Statistics & Probability Letters 67 (2004) 47–55

Optimal diallel cross designs for the interval estimation of heredity

Himadri Ghosh[a], Ashish Das[b,*]

[a] *Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110 012, India*
[b] *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 7, S.J.S. Sansanwal Marg, New Delhi 110 016, India*

Abstract

Available results on optimal block designs for diallel crosses are based on standard linear model assumptions where the general combining ability effects are taken as fixed. In many practical situations, this assumption may not be tenable since often one studies only a sample of inbred lines from a possibly large (hypothetical) population. In this paper, a random effects model is proposed that allows us to obtain an interval estimate of the ratio of variance components. We address the issue of optimal designs by considering the L-optimality criteria. Designs that are L-optimal for the estimation of heredity are obtained in the sense that the designs minimize the maximum expected normalized length of confidence intervals. The approach leads to certain connections with an optimization problem under the fixed effects model.

Keywords: L-optimality; Variance components; Interval estimation; Heredity

1. Introduction

Diallel crosses as mating designs are used to study the genetic properties of inbred lines in plant breeding experiments. Plant breeders frequently need overall information on average performance of individual inbred lines in crosses, known as general combining ability, for subsequent choosing of the best amongst them for further breeding.

Consider a hypothetical population involving a large number of lines and crosses so that all means are estimated without error. Crossing a line to several others provides the mean performance of the line in all its crosses. This mean performance, when expressed as a deviation from the mean of all crosses, is called the general combining ability of the line. Any particular cross, then, has an expected value which is the sum of the general combining abilities of its two parental lines. The cross may, however, deviate from this expected value to a greater or lesser extent. This deviation is called the specific combining ability of the two lines in combination. In statistical terms, the general combining abilities are main effects and the specific combining ability is an interaction. Griffing (1956) defines diallel crosses in terms of genotypic values where the sum of general combining abilities for the two gametes is the breeding value of the cross (i,j). Similarly, specific combining ability represents the dominance deviation value in the simplest case ignoring epistatic deviation; see Kempthorne (1969) and Mayo (1980) for details.

In practice, often a plant breeder carries out a diallel cross experiment by selecting p lines randomly from a population consisting of a large number of lines. In such a case, the expected value of an observation Y_{ij}, conditional on the realized value of the general combining ability and specific combining ability, arising out of cross (i,j) involving lines i and $j, i<j$; $i,j=1,\ldots,p$ can be modeled as

$$E(Y_{ij}) = \mu + g_i^* + g_j^* + s_{ij}^*, \tag{1.1}$$

where μ is the general mean, g_i^* (g_j^*) is the realized value of g_i (g_j), the general combining ability effect of sampled ith (jth) line and s_{ij}^* is the realized value of s_{ij}, the specific combining ability effect of cross (i,j).

Accordingly, in experimental mating design, the analysis of the observations arising out of n crosses involving p lines may be carried out based on a model

$$Y_{ijl} = \mu + g_i + g_j + e_{ijl}, \quad i<j, \tag{1.2}$$

where Y_{ijl} is the observation arising out of the lth replication of the cross (i,j), g_i is the ith line effect with $E(g_i)=0$, $\mathrm{Var}(g_i)=\sigma_g^2 \geqslant 0$, $\mathrm{Cov}(g_i,g_j)=0$, μ is the general mean and e_{ijl} is the random error component, uncorrelated with g_i, with expectation zero and variance $\sigma_e^2>0$, $1 \leqslant i<j \leqslant p$. Here μ, σ_e^2 and σ_g^2 are unknown parameters. Also, the specific combining ability effects are assumed to be negligible and have been absorbed in the error component; see Hinkelmann (1975) and Hinkelmann and Kempthorne (1963) for a discussion on this assumption. In model (1.2), μ is a fixed effect while g_i, g_j $(i<j)$ and e_{ijl} are random effects.

The basic idea in the study of variation among observations arising out of crosses is its partitioning into components attributed to different causes like additive value, dominance deviation and epistatic deviation; see Falconer (1991). The relative magnitude of these components determines the genetic properties of the population. One such property is *heredity* which is of paramount interest to plant breeders. The ratio $4\sigma_g^2/\sigma_p^2 = h^2$ gives a measure of heredity, where $\sigma_p^2 = 2\sigma_g^2 + \sigma_e^2$ is the phenotypic variance and σ_g^2 is the genotypic variance. Such a measure expresses the extent to which individual's phenotypes are determined by the genotypes.

Our primary interest is thus in $h^2 = 4\sigma_g^2/(2\sigma_g^2+\sigma_e^2)$. In order to get a good estimate of h^2 we propose optimal designs for interval estimation of σ_g^2/σ_e^2 since $h^2 = 4\sigma_g^2/(2\sigma_g^2+\sigma_e^2) = 4(\sigma_g^2/\sigma_e^2)/(2(\sigma_g^2/\sigma_e^2)+1)$. Let T be an estimator of σ_g^2/σ_e^2. Then an asymptotically unbiased estimator of h^2 is $4T/(2T+1)$. Hence an interval estimate of σ_g^2/σ_e^2 will lead to a meaningful interval estimate of h^2. The problem confronted in constructing a confidence interval on either $\sigma_g^2/(\sigma_g^2+\sigma_e^2)$ or σ_g^2/σ_e^2 has been referred in Burdick and Graybill (1992). An approximate solution to this interval estimation problem is given by Burdick et al. (1986) by employing Thomas–Hultquist (1978) approximation of χ^2 distributions under certain parameter values of σ_g^2/σ_e^2. The only exact interval estimate of σ_g^2/σ_e^2 is due to Wald (1940), which is based on iterative solutions of non-linear equations. We give a non-iterative method of constructing exact confidence interval of σ_g^2/σ_e^2 and study their expected normalized length.

An experiment is carried out using a diallel cross design with p lines and n crosses. A diallel cross experiment is said to be complete, if each of the $\binom{p}{2}$ crosses appears at least once in the experiment, otherwise it is said to be a partial diallel cross experiment and then necessarily $n < \binom{p}{2}$. Most of the available literature on optimal designs for diallel crosses is based on standard linear model assumptions where the general combining ability effects are taken as fixed and the primary interest lies in comparing the lines with respect to their general combining ability effects. Under such a model, among others, Gupta and Kageyama (1994), Dey and Midha (1996), Mukerjee (1997), Das et al. (1998b) and Das et al. (1998a) have characterized and obtained optimal completely randomized designs and incomplete block designs for diallel crosses. When one is studying only a sample of inbred lines from a possibly large hypothetical population the fixed effects assumption may not be tenable. A random effects model is proposed in this paper that allows us to obtain an interval estimate of the ratio of variance components. We address the issue of optimal designs by considering the L-optimality criteria. We obtain designs that are L-optimal for the estimation of heredity in the sense that the designs minimize the maximum expected normalized length of the h confidence intervals based on h distinct eigenvalues of the information matrix. The approach leads to certain connections with an optimization problem under the fixed effects model.

In Section 2, under blocked and unblocked models, we first obtain the interval estimate of σ_g^2/σ_e^2. Subsequently, we obtain suitable bounds of the expected normalized length of the confidence interval. In Section 3, we characterize L-optimal designs.

2. Confidence intervals in diallel cross experiments

Consider an experiment carried out using a diallel cross design with p lines and b blocks each having k crosses ($n = bk$). Here our model is

$$Y = \mu 1_n + D_2'\beta + D_1'g + e, \tag{2.1}$$

where Y is the vector of n observations, g is the $p \times 1$ vector of general combining ability effects with $E(g) = 0$ and $\mathrm{Var}(g) = \sigma_g^2 I$, β is the fixed effect due to blocks and e is the error vector with $E(e) = 0$ and $\mathrm{Var}(e) = \sigma_e^2 I$. Also, $D_1 = (d_{uv}^{(1)})$ is the $p \times n$ line versus observation incidence matrix with $d_{uv}^{(1)} = 1$ if vth observation is out of a cross involving the uth line and $d_{uv}^{(1)} = 0$ otherwise. Similarly, $D_2 = (d_{uv}^{(2)})$ is the $b \times n$ block versus observation incidence matrix with $d_{uv}^{(2)} = 1$ if the vth observation arise from the uth block and $d_{uv}^{(2)} = 0$ otherwise. Equivalently, we can write (2.1) as

$$Y = X \begin{pmatrix} \mu \\ \beta \\ g \end{pmatrix} + e,$$

where $X = (1\ D_2'\ D_1')$. Here, $E(Y) = \mu 1_n, \mathrm{Var}(Y|\sigma_g^2, \sigma_e^2) = \sigma_g^2 D_1' D_1 + \sigma_e^2 I_n$. We assume that $Y \sim N_n(\mu 1_n + D_2'\beta, \sigma_g^2 D_1' D_1 + \sigma_e^2 I_n)$, where $N_n(\mu, \Sigma)$ denotes n-variate normal distribution with mean vector $\boldsymbol{\mu}$ and dispersion matrix Σ. Let $G = D_1D_1' = (g_{ij})$ and $s = D_1 1$. Using the definition of D_1 it can be verified that for $i \neq j$, g_{ij} gives the number of times cross (i,j) appears in the design, $g_{ii} = s_i$ where s_i is the replication of the ith line. Let $N = D_1D_2' = (n_{ij})$ be the incidence matrix with n_{ij} indicating the number of times the ith line occurs in the jth block. Also, let $C = G - k^{-1}NN'$. Then $\mathrm{Rank}(C) \leqslant p-1$ and throughout we assume $\mathrm{Rank}(C) = p-1$. Here 1_t represents a $t \times 1$ column vector of all ones and I_t denotes an identity matrix of order t. In situations where the order is evident from the context, we write, respectively, 1 and I instead of 1_t and I_t.

Let H be an $n \times (n-b)$ matrix such that the columns of H form an orthonormal basis of the orthocomplement of the space spaned by $(1 \;\; D_2')$ in $\mathscr{R}^n$. Thus $H'H = I_{n-b}$ and $HH' = I - (1 \;\; D_2')[(1 \;\; D_2')'(1 \;\; D_2')]^{-}(1 \;\; D_2')'$ where T^- is a generalized-inverse of a matrix T. Note that $D_2H = 0$ and $1_n'H = 0$. Hence, $Z = H'Y \sim N_{n-b}(0, \sigma_g^2 H'D_1'D_1H + \sigma_e^2 I_{n-b})$.

We observe that the non-zero eigenvalues of $H'D_1'D_1H$ are the same as the non-zero eigenvalues of $D_1HH'D_1' = D_1(I - (1 \;\; D_2')\begin{pmatrix} 0 & 0 \\ 0 & k^{-1}I_b \end{pmatrix}(1 \;\; D_2')')D_1' = D_1D_1' - (1/k)NN' = G - (1/k)NN' = C$. This implies that the eigenvalues of $H'D_1'D_1H$ are zero with multiplicity $((n-b)-(p-1)) = n_e$ and the remaining $(p-1)$ eigenvalues are identical to the non-zero eigenvalues of the C-matrix.

Define $0 = \lambda_0^* < \lambda_1^* < \cdots < \lambda_h^*$ as the $h+1$ distinct eigenvalues of $H'D_1'D_1H$ with multiplicities $m_0^* = n_e, m_1^*, \ldots, m_h^*$, respectively. Now there exits an orthogonal matrix $P = (P_{(0)} \; P_{(1)} \; \cdots \; P_{(h)})$ of order $n-b$ such that $H'D_1'D_1H = P\begin{pmatrix} 0 & 0 \\ 0 & \Delta \end{pmatrix}P'$ where $\Delta = \mathrm{diag}(\lambda_1^* I_{m_1^*}, \lambda_2^* I_{m_2^*}, \ldots, \lambda_h^* I_{m_h^*})$. Consider the transformation $Z_{(i)}^* = P_{(i)}'Z$, $i = 0, \ldots, h$. Then it follows that $Q_i = Z_{(i)}^{*\prime}Z_{(i)}^*$ are independent and

$$(\sigma_g^2\lambda_i^* + \sigma_e^2)^{-1}Q_i \tag{2.2}$$

follows a χ^2 distribution with m_i^* degrees of freedom, $i = 0, \ldots, h$. We now construct the confidence interval of σ_g^2/σ_e^2 with confidence coefficient $1-\alpha$. From (2.2) we get that for $i = 0, 1, 2, \ldots, h$, $(\sigma_g^2\lambda_i^* + \sigma_e^2)^{-1}Q_i$ follows a χ^2-distribution with m_i^* degrees of freedom and furthermore, they are independently distributed.

For $i = 1, \ldots, h$, let $L_i = F_{1-\alpha_1, m_i^*, n_e}$ and $U_i = F_{\alpha_2, m_i^*, n_e}$, where $\alpha_1 + \alpha_2 = \alpha$ and F_{α, p_1, p_2} is the upper $100\alpha\%$ cut-off point of the central F-distribution with p_1 and p_2 degrees of freedom. Then

$$\Pr\left[L_1 \leqslant \frac{m_1^{*-1}(\lambda_1^*\sigma_g^2 + \sigma_e^2)^{-1}Q_1}{n_e^{-1}\sigma_e^{-2}Q_0} \leqslant U_1\right] = 1-\alpha$$

$$\Leftrightarrow \; Pr\left[\frac{n_eQ_1}{m_1^*U_1\lambda_1^*Q_0} - \frac{1}{\lambda_1^*} \leqslant \frac{\sigma_g^2}{\sigma_e^2} \leqslant \frac{n_eQ_1}{m_1^*L_1\lambda_1^*Q_0} - \frac{1}{\lambda_1^*}\right] = 1-\alpha,$$

giving a confidence interval

$$I_1 = \left(\frac{n_eQ_1}{m_1^*U_1\lambda_1^*Q_0} - \frac{1}{\lambda_1^*}, \quad \frac{n_eQ_1}{m_1^*L_1\lambda_1^*Q_0} - \frac{1}{\lambda_1^*}\right) \tag{2.3}$$

of σ_g^2/σ_e^2 with confidence coefficient $1-\alpha$.

For $I = \{x\colon a \leqslant x \leqslant b\}$, we define $l(I) = b - a$. Now, using the result $E(F_{p_1,p_2}) = p_2/(p_2 - 2)$, where F_{p_1,p_2} follows an F-distribution with p_1 and $p_2(>2)$ degrees of freedom, we get

$$E(l(I_1)) = E\left(\frac{n_e Q_1}{m_1^* L_1 \lambda_1^* Q_0} - \frac{n_e Q_1}{m_1^* U_1 \lambda_1^* Q_0}\right)$$

$$= E\left(\frac{Q_1(\lambda_1^* \sigma_g^2 + \sigma_e^2)^{-1} m_1^{*^{-1}}}{Q_0 \sigma_e^{-2} n_e^{-1}} \left(\frac{\sigma_e^2 + \lambda_1^* \sigma_g^2}{\sigma_e^2}\right) \frac{1}{L_1 \lambda_1^*}\right)$$

$$- E\left(\frac{Q_1(\lambda_1^* \sigma_g^2 + \sigma_e^2)^{-1} m_1^{*^{-1}}}{Q_0 \sigma_e^{-2} n_e^{-1}} \left(\frac{\sigma_e^2 + \lambda_1^* \sigma_g^2}{\sigma_e^2}\right) \frac{1}{U_1 \lambda_1^*}\right)$$

$$= \left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right) \frac{1}{L_1} E(F_{m_1^*, n_e}) - \left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right) \frac{1}{U_1} E(F_{m_1^*, n_e}).$$

Thus,

$$E(l(I_1)) = \left(\frac{n_e}{n_e - 2}\right) \left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right) \left(\frac{1}{L_1} - \frac{1}{U_1}\right). \tag{2.4}$$

Now, the pair (L_1, U_1) is not unique for setting up the confidence interval with confidence coefficient $1 - \alpha$. Hence we normalize the expected length of the confidence interval I_1 by dividing the distance between L_1 and U_1 defined by $1/L_1 - 1/U_1$. Note that the distance $d(a,b) = |1/a - 1/b|$, $a > 0, b > 0$, satisfies the three properties of the distance function since (i) $d(a,b) = |1/a - 1/b| \geqslant 0$, $d(a,b) = 0$ if and only if $a = b$; (ii) $d(a,b) = |1/a - 1/b| = |1/b - 1/a| = d(b,a)$; (iii) for $c > 0$, $d(a,b) = |1/a - 1/b| = |1/a - 1/c + 1/c - 1/b| \leqslant |1/a - 1/c| + |1/c - 1/b| = d(a,c) + d(c,b)$. Hence, the expected normalized length comes out as

$$E(l_N(I_1)) = E(l(I_1)) \Big/ \left(\frac{1}{L_1} - \frac{1}{U_1}\right) = \left(\frac{n_e}{n_e - 2}\right) \left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right). \tag{2.5}$$

The other $h - 1$ confidence intervals of σ_g^2/σ_e^2 are constructed, on similar lines, and are given by

$$I_i = \left(\frac{n_e Q_i}{m_i^* U_i Q_0 \lambda_i^*} - \frac{1}{\lambda_i^*},\quad \frac{n_e Q_i}{m_i^* L_i Q_0 \lambda_i^*} - \frac{1}{\lambda_i^*}\right), \quad i = 2, \ldots, h \tag{2.6}$$

each with confidence coefficient $1 - \alpha$. Then the expected normalized length of the ith confidence interval is

$$E(l_N(I_i)) = E(l(I_i)) \Big/ \left(\frac{1}{L_i} - \frac{1}{U_i}\right) = \left(\frac{1}{\lambda_i^*} + \frac{\sigma_g^2}{\sigma_e^2}\right) \left(\frac{n_e}{n_e - 2}\right), \quad i = 2, \ldots, h. \tag{2.7}$$

Let $\alpha_1^*, \alpha_2^*, \ldots, \alpha_h^*$ be positive numbers such that $\sum_{i=1}^h \alpha_i^* = \alpha$. On lines similar to (2.3) and (2.6), for $i = 1, \ldots, h$ define I_i^* as a confidence interval of σ_g^2/σ_e^2 with confidence coefficient $1 - \alpha_i^*$. Then by applying Bonferroni's inequality we have,

$$Pr\left[Y:\ \frac{\sigma_g^2}{\sigma_e^2}\in\bigcap_{i=1}^{h} I_i^*\right] = Pr\left[\bigcap_{i=1}^{h}\left\{Y:\ \frac{\sigma_g^2}{\sigma_e^2}\in I_i^*\right\}\right] \geqslant \sum_{i=1}^{h}(1-\alpha_i^*)-(h-1)=1-\alpha,$$

which gives the confidence interval $I^*=\bigcap_{i=1}^{h} I_i^*$ of σ_g^2/σ_e^2 with confidence coefficient at least $1-\alpha$. Thus we may construct infinitely many confidence intervals of σ_g^2/σ_e^2. We define the expected normalized length of the confidence interval I^* as $E(l_N(I^*))=\max_{1\leqslant i\leqslant h} E(l(I^*))/(1/L_i^*-1/U_i^*)$. Here we have normalized by the respective distances between L_i^* and U_i^* which are the lower and upper cut off points of the F-distribution with probability α_i^*. Now using the fact $E(l_N(I_i^*))=E(l_N(I_i))$, $i=1,\dots,h$, for any choice of $\alpha_1^*,\alpha_2^*,\dots,\alpha_h^*$ and observing that the confidence interval I^* being subset of I_i^* for each $i=1,\dots,h$, we have $E(l(I^*))/(1/L_i^*-1/U_i^*)\leqslant E(l_N(I_i^*))=E(l_N(I_i))$, $i=1,\dots,h$. Now, defining $E(l_{N_i}(I^*))=E(l(I^*))/(1/L_i^*-1/U_i^*)$, we get for all I^*, $E(l_N(I^*))=\max_{1\leqslant i\leqslant h}E(l_{N_i}(I^*))\leqslant \max_{1\leqslant i\leqslant h}E(l_N(I_i))=(n_e/(n_e-2))(1/\lambda_1^*+\sigma_g^2/\sigma_e^2)=E(l_N(I_1))$.

Now, we define $\phi=\max_{1\leqslant i\leqslant h}E(l_N(I_i))$ which represents the maximum loss due to h individual confidence intervals with confidence coefficient $1-\alpha$. Furthermore, for every $i=1,\dots,h$

$$E(l_N(I_i))\leqslant\phi=\left(\frac{n_e}{n_e-2}\right)\left(\frac{1}{\lambda_1^*}+\frac{\sigma_g^2}{\sigma_e^2}\right)=E(l_N(I_1)). \tag{2.8}$$

It is to be noted that the upper and lower confidence limits of I^* comes out as the order statistic of the upper and lower confidence limits of I_i^*, $i=1,\dots,h$. Further the order statistic is based on two sets of h random variables which are neither independently nor identically distributed. We have taken this detour in order to set a well defined criterion to carry out the design optimization. For comparing designs we set an upper bound of the expected normalized length of I^*. One of the possible ways to achieve this is provided by our workable measure ϕ.

The results for interval estimation of σ_g^2/σ_e^2 under unblocked diallel cross experiments can be obtained as a special case of the above results by taking the number of blocks as one. The maximum expected normalized length of the interval estimate of σ_g^2/σ_e^2 under an unblocked model reduces to $\phi_0=((n-p)/(n-p-2))(1/\lambda_1+\sigma_g^2/\sigma_e^2)$, where λ_1 is the minimum non-zero eigenvalue of $C_0=G-(1/n)ss'$, the C-matrix under the unblocked setup.

3. Optimal designs

In the previous section, we have explicitly obtained the maximum expected normalized length (ϕ and ϕ_0) of the interval estimate of σ_g^2/σ_e^2 under a blocked and an unblocked model. Our objective in obtaining an optimal design would be to minimize $\phi_0=((n-p)/(n-p-2))(1/\lambda_1+\sigma_g^2/\sigma_e^2)$ in case of an unblocked model and to minimize $\phi=(n_e/(n_e-2))(1/\lambda_1^*+\sigma_g^2/\sigma_e^2)$ in case of a blocked model. Let $\mathscr{D}(p,n)$ be the class of unblocked diallel cross designs involving p lines and n crosses and $\mathscr{D}(p,b,k)$, the class of diallel cross designs with p lines arranged in b blocks of k crosses each. For a design d, let the non-zero eigenvalues of C_{0d} (C_d) be $\lambda_{1d}<\lambda_{2d}<\cdots<\lambda_{hd}$ ($\lambda_{1d}^*<\lambda_{2d}^*<\cdots<\lambda_{hd}^*$) with respective multiplicities $m_{1d},m_{2d},\dots,m_{hd}$ ($m_{1d}^*,m_{2d}^*,\dots,m_{hd}^*$). A design d^* will be said to be L-optimal if, among all designs in $\mathscr{D}$, d^* minimizes ϕ_{0d} (or ϕ_d). Thus we see a connection between L-optimal designs in our set-up and E-optimal diallel cross designs under a fixed effects model.

It is well known that under fixed effects model, a complete diallel cross design is universally optimal in $\mathscr{D}(p,n)$. Since a universally optimal design is E-optimal as well, it follows that complete diallel cross designs are L-optimal in $\mathscr{D}(p,n)$ under our setup.

Under the fixed effects model, Gupta and Kageyama (1994), Dey and Midha (1996) and Das et al. (1998b) have obtained universally optimal (and hence E-optimal) diallel cross designs. It thus follows that their designs are L-optimal under our setup.

The close connection between nested balanced incomplete block design of Preece (1967) and optimal designs for diallel crosses under a fixed effects model was first observed by Gupta and Kageyama (1994). A nested balanced incomplete block design with parameters $(v, b_1, k_1, r, \mu_1, b_2, k_2, \mu_2, m)$ 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 usual parameters v, b_1, k_1, r, μ_1; (c) ignoring the first system leaves a balanced incomplete block design with parameters v, b_2, k_2, r, μ_2.

Consider now a nested balanced incomplete block 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 cross experiment and perform crosses among the lines appearing in the same sub-block of d, we get a block design d^* for a diallel cross experiment involving p lines with $v_c = p(p-1)/2$ crosses, each replicated $r = 2b_2/\{p(p-1)\}$ times, and $b=b_1$ blocks, each of size $k = k_1/2$. Such a design $d^* \in \mathscr{D}(p,b,k)$ and is universally optimal in $\mathscr{D}(p,b,k)$ under the fixed effects model. Summarizing, therefore, we have

Theorem 3.1. *The existence of a nested balanced incomplete block design d with parameters* $v = p, b_1 = b, b_2 = bk, k_1 = 2k, k_2 = 2$ *implies the existence of a L-optimal incomplete block design* d^* *for diallel crosses.*

The construction methods and elaborate tables of nested balanced incomplete block designs are available in a recent review paper by Morgan et al. (2001). The tables in their paper provide solutions to our L-optimal diallel cross designs within the parametric range $2k < p < 16$, $s \leqslant 30$. The case $2k = p$ is dealt in Gupta and Kageyama (1994). The nested balanced incomplete block designs have been extended to nested balanced block designs and a series of designs, L-optimal under our set-up, is given in Das et al. (1998b).

Mukerjee (1997) has obtained E-optimal partial diallel cross designs under the fixed effects model. Following Mukerjee (1997) we now present L-optimal designs for the unblocked case.

Let $p = n_1 n_2$ where $n_1 \geqslant 2$, $n_2 \geqslant 3$. Partition the set $\{1, \ldots, p\}$ into n_1 mutually exclusive and exhaustive subsets $S_1, \ldots, S_{n_1}$ each of cardinality n_2. Let

$$d_1^* = \{(i,j) : 1 \leqslant i < j \leqslant p \quad \text{and } i,j \in S_u \text{ for some } u\}. \tag{3.1}$$

Then $d_1^* \in \mathscr{D}(p,n)$, where $n = \frac{1}{2} n_1 n_2 (n_2 - 1)$, and D_{1d^*} is the incidence matrix of a group divisible design with the usual parameters $p = n_1 n_2, k = 2, \lambda_1 = 1, \lambda_2 = 0$.

Theorem 3.2. *For each* $n_1 \geqslant 2$ *and* $n_2 \geqslant 3$, *up to isomorphism, the design* d_1^* *is uniquely L-optimal in* $\mathscr{D}(p,n)$, *where* $p = n_1 n_2$ *and* $n = \frac{1}{2} n_1 n_2 (n_2 - 1)$.

Example 3.1. Suppose we have $p = 12$ lines and $n = 18$ crosses. Then $n_1 = 3$, $n_2 = 4$ and the subsets are $S_1 = \{1,2,3,4\}$, $S_2 = \{5,6,7,9\}$, $S_3 = \{9,10,11,12\}$. Consider the following design: $\{(1,2);(1,3);(1,4);(2,3);(2,4);(3,4);(5,6);(5,7);(5,8);(6,7);(6,8);(7,8);(9,10);(9,11);(9,12);(10,11);(10,12);(11,12)\}$. Following Theorem 3.2, this design is L-optimal in $\mathscr{D}(12,18)$.

Let $p = n_1n_2 + t$, where $n_1 \geqslant 2$, $n_2 \geqslant 3$ and t $(1 \leqslant t \leqslant n_1 - 1)$ are positive integers. Partition $\{1,\dots,p\}$ into n_1 mutually exclusive and exhaustive subsets $S_1,\dots,S_{n_1}$ such that $S_1,\dots,S_{n_1-t}$ have cardinality n_2 and $S_{n_1-t+1},\dots,S_{n_1}$ have cardinality $n_2 + 1$. Analogous to (3.1), let

$$d_2^* = \{(i,j) : 1 \leqslant i < j \leqslant p \quad \text{and } i,j \in S_u \text{ for some } u\}. \tag{3.2}$$

Then $d_2^* \in \mathscr{D}(p,n)$, where $n = \frac{1}{2}n_1n_2(n_2 - 1) + n_2t$.

Theorem 3.3. *For $n_1 \geqslant 2$, $n_2 \geqslant 3$, $p = n_1n_2 + t$, $n = \frac{1}{2}n_1n_2(n_2 - 1) + n_2t$ and $1 \leqslant t \leqslant n_1 - 1$, the design d_2^* is L-optimal in $\mathscr{D}(p,n)$, provided $(n_1 - t)n_2f > 1$, where $f = n^{-1}(n_2 - 1)^2 - p^{-1}(n_2 - 2)$.*

Example 3.2. Suppose we have $p = 13$ lines and $n = 22$ crosses. Then $n_1 = 3$, $n_2 = 4$, $t = 1$ and the subsets are $S_1 = \{1,2,3,4\}$, $S_2 = \{5,6,7,9\}$, $S_3 = \{9,10,11,12,13\}$. Consider the following design: $\{(1,2);\ (1,3);\ (1,4);\ (2,3);\ (2,4);\ (3,4);\ (5,6);\ (5,7);\ (5,8);\ (6,7);\ (6,8);\ (7,8);\ (9,10);\ (9,11);\ (9,12);\ (9,13);\ (10,11);\ (10,12);\ (10,13);\ (11,12);\ (11,13);\ (12,13)\}$. Following Theorem 3.3, since $(n_1 - t)n_2f = 2.04 > 1$, this design is L-optimal in $\mathscr{D}(13,22)$.

The condition in Theorem 3.3 holds a large number of cases over a practicable range. Thus, among the 79 cases of (n_1,n_2,t) satisfying $n_1 \geqslant 2$, $n_2 \geqslant 3$, $1 \leqslant t \leqslant n_1 - 1$, $p = n_1n_2 + t \leqslant 30$, there are as many as 57 where the condition holds and hence d_2^* is L-optimal.

The notion of orthogonal blocking for designs in diallel cross experiments was introduced by Gupta et al. (1995). The blocking of optimal designs of Theorems 3.2 and 3.3 is given in Mukerjee (1997) where orthogonal blocking has been achieved for designs corresponding to Theorem 3.2. Thus, Mukerjee's method of construction of orthogonal block designs lead to L-optimal diallel cross block designs in $\mathscr{D}(p,b,k)$.

Example 3.3. Consider the following design (rows are blocks) with parameters $p = 12$, $b = 3$ and $k = 6$:

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

This design is L-optimal in $\mathscr{D}(12,3,6)$.

In our model (2.1) we may consider β to be a random effects block parameter. Such a consideration do not alter the optimality results obtained here. With the increase in the number of lines, the optimality criteria based on the interval estimation of $h^2 = 4\sigma_g^2/(2\sigma_g^2 + \sigma_e^2)$ is same as that obtained for the interval estimation of σ_g^2/σ_e^2. Thus the design optimality results obtained here would remain valid for estimation of heredity.

Acknowledgements

The authors thank the referees for useful comments on an earlier draft.

References

Burdick, R.K., Graybill, F.A., 1992. Confidence Intervals on Variance Components. Marcel Dekker, New York.
Burdick, R.K., Maqsood, F., Graybill, F.A., 1986. Confidence intervals on the intraclass correlation in the unbalanced one-way classification. Comm. Statist. Theory Methods 15, 3353–3378.
Das, A., Dean, A.M., Gupta, S., 1998a. On optimality of some partial diallel cross designs. Sankhyā Ser. B 60, 511–524.
Das, A., Dey, A., Dean, A.M., 1998b. Optimal block designs for diallel cross experiments. Statist. Probab. Lett. 36, 427–436.
Dey, A., Midha, C.K., 1996. Optimal block designs for diallel crossses. Biometrika 83, 484–489.
Falconer, D.S., 1991. Introduction to Quantitative Genetics. Oliver & Boyd, Edinburgh.
Griffing, B., 1956. Concepts of general and specific combining ability in relation to diallel crossing systems. Austral. J. Biol. Sci. 9, 463–493.
Gupta, S., Kageyama, S., 1994. Optimal complete diallel crosses. Biometrika 81, 420–424.
Gupta, S., Das, A., Kageyama, S., 1995. Single replicate orthogonal block designs for circulant partial diallel crosses. Comm. Statist. Theory Methods 24, 2601–2607.
Hinkelmann, K., 1975. Design of genetical experiments. In: Srivastava, J.N. (Ed.), A Survey of Statistical Design and Linear Models. pp. 243–266. Amsterdam, North-Holland.
Hinkelmann, K., Kempthorne, O., 1963. Two classes of group divisible partial diallel crosses. Biometrika 50, 281–291.
Kempthorne, O., 1969. An Introduction to Genetic Statistic. Iowa State University Press, Iowa.
Mayo, O., 1980. The Theory of Plant Breeding. Clarendon Press, Oxford.
Morgan, J.P., Preece, D.A., Rees, D.H., 2001. Nested balanced incomplete block designs. Discrete Math. 231, 351–389.
Mukerjee, R., 1997. Optimal partial diallel crosses. Biometrika 84, 939–948.
Preece, D.A., 1967. Nested balanced incomplete block designs. Biometrika 54, 479–486.
Thomas, J.D., Hultquist, R.A., 1978. Interval estimation for the unbalanced case of the one way random effects model. Ann. Statist. 6, 582–587.
Wald, A., 1940. A note on the analysis of variance with unequal class frequencies. Ann. Math. Statist. 11, 96–100.