) Ignoring the first system leaves a balanced incomplete block design with parameters $v, b_2, k_2, r^*, \lambda_2.$

Clearly, the parameters of a NBIB design satisfy the following relations:

$$vr^* = b_1 k_1 = m b_1 k_2 = b_2 k_2, (v-1)\lambda_1 = (k_1 - 1) r^*, (v-1)\lambda_2 = (k_2 - 1) r^*.$$

Here is an example of NBIB design; here the square brackets indicate blocks and the parentheses indicate sub-blocks

$$[(2,3),(6,8)];\ [(1,3),(4,9)];\ [(1,2),(5,7)];\ [(5,6),(2,9)];\ [(4,6),(3,7)];\ [(4,5),(1,8)];$$
$$[(8,9),(3,5)];\ [(7,9),(1,6)];\ [(7,8),(2,4)];\ [(4,7),(5,9)];\ [(5,8),(6,7)];\ [(6,9),(4,8)];$$
$$[(1,7),(3,8)];\ [(2,8),(1,9)];\ [(3,9),(2,7)];\ [(1,4),(2,6)];\ [(2,5),(3,4)];\ [(3,6),(1,5)].$$

Here $v = 9, b_1 = 18, k_1 = 4, r^*=8, k_2 = 2, b_2=36, \lambda_1 = 3, \lambda_2 = 1, m=2.$

Consider now an NBIB design d with parameters $v = p, b_1, k_1, k_2 = 2, r^*$. If we identify the treatments of d as lines of a diallel experiment and perform crosses among the lines appearing in the same sub-block of d, (which is of size 2) we get a block design d^* for a complete diallel experiment involving p lines with each cross replicated $r = 2b_2/\{p(p-1)\}$ times, and $b = b_1$ blocks, each of size $k = k_1/2$. Such a design d^* is connected; also, for such a design,

$$\mathbf{C}_{d^*} = (p-1)^{-1}\, 2b(k-1)(\mathbf{I}_p - p^{-1}\mathbf{J}_{pp}), \qquad (2.4)$$

Clearly $\mathbf{C}_{d^*}$ given by (2.4) is completely symmetric and $tr\ (\mathbf{C}_{d^*}) = 2b(k-1)$ which equals the upper bound for $tr\ (\mathbf{C}_d)$ given by (2.3). Thus, one concludes that the design d^* is universally optimal in $\mathbf{D}\ (p, b, k)$. It is also easy to see that using d^*, each elementary contrast among general combining ability effects is estimated with a variance

$$(p-1)\sigma^2 / \{b(k-1)\}. \qquad (2.5)$$

Further, if the NBIB design with parameters $v = p, b_1, k_1, b_2 = b_1 k_1/2, k_2 = 2$ is such that $\lambda_2 = 1$ or equivalently,

$$b_1 k_1 = p(p-1) \qquad (2.6)$$

then the optimal design d^* for diallel crosses derived from this design has each cross replicated just once and hence uses the minimal number of experimental units. summarizing, therefore, we have the following result due to Das, Dey and Dean (1998).

Theorem 2.3: The existence of an NBIB design d with parameters $v = p, b_1 = b, b_2 = bk;$ $k_1 = 2k, k_2 = 2$ implies the existence of a universally optimal incomplete block design d^* for diallel crosses. Further, if the parameters of d satisfy (2.6), then d^* has the minimal number of experimental units.

Several families of NBIB designs, leading to optimal block designs for diallel crosses are available. See Gupta and Kageyama (1994), Das *et al.* (1998) and Parsad *et al.* for details on these.

3. Optimal Designs Based on Triangular PBIB Designs

A class of two associate class PBIB designs, called triangular incomplete block designs, can be used to derive optimal incomplete block designs for complete diallel crosses. To begin with let us recall the definition of a triangular design.

Definition 2. A binary block design with $v = p(p-1)/2$ treatments and b blocks, each of size k is called a triangular design if

(i) each treatment is replicated r times,

(ii) the treatments can be indexed by a set of two labels $(i, j), i < j, i, j = 1,2,\ldots, p$; two treatments, say (α, β) and (γ, δ) occur together in λ_i blocks if $(\alpha, \beta)\cap(\gamma, \delta)=2-i$, $i = 1, 2.$

We consider the same model as in the previous section, namely, (2.1). The results described below are due to Dey and Midha (1996) and Das *et al.* (1998).

A block design $d \in \mathbf{D}(p, b, k)$ for diallel crosses can be derived from a triangular design d_1 with parameters $v = p(p-1)/2, b, r, k, \lambda_1, \lambda_2$, by replacing a treatment (i, j) in d_1 with the cross $(i \times j), i < j, i, j = 1,2,\ldots, p.$ It can be shown that

$$\mathbf{C}_d = \theta(\mathbf{I}_p - p^{-1}\mathbf{J}_p) \qquad (3.1)$$

where $\theta = pk^{-1}\{r(k-1)-(p-2)\lambda_1\}$. Therefore, using the design d, any elementary comparison among general combining ability effects is estimated with a variance $2\sigma^2/\theta$, and the efficiency factor of the design relative to a randomized complete block design is $\theta/\{r(p-2)\}$. We now have the following result

Theorem 3.1: A block design for diallel crosses derived from a triangular design with parameters $v = p(p-1)/2, b, r, k, \lambda_1, \lambda_2$ is universally optimal over $\mathbf{D}(p, b, k)$ if

$$p(p-1)(p-2)\lambda_1 = bx\{4k - p(x+1)\} \qquad (3.2)$$

where $x = [2k/p]$. Further, when the condition in (3.2) holds, the efficiency factor is given by

$$e = p\{2k(k-1-2x) + px(x+1)\}/\{2k^2(p-2)\}. \qquad (3.3)$$

Corollary: A triangular design with parameters $v = p(p-1)/2, b, r, k, \lambda_1 = 0, \lambda_2$, leads to a universally optimal design for diallel crosses.

Several other designs, not satisfying the condition of the above corollary are also optimal, as these satisfy the condition (3.2). See Table in Das *et al.* (1998) and Table 1 in Dey and Midha (1996).

The results discussed thus far are derived under a model with no specific combining abilities. Now, suppose that the model includes specific combining ability effects as well. The interest of the experimenter may still be in estimating optimally contrast among general combining ability effects, but *in the presence of specific combining ability effects*. Chai and Mukerjee (1999) have shown that diallel cross designs derived from triangular designs satisfying (3.2), remain optimal for the general combining ability effects even when the specific combining ability effects are included in the model. Thus, the findings in Das *et al.* (1998) and Dey and Midha (1996), on triangular designs satisfying (3.2) remain *robust* under a model including specific combining ability effects.

Turning to the issue of optimality for the specific combining ability effects themselves, we have the following result due to Chai and Mukerjee (1999).

Theorem 3.2: Let d_0 be a triangular design with usual parameters $v = \binom{p}{2}, b, r, k, \lambda_1 > 0, \lambda_2 = 0$. Then d_0 is connected and is universally optimal in **D** *(p, b, k)* for any complete set of orthonormal contrasts representing the specific combining ability effects.

4. Optimal Partial Diallel Crosses

With a large number of lines *p*, a complete diallel cross may become prohibitively large and the use of a partial diallel cross is necessitated. Even in the unblocked situation, the issue of finding an optimal partial diallel cross plan has received relatively less attention. Having chosen an optimal partial diallel cross plan, further blocking of the crosses might be necessary to control the error and the literature on this aspect is also scanty. Some aspects of finding optimal partial diallel cross plans and their blocking are now discussed.

The model considered is one in (2.1). To begin with, we consider an unblocked situation and thus, the block parameters in (2.1) are assumed to be absent. Subsequently, results under a model with block effects will be discussed.

Suppose $p = mn$, where $m \geq 2$, $n \geq 3$ are integers. Let us partition the set $\{1, 2, \cdots, p\}$ into *m* mutually exclusive and exhaustive subsets $S_1, \ldots, S_m$, each having *n* elements. Let

$$d^* = \{(i \times j) : 1 \leq i \leq j \leq p, \text{ and } i, j \in S_u \text{ for some } u\} \qquad (4.1)$$

if $D(N, p)$ denotes the class of all *N*-observation partial diallel cross plans with $N < \binom{p}{2}$, then clearly, $d^* \in D(N, p)$, where $N = \frac{1}{2}mn(n-1)$. Mukerjee (1997) proved the following result.

Theorem 4.1. For each $m \geq 2$ and $n \geq 3$, the plan d^* is uniquely (up to isomorphism) E-optimal in $D(N, p)$, where $N = \frac{1}{2}mn(n-1)$. Furthermore, the plan d^* is uniquely D- and A- optimal in $D(N, p)$, for $n = 3$.

Recall that an E-optimal plan is one that, over a relevant class of competing plans, minimizes the maximum possible variance of a normalized contrast among the general combining ability effects. Similarly, a D- or A- optimal plan can be interpreted in the present context. Mukerjee (1997) also showed that for $n \geq 4$, the D- and A-efficiencies of d^* are quit high. Thus, for instance, among the 29 pairs (m, n) satisfying $m \geq 2, n \geq 4, p = mn \leq 30$, there are 24 for which the D-efficiency is as least as large as 0.90 and 15 for which this is at least 0.95; similarly in 19 cases, the A-efficiency is at least 0.90 and in 10 cases, the same is at least 0.95. From these numerical results, one might like to conjecture that d^* is indeed D- and A-optimal in $D\ (N,\ p)$ for all $n \geq 3$. However this conjecture remains to be proved and the proof of its truth (or, its falsity) appears to be highly non-trivial.

We now turn to the question of optimal and efficient blocking of partial diallel cross plans. Let $\overline{\mathbf{C}}_d$ be the information matrix of a plan d under a model that does not include block effects and $\overline{\mathbf{C}}_d$ be as in (2.2). Furthermore, let $\mathbf{D}\ (b, k, p)$ denote the class of partial diallel cross plans involving p parental lines and bk experimental units, arranged in b blocks of size k each. Then, we have the following result due to Gupta *et al.* (1995).

Lemma 4.1: For any $d \in \mathbf{D}\ (b, k, p)$,. $\overline{\mathbf{C}}_d - \mathbf{C}_d$ is nonnegative definite and $\overline{\mathbf{C}}_d = \mathbf{C}_d$ is and only if the condition

$$\mathbf{N}_d = b^{-1}\mathbf{w}_d \mathbf{1}'_b \tag{4.2}$$

holds, where $\mathbf{w}_d = \left(w_{d1}, ..., w_{dp}\right)'$.

This lemma has useful implications in the construction of optimal or efficient block designs for partial diallel cross plans. Mukerjee (1997) gives two classes of E-optimal block designs for the following two cases: (i) $n \geq 5$ is odd and (ii) $n \geq 4$ even. We do not provide details here and refer the interested readers to the original source. For some more results on partial diallel cross designs, see Singh and Hinkelmann (1995) and Das, Dean and Gupta (1998).

References

Ceranka, B. and Mejza, S. (1988). Analysis of diallel table for experiments carried out in BIB designs-mixed model. *Biomet. J.*, **30,** 3-16**.**

Chai,F.S. and Mukerjee,R. (1999). Optimal designs for diallel crosses with specific combining abilities. *Biometrika,* **86**, 453-458.

Das,A., Dey,A. and Dean,A.M. (1998). Optimal designs for diallel cross experiments. *Statistics and Probability Letters*, **36,** 427-436.

Das,M.N. and Giri,N.C. (1986). *Design and Analysis of Experiments* 2nd Ed. Wiley Eastern, New Delhi.

Dey, A. and Midha, C.K. (1996). Optimal block designs for diallel crosses. *Biometrika*, **83,** 484-489.

Ghosh,D.K. and Divecha,J. (1997). Two-associate class partially balanced incomplete block designs and partial diallel crosses. *Biometrika,* **84,** 245-248**.**

Griffing, B.(1956). Concepts of general and specific combining ability in relation to diallel crossing systems. *Aust. J. Biol. Sci.*, **9**, 463-493**.**

Gupta, S., Das,A. and Kageyama,S.(1995). Single replicate orthogonal block designs for circulant partial diallel crosses. *Commun. Statist.: Theor. Meth.*, **A24**, 2601-2607.

Gupta, S. and Kageyama, S.(1994). Optimal complete diallel crosses. *Biometrika,* **81,** 420-424.

Hinkelmann,K. (1975). Design of genetical experiments. In *A Survey of Statistical Design and Linear Models* (Ed. J. N. Srivastava), North Holland, Amsterdam, 243-269.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models* (Ed. J.N.Srivastava), North Holland, Amsterdam, 333-353.

Mukerjee,R. (1997). Optimal partial diallel crosses. *Biometrika*, **84**, 939-948.

Parsad,R., Gupta,V.K. and Srivastava,R. (1999). Universally optimal block designs for diallel crosses. *Statist. Appl.*, **11**, 35-52.

Preece, D.A. (1967). Nested balanced incomplete block designs**.** *Biometrika*, **54**,479-486.

Sharma,M.K. (1998). Partial diallel crosses through circular designs. *J. Indian Soc. Agric. Statist.,* **51(1)**, 17-27.

Singh, M. and Hinkelmann, K.(1995). Partial diallel crosses in incomplete blocks. *Biometrics,* **51**, 1302-1314**.**