Statistica Sinica **25** (2015), 537-549
doi:http://dx.doi.org/10.5705/ss.2012.042

OPTIMAL ALLOCATION TO TREATMENT GROUPS UNDER VARIANCE HETEROGENEITY

Ellinor Fackle-Fornius and Hans Nyquist

Stockholm University

Abstract: The problem of allocating experimental units to treatment groups when variance heterogeneity over treatment groups is present is considered. A_A- and D_A-optimal allocations are derived for estimation of linear combinations of treatment means. Explicit expressions for the design weights are provided for the A_A-optimal design. The minimax strategy is introduced as an approach to handle unknown variances. Efficiencies of minimax allocations are evaluated.

Key words and phrases: A-optimality, D-optimality, minimax designs, factorial experiments.

1. Introduction

This paper concerns optimal experimental design when the aim of the experiment is to compare one or more treatments and the variance of the outcome varies across groups. The number of experimental groups is fixed in advance and the design problem concerns allocating experimental units to experimental groups. An optimal allocation is sought since a careless allocation may result in inaccurate estimates of important effects and inefficient use of resources.

Our approach to optimal design is the theory for continuous designs, which originated in works by Kiefer (e.g., Kiefer (1974)) and is further discussed in e.g., Atkinson, Donev, and Tobias (2007) and Silvey (1980). Deriving an optimal design involves choosing an appropriate optimality criterion that will determine in what sense the design is optimal. If estimation is of primary concern, two criteria arise as natural candidates, the A criterion, which minimizes the sum of the variances of the estimators, and the D criterion, which minimizes the confidence region for the estimators. When interest is in estimation of linear combinations of parameters, the A_A- and D_A-criteria correspond to the A- and D-criteria, respectively.

In many experiments the assumption of homogenous variances across treatment groups is not realistic. Response variables with a binomial or a Poisson distribution are obvious examples of this. Different group variances have an important and essential impact on optimal allocation. However, this implies that optimal allocations are in general possible to construct only if these variances are known. Several approaches to deal with the allocation problem when the variances are unknown have been proposed. One approach is to guess the unknown group variances, yielding a locally optimal design. Locally optimal A_A- and D_A-designs for control group experiments are given in Wong and Zhu (2008). The case of generalized linear models was studied by Arnoldsson (1996) for 2^2 and by Yang, Mandal, and Majumdar (2012) for 2^k factorial experiments. If guesses of group variances are specified in terms of a probability distribution an optimum on-the-average (Bayesian) design may be obtained, see Pettersson and Nyquist (2003). A third approach is to minimize the maximum of the criterion function taken over a specified region of possible group variances, yielding a minimax design. The goal here is to give formulae for locally optimal designs and minimax designs when linear combinations of treatment effects are to be estimated and the D_A - or A_A-criterion is used. We generalize current research by considering any linear combination of treatment group means and the minimax allocations of subjects to treatment groups. A consideration of general linear combinations extends the application of the theory from control group experiments with one control group to a large range of important applications including, control group experiments with several control groups (see e.g., Hedayat, Jacroux, and Majumdar (1988) and the discussion therein) and factorial experiments where estimation of contrasts is a main issue (see e.g. the agricultural experiments on mangold and sugar beets reported in Rothamsted Report (1936, 1937) and further discussed in Snedecor and Cochran (1989)).

The setup of the experiments we consider is defined in the next section, while local and minimax optimality are presented in the two subsequent sections. Some efficiency comparisons are presented in Section 5. The paper ends with concluding remarks in Section 6.

2. Preliminaries

Suppose there are m treatment groups and that observations from treatment group j are stochastically independent observations on a random variable Y_j, $j = 1, \ldots, m$. The expectation and variance of Y_j are denoted by μ_j and $v_j(\theta_j)$, respectively, where θ_j is a vector of parameters, $(\theta_1, \theta_2, \ldots, \theta_m) \in \Theta$, Θ being the parameter space.

The covariance matrix for the group averages $\overline{Y}_j$ if n_j observations are assigned to the treatment groups is an $m \times m$ diagonal matrix M^{-1} with elements of the form $\rho_j = v_j/n_j$. Assume that interest is in inference about p linear combinations $A^T\mu$, where A is an $m \times p$ matrix of constants a_{jk}, $j = 1, \ldots, m$, $k = 1, \ldots, p$. The covariance matrix for the linear combinations is then $C = \left(A^T M^{-1} A\right)$.

Example 1 (Control group experiments)**.** In control group experiments, the first treatment group receive a control or placebo treatment and we wish to make inference about the differences in expected responses $\mu_1 - \mu_j$, $j = 2, \ldots, m$. The vector of differences is estimated by $A^T\overline{Y}$, where A^T is the $(m-1) \times m$ matrix $A^T = (1_{m-1} \quad -I_{m-1})$ with 1_{m-1} a vector of ones and I_{m-1} the identity matrix of order $m-1$. The covariance matrix associated to the vector of differences is

$$C = \begin{pmatrix} \rho_1+\rho_2 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & \rho_1+\rho_3 & \cdots & \rho_1 \\ \vdots & \vdots & & \vdots \\ \rho_1 & \rho_1 & \cdots & \rho_1+\rho_m \end{pmatrix}.$$

Example 2 (2×2 factorial experiment)**.** Let the two treatments be represented by a and b and assign the treatment combinations (not a, not b), (a, not b), (not a, b), and (a, b) to the four treatment groups. The interaction effect is estimated by $A^T\overline{Y}$, with $A^T = \begin{pmatrix} 1 & -1 & -1 & 1 \end{pmatrix}$ with variance

$$C = \rho_1 + \rho_2 + \rho_3 + \rho_4 = \frac{v_1}{n_1} + \frac{v_2}{n_2} + \frac{v_3}{n_3} + \frac{v_4}{n_4}.$$

If interest is in inference about the main effects a and b, as well as the interaction effect, the matrix defining the linear combinations is

$$A^T = \begin{pmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

and the covariance matrix is

$$C = \begin{pmatrix} \rho_1+\rho_2 & \rho_1 & -\rho_1-\rho_2 \\ \rho_1 & \rho_1+\rho_3 & -\rho_1-\rho_3 \\ -\rho_1-\rho_2 & -\rho_1-\rho_3 & \rho_1+\rho_2+\rho_3+\rho_4 \end{pmatrix}.$$

3. Local $\mathbf{A}_A$- and $\mathbf{D}_A$−optimality

The A_A criterion Ψ_A minimizes the sum of variances of the estimated statistics while the $\mathrm{D_A}$ criterion Ψ_D minimizes (log of) the determinant of the covariance matrix. The latter is equivalent to minimizing the volume of a confidence ellipsoid and is sometimes called (the log of) the generalized variance.

Let S be the set of subsets of p elements from the set $\{1,\ldots,m\}$, $\binom{m}{p}$ in all. Let $A_{[s]}$ be the $p \times p$ matrix obtained from selecting the rows $s \in S$ from the matrix A and $M_{[s]}$ be the $p \times p$ matrix obtained from selecting rows and columns $s \in S$ from the matrix M. For example, for $m = 3$, the matrix $M_{[1,3]}$ is the 2×2 matrix obtained when deleting the second row and second column of M.

Theorem 1. *When estimating* $A^T\mu$*, the criteria for* A_A*-optimality and* D_A *-optimality are*

$$\Psi_A = tr\,(C) = \sum_{k=1}^{p}\sum_{j=1}^{m} a_{jk}^2 \rho_j,$$
$$\Psi_D = \ln\det\,(C) = \ln\sum_{s\in S}\left(\det A_{[s]}\right)^2 \prod_{j\in s}\rho_j.$$

Corollary 1. *In the control group experiment the criteria for* A_A*-optimality and* D_A*-optimality are*

$$\Psi_A = tr\,(C) = (m-1)\,\rho_1 + \sum_{j=2}^{m}\rho_j,$$
$$\Psi_D = \ln\det\,(C) = \ln\sum_{j=1}^{m}\prod_{\substack{k=1\\ k\neq j}}^{m}\rho_k.$$

Example 1 (continued). If $m = 3$, then

$$\Psi_A = 2\frac{v_1}{n_1} + \frac{v_2}{n_2} + \frac{v_3}{n_3},$$
$$\Psi_D = \ln\left\{\left(\frac{v_2}{n_2}\frac{v_3}{n_3}\right) + \left(\frac{v_1}{n_1}\frac{v_3}{n_3}\right) + \left(\frac{v_1}{n_1}\frac{v_2}{n_2}\right)\right\}.$$

Example 2 (continued). The criterion functions when interest is in inference about the interaction effect in a 2×2 experiment are trivially $\Psi_A = v_1/n_1 + v_2/n_2 + v_3/n_3 + v_4/n_4$ and $\Psi_D = \ln\Psi_A$, respectively. For inference about main effects as well as the interaction effect, the criterion functions are obtained from Theorem 1 as

$$\Psi_A = 3\frac{v_1}{n_1} + 2\frac{v_2}{n_2} + 2\frac{v_3}{n_3} + \frac{v_4}{n_4},$$
$$\Psi_D = \ln\left(\frac{v_1}{n_1}\frac{v_2}{n_2}\frac{v_3}{n_3} + \frac{v_1}{n_1}\frac{v_2}{n_2}\frac{v_4}{n_4} + \frac{v_1}{n_1}\frac{v_3}{n_3}\frac{v_4}{n_4} + \frac{v_2}{n_2}\frac{v_3}{n_3}\frac{v_4}{n_4}\right).$$

For finding the A_A- and D_A-optimal designs we take $n_j = \omega_j N$, where N is the total number of observations so the ω_j, $j = 1,\ldots,m$, are design weights. Denote optimal design weights by $\omega^* = (\omega_1^*,\ldots,\omega_m^*)^T$.

Theorem 2. *For* A_A*-optimal and* D_A*-optimal designs, weights are given by, respectively,*

$$\omega_j^* = \frac{\left(\sum\limits_{k=1}^{p} a_{jk}^2\right)^{1/2} v_j^{1/2}}{\sum\limits_{r=1}^{m}\left(\sum\limits_{k=1}^{p} a_{rk}^2\right)^{1/2} v_r^{1/2}},\ j = 1,\ldots,m, \tag{3.1}$$

$$\omega_j^* = p^{-1}\left(\frac{\sum_{r\in R_j}\left(\det A_{[r]}\right)^2\prod_{k\in r}\left(\frac{\upsilon_k}{\omega_k^*}\right)}{\exp\left(\Psi_D\left(\omega^*\right)\right)}\right),\ j=1,\ldots,m, \tag{3.2}$$

where $R_j \subset S$ is the set of all $s \in S$ that contain j.

For $p = 1$ and homogenous variances, this result reduces to Corollary 1 of Pukelsheim and Torsney (1991). The weights for A_A-optimality are explicitly given, while, in general, a system of non-linear equations needs to be solved to obtain the weights for $\mathrm{D_A}$-optimality. Here the Neyman allocation formula, used in stratified sampling for estimating a population mean, coincides with the A_A-optimal design weights with $A^T = \begin{pmatrix} a_1\, a_2 \cdots a_m\end{pmatrix}$, where the constant a_j is interpreted as the relative size of stratum j.

Corollary 2. *In the control group experiment, the A_A-optimal and D_A-optimal design weights are given by, respectively,*

$$\omega_1^* = \frac{\sqrt{(m-1)\,\upsilon_1}}{\sqrt{(m-1)\,\upsilon_1}+\sum_{r=2}^{m}\sqrt{\upsilon_r}},$$

$$\omega_j^* = \frac{\sqrt{\upsilon_j}}{\sqrt{(m-1)\,\upsilon_1}+\sum_{r=2}^{m}\sqrt{\upsilon_r}},\ j=2,3,\ldots,m,$$

$$\omega_j^* = (m-1)^{-1}\left(1-\frac{\prod_{\substack{k=1\\k\neq j}}^{m}\left(\upsilon_k/\omega_k^*\right)}{\exp\left(\Psi_D\left(\omega^*\right)\right)}\right),\ j=1,\ldots,m.$$

The weights in the corollary and an approximation to the $\mathrm{D_A}$ -optimal weights are also in Wong and Zhu (2008).

Example 1 (continued). In a control group study with $m = 2$, Corollary 2 gives

$$\omega_1^* = \frac{\upsilon_1^{1/2}}{\upsilon_1^{1/2}+\upsilon_2^{1/2}}, \qquad \omega_2^* = 1-\omega_1^*,$$

for A_A- and D_A-optimality. With $r = \upsilon_2/\upsilon_1$, this is

$$\omega_1^* = \frac{1}{1+\sqrt{r}}.$$

For equal variances a uniform allocation is optimal, while the treatment weight $\omega_2^* \to 1$ as $r \to \infty$, and $\omega_2^* \to 0$ as $r \to 0$.

When there are two treatment groups and one control group, the A_A -optimal weights are $\omega_1^* = \sqrt{2\upsilon_1}/D$, $\omega_2^* = \sqrt{\upsilon_2}/D$, $\omega_3^* = \sqrt{\upsilon_3}/D$, where $D = \sqrt{2\upsilon_1} + \sqrt{\upsilon_2} + \sqrt{\upsilon_3}$. By defining the variance ratios $r_2 = \upsilon_2/\upsilon_1$ and $r_3 = \upsilon_3/\upsilon_1$ these are

$$\omega_1^* = \frac{1}{1+\sqrt{r_2}+\sqrt{r_3}},\ \omega_2^* = \frac{\sqrt{r_2}}{1+\sqrt{r_2}+\sqrt{r_3}},\ \omega_3^* = 1-\omega_1^*-\omega_2^*.$$

The $\mathrm{D_A}$-optimal weights for a three group experiment satisfy

$$\begin{aligned}
\omega_1^* &= \frac{1}{2}\left(1-\frac{(\upsilon_2/\omega_2^*)(\upsilon_3/\omega_3^*)}{\exp\Psi_D(\omega^*)}\right),\\
\omega_2^* &= \frac{1}{2}\left(1-\frac{(\upsilon_1/\omega_1^*)(\upsilon_3/\omega_3^*)}{\exp\Psi_D(\omega^*)}\right),\\
\omega_3^* &= \frac{1}{2}\left(1-\frac{(\upsilon_2/\omega_2^*)(\upsilon_3/\omega_3^*)}{\exp\Psi_D(\omega^*)}\right).
\end{aligned}$$

In general, the D_A-optimal weights need to be obtained numerically. However, for the special case of equal treatment group variances, $r = r_2 = r_3$,

$$\begin{aligned}
\omega_1^* &= \frac{3-\sqrt{1+8r}}{4\,(1-r)}, \quad r \neq 1,\\
\omega_2^* = \omega_3^* &= \frac{1-\omega_1^*}{2}.
\end{aligned}$$

The A_A-optimal allocation is generally more concentrated to the control group than the D_A-optimal allocation.

Example 2 (continued). For inference about the interaction effect in a 2×2 experiment Theorem 2 yields

$$\omega_j^* = \frac{\sqrt{\upsilon_j}}{\sum\limits_{r=1}^{4}\sqrt{\upsilon_r}}, \quad j = 1,2,3,4,$$

for both the A_A- and $\mathrm{D_A}$-criteria. Similarly, for inference about main effects as well as the interaction effect, the optimal design weights are $\omega_1^* = \sqrt{3\upsilon_1}/D, \omega_2^* = \sqrt{2\upsilon_2}/D, \omega_3^* = \sqrt{2\upsilon_3}/D, \omega_4^* = \sqrt{\upsilon_4}/D$, when using the A_A-criterion, where $D = \sqrt{3\upsilon_1} + \sqrt{2\upsilon_2} + \sqrt{2\upsilon_3} + \sqrt{\upsilon_4}$, and

$$\omega_j^* = \frac{1}{3}\left(1-\prod_{\substack{k=1\\k\neq j}}^{4}\frac{\upsilon_k/\omega_k^*}{\Psi_D(\omega^*)}\right)$$

when using the D_A-criterion.

The optimal design weights depend on the variances so if they are unknown, one cannot compute the optimal design weights. If the variances are unknown but their ratios are known, say $\upsilon_j = \gamma_j \upsilon_1$ for $j = 2, \ldots, m$, Theorem 2 applies with υ_j replaced by γ_j.

4. Minimax Optimal Allocations

For the minimax approach we define a region $\Theta_0 \subset \Theta$ of assumed values for the parameters defining the variances, $\upsilon(\theta)$, and compute a design that minimizes

$$\max_{\theta \in \Theta_0} \Psi(\omega, \upsilon(\theta)).$$

In many applications Θ_0 is a Cartesian product $\Theta_0 = \Theta_{01} \times \Theta_{02} \times \ldots \times \Theta_{0m}$, where Θ_{0j} is the set of assumed parameter values θ_j in group j. This implies that the assumptions on the variance in one group are made independently of the assumptions on the variance in another group. With this restriction, the maximization of the criterion function is considerably simplified.

Theorem 3. *Let A be a $m \times p$ matrix, let Θ_0 be the Cartesian product $\Theta_0 = \Theta_{01} \times \Theta_{02} \times \ldots \times \Theta_{0m}$, and consider estimation of $A^T\mu$. Then, $\Psi_A(\omega, \upsilon(\theta))$ and $\Psi_D(\omega, \upsilon(\theta))$ are maximized over $\theta \in \Theta_0$ at $\upsilon(\theta) = \upsilon^*(\theta)$, independently of ω, where $\upsilon^*(\theta) = (\upsilon_1^*, \upsilon_2^*, \ldots, \upsilon_m^*)^T$ and*

$$\upsilon_j^* = \max_{\theta_j \in \Theta_{0j}} \upsilon_j(\theta_j). \tag{4.1}$$

Without the restriction of a Cartesian parameter space Theorem 3 does not hold in general.

Corollary 3. *A_A- and D_A-minimax designs are obtained if υ_j, $j = 1, \ldots, m$ in (3.1) and (3.2) are replaced by υ_j^*, $j = 1, \ldots, m$, defined in (4.1).*

According to Corollary 3, A_A- and D_A-minimax designs are based on the largest possible variance in each group, which corresponds to the largest possible uncertainty in the group means and is different from the largest possible uncertainty in parameter values. An illustration of this is when responses are binary with group variances $\upsilon_j(\theta_j) = \theta_j(1 - \theta_j)$, where $\theta_j \in \Theta_{0j}$ is the response probability in group j and υ_j^* is the variance associated to the value of θ_j being closest to 0.5. Hence, two regions $\Theta_{01} = [0, 1]$, with complete ignorance about the parameter, and $\Theta_{02} = [0.5 - \alpha, 0.5 + \alpha]$ for some small $\alpha > 0$, with fairly precise knowledge about the parameter value, yield the same maximum variance, $\upsilon^* = 0.25$, and thereby yield identical minimax designs. On the other hand, a region of the form $\Theta_{03} = [0.1 - \alpha, 0.1 + \alpha]$ would yield completely different minimax designs than those for Θ_{01} (and Θ_{02}). Further, if concern is on efficiency one would consider a standardized maximin criterion (see e.g., Dette et al. (2006)).

5. Efficiency Comparisons

In this section efficiency evaluations are made to clarify (i) robustness of the minimax allocations to erroneous prior assumptions and (ii) the potential gains from using an optimal allocation over a uniform allocation. For this D_A-efficiency of the minimax allocation ω_{Minimax} is taken as

$$eff_D\left(\omega_{\text{Minimax}}, \theta\right) = \left[\frac{\Psi_D\left(\omega^*, \upsilon(\theta)\right)}{\Psi_D\left(\omega_{\text{Minimax}}, \upsilon(\theta)\right)}\right]^{1/m},$$

and the corresponding A_A-efficiency as

$$eff_A\left(\omega_{\text{Minimax}}, \theta\right) = \frac{\Psi_A\left(\omega^*, \upsilon\left(\theta\right)\right)}{\Psi_A\left(\omega_{\text{Minimax}}, \upsilon\left(\theta\right)\right)},$$

where ω^* is the locally optimal allocation given the variances $\upsilon\left(\theta\right)$.

Example 1 (continued). For the control group experiment with $m = 2$ the A_A- and D_A-criteria result in the same allocations and both efficiences are

$$eff\left(\omega_{\text{Minimax}}, r\right) = \frac{1/\omega^* + r/(1-\omega^*)}{1/\omega_{\text{minimax}} + r/(1-\omega_{\text{minimax}})}.$$

Figure 1 displays the efficiencies for the minimax designs for $0 \le r \le 20$. The assumed maximum variance ratio defining the minimax design is denoted by $r^{\max}$ to avoid confusion with true r. The minimax designs are fairly robust to moderate deviations from the prior assumption about the variance ratio. For example when $r^{\max} = 5$ the efficiency of the corresponding minimax design exceeds 0.95 for values of r between 2 and 15, approximately. If the variances are assumed equal, the efficiency is close to 1 for values of r between 0.5 and 2, and one has $eff \to \omega_{\text{minimax}}$ as $r \to 0$ and $eff \to 1 - \omega_{\text{minimax}}$ as $r \to \infty$.

Figure 2 shows the contours of eff_D and eff_A for minimax allocations when $m = 3$. As long as the two variance ratios are roughly equal and neither one is very large or very small the A_A-efficiency of the minimax design (panel a) is quite high. On the other hand, the A_A -efficiency drops substantially, especially for combinations of relatively high and low values of r_2 and r_3. The D_A-efficiency (panel b) exceeds 0.9 for most values of the variance ratios. Examples with other values on r_2 and r_3 may be obtained from the authors upon request.

The uniform allocation is often chosen in applications. To compare the performance of the minimax allocation with the uniform allocation we evaluate the relative efficiencies

$$eff_D^U\left(\omega_{\text{Minimax}}, \theta\right) = \left[\frac{\Psi_D\left(\omega_{\text{Uniform}}, \upsilon\left(\theta\right)\right)}{\Psi_D\left(\omega_{\text{Minimax}}, \upsilon\left(\theta\right)\right)}\right]^{1/p},$$

$$eff_A^U\left(\omega_{\text{Minimax}}, \theta\right) = \frac{\Psi_A\left(\omega_{\text{Uniform}}, \upsilon\left(\theta\right)\right)}{\Psi_A\left(\omega_{\text{Minimax}}, \upsilon\left(\theta\right)\right)}.$$

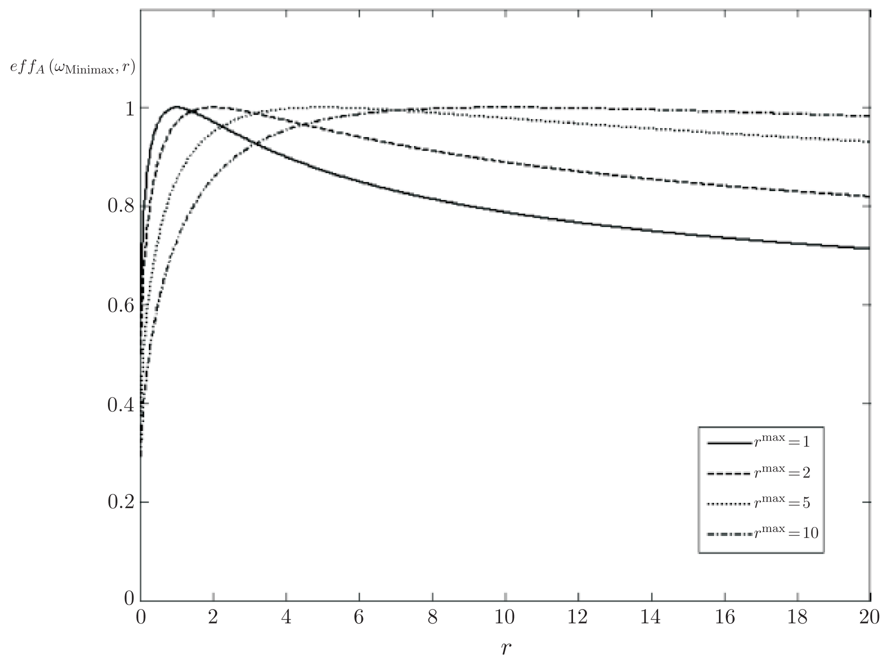
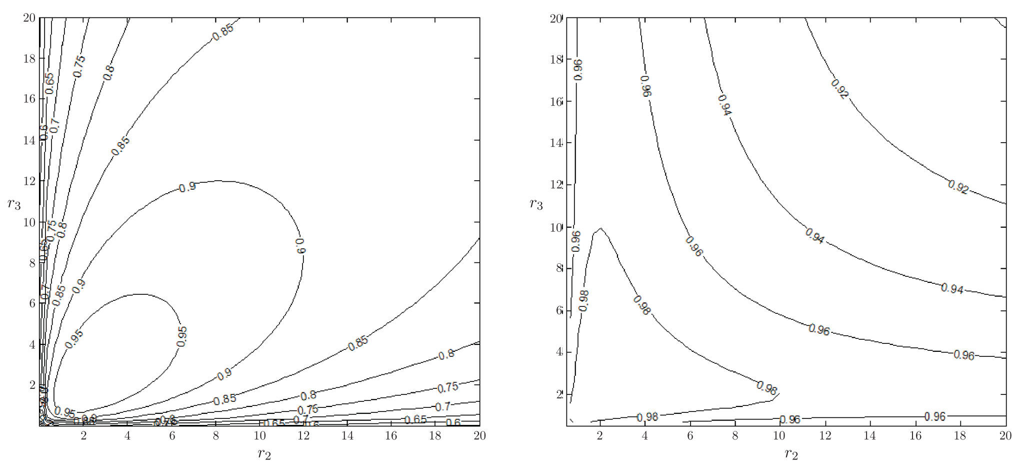


Figure 1. A_A/D_A-efficiencies for minimax allocations.

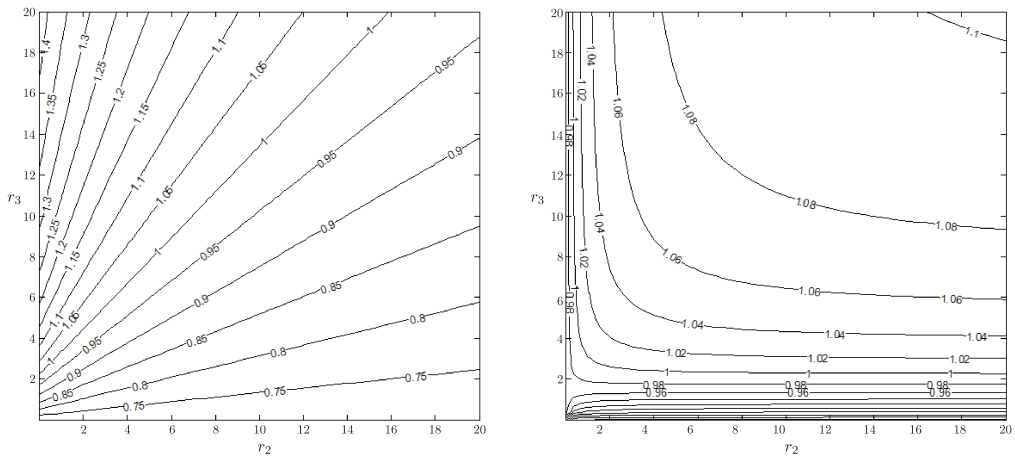


(a) A_A-minimax allocation. (b) D_A-minimax allocation.

Figure 2. Efficiency contour plots when $r_2^{\max} = r_3^{\max} = 2$.

When $m = 2$, $eff_D^U = eff_A^U$ and is

$$eff^U(\omega_{\text{Minimax}}, r) = \frac{2 + 2r}{1/\omega_{\text{minimax}} + r/(1 - \omega_{\text{minimax}})}.$$



(a) A_A-minimax allocation. (b) D_A-minimax allocation.

Figure 3. Contour plots of efficiencies for the minimax allocations compared to a uniform allocation when $r_2^{\max} = 2$ and $r_3^{\max} = 10$.

The minimax allocation is more efficient than the uniform allocation as long as the variance ratio r exceeds $\sqrt{r^{\max}}$. If it is believed that the variances are different in the groups the minimax allocation offers an improvement compared to the uniform allocation. Note that $eff^U \to 2\omega_{\text{minimax}}$ as $r \to 0$ and $eff^U \to 2\left(1 - \omega_{\text{minimax}}\right)$ as $r \to \infty$.

When $r_2^{\max} = r_3^{\max} = 2$ the minimax A_A-optimal allocation coincides with the uniform allocation and the relative A_A-efficiency is 1. Panel a) of Figure 3 shows the relative A_A-efficiency when $r_2^{\max} = 2$ and $r_3^{\max} = 10$. Apparently, the relative A_A -efficiency favors the minimax allocation in a region where r_3 exceeds r_2. The relative D_A-efficiencies are shown in panel b). The relative D_A-efficiency is slightly larger than 1 when the variance ratios are in the neighborhood of the assumed values, and increases as the variance ratios increases. For the minimax allocation to be superior to the uniform, r_3 cannot be too small.

6. Conclusions

General expressions for optimal allocations when estimating linear combinations of treatment means using the D_A- and A_A-criteria are derived with explicit expressions obtained for the A_A-optimal design weights.

With model parameters and group variances rarely known beforehand, a minimax strategy can be used. Minimax allocations based on the D_A- and A_A -criteria are shown to be particularly simple when the parameter space is the Cartesian of parameter spaces associated to each treatment group.

Efficiencies for minimax designs are quite robust and more efficient than those for the uniform designs in control group experiments, as long as the prior information is fairly accurate. With a control group and two treatment groups, the A-efficiency becomes low only when the variances in the two treatment groups are opposite and differs from what was specified. When information about true parameter values is vague we recommend using a pilot study to gain more information and thereby decrease the uncertainty about the treatment group variances. When efficiency is the primary concern, it would be interesting to consider standardized maximin efficient designs, as in Dette et al. (2006) and Dette, Trampisch, and Hothorn (2007).

Acknowledgements

Research reported in this article has been financially supported by the Swedish Research Council. The authors are grateful to Professor Subir Gosh for discussions on applications of the theory presented and to an associate editor and the referees for their valuable comments that substantially improved the presentation.

Appendix

Proof of Theorem 1. The criterion for A_A-optimality is obtained by evaluating the matrix multiplications in $C = A^T M^{-1} A$ and the stated result follows immediately.

The criterion for D_A-optimality is obtained by applying the Cauchy-Binet theorem for the matrices $A^T M^{-1}$ and A. With it, $\det\big((A^T M^{-1})A\big) = \sum_{s \in S} \det\Big(\big(A^T M^{-1}\big)_{[s]} A_{[s]}\Big)$. For $s = (s_1, s_2, \ldots s_p)$ we find $\det A^T M^{-1} A = \det \sum_{s \in S} A^T_{[s]} M^{-1}_{[s]} A_{[s]}$ and since $\det M^{-1}_{[s]} = \prod_{j \in s} \rho_j$, the desired result is obtained.

Proof of Theorem 2. For A_A-optimality we differentiate the Lagrange function, equate the derivatives to zero, multiply the first m equations by ω_j^{*2}, take the square root of both sides and sum over all j. This yields

$$\lambda^{1/2} \sum_{j=1}^{m} \omega_j^* = \sum_{j=1}^{m} \left(\sum_{k=1}^{p} a_{jk}^2 v_j \right)^{1/2}.$$

Since ω_j^* sum to one the stated first order condition follows.

The condition for D-optimality is obtained similarly. Differentiating the Lagrange function, equating the derivatives to zero, multiplying the first m equations by ω_j^*, and summing yields

$$\sum_{j=1}^{m} \frac{\sum_{r\in R_j} \left(\det A_{[r]}\right)^2 \prod_{k\in r} \rho_k^*}{\exp\left(\Psi_D\left(\omega^*\right)\right)} = \lambda \sum_{j=1}^{m} \omega_j^*.$$

The summation over R_j in the left hand side is over the p -combinations that contain j. Each unique p-combination which contains j is included in the summation as j runs from 1 to m. Hence, each p -combination appears p times and sum to $p\exp\left(\Psi_D\left(\omega^*\right)\right)$. It follows that $\lambda = p$ and the stated first order condition is obtained.

Proof of Corollary 2. For A_A-optimality note that $\sum_{k=1}^{p} a_{1k}^2 = (m-1)$ and $\sum_{k=1}^{p} a_{jk}^2 = 1$ for $j = 2, 3, \ldots, m$. Application of Theorem 2 yields the stated first order condition.

For $\mathrm{D_A}$-optimality note that $\left(\det A_{[r]}\right)^2 = 1$ for all $r \in R$. With $p = m-1$, Theorem 2 implies

$$\begin{aligned}\omega_j^* &= (m-1)^{-1} \sum_{r\in R_j} \prod_{k\in r} \left(\frac{\upsilon_k}{\omega_k^*}\right) / \exp\left(\Psi_D\left(\omega^*\right)\right) \\ &= (m-1)^{-1} \left\{ \sum_{s=1}^{m} \prod_{\substack{k=1 \\ k\neq s}}^{m} \left(\frac{\upsilon_k}{\omega_k^*}\right) - \prod_{\substack{k=1 \\ k\neq j}}^{m} \left(\frac{\upsilon_k}{\omega_k^*}\right) \right\} / \exp\left(\Psi_D\left(\omega^*\right)\right)\end{aligned}$$

and the stated first order condition is obtained.

Proof of Theorem 3. It follows from Theorem 1 that $\Psi_A\left(\omega, \upsilon\left(\theta\right)\right)$ is a sum of positive constants times $\rho_j = \upsilon_j\left(\theta_j\right)/\omega_j$. Hence, $\Psi_A\left(\omega, \upsilon\left(\theta\right)\right)$ is maximized over Θ_0 when each $\upsilon_j\left(\theta_j\right)$ is maximized. Similarly, from Theorem 1, $\exp\left(\Psi_D\left(\omega, \upsilon\left(\theta\right)\right)\right)$ is a sum of positive constants times $\det M_{[s]}^{-1} = \prod_{j\in s} \upsilon_j/\omega_j$ and it follows that $\Psi_D\left(\omega, \upsilon\left(\theta\right)\right)$ is maximized over Θ_0 when $\upsilon_j\left(\theta_j\right)$ is maximized.

References

Arnoldsson, G. (1996). Optimal allocation for linear combinations of linear predictors in generalized linear models. *J. Appl. Statist.* **23**, 493-505.

Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS.* Oxford University Press, New York.

Dette, H., Lopez, I. M., Rodriguez, I. M. and Pepelyshev, A. (2006). Maximin effect design of experiment for exponential regression models. *J. Statist. Plann. Inference* **136**, 4397-4418.

Dette, H., Trampisch, M. and Hothorn, L.A. (2007). Robust designs in non-inferiority three arm clinical trials with presence of heteroscedasticity. *Statist. Biopharmaceutical Research* **1**, 268-278.

Hedayat, A. S., Jacroux, M. and Majumdar, D. (1988). Optimal designs for comparing test treatments with controls. *Statist. Sci.* **3**, 462-491.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* **2**, 849-879.

Pettersson, H. and Nyquist, H. (2003). Computation of optimum in average designs for experiments with finite design space. *Comm. Statist. Simulation Comput.* **32**, 205-221.

Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19**, 1614-1625.

Rothamsted Experimental Station Report (1936). p. 289.

Rothamsted Experimental Station Report (1937). p. 212.

Silvey, S. D. (1980). *Optimal Design*. Chapman and Hall, London.

Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods*. 8th edition. Iowa State University Press, Ames, Iowa.

Wong, W. K. and Zhu, W. (2008). Optimum treatment allocation rules under a variance heterogeneity model. *Statist. Medicine* **27**, 4581-4595.

Yang, J., Mandal, A. and Majumdar, D. (2012). Optimal designs for two-level factorial experiments with binary response. *Statist. Sinica* **22**, 885-907.

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden.

E-mail: Ellinor.Fackle-Fornius@stat.su.se

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden.

E-mail: Hans.Nyquist@stat.su.se

(Received February 2012; accepted February 2014)