Optimal Diallel Cross Designs for the Interval Estimation of Heredity

Himadri Ghosh
and
Ashish Das

March 15, 2002
isid/ms/2002/02

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

Optimal Diallel Cross Designs for the Interval Estimation of Heredity

By HIMADRI GHOSH
Indian Agricultural Statistics Research Institute, Library Avenue,
New Delhi 110 012, India
him_adri@iasri.delhi.nic.in

and ASHISH DAS
Statistics and Mathematics Unit, Indian Statistical Institute,
New Delhi 110 016, India
ashish@isid.ac.in

Summary

The results on optimal diallel cross designs 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. A random effects model is proposed in this paper that allows us to obtain an interval estimate of a ratio of the variance components. We address the issue of optimal designs by considering the D_l-optimality criteria. Designs that are D_l-optimal for the estimation of heredity are obtained in the sense that the designs minimize the maximum expected length of the h confidence intervals. The approach leads to certain connections with the optimization problem under the fixed effects model.

Some key words : D_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 the best amongst them for further breeding. For this purpose diallel crossing techniques are employed.

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. Since we are observing a sample from a large hypothetical population of lines and crosses, 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 $j, i < j;\ i, j = 1, \ldots, p$ is 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 i-th (j-th) 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 will be carried out based on a model

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

where Y_{ijl} is the observation arising out of the l-th replication of the cross (i, j), g_i is the i-th line effect with $E(g_i) = 0,\ Var(g_i) = \sigma_g^2 \geq 0,\ Cov(g_i, g_j) = 0$, μ is the general mean and e_{ijl} is the random error component, uncorrelated with g_i, with expectation zero and variance $\sigma_e^2 > 0,\ 1 \leq i < j \leq p$. 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 & Kempthorne (1963). This allows us to estimate the genetic variance components leading to general combining ability analysis in random effects. In addition to the fact that our model simplifies the estimation process, we observe that Kempthorne & Curnow (1961) suggested to ignore the specific combining ability effects in the model when the purpose of the study is to estimate the yielding capacities. In the model, as given in (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 of 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 $h^2 = \frac{4\sigma_g^2}{2\sigma_g^2+\sigma_e^2} = \frac{4(\sigma_g^2/\sigma_e^2)}{2(\sigma_g^2/\sigma_e^2)+1}$. Let T be an estimator of σ_g^2/σ_e^2. Then an asymptotically unbiased estimator of h^2 is $\frac{4T}{2T+1}$. 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 refered in Burdick & Graybill (1992). An approximate solution to this interval estimation problem is given by Burdick, Maqsood & Graybill (1986) by employing Thomas-Hultquist 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 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 appear atleast 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 theory of optimal diallel cross designs 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 & Kageyama (1994), Dey & Midha (1996), Mukerjee (1997), Das, Dey & Dean (1998) and Das, Dean & Gupta (1998) have characterised and obtained optimal completely randomised designs and incomplete block designs for diallel crosses. In many practical situations, the fixed effects assumption may not be tenable when one is studying only a sample of inbred lines from a possibly large hypothetical population. A random effects model is proposed in this paper that allows us to obtain an interval estimate of the ratio of the variance components. We address the issue of optimal designs by considering the D_l-optimality criteria. We obtain designs that are D_l-optimal for the estimation of heredity in the sense that the designs minimize the maximum expected lengths of the k confidence intervals based on k distinct eigenvalues of the information matrix. The approach leads to certain connections with the optimaztion problem under the fixed effects model.

In §2 and §3, under unblocked and blocked models, we first obtain the interval estimate of σ_g^2/σ_e^2 and then obtain suitable bounds of the expected confidence interval of these estimates. In §4 we characterize D_l-optimal designs.

2. Unblocked Diallel Cross Experiments

Consider p inbred lines sampled from a hypothetical population involving a large number of lines. The yield Y_{ijl} arising out of the lth replication of cross (i, j) involving ith and jth sampled lines is modeled as in (1.2).

When an experiment is carried out using unblocked diallel cross design with p lines and n crosses, we can represent the model (1.2) in matrix notation as

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

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 $Var(g)=\sigma_g^2 I_p$, e is the error vector with $E(e)=0$ and $Var(e)=\sigma_e^2 I_n$, and $D_1=(d_{uv}^{(1)})$ is the $p\times n$ line versus observation incidence matrix with $d_{uv}^{(1)}=1$ if v-th observation is out of a cross involving the u-th line and $d_{uv}^{(1)}=0$ otherwise. 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. We assume that D_1 has full row rank. Equivalently, (2.1) can be written as

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

where $X=(1_n\ \ D_1')$. Here,

$$E(Y)=\mu 1_n,\ \text{Var }(Y|\sigma_g^2,\sigma_e^2)=\sigma_g^2 D_1'D_1+\sigma_e^2 I_n. \tag{2.2}$$

As usual, we assume that

$$Y\sim N_n(\mu 1_n,\sigma_g^2 D_1'D_1+\sigma_e^2 I_n). \tag{2.3}$$

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 i-th line. Also, since we assume $Rank(D_1)=p$, G is symmetric with $Rank(G)=p$ and $\text{tr}(G)=2n$ where for a square matrix A, $\text{tr}(A)$ stands for the trace. Let $C_0=G-\frac{1}{n}ss'$ where $s=(s_1,s_2,\ldots,s_p)'$. Then, $C_0 1=0$ and $Rank(C_0)\leq p-1$. However, since $Rank(D_1)=p$, it follows that $Rank(C_0)=p-1$.

Let H be an $n\times(n-1)$ matrix such that the columns of H form an orthonormal basis of the orthocomplement of the space spaned by 1_n in $\mathcal{R}^n$. Thus $H'H=I_{n-1}$, $HH'=I-11'/n$ and $Z=H'Y\sim N_{n-1}(0,\sigma_g^2 H'D_1'D_1H+\sigma_e^2 I_{n-1})$.

We observe that the non-zero eigenvalues of $H'D_1'D_1H$ are the same as the non-zero eigenvalues of $D_1HH'D_1'=D_1(I_n-\frac{1}{n}11')D_1'=D_1D_1'-\frac{1}{n}ss'=G-\frac{1}{n}ss'=C_0$. This implies that the eigenvalues of $H'D_1'D_1H$ are zero with multiplicity $(n-p)$ and the remaining $(p-1)$ eigenvalues are identical to the eigenvalues of C_0-matrix.

Let there be $h+1$ distinct eigenvalues of $H'D_1'D_1H$ denoted by $0=\lambda_0<\lambda_1<\lambda_2<\cdots<\lambda_h$ and let their respective multiplicities be $m_0=n-p,m_1,m_2,\ldots,m_h$. Note that, as a consequence of the argument in the previous paragraph, the non-zero eigenvalues of C_0 are $\lambda_1<\lambda_2<\cdots<\lambda_h$ with respective multiplicities $m_1,m_2,\ldots,m_h$.

Now there exits a matrix P such that $H'D_1'D_1H=P\Delta P'$ where P is a $(n-1)\times(p-1)$ matrix such that $P'P=I_{p-1}$ and $\Delta=diag(\lambda_1 I_{m_1},\lambda_2 I_{m_2},\ldots,\lambda_h I_{m_h})$. Now, there exits a matrix $\bar{P}$ of order $(n-1)\times(n-p)$ such that $P^*=(\bar{P}\ \ P)$ is an orthogonal matrix of order $n-1$. Applying the transformation $Z^*=P^{*'}Z$, it is easy to see that $Z^*=P^{*'}Z\sim N_{n-1}(0,\Sigma)$ where $\Sigma=\begin{pmatrix}\sigma_e^2 I_{n-p} & 0\\ 0 & \sigma_g^2\Delta+\sigma_e^2 I_{p-1}\end{pmatrix}$.

Partitioning P, we write $P=(P_1\ \cdots\ P_h)$ where $P_i, i=1,\ldots,h$, corresponds to an $(n-1)\times m_i$ matrix whose columns are orthogonal eigenvectors of $H'D_1'D_1H$ corresponding to eigenvalue

λ_i. Also, let $P_0 = \bar{P}$ and $Z_i^* = P_i' Z,\ i = 0, \ldots, h$. Then $Q_i = Z_i^{*'} Z_i^* = Z' P_i P_i' Z,\ i = 0, 1, \ldots, h$ are independent and

$$(\sigma_g^2 \lambda_i + \sigma_e^2)^{-1} Q_i, \tag{2.4}$$

follows a χ^2-distribution with m_i degrees of freedom, $i = 0, 1, \ldots, h$.

We now construct the confidence interval of σ_g^2/σ_e^2 with confidence coefficient $1-\alpha$. From (2.4) 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 $L_i = F_{1-\alpha/2, m_i, n-p}$ and $U_i = F_{\alpha/2, m_i, n-p}$. Then

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

$$\Leftrightarrow Pr\left[\frac{L_1 m_1 Q_0}{(n-p)Q_1} \le \frac{\sigma_e^2}{\lambda_1\sigma_g^2 + \sigma_e^2} \le \frac{U_1 m_1 Q_0}{(n-p)Q_1}\right] = 1-\alpha$$

$$\Leftrightarrow Pr\left[\frac{(n-p)Q_1}{U_1 m_1 Q_0} \le 1 + \lambda_1 \frac{\sigma_g^2}{\sigma_e^2} \le \frac{(n-p)Q_1}{L_1 m_1 Q_0}\right] = 1-\alpha$$

$$\Leftrightarrow Pr\left[\frac{(n-p)Q_1}{U_1 m_1 \lambda_1 Q_0} - \frac{1}{\lambda_1} \le \frac{\sigma_g^2}{\sigma_e^2} \le \frac{(n-p)Q_1}{L_1 m_1 \lambda_1 Q_0} - \frac{1}{\lambda_1}\right] = 1-\alpha, \tag{2.5}$$

and one has a confidence interval

$$I_1 = \left(\frac{(n-p)Q_1}{U_1 m_1 \lambda_1 Q_0} - \frac{1}{\lambda_1},\ \frac{(n-p)Q_1}{L_1 m_1 \lambda_1 Q_0} - \frac{1}{\lambda_1}\right) \tag{2.6}$$

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

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

$$\begin{aligned}
E(l(I_1)) &= E\left(\frac{(n-p)Q_1}{L_1 m_1 \lambda_1 Q_0} - \frac{(n-p)Q_1}{U_1 m_1 \lambda_1 Q_0}\right) \\
&= E\left(\frac{Q_1(\lambda_1\sigma_g^2 + \sigma_e^2)^{-1} m_1^{-1}}{Q_0 \sigma_e^{-2}(n-p)^{-1}} \left(\frac{\sigma_e^2 + \lambda_1\sigma_g^2}{\sigma_e^2}\right) \frac{1}{L_1\lambda_1})\right) \\
&\quad - E\left(\frac{Q_1(\lambda_1\sigma_g^2 + \sigma_e^2)^{-1} m_1^{-1}}{Q_0 \sigma_e^{-2}(n-p)^{-1}} \left(\frac{\sigma_e^2 + \lambda_1\sigma_g^2}{\sigma_e^2}\right) \frac{1}{U_1\lambda_1}\right) \\
&= \left(\frac{1}{\lambda_1} + \frac{\sigma_g^2}{\sigma_e^2}\right) \frac{1}{L_1} E(F_{m_1,n-p}) - \left(\frac{1}{\lambda_1} + \frac{\sigma_g^2}{\sigma_e^2}\right) \frac{1}{U_1} E(F_{m_1,n-p}).
\end{aligned}$$

Thus,

$$E(l(I_1)) = \left(\frac{n-p}{n-p-2}\right)\left(\frac{1}{\lambda_1} + \frac{\sigma_g^2}{\sigma_e^2}\right)\left(\frac{1}{L_1} - \frac{1}{U_1}\right). \tag{2.7}$$

Now, the pair (L_1, U_1) is not unique as a confidence interval for σ_g^2/σ_e^2 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 $\frac{1}{L_1} - \frac{1}{U_1}$. Note that the distance $d(a,b) = |\frac{1}{a} - \frac{1}{b}|,\ a > 0, b > 0$, satisfies the three properties of the distance function since (i) $d(a,b) = |\frac{1}{a} - \frac{1}{b}| \geq 0,\ d(a,b) = 0$ if and only if $a = b$; (ii) $d(a,b) = |\frac{1}{a} - \frac{1}{b}| = |\frac{1}{b} - \frac{1}{a}| = d(b,a)$; (iii) for $c > 0,\ d(a,b) = |\frac{1}{a} - \frac{1}{b}| = |\frac{1}{a} - \frac{1}{c} + \frac{1}{c} - \frac{1}{b}| \leq |\frac{1}{a} - \frac{1}{c}| + |\frac{1}{c} - \frac{1}{b}| = d(a,c) + d(c,b)$. Hence, the normalized expected length comes out as

$$E_N(l(I_1)) = E(l(I_1))/(\frac{1}{L_1} - \frac{1}{U_1}) = (\frac{n-p}{n-p-2}) \left(\frac{1}{\lambda_1} + \frac{\sigma_g^2}{\sigma_e^2}\right). \tag{2.8}$$

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

$$I_i = \left(\frac{(n-p)Q_i}{m_i U_i Q_0 \lambda_i} - \frac{1}{\lambda_i},\ \frac{(n-p)Q_i}{m_i L_i Q_0 \lambda_i} - \frac{1}{\lambda_i}\right),\ i = 2, \ldots, h \tag{2.9}$$

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

$$E_N(l(I_i)) = \left(\frac{n-p}{n-p-2}\right) \left(\frac{1}{\lambda_i} + \frac{\sigma_g^2}{\sigma_e^2}\right),\ i = 2, \ldots, h. \tag{2.10}$$

The motivation for quantifying the loss function of interval estimation is easy to see since we have constructed the h confidence intervals I_i (each with confidence coefficient $1 - \alpha$), $i = 1, \ldots, h$, and then considered the normalized expected length(s), $E_N(l(I_i)),\ i = 1, \ldots, h$. Define $\phi_0 = \max_{1 \leq i \leq h} E_N(l(I_i))$ which represents the maximum loss due to h individual confidence intervals with confidence coefficient $1 - \alpha$. Now, since for every s ($s = 1, \ldots, h-1$), $E_N(l(I_{s+1})) < E_N(l(I_s))$, it follows that for every $i = 1, \ldots, h$

$$E_N(l(I_i)) \leq \phi_0 = \left(\frac{n-p}{n-p-2}\right) \left(\frac{1}{\lambda_1} + \frac{\sigma_g^2}{\sigma_e^2}\right) = E_N(l(I_1)). \tag{2.11}$$

The performance criterion of a design has been taken as the maximum loss which is the maximum normalized expected lengths of $I_i,\ i = 1, \ldots, h$. Hence the loss function for interval estimation of the ratio of variance components has been taken as ϕ_0 and our interest in this study is to discriminate among designs with respect to this performance criterion.

Remark 2.1 Let $\alpha_1^*, \alpha_2^*, \ldots, \alpha_h^*$ be h positive numbers such that $\sum_{i=1}^h \alpha_i^* = \alpha$. On lines similar to (2.6) and (2.9), 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,

$$\begin{aligned} Pr\left[Y :\ \frac{\sigma_g^2}{\sigma_e^2} \in \cap_{i=1}^h I_i^*\right] &= Pr\left[\cap_{i=1}^h \{Y :\ \frac{\sigma_g^2}{\sigma_e^2} \in I_i^*\}\right] \\ &\geq \sum_{i=1}^h (1 - \alpha_i^*) - (h-1) = 1 - \alpha, \end{aligned}$$

which gives the confidence interval $I^* = \cap_{i=1}^h I_i^*$ of σ_g^2/σ_e^2 with confidence coefficient $1 - \alpha$. Thus we may construct infinitely many confidence intervals of σ_g^2/σ_e^2. We define the normalized expected lengths of the confidence interval I^* by $E_{N_i}^*(l(I^*)),\ i = 1, \ldots, h$, 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 $E_N(l(I_i^*)) = E_N(l(I_i)),\ i = 1,\ldots,h$, for any choice of $\alpha_1^*, \alpha_2^*, \ldots, \alpha_h^*$ and observing that the confidence interval, I^* being subset of I_i^* for each $i = 1,\ldots,h$, we have $E_{N_i}^*(l(I^*)) \leq E_N(l(I_i^*)) = E_N(l(I_i)),\ i = 1,\ldots,h$. This gives $E_N^*(l(I^*)) = \max_{1\leq i\leq h} E_{N_i}^*(l(I^*)) \leq \max_{1\leq i\leq h} E_N(l(I_i)) = \phi_0$ for all I^*.

Remark 2.2 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 $I_i^*,\ i = 1,\ldots,h$. 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 loss function to carry out the design optimization. It is interesting to note that the method of normalization disentangles the effect of nuisance parameters, i.e., the lower and upper cutoff values, from the effect of design parameters. In our approach a natural way to set the loss function in terms of the maximum normalized expected length of the interval estimate has been adopted which is in agreement with the minimax principle.

3. Blocked Diallel Cross Experiments

The study with respect to interval estimation of ratio of variance components for a diallel cross experiment in blocks can be carried out on lines similar to the designs in an unblocked diallel cross experment. 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

$$Y = \mu 1_n + D_2'\beta + D_1'g + e, \tag{3.1}$$

where as before, Y is the vector of n observations, g is the $p \times 1$ vector of general combining ability effects with $E(g) = 0$ and $Var(g) = \sigma_g^2 I$, β is the fixed effect due to blocks and e is the error vector with $E(e) = 0$ and $Var(e) = \sigma_e^2 I$. Also, $D_1 = (d_{uv}^{(1)})$ is the $p \times n$ line versus observation incidence matrix, as mentioned earlier, and $D_2 = (d_{uv}^{(2)})$ is the $b \times n$ block versus observation incidence matrix with $d_{uv}^{(2)} = 1$ if the v-th observation arise from the u-th block and $d_{uv}^{(2)} = 0$ otherwise. Equivalently, we can write (3.1) as

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

where $X = (1\ \ D_2'\ \ D_1')$. Here, $E(Y) = \mu 1_n$, Var $(Y|\sigma_g^2, \sigma_e^2) = \sigma_g^2 D_1' D_1 + \sigma_e^2 I_n$. Again, as in the unblocked case, we assume that $Y \sim N_n(\mu 1_n, \sigma_g^2 D_1' D_1 + \sigma_e^2 I_n)$. Let $N = (n_{ij})$ be the incidence matrix with n_{ij} indicating the number of times the i-th line occurs in the j-th block. Also, let $C = G - k^{-1}NN'$. It is easy to see that $Rank(C) \leq p-1$. In our model (3.1) we may consider β to be a random effects block parameter. Such a consideration do not alter the results obtained here.

Let H_B be an $n \times (n-b)$ matrix such that the columns of H_B form an orthonormal basis of the orthocomplement of the space spaned by $(1 \;\; D_2')$ in $\mathcal{R}^n$. Thus $H_B'H_B = I_{n-b}$ and $H_BH_B' = I-(1 \;\; D_2')[(1 \;\; D_2')'(1 \;\; D_2')]^-(1 \;\; D_2')'$ where T^- is a generalized-inverse of a matrix T. Note that $D_2H_B = 0$ and $1_n'H_B = 0$. Hence, $Z_B^{(n-b)\times 1} = H_B'Y \;\sim\;\; N_{n-b}(0, \sigma_g^2 H_B'D_1'D_1H_B + \sigma_e^2 I_{n-b})$.

We observe that the non-zero eigenvalues of $H_B'D_1'D_1H_B$ are the same as the non-zero eigenvalues of $D_1H_BH_B'D_1' = D_1(I-(1 \;\; D_2')\begin{pmatrix} 0 & 0 \\ 0 & k^{-1}I_b \end{pmatrix}(1 \;\; D_2')')D_1' = D_1D_1' - \frac{1}{k}NN' = G - \frac{1}{k}NN' = C$. This implies that the eigenvalues of $H_B'D_1'D_1H_B$ are zero with multiplicity $((n-b)-(p-1)) = n_e$ and the remaining $(p-1)$ eigenvalues are identical to the eigenvalues of the C-matrix.

Define $0 = \lambda_0^* < \lambda_1^* < \cdots < \lambda_h^*$ as the $h+1$ distinct eigen values of $H_B'D_1'D_1H_B$ with multiplicities $m_0^* = n_e, m_1^*, \ldots, m_h^*$ respectively. On lines similar to Section 2 we observe that there exits an orthogonal matrix $P_B^* = (P_{(0)}^* \;\; P_{(1)}^* \;\; \cdots \;\; P_{(h)}^*)$ of order $n-b$ such that $H_B'D_1'D_1H_B = P_B^*\begin{pmatrix} \Delta_B & 0 \\ 0 & 0 \end{pmatrix}P_B^{*'}$ where $\Delta_B = diag(\lambda_1^*I_{m_1^*}, \lambda_2^*I_{m_2^*}, \ldots, \lambda_h^*I_{m_h^*})$. Define $Z_{(i)}^* = P_{(i)}'Z_B$, $i = 0, \ldots, h$. Then $Q_i^* = Z_{(i)}^{*'}Z_{(i)}^*$ are independent and

$$(\sigma_g^2\lambda_i^* + \sigma_e^2)^{-1}Q_i^* \tag{3.2}$$

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 (3.2) 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 $L_i' = F_{1-\alpha/2, m_i^*, n_e}$ and $U_i' = F_{\alpha/2, m_i^*, n_e}$. Then

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

$$\Leftrightarrow Pr\left[\frac{n_eQ_1^*}{U_1'm_1^*\lambda_1^*Q_0^*} - \frac{1}{\lambda_1^*} \leq \frac{\sigma_g^2}{\sigma_e^2} \leq \frac{n_eQ_1^*}{L_1'm_1^*\lambda_1^*Q_0^*} - \frac{1}{\lambda_1^*}\right] = 1-\alpha, \tag{3.3}$$

giving a confidence interval

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

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

Now, on lines similar to (2.7), we get

$$E(l(I_1')) = E\left(\frac{n_eQ_1^*}{L_1'm_1^*\lambda_1^*Q_0^*} - \frac{n_eQ_1^*}{U_1'm_1^*\lambda_1^*Q_0^*}\right) = \left(\frac{n_e}{n_e-2}\right)\left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right)\left(\frac{1}{L_1'} - \frac{1}{U_1'}\right). \tag{3.4}$$

Now, the pair (L_1', U_1') is not unique for setting up the confidence interval with confidence coefficient $1-\alpha$. As before, the normalized expected length comes out as

$$E_N(l(I_1')) = E(l(I_1'))/(\frac{1}{L_1'} - \frac{1}{U_1'}) = (\frac{n_e}{n_e-2})\left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right). \tag{3.5}$$

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

$$I_i' = \left(\frac{n_e Q_i^*}{m_i^* U_i' Q_0^* \lambda_i^*} - \frac{1}{\lambda_i^*},\ \frac{n_e Q_i^*}{m_i^* L_i' Q_0^* \lambda_i^*} - \frac{1}{\lambda_i^*} \right),\ i = 2, \ldots, h \tag{3.6}$$

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

$$E_N(l(I_i')) = \left(\frac{1}{\lambda_i^*} + \frac{\sigma_g^2}{\sigma_e^2} \right) \left(\frac{n_e}{n_e - 2} \right),\ i = 2, \ldots, h. \tag{3.7}$$

Now as in §2 we define $\phi = \max_{1 \le i \le h} E_N(l(I_i'))$ which represents the maximum loss due to h individual confidence intervals with confidence coefficient $1-\alpha$. Furthermore, for every $i = 1, \ldots, h$

$$E_N(l(I_i')) \le \phi = \left(\frac{n_e}{n_e - 2} \right) \left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2} \right) = E_N(l(I_1')). \tag{3.8}$$

It is to be noted that we may construct infinitely many confidence intervals, I_B^* of σ_g^2/σ_e^2 by taking the intersection of $I_i'^*$, $i = 1, \ldots, h$ where $I_i'^*$ is the confidence interval of σ_g^2/σ_e^2 with confidence coefficient $1 - \alpha_i^*$, $\alpha_i^* > 0$, $i = 1, \ldots, h$ such that $\sum_{i=1}^h \alpha_i^* = \alpha$. After normalization it can be seen that $E_N^*(l(I_B^*)) = \max_{1 \le i \le h} E_{N_i}^*(l(I_B^*)) \le \max_{1 \le i \le h} E_N(l(I_i'^*)) = \phi$.

4. Optimal Designs

In the previous sections we have explicitly obtained the maximum normalized expected length (ϕ_0 and ϕ) of the interval estimate of σ_g^2/σ_e^2 under an unblocked and a blocked model. Our objective in obtaining an optimal design would be to minimize the loss function $\phi_0 = \left(\frac{n-p}{n-p-2}\right)\left(\frac{1}{\lambda_1} + \frac{\sigma_g^2}{\sigma_e^2}\right)$ in case of an unblocked model and to minimize $\phi = \left(\frac{n_e}{n_e-2}\right)\left(\frac{1}{\lambda_1^*} + \frac{\sigma_g^2}{\sigma_e^2}\right)$ in case of a blocked model. Let $\mathcal{D}(p,n)$ be the class of diallel cross unblocked designs involving p lines and n crosses and $\mathcal{D}(p,b,k)$, 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 $m_{1d}, m_{2d}, \ldots, m_{hd}$ ($m_{1d}^*, m_{2d}^*, \ldots, m_{hd}^*$). The corresponding loss functions are ϕ_{0d} and ϕ_d. A design d^* will be said to be D_l-optimal if, among all designs in $\mathcal{D}$, d^* minimizes ϕ_{0d} (or ϕ_d). It is easy to see that a D_l-optimal design maximizes λ_{1d} (λ_{1d}^*) within the competing class of designs. Thus we see a connection between D_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 $\mathcal{D}(p,n)$. Since a universally optimal design is E-optimal as well, it follows that complete diallel cross designs are D_l-optimal in $\mathcal{D}(p,n)$ under our setup.

Under the fixed effects model, Gupta & Kageyama (1994), Dey & Midha (1996) and Das, Dey & Dean (1998) have obtained universally optimal (and hence E-optimal) diallel cross designs. It thus follows that their designs are D_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 &

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 $d^* \in \mathcal{D}(p, b, k)$ and is universally optimal in $\mathcal{D}(p, b, k)$ under the fixed effects model. Summarizing, therefore, we have

THEOREM 4.1 *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 D_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, Preece & Rees (2001). The tables in their paper provide solutions to our D_l-optimal diallel cross designs within the parametric range $2k < p < 16,\ s \leq 30$. The case $2k = p$ is dealt in Gupta & Kageyama (1994). The nested balanced incomplete block designs have been extended to nested balanced block designs and a series of designs, D_l-optimal under our set-up, is given in Das, Dey & Dean (1998).

Mukerjee (1997) has obtained E-optimal partial diallel cross designs under the fixed effects model. Following Mukerjee (1997) we have the following results on D_l-optimal designs for the estimation of σ_g^2/σ_e^2.

Let $p = n_1 n_2$ where $n_1 \geq 2,\ n_2 \geq 3$. 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

$$d_1^* = \{(i, j) : 1 \leq i < j \leq p \text{ and } i, j \in S_u \text{ for some } u\}. \tag{4.1}$$

Then $d_1^* \in \mathcal{D}(p, n)$, 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 4.2 *For each $n_1 \geq 2$ and $n_2 \geq 3$, up to isomorphism, the design d_1^* is uniquely D_l-optimal in $\mathcal{D}(p, n)$, where $p = n_1 n_2$ and $n = \frac{1}{2} n_1 n_2 (n_2 - 1)$.*

EXAMPLE 4.1 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 4.2, this design is D_l-optimal in $\mathcal{D}(12, 18)$.

Let $p = n_1n_2 + t$, where $n_1 \geq 2$, $n_2 \geq 3$ and t $(1 \leq t \leq n_1 - 1)$ are positive integers. Partition $\{1, \ldots, p\}$ into n_1 mutually exclusive and exhaustive subsets $S_1, \ldots, S_{n_1}$ such that $S_1, \ldots, S_{n_1-t}$ have cardinality n_2 and $S_{n_1-t+1}, \ldots, S_{n_1}$ have cardinality $n_2 + 1$. Analogous to (4.1), let

$$d_2^* = \{(i,j) : 1 \leq i < j \leq p \text{ and } i,j \in S_u \text{ for some } u\}. \tag{4.2}$$

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

THEOREM 4.3 *For* $n_1 \geq 2$, $n_2 \geq 3$, $p = n_1n_2+t$, $n = \frac{1}{2}n_1n_2(n_2-1)+n_2t$ *and* $1 \leq t \leq n_1-1$, *the design* d_2^* *is* D_l*-optimal in* $\mathcal{D}(p,n)$, *provided* $(n_1-t)n_2f > 1$ *where* $f = n^{-1}(n_2-1)^2 - p^{-1}(n_2-2)$.

EXAMPLE 4.2 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 4.3, since $(n_1 - t)n_2f = 2.04 > 1$, this design is D_l-optimal in $\mathcal{D}(13,22)$.

The condition (4.2) holds in a large number of cases over a practicable range. Thus, among the 79 cases of (n_1, n_2, t) satisfying $n_1 \geq 2$, $n_2 \geq 3$, $1 \leq t \leq n_1 - 1$, $p = n_1n_2 + t \leq 30$, there are as many as 57 where the condition holds and hence d_0 is D_l-optimal.

The blocking of optimal designs of Theorems 4.2 and 4.3 is given in Mukerjee (1997) where orthogonal blocking has been achieved for designs corresponding to Theorem 4.2. Thus, Mukerjee's method of constuction of orthogonal block designs lead to D_l-optimal diallel cross block designs in $\mathcal{D}(p,b,k)$.

EXAMPLE 4.3 Consider the following design (rows are blocks) with parameters $p = 12$, $b = 3$ and $k = 6$.

$$\begin{array}{cccccc}
(1,2) & (3,4) & (5,6) & (7,8) & (9,10) & (11,12) \\
(1,3) & (2,4) & (5,7) & (6,8) & (9,11) & (10,12) \\
(1,4) & (2,3) & (5,8) & (6,7) & (9,12) & (10,11)
\end{array}$$

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

In our model (3.1) 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.

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. -Theor. Meth.* **15**, 3353-78.

Das, A., Dean, A. M. & Gupta, S. (1998). On optimality of some partial diallel cross designs. *Sankhyā* B **60**, 511-24.

Das, A., Dey, A. & Dean, A. M. (1998). Optimal block designs for diallel cross experiments. *Statist. Prob. Letters* **36**, 427-36.

Dey, A. & Midha, C. K. (1996). Optimal block designs for diallel crossses. *Biometrika* **83**, 484-89.

Falconer, D. S. (1991). *Introduction to Quantitative Genetics.* Oliver and Boyd, Edinburgh.

Griffing, B. (1956). Concepts of general and specific combining ability in relation to diallel crossing systems. *Aust. Jour. Bio. Sci.* **9**, 463-93.

Gupta, S. & Kageyama, S. (1994). Optimal complete diallel crosses. *Biometrika* **81**, 420-24.

Hinkelmann, K. (1975). Design of Genetical experiments. In *A survey of statistical design and linear models*. Ed. J.N. Srivastava, 243-66.

Hinkelmann, K. & Kempthorne, O. (1963). Two classes of group divisible partial diallel crosses. *Biometrika*, **50**, 281-91.

Kempthorne, O. (1969). *An Introduction to Genetic Statistic.* The Iowa State University Press, Iowa.

Kempthorne, O. & Curnow, R. N. (1961). The partial diallel cross. *Biometrics* **17**, 229-250.

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**, 351-89.

Mukerjee, R. (1997). Optimal partial diallel crosses. *Biometrika* **84**, 939-48.

Preece, D. A. (1967). Nested balanced incomplete block designs. *Biometrika* **54**, 479-86.

Thomas, J.D. & Hultquist, R. A. (1978). Interval estimation for the unbalanced case of the one way random effects model. *Ann. Statist.* **6**, 582-87.

Wald, A. (1940). A note on the analysis of variance with unequal class frequencies. *Ann. Math. Statist.* **11**, 96-100.