Stockholms universitet

Research Report

Department of Statistics

No. 2010:3

Optimal Allocation for Comparing Treatment Effects

Ellinor Fackle-Fornius
Hans Nyquist

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

Optimal Allocation for Comparing Treatment Effects

Ellinor Fackle-Fornius and Hans Nyquist
Stockholm University

To be presented at
Prague Stocatics 2010, August 30-September 3

Abstract

Suppose the mean responses from m-1 treatment groups in an experiment are to be compared to the mean of a control group. A uniform allocation of observations over the treatment groups is then often used. However, other allocation schemes can give a better precision in the inference. This is particularly emphasised when the variances of the responses are different in different treatment groups. Here we consider optimal allocation according to the A- and D_A-criteria for the cases of equal as well as different variances of the responses in the treatment groups. We also consider the case when costs for observations are different in different treatment groups. As expected, optimal allocation depends on the variances of the responses in the treatment groups. If the variances are unknown, a minimax strategy can be used. This means that allocation is made subject to the worst case as the variances are varied within specified intervals. In general, minimax designs are difficult to find. However, for the case of treatment groups, as is considered here, we show that the minimax strategy is very simple to apply. The efficiency of allocations according to the minimax strategy is compared to the uniform as well as the locally optimal D_A- and A-optimal allocations.

1 Introduction

This paper concerns optimal experimental design when the aim of the experiment is to compare treatment effects. Such situations are commonly encountered e.g. in clinical trials. In this setting interest is often in making comparisons of the average response in the treatment group(s) with the average response in a control group. Hence, the number of experimental groups is fixed in advance. The design problem then becomes a problem of allocating experimental units to experimental groups, i.e. to find an optimal allocation.

The general theory of optimal design allows to find an experimental design that is optimal with respect to a specific inferential goal, see for example Atkinson and Donev (1992), Atkinson, Donev and Tobias (2007), or Silvey (1980). Deriving an optimal design involves choosing an appropriate optimality criterion which will determine in what sense the design is optimal. If the inferential goal is to make estimation as precisely as possible two criteria arise as natural candidates. One is the A-optimality criterion which minimizes the sum of the variances of the estimated statistics. Another one is the D-optimality criterion which minimizes the size of the confidence region for the estimators. Some results on D-optimality when using a generalized linear model are presented in Arnoldsson (1996). Both the A- and D-optimality criteria are considered here. Comparing treatment effects means estimating a number of linear combinations of average responses resulting in the special case of D_A-optimality. No analogous term A_A-optimality has been established so we will simply use A-optimality to include all cases. It is demonstrated that the optimal allocation depends on the optimality criterion. For example, if the variances in all groups are equal the D_A-optimal criterion yields a uniform allocation, while the A-optimality criterion allocates more observations to the control group.

The results are not restricted to the case of equal group variances, it generalizes to cases with unequal group variances as well. However, the optimal allocation is a function of the group variances which are generally unknown. Several approaches to solve this issue have been proposed. One is to simply make a (hopefully educated) guess of the unknown group variances, yielding a locally optimal design. An elaboration of this idea is to specify the guess in terms of a probability distribution over some region of possible variances, yielding an optimum on average design. Alternatively, a minimax design is obtained by minimizing the maximum of the criterion function

taken over the region of possible variances. In this paper it is illustrated how to derive an optimal allocation based on minimax designs. The results are applied to experiments with binary responses. Binary responses frequently occur in clinical trials, for example the "success" outcome of an experiment may be that a test subject is cured.

In a practical situation there is most certainly a cost restriction that has to be met and the costs may differ between the experimental groups. It is also shown how such cost restrictions can be taken into account when deriving the optimal allocation.

The paper is organized as follows. In section 2 we state the model for comparing treatment groups and give a general theorem on the criterion functions for A- and D_A-optimality. These results are then specialized to the case of control group experiments in section 3, including the particular case of cost restrictions. Specifically, section 3 covers the case of equal group variances as well as the minimax approach for dealing with the problem of unknown and possibly unequal group variances. Section 4 contains some numerical examples of the minimax designs and evaluations of their efficiencies.

2 Model

Suppose there are m treatment groups and let observations from treatment group j be stochastically independent realizations of a random variable Y_j, $j = 1, 2, \ldots, m$. The variance of Y_j is denoted by v_j, and possibly dependent on a (p-vector) of parameters θ_j, so that

$$v_j = v_j(\theta_j), \quad \theta_j \in \Theta \subset R^p, \quad j = 1, 2, ..., m,$$

where Θ is the parameter space.

The information matrix obtained about the group averages if n_j observations are assigned to treatment group j is a $m \times m$ diagonal matrix M with elements of the form $\rho^{-1} = n_j/v_j$.

In the general case we assume that interest is in inference about p linear combinations $A^T\overline{Y}$, where A is an $m \times p$ matrix of constants a_{jk}, $j = 1, 2, \ldots, m$, $k = 1, 2, \ldots, p$. The information matrix of the linear combinations is $C^{-1} = \left(A^T M^{-1} A\right)^{-1}$.

The A-optimality criterion seeks to minimize the sum of variances of the estimated statistics while the D-optimality criterion seeks to minimize (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. For this particular case of estimating $A^T\overline{Y}$ D-optimality is termed D_A-optimality.

The following theorem gives general expressions for the criterion functions associated with these two criteria. For the expression for the D_A-criterion we define S as the set of all p-combinations of the set $\{1, 2, \ldots, m\}$ so that there are $\left(\begin{array}{c} m \\ p \end{array} \right)$ elements in S. We further define $A_{[s]}$ as the matrix obtained from selecting the rows $s \in S$ from the matrix A and $M_{[s]}$ as the diagonal matrix obtained from selecting the rows and columns $s \in S$ from the matrix M.

Theorem 1 *When estimating $A^T\overline{Y}$, the criterion for A-optimality is*

$$\Psi_A = tr\,(C) = \sum_{k=1}^{p} \sum_{j=1}^{m} a_{jk}^2 \rho_j,$$

and the criterion for D_A-optimality is

$$\Psi_D = \ln \det\,(C) = \ln \sum_{s \in S} \left(\det A_{[s]} \right)^2 \prod_{j \in s} \rho_j.$$

Proof. The criterion for A-optimality is obtained by evaluating the matrix multiplications, yielding

$$C^{-1} = \left(\begin{array}{ccc} \sum_{j=1}^{m} a_{j1}^2 \rho_j & \sum_{j=1}^{m} a_{j1} a_{j2} \rho_j & \cdots & \sum_{j=1}^{m} a_{j1} a_{jp} \rho_j \\ \sum_{j=1}^{m} a_{j1} a_{j2} \rho_j & \sum_{j=1}^{m} a_{j2}^2 \rho_j & \cdots & \sum_{j=1}^{m} a_{j2} a_{jp} \rho_j \\ \vdots & \vdots & & \vdots \\ \sum_{j=1}^{m} a_{j1} a_{jp} \rho_j & \sum_{j=1}^{m} a_{j2} a_{jp} \rho_j & \cdots & \sum_{j=1}^{m} a_{jp}^2 \rho_j \end{array} \right),$$

and the stated result follows immediately. The criterion for D_A-optimality is obtained by applying the Cauchy-Binet theorem. This theorem states that the determinant of a product of rectangular matrices equals the sum of determinants of products of square matrices formed by all combinations of columns and corresponding rows of the original matrices. Noting that $\det M_{[s]}^{-1} = \prod_{j \in s} \rho_j$ yields the result. ■

3 Design of control group experiments

3.1 Optimal design without cost restrictions

Suppose now that the first group is a control group and we wish to compare each of the treatment groups with the control group, i.e. we wish to make inference about the differences $\overline{Y}_1 - \overline{Y}_j$, $j = 2, \ldots, m$. The vector of differences of averages is obtained as $A^T\overline{Y}$, where A^T is a $(m-1) \times m$ matrix defined as

$$A^T = \left(\mathbf{1}_{m-1} \quad -I_{m-1}\right),$$

$\mathbf{1}_{m-1}$ being a vector of ones, I_{m-1} being the identity matrix of order $m-1$, and $\overline{Y} = \left(\overline{Y}_1, \overline{Y}_2, \ldots, \overline{Y}_m\right)^T$ is the vector of averages. The information matrix associated to the vector of differences is $C^{-1} = \left(A^T M^{-1} A\right)^{-1}$, and the covariance matrix of the vector of differences is therefore

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

Thus, when there is a control group and only one treatment group the variance of the difference is

$$V\left(\overline{Y}_1 - \overline{Y}_2\right) = \frac{v_1}{n_1} + \frac{v_2}{n_2},$$

while if there are three groups and we wish to make inference about $\left(\overline{Y}_1 - \overline{Y}_2, \overline{Y}_1 - \overline{Y}_3\right)^T$ we have the covariance matrix

$$\begin{pmatrix} \frac{v_1}{n_1} + \frac{v_2}{n_2} & \frac{v_1}{n_1} \\ \frac{v_1}{n_1} & \frac{v_1}{n_1} + \frac{v_3}{n_3} \end{pmatrix}.$$

The following theorem gives expressions for the A- and D_A-criterion functions for control group experiments.

Theorem 2 *When estimating $A^T\overline{Y}$, the criterion for A-optimality is*

$$\Psi_A = tr\left(C\right) = \left(m-1\right)\rho_1 + \sum_{j=2}^{m} \rho_j,$$

and the criterion for D_A-optimality is

$$\Psi_D = \ln\det(C) = \ln\sum_{j=1}^{m}\prod_{\substack{1\le k\le m\\ k\neq j}}\rho_k.$$

Proof. The theorem follows from application of Theorem 1 with $A = (\mathbf{1}_{m-1} \quad -I_{m-1})^T$ and $p = m-1$. Alternatively, the criterion for A-optimality is obtained by summation of the diagonal elements of C and the criterion for D_A-optimality can be found by a cofactor expansion of C. ■

For an application of this theorem to the three group example above, we find that

$$\Psi_A = 2\frac{v_1}{n_1} + \frac{v_2}{n_2} + \frac{v_3}{n_3}$$

and

$$\begin{aligned}\Psi_D &= \ln\left\{\left(\frac{v_2}{n_2}\frac{v_3}{n_3}\right) + \left(\frac{v_1}{n_1}\frac{v_3}{n_3}\right) + \left(\frac{v_1}{n_1}\frac{v_2}{n_2}\right)\right\}\\ &= \ln\left\{\left(\frac{v_1}{n_1} + \frac{v_2}{n_2}\right)\left(\frac{v_1}{n_1} + \frac{v_3}{n_3}\right) - \left(\frac{v_1}{n_1}\right)^2\right\}.\end{aligned}$$

For finding the A- and D_A-optimal designs we first change notations so that $n_j = \omega_j N$, where N is the total number of observations and ω_j, $j = 1, 2, \ldots, m$ are so called design weights, representing the proportion of the total number of observations that is allocated to treatment group j. Optimal designs are now obtained by minimizing the criteria under the restriction that the design weights sum to unity. Lagrange functions are obtained as

$$L(\omega) = \Psi(\omega) + \lambda\left(\sum_{j=1}^{m}\omega_j - 1\right),$$

where λ is a Lagrange multiplier. First order conditions for the optimal designs are obtained by differentiating the Lagrange function with respect to the design weights and with respect to the Lagrange multiplier and equating the derivatives to zero. The results are presented in the following theorem.

Theorem 3 *Denote optimal design weights by $\omega^* = (\omega_1^*, \omega_2^*, \ldots, \omega_m^*)^T$ and let $\rho_j^* = v_j/\omega_j^*$, $j = 1, 2, \ldots, m$. For an A-optimal design, the optimal design*

weights satisfy

$$\begin{aligned}\omega_1^* &= (m-1)\,\rho_1^*/\Psi_A(\omega^*)\\ \omega_j^* &= \rho_j^*/\Psi_A(\omega^*),\ j=2,3,\ldots,m.\end{aligned}$$

For a D_A-optimal design the design weights satisfy

$$\omega_j^* = (m-1)^{-1}\left(1-\prod_{\substack{k=1\\k\neq j}}^{m}\rho_k^*/\exp(\Psi_D(\omega^*))\right),\ j=1,2,\ldots,m.$$

Proof. For obtaining the condition for A-optimality we differentiate the Lagrange function

$$\begin{aligned}\frac{\partial L}{\partial \omega_1} &= -(m-1)\frac{v_1}{\omega_1^2}+\lambda\\ \frac{\partial L}{\partial \omega_j} &= -\frac{v_j}{\omega_j^2}+\lambda,\ j=2,3,\ldots,m\\ \frac{\partial L}{\partial \lambda} &= \sum_{j=1}^{m}\omega_j-1\end{aligned}$$

Equating these derivatives to zero, multiplying the first m by ω_j^* and summing yields

$$\lambda\sum_{j=1}^{m}\omega_j^* = (m-1)\frac{v_1}{\omega_1^*}+\sum_{j=2}^{m}\frac{v_j}{\omega_j^*},$$

hence, we conclude that $\lambda=\Psi_A(\omega^*)$ and the stated first order condition follows.

The condition for D_A-optimality is obtained similarly. Differentiating the Lagrange function yields

$$\begin{aligned}\frac{\partial L}{\partial \omega_j} &= -\frac{1}{\omega_j}\sum_{\substack{k=1\\k\neq j}}^{m}\prod_{\substack{1\le\ell\le m\\ \ell\neq k}}\frac{v_\ell}{\omega_\ell}/\exp(\Psi_D(\omega))+\lambda\\ &= \frac{1}{\omega_j}\left(\prod_{\substack{k=1\\k\neq j}}^{m}\left(\frac{v_j}{\omega_j}\right)/\exp(\Psi_D(\omega))-1\right)+\lambda,\ j=1,2,\ldots,m\\ \frac{\partial L}{\partial \lambda} &= \sum_{j=1}^{m}\omega_j-1.\end{aligned}$$

Equating the derivatives to zero, multiplying with ω_j^* and summing yields $\lambda = m - 1$ and the stated first order condition is obtained. ■

3.2 Optimal design with cost restrictions

Suppose now that there is a unit cost κ_j for taking observations in treatment group j, such that the cost for all observations in treatment group j is $c_j = \kappa_j n_j$. Collect the unit costs in the $m \times 1$ vector κ, so that the total cost for the experiment is $\kappa^T n$, where $n = (n_1, n_2, \ldots, n_m)^T$ is the vector of replicates to be allocated to each of the experimental groups. Suppose also that there is a cost restriction such that the total cost for the experiment may not exceed the amount G. Hence, the relative cost for observations in group j is $\tau_j = c_j/G$. The problem we consider in this section is how to allocate observations to the experimental groups so that the A- or D-criterion is minimized subject to the cost restriction

$$\begin{aligned} &\min_{n} \Psi \\ &\text{subject to } \kappa^T n = G. \end{aligned}$$

In this case optimal designs can be obtained by defining the Lagrange function

$$L(n) = \Psi(n) + \lambda\left(\kappa^T n - G\right),$$

where λ is, as before, a Lagrange multiplier.

Theorem 4 *Denote the optimal allocation vector by $n^* = (n_1^*, n_2^*, \ldots, n_m^*)^T$ and let $\rho_j^* = \upsilon_j/n_j^*$, and $\tau_j = \kappa_j/G$, $j = 1, 2, \ldots, m$. For an A-optimal design under cost restriction, the optimal allocation satisfies*

$$\begin{aligned} n_1^* &= (m-1)\,\rho_1^*/\left(\tau_1 \Psi_A(n^*)\right) \\ n_j^* &= \rho_j^*/\left(\tau_j \Psi_A(n^*)\right),\ j = 2, 3, \ldots, m. \end{aligned}$$

For a D_A-optimal design under cost restriction, the optimal allocation satisfies

$$n_j^* = \left(\tau_j (m-1)\right)^{-1} \left(1 - \prod_{\substack{k=1 \\ k \neq j}}^{m} \rho_k^*/\exp\left(\Psi_D(n^*)\right)\right),\ j = 1, 2, \ldots, m.$$

Proof. Similarly as for the case without cost restrictions, we find the derivatives of the Lagrange function. For the A-optimality criterion, the derivatives are

$$\begin{aligned}
\frac{\partial L}{\partial n_1} &= -(m-1)\frac{\upsilon_1}{n_1^2} + \lambda\kappa_1 \\
\frac{\partial L}{\partial n_j} &= -\frac{\upsilon_1}{n_j^2} + \lambda\kappa_j,\ j = 2, 3, \ldots, m, \\
\frac{\partial L}{\partial \lambda} &= \kappa^T n
\end{aligned}$$

Equating these derivatives to zero, multiplying the first m by n_j^* and summing yields

$$\lambda\kappa^T n^* = (m-1)\frac{\upsilon_1}{n_1^*} + \sum_{j=2}^{m}\frac{\upsilon_j}{n_j^*},$$

and it follows that $\lambda = \Psi_A(n^*)/G$. The stated first order condition follows immediately.

For the D_A-optimality criterion the derivatives are

$$\begin{aligned}
\frac{\partial L}{\partial n_j} &= -\frac{1}{n_j}\sum_{\substack{k=1\\k\neq j}}^{m}\prod_{\substack{1\le\ell\le m\\ \ell\neq k}}\frac{\upsilon_\ell}{n_\ell} + \lambda\kappa_j = \frac{1}{n_j}\left(\prod_{\substack{k=1\\k\neq j}}^{m}\left(\frac{\kappa_j}{n_j}\right)/\exp\left(\Psi_D(\omega)\right) - 1\right) + \lambda\kappa_j, \\
j &= 1, 2, \ldots, m \\
\frac{\partial L}{\partial \lambda} &= \kappa^T n - G.
\end{aligned}$$

Equating the derivatives to zero, multiplying with n_j^* and summing yields $\lambda = (m-1)/G$ and the stated first order condition is obtained. ■

3.3 Equal group variances

It is seen that the optimal design weights depend on the variances υ_j so that if the variances are unequal and unknown, it is impossible to compute the optimal design weights. However, in the classical model it is assumed that the variances are equal across the groups, i.e. $\upsilon_j = \sigma^2$ for all j. In this case, the unknown group variances cancel out in the formulae and optimum design weights are possible to compute as is stated in the following corollary.

Corollary 5 *If variances in all groups are equal, the A-optimal design weights are*

$$\begin{aligned}\omega_1^* &= \begin{cases} 0.5 \text{ if } m = 2 \\ \frac{\sqrt{m-1}-1}{m-2} \text{ if } m \geq 3 \end{cases} \\ \omega_j^* &= \begin{cases} 0.5 \text{ if } m = 2 \\ \frac{m-1-\sqrt{m-1}}{(m-1)(m-2)} \text{ if } m \geq 3 \end{cases} \text{, for } j = 2, 3, \ldots, m,\end{aligned}$$

and the D_A-optimal weights are

$$\omega_j^* = m^{-1}, \text{ for } j = 1, 2, \ldots, m.$$

It is here interesting to note that an A-optimal design allocates more observations to the control group while the D_A-optimal design utilizes a uniform design. For $m = 5$, for example, A-optimal design weights are $\omega_1^* = 1/3$, $\omega_2^* = \ldots = \omega_5^* = 1/6$, while D_A-optimal design weights are $\omega_1^* = \ldots = \omega_5^* = 1/5$.

When the variances are unknown but their ratios are known, say $\upsilon_j = \gamma_j \upsilon_1$ for $j = 2, 3, \ldots, m$, where γ_j are known constants, then Theorem 3 applies with $\rho_j = \gamma_j / \omega_j$.

3.4 Minimax designs

In the general case with possibly different and unknown variances across the groups, it is not possible to find A- or D-optimal designs. In this case one can guess the variances and compute locally optimal designs or one can utilize a prior distribution over the variances and compute a design that is optimum on average, the average taken with respect to the prior distribution. Another possibility is to define a region $\Theta_0 \subset \Theta$ of possible values for the parameters defining the variances, $\upsilon(\theta)$, and compute a design that is minimax over Θ_0. Hence, we seek a design that minimizes

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

Here we let $\upsilon(\theta) = (\upsilon_1(\theta_1), \upsilon_2(\theta_2), \ldots, \upsilon_m(\theta_m))^T$ be in the argument of Ψ to emphasize its dependence on unknown parameters.

We consider the case in which the parameter vector in group j is believed to belong to a set Θ_{0j} and the set Θ_0 is the Cartesian product $\Theta_0 = \Theta_{01} \times \Theta_{02} \times \ldots \times \Theta_{0m}$. The next theorem proves that maximization of the criterion Ψ over $\theta \in \Theta_0$ is obtained by maximizing each variance υ_j as the parameters θ_j are varied over Θ_{0j}. Furthermore, this result holds when considering a general linear combination of averages $A^T\overline{Y}$.

Theorem 6 *Let A^T be a $p \times m$ matrix and consider estimation of $A^T\overline{Y}$. Then $\delta = (\delta_1, \delta_2, \ldots, \delta_m)^T$, where*

$$\delta_j = \max_{\theta_j \in \Theta_{0j}} \upsilon_j(\theta_j)$$

maximize the criterion functions $\Psi_A(\omega, \upsilon(\theta))$ and $\Psi_D(\omega, \upsilon(\theta))$ over $\theta \in \Theta_0$.

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

It is here important to note that it is easy to find examples in which the response variable is dependent between groups and the results in Theorem 6 do not hold. Hence, the assumption of independence between groups is essential for Theorem 5. With independence between groups we immediately obtain the following corollary.

Corollary 7 *Theorem 3 with $\rho_j^* = \delta_j/\omega_j^*$ yields the A-minimax and D_A-minimax designs for control group experiments, respectively.*

4 Numerical examples and efficiency comparisons

4.1 Numerical examples

The results in the previous section about minimax designs for control group experiments are illustrated for an experiment with binary responses. In such case the group variances are $\upsilon_j(\theta_j) = \theta_j(1 - \theta_j)$, where θ_j is the response probability in group j. Θ_{0j} are sets of possible values for θ_j, which are

generally different across the groups. For each treatment group δ_j is simply given by the variance associated with the value of θ_j being closest to 0.5. It can be noted here that two regions $\Theta_{0j} = [0, 1]$ and $\Theta_{0j} = [0.45, 0.55]$ will yield the same maximum variance $\delta_j = 0.25$ since both include $\theta_j = 0.5$ and thereby also yield identical designs. It is noteworthy that even though the second region is smaller, reflecting a more precise prior information, it does not affect the minimax design. This is because the minimax design is entirely determined by the maximum variance, irrespective of the size of the region.

In the case of one treatment group, that is $m = 2$, applying Theorem 3 and Corollary 7 gives

$$\begin{aligned} \omega_1^* &= \frac{\rho_1^*}{\rho_1^* + \rho_2^*} = \frac{\frac{\delta_1}{\omega_1^*}}{\frac{\delta_1}{\omega_1^*} + \frac{\delta_2}{1-\omega_1^*}} \\ \omega_2^* &= 1 - \omega_1^* \end{aligned}$$

for A-optimality as well as for D_A-optimality. δ_1 is the maximum variance in the control group and δ_2 is the maximum variance in the treatment group. The expression can be rewritten in terms of the ratio between the maximum variances, $r = \delta_2/\delta_1$, as

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

The optimal control group weight ω_1^* is then given by

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

since ω_1^* is restricted to $0 < \omega_1^* < 1$. Some examples are given in Table 1. When $r = 2$, the maximum variance in the treatment group is twice as large as in the control group and the optimal allocation accordingly assigns more weight to the treatment group. When r gets larger the weight ω_2^* gets even larger. The optimal allocation for $r < 1$ is not given in the table. For example, when $r = 0.5$ the optimal allocation is $\omega_1^* = 0.586$ and $\omega_2^* = 0.414$ which are the weights for $r = 2$ but in reverse order. This is because

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

Table 1: Minimax allocation, $m = 2$ groups

r	ω_1^*	ω_2^*
1	0.5	0.5
2	0.414	0.586
5	0.309	0.691
10	0.24	0.76

When there are two treatment groups and one control group, i.e. $m = 3$, analytical solutions for the optimal weights may be obtained but the resulting expressions would be rather impractical. However, for the special case when the maximum variances in the treatment groups are equal the formulas for the optimal weights are simplified. Some algebra yields the following expressions for the A-optimal weights

$$\begin{aligned} \omega_1^* &= \frac{1}{1+\sqrt{2r}}, \\ \omega_2^* &= \omega_3^*, \end{aligned}$$

where $r = \delta_2/\delta_1 = \delta_3/\delta_1$. Similarly, the D_A-optimal weights are found to be

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

Table 2 displays different combinations of variance ratios $r_2 = \delta_2/\delta_1$ and $r_3 = \delta_3/\delta_1$. For example, $r_2 = r_3 = 1$ means that the maximum variances are equal in the three groups and the optimal allocations are as given by Corollary 5. As another example, when $r_2 = 5$ and $r_3 = 10$ the maximum variances in the first and second treatment groups are five and ten times the maximum variance in the control group, respectively. Generally, the A-optimal allocation is more concentrated to the control group compared to the D_A-optimal allocation.

4.2 Efficiency comparisons

Two relevant questions appear in this context: what are the potential gains from using an optimal allocation over a uniform allocation and how large are the potential losses if the prior assumptions about the group variances are

Table 2: Minimax allocation, $m = 3$ groups

		A-optimal allocation			D_A-optimal allocation		
r_2	r_3	ω_1^*	ω_2^*	ω_3^*	ω_1^*	ω_2^*	ω_3^*
1	1	0.414	0.293	0.293	0.333	0.333	0.333
2	2	0.333	0.333	0.333	0.28	0.36	0.36
5	5	0.24	0.38	0.38	0.213	0.394	0.394
10	10	0.183	0.409	0.409	0.167	0.417	0.417
2	5	0.279	0.279	0.442	0.25	0.333	0.417
2	10	0.236	0.236	0.528	0.233	0.318	0.449
5	10	0.208	0.328	0.464	0.191	0.378	0.431

erroneous? Therefore, efficiency comparisons are made in order to shed some light over these issues The comparisons are made in terms of the relative sizes of the criterion functions.

4.2.1 Robustness evaluations

D_A-efficiency is defined by

$$eff_D(\omega, \theta) = \left(\frac{\Psi_D(\omega, \upsilon(\theta))}{\Psi_D(\omega^*, \upsilon(\theta))} \right)^{1/p}$$

and A-efficiency is defined by

$$eff_A(\omega, \theta) = \frac{\Psi_A(\omega, \upsilon(\theta))}{\Psi_A(\omega^*, \upsilon(\theta))},$$

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

For $m = 2$ the A- and D_A-criteria result in the same allocations and therefore the efficiencies eff_A and eff_D coincide. Figure 1 displays the efficiencies for the minimax designs given in Table 1 for ratios r in the range $0 < r < 20$. The assumed maximum variance ratio underlying the minimax design is here denoted by $r^{\max}$ to avoid confusion with true r. It can be seen from the graph that the minimax designs are fairly robust to moderate deviations from the prior assumption about the variance ratio. Consider for example the assumption that $r^{\max} = 5$ illustrated by the dotted line. The efficiency of the corresponding minimax design is over 0.95 for values of r between 2 and 15, approximately. If the variances are assumed to be equal,

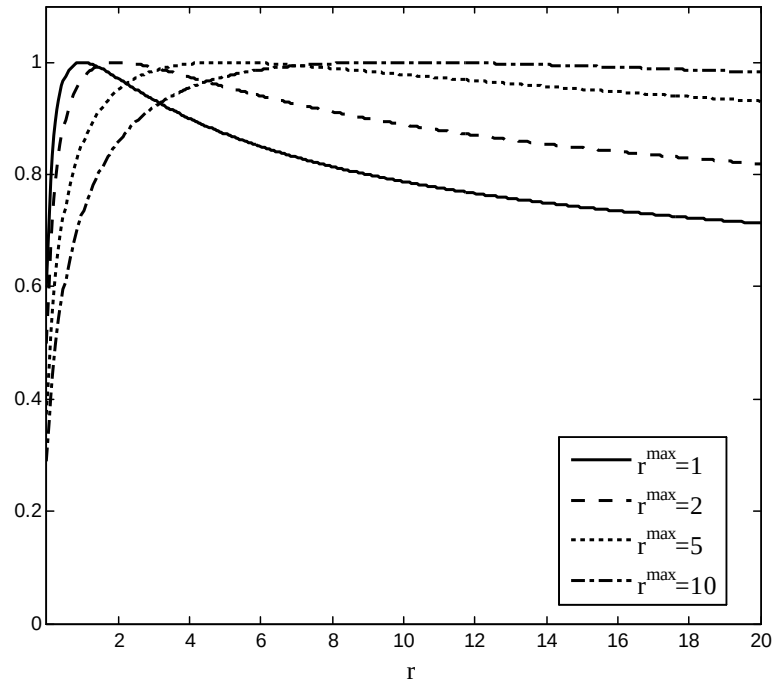
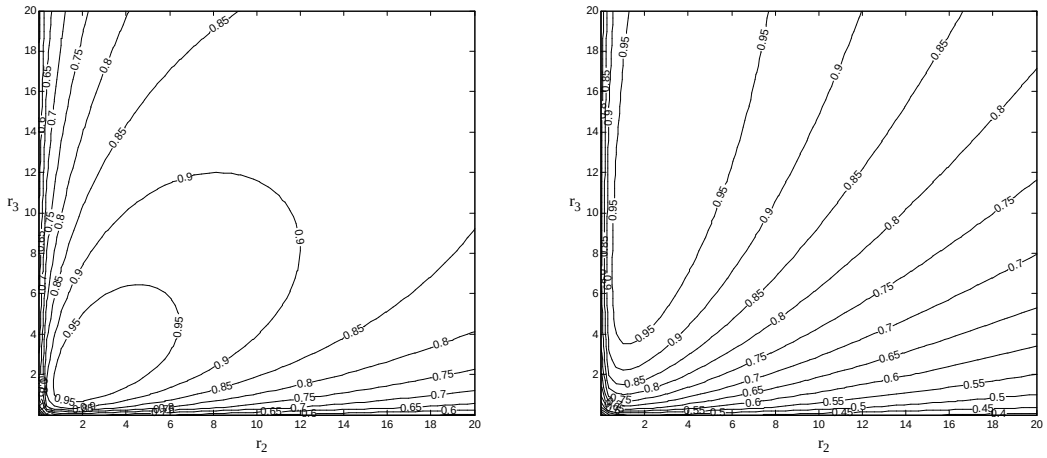


Figure 1: Efficiencies for the minimax designs given in Table 1.

that is $r^{\max} = 1$, the efficiency is close to 1 for values of r between 0.5 and 2. However, for large deviations from the prior assumption the efficiency loss can be substantial, especially when r is close to zero.

Figures 2 and 3 shows the contours of eff_A and eff_D for two minimax designs when $m = 3$. Several other examples given in Table 2 display similar patterns and are not presented here to save space.

In panel a) in Figure 2 the treatment group variances are assumed to be twice as large as the control group variance. As long as the two variance ratios are roughly equal and neither one is very large or very small the A-efficiency is quite high. On the other hand, there are areas where the A-efficiency drops a lot, especially for combinations of relatively high and low values of r_2 and r_3. In panel b) the variances in the treatment groups are not assumed to be equal and consequently the efficiency contours are not symmetrical. If the assumed ordering of the variances is correct the A-efficiency is generally



a) $r_2^{\max} = r_3^{\max} = 2$ b) $r_2^{\max} = 2,\ r_3^{\max} = 10$

Figure 2: A-efficiency contour plots.

higher. There is a region concentrated to the upper-left part of the graph were the A-efficiency is at least 0.95. There are however regions of considerable efficiency losses, particularly in the lower-right corner of the graph where r_2 is high and r_3 is low, which is contrary to the assumptions.

D_A-efficiencies are given in Figure 3. In both graphs the D_A-efficiency is above 0.9 for most values of the variance ratios. Also, there is an asymmetry in panel b) since the maximum treatment variances are assumed to be unequal.

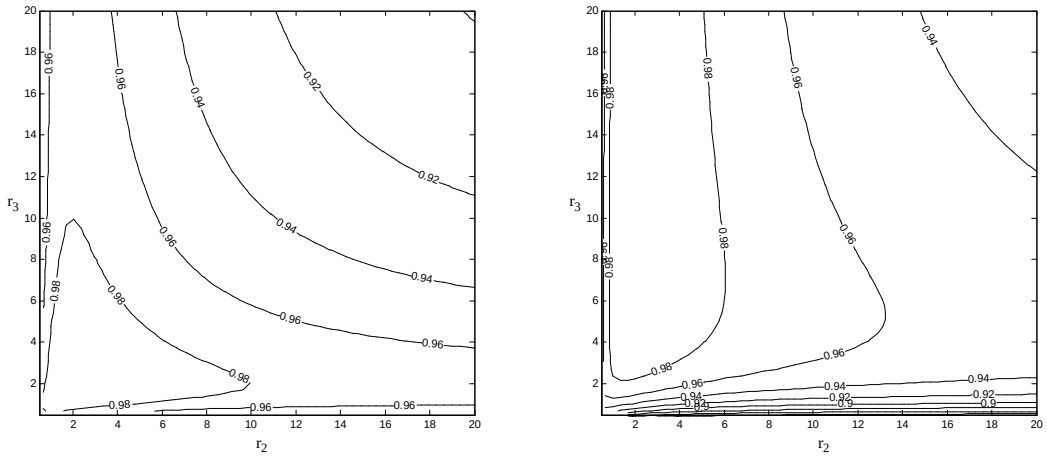
4.2.2 Efficiency over a uniform allocation

To compare the performance of the minimax allocation with a uniform allocation we evaluate the relative efficiencies

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

and

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



a) $r_2^{\max} = r_3^{\max} = 2$ b) $r_2^{\max} = 2,\ r_3^{\max} = 10$

Figure 3: D-efficiency contour plots

Figure 4 shows these efficiencies over the range $0 < r < 20$, again $eff_D = eff_A$ when $m = 2$. When the efficiency is larger than 1 the uniform allocation requires more observations to achieve the same precision and the minimax allocation is preferred. The minimax allocation for $r^{\max} = 2$ is more efficient than the uniform allocation when the variance ratio is over 1.4, approximately. The minimax allocations for $r^{\max} = 5$ and $r^{\max} = 10$ are superior to the uniform allocation when the variance ratios exceed 2.2 and 3.2 (roughly). If one has reason to believe that the variances are different in the groups the minimax allocation offers an improvement compared to the uniform allocation.

Figure 5 shows the relative A-efficiency of the minimax allocation when $r_2^{\max} = 2$ and $r_3^{\max} = 10$. When $r_2^{\max} = r_3^{\max} = 2$ the minimax A-optimal allocation is the uniform allocation and the relative efficiency will of course be 1. It can be seen that the relative A-efficiency is in favor for the minimax allocation in a region where r_3 exceeds r_2. The relative D_A-efficiencies are shown in Figure 6. The relative D_A-efficiency is slightly larger than 1 when the variance ratios are in the neighbourhood of the as assumed values and gets larger as the assumed variance ratios increases.

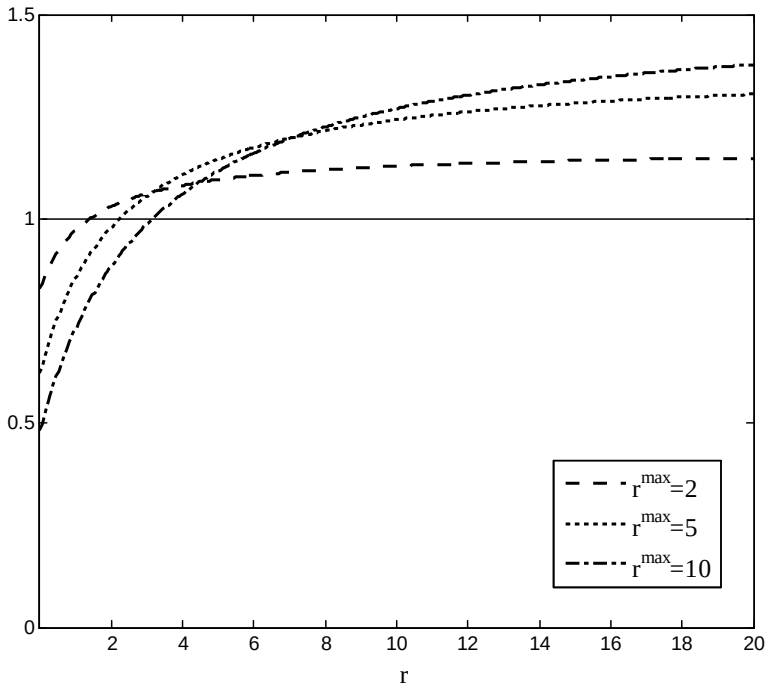
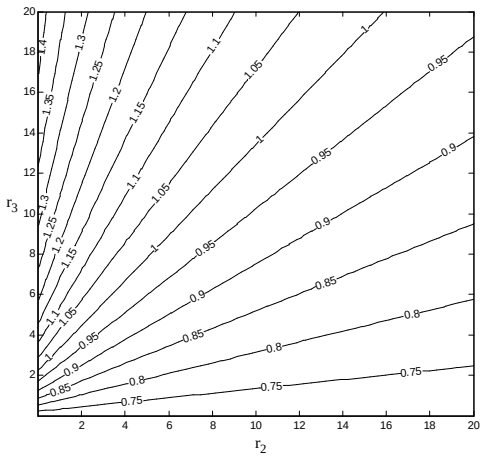
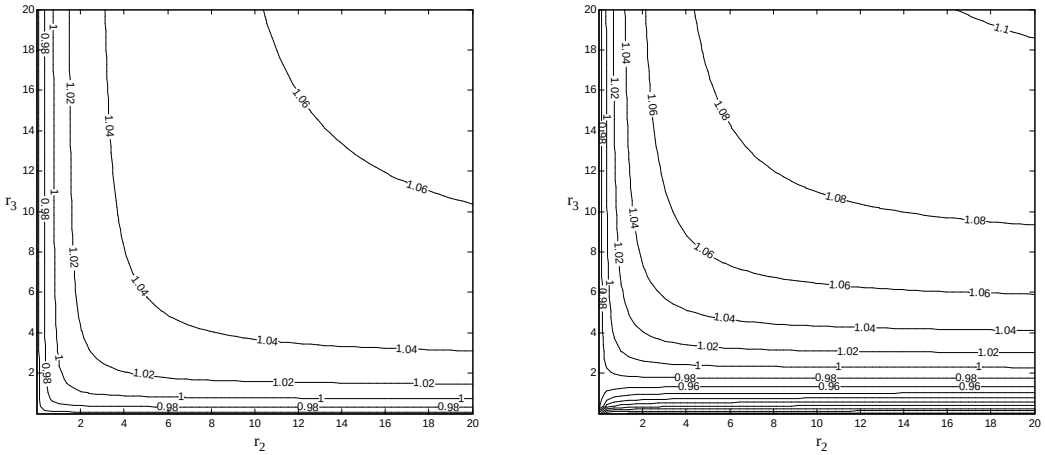


Figure 4: Efficiencies for the minimax designs given in Table 1 compared to a uniform allocation.



$r_2^{\max} = 2,\ r_3^{\max} = 10$

Figure 5: A-efficiency compared to a uniform allocation.



a) $r_2^{\max} = r_3^{\max} = 2$ b) $r_2^{\max} = 2,\ r_3^{\max} = 10$

Figure 6: D-efficiencies compared to a uniform allocation.

5 Conclusions

In this paper we derived optimal allocations for comparing the mean responses in a number of treatment groups to the mean response in a control group according to the D_A- and A-criteria. If the variances are equal in all groups the D_A-optimal allocation assign observations uniformly over the groups while the A-optimal allocation assigns more observations to the control group. When the variances are unequal more observations are assigned to groups with larger variances. The A-optimal allocation is generally more concentrated to the control group. We also considered how to adjust the optimal allocation to take into account when the costs of observations are varying between the groups. The effect of varying costs acts as a counterbalance to the effect of different variances.

The model parameters and group variances are generally unknown beforehand in which case the minimax strategy can be used. Minimax designs based on the D_A- and A-criteria were proved to be simple to derive assuming independence between the groups and the result holds for any linear combination of group averages. The minimax design is obtained as the locally optimal design based on the maximum variance in each group, the maximum taken over a specified region of possible parameter values.

Efficiency evaluations were performed which revealed that there could be substantial efficiency losses if the prior assumptions are far off. In light of these results it would be interesting to consider the so called maximin efficient designs as in Dette et al. (2006). A maximin efficient design is determined by maximizing the minimum of the D-efficiency over a specified range, thus maximin efficient designs provides a means to protect against the worst cases of efficiency losses. On the other hand, if the prior information is fairly accurate the minimax designs are quite robust and more efficient than the uniform design.

References

Arnoldsson, G. (1996) Optimal Allocation for Linear Combinations of Linear Predictors in Generalized Linear Models. *Journal of Applied Statistics,* **23**, 493-505.

Atkinson, A. C. (1982) Optimum Biased Coin Design for Sequential Clinical Trials with Prognostic Factors. *Biometrika* **69**, 61-67.

Atkinson, A.C. and Donev, A.N. (1992). *Optimum Experimental Designs*. Oxford University Press, New York.

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

Begg, C. and Kalish, L. (1984). Treatment Allocation for Nonlinear Models in Clinical Trials: The Logistic Model. *Biometrics* **40**, 409-420.

Dette, H, Lopez, I.M., Rodriguez, I.M. and Pepelyshev, A. (2006) Maximin Efficient Design of Experiment for Exponential Regression Models. *Journal of Statistical Planning and Inference* **136**, 4397-4418.

Pettersson, H. and Nyquist, H. (2003) Computation of Optimum in Average Designs for Experiments with Finite Design Space. *Communications in Statistics - Simulation and Computation* **32**, 205-221.

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