OPTIMUM DESIGNS FOR TWO TREATMENTS

WITH

UNEQUAL VARIANCES

Anthony Atkinson, August 2014

Department of Statistics
London School of Economics
London WC2A 2AE, UK

a.c.atkinson@lse.ac.uk

Introduction

- Clinical trials are often designed to evaluate two treatments in the presence of covariates (prognostic factors).
- Design problem: proportional allocation to treatments when the variances of response to the two treatments are different.
- If the focus is the treatment difference and there are no covariates, unequal (Neyman) allocation minimizes the variance of the estimated treatment difference.
- Contrariwise, if estimation of both treatment effects (rather than the difference) is the aim, equal allocation is optimum, however unequal the variances.
- However, surprisingly, these results no longer hold if the effects of the covariates are also of interest.

Personalised Medicine

- In the design of clinical trials often randomize over covariates, leading to a similar structure of the covariates for the treatments.
- In sequential trials restricted randomization can avoid selection bias. Here only consider trials in which all patients are recruited before the trial starts.
- Randomization assumes effect of prognostic factors is not itself of interest; the parameters for the effects of the covariates are ignored.
- In personalized medicine imperative to estimate not only the treatment effects but also the effects of the covariates, so that the appropriate treatment can be chosen for each patient.

The Model

- The two-treatment model with covariates for observation i is

$$y_i = \beta^T f(x_i) + \sigma_\delta \epsilon_i = \alpha_1 \delta_i + \alpha_2 (1 - \delta_i) + \sum_{j=1}^{k} \gamma_j x_{ij} + \sigma_{\delta_i} \epsilon_i .$$

- The treatment indicator δ_i is one if treatment 1 is allocated and is zero otherwise.
- The heterogeneity of variance is modelled by σ_δ^2 with:

$$\sigma_\delta^2 = \begin{cases} \sigma_1^2 & (\delta_i = 1) \\ \sigma_2^2 & (\delta_i = 0). \end{cases}$$

- The errors ϵ_i are independent with a standard normal distribution $\mathcal{N}(0, 1)$.
- As well as the treatment effect α_j, the response depends linearly on the values of k covariates $x_1, \ldots, x_k$. There is no constant term in the model. The total number of parameters to be estimated is $p = k + 2$.

Models and D-optimality

- In the regression model

$$\mathsf{E}\, Y = F\beta,$$

  F is an $n \times p$ matrix of known constants which may include powers and products of the k covariates x. The ith row of f is $f^T(x_i)$. The additive normal errors for observation i have variance σ_i^2. With

$$\Sigma = \text{diag}\ \sigma_i^2$$

  the (weighted) least squares estimate of β is

$$\hat{\beta} = (F^T\Sigma^{-1}F)^{-1}F^T\Sigma^{-1}y,$$

  with information matrix $F^T\Sigma^{-1}F$ and covariance matrix $(F^T\Sigma^{-1}F)^{-1}$.
- The volume of the confidence region for all p parameters is inversely proportional to the square root of the determinant $|F^T\Sigma^{-1}F|$. Designs which maximize this determinant are called D-optimum; they minimize the generalized variance of $\hat{\beta}$.

Continuous and Exact Designs

- It is helpful to avoid dependence of the design on n.
- Continuous designs are represented by the measure ξ over the design region $\mathcal{X}$. If the design has trials at n distinct points in $\mathcal{X}$,

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & \dots & x_n \\ w_1 & w_2 & \dots & w_n \end{array} \right\}.$$

- The information matrix for the continuous design ξ with the heteroskedastic model

$$M(\xi) = \sum_{i=1}^{t} w_i f(x_i) f^{\mathrm{T}}(x_i)/\sigma_i^2.$$

- The *standardized* variance of the predicted response for continuous designs is

$$d(x,\xi) = f^{\mathrm{T}}(x) M^{-1}(\xi) f(x)/\sigma_i^2.$$

- The D-optimality of a proposed optimum continuous design can be checked using the General Equivalence Theorem relating maximization of $\log |M(\xi)|$, or equivalently $|M(\xi)|$, to properties of $d(x,\xi)$.
- Let the maximum over $\mathcal{X}$ of $d(x,\xi)$ be $\bar{d}(\xi)$ and suppose ξ^* maximizes $\log |M(\xi)|$. Then $\bar{d}(\xi^*) = p$, the dimension of β.

Background to the Design

- With constant error variance σ^2, the optimum design does not depend on the value of σ^2
- With two variances, the optimum design depends on the ratio $\tau = \sigma_2^2/\sigma_1^2$. Without loss of generality we take $\sigma_1^2 = 1$. Assume τ known.
- The designs put weight w on treatment 1 and weight $1 - w$ on treatment 2.
$$\xi = \left\{ \begin{array}{cc} T_1 & T_2 \\ w & 1 - w \end{array} \right\}.$$
- There is a symmetry between optimum designs for τ and those calculated for the ratio $\sigma_1^2/\sigma_2^2 = 1/\tau$. If w^* is the optimum weight for τ, the optimum weight for $1/\tau$ is $1 - w^*$.

The Information Matrix

- Proceed by assuming that a particular class of designs is optimum, and then show that it satisfies the equivalence theorem for a much wider class, and so is optimum in that class too.
- The optimum design does not depend on the scaling of the k linear factors. Take the design region $\mathcal{X}$ as the cube for which $-1 \leq x_j \leq 1, \quad (j = 1, \ldots, k)$.
- Assume (test this) the optimum design for the k linear factors is a 2^k factorial at the points ± 1, with complete factorials at both treatment levels. (Interactions?)
- Calculation of the information matrix for the design requires

$$\sum_{i=1}^{n} x_{ij}^2/\sigma_i^2 = w + (1-w)/\tau \quad (j = 1, \ldots, k).$$

- For the general design putting weight w on the trials of the 2^k factorial for treatment 1, the information matrix is

$$M(\xi) = \begin{pmatrix} w & 0 & 0 & \cdots & 0 \\ 0 & (1-w)/\tau & 0 & \cdots & 0 \\ 0 & 0 & w + (1-w)/\tau & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & w + (1-w)/\tau \end{pmatrix}.$$

No or One Covariate

- The diagonal information matrix has determinant

$$
\begin{aligned}
|M(\xi)| &= w(1-w)[\{w+(1-w)/\tau\}^k]/\tau \\
&= w(1-w)[\{1+w(\tau-1)\}^k]/\tau^{k+1},
\end{aligned}
$$

  which is to be maximized as a function of w. The term in τ^{k+1} does not affect the optimum value of w.

-
  1. Find optimum weight w^*.
  2. Prove it is optimum not just for 2^k factorial

- No covariate: $k=0$. $|M(\xi)| \propto w(1-w)$, so that the optimum value $w^* = 1/2$, even though the variances are unequal. ✓

- One covariate: $k=1$. Find w to maximize

$$h = w(1-w)(w\tau + 1 - w),$$

$$\text{so that} \quad w^* = \frac{(2-\tau) - \surd(1-\tau+\tau^2)}{3(1-\tau)}.$$

  Initial check: $w^* = 0.5$ when $\tau = 1$ (L'Hôpital's rule).

One Covariate: Optimality

- Limits. Rearranging the derivative of h

$$(3w^* - 1)(w^* - 1) - \tau w^*(3w^* - 2) = 0.$$

  As $\tau \to 0, w^* \to 1/3$. As $\tau \to \infty$, $w^* \to 2/3$.
- Symmetric in τ and $1/\tau$. Much less extreme than Neyman allocation (no covariate), when the limits of w^* are 0 and 1.
- Check Optimality. Use Equivalence Theorem.

$$M^{-1}(\xi) = \text{diag}\left(\frac{1}{w} \quad \frac{\tau}{1-w} \quad \frac{\tau}{1+w(\tau-1)}\right).$$

- The check of optimality requires $\bar{d}(\xi)$, the maximum of $d(x,\xi)$, over the whole of the design region $\mathcal{X}$. We consider separately the allocations of treatments 1 and 2.

One Covariate: Equivalence Theorem

- Treatment 1. For treatment 1, Var(y) = 1 and

$$f(x) = (1 \quad 0 \quad x)^T,$$

$$d_1(x, \xi) = \frac{1}{w} + \frac{\tau x^2}{1 + w(\tau - 1)}.$$

A maximum for $x = \pm 1$, so it is only necessary to check at these points. For the optimum design the value of $\bar{d}_1(\xi^*)$ will be 3. Need to verify that

$$\frac{1}{w^*} + \frac{\tau}{1 + w^*(\tau - 1)} = \frac{2\tau w^* + (1 - w^*)}{w^*\{\tau w^* + (1 - w^*)\}} = 3,$$

which follows by simplification from expression for w^*.

- Treatment 2. For treatment 2, Var$(y) = \tau$ and

$$f(x) = (0 \quad 1 \quad x)^T,$$

Divide by τ to allow for heterogeneity:

$$d_2(x, \xi) = \frac{1}{1 - w} + \frac{x^2}{1 + w(\tau - 1)},$$

again a maximum at $x = \pm 1$. Similar arguments show the maximum value, $\bar{d}_2(\xi^*) = 3$. Thus the design is indeed D-optimum over $\mathcal{X}$.

Several Covariates

- The arguments are similar
- With k covariates the D-optimum design now maximizes the information matrix

$$h = w(1-w)(w\tau + 1 - w)^k\}.$$

  Differentiation and putting the derivative equal to zero yields a polynomial of degree $k+1$ to be solved for the optimum weight. It is easiest directly to maximize h numerically.
- Limits. Rearranging the derivative in terms including and excluding τ

$$[\{(k+2)w - 1\}(w-1) - \tau w\{(2+k)w - (k+1)\}](\tau w + 1 - w)^{k-1} = 0.$$

  As $\tau \rightarrow 0$, the equation becomes $\{(k+2)w - 1\}(1-w)^k = 0$. So the limiting value of w^* is $1/(k+2)$. Likewise, for large τ the dominant equality becomes $(2+k)w - (k+1) = 0$, so that $w^* = (k+1)/(k+2)$. Again, the weights for τ and $1/\tau$ sum to one.
- The designs for extreme τ become less balanced as the number of covariates k increases.

Several Covariates: Optimality

- Generalization of the arguments for $k = 1$ shows that the optimum design for k factors is of the 2^k factorial form. Namely:
- treatment 1.

$$f(x) = (1 \quad 0 \quad x_1 \quad \dots \quad x_k)^T,$$

At the factorial points (the maximum),

$$d_1(x, \xi) = \frac{1}{w} + \frac{k\tau}{1 + w(\tau - 1)}.$$

- Setting this value of $\bar{d}_1(\xi)$, i.e. $\bar{d}_1(w^*)$, equal to the value of $k + 2$ for the optimum design, yields the quadratic

$$w^*(w^* - 1)(k + 2)(1 - \tau) - w^*(1 + \tau) + 1 = 0.$$

- Similar result when allocating treatment 2.
- Solving this equation also yields the value of w^*, more easily than numerical maximization.

Numerical

Variance ratio τ	Number of variables k 1	2	3	5	7	10
0	0.3333	0.25	0.2	0.1429	0.1111	0.0833
0.2	0.3681	0.2873	0.2347	0.1712	0.1346	0.1018
0.4	0.4046	0.3333	0.2812	0.2124	0.1700	0.1305
0.6	0.4402	0.3876	0.3432	0.2756	0.2284	0.1808
0.8	0.4725	0.4458	0.4202	0.3735	0.3333	0.2843
1	0.5	0.5	0.5	0.5	0.5	0.5
1.25	0.5275	0.5542	0.5798	0.6265	0.6667	0.7157
1.667	0.5598	0.6124	0.6568	0.7244	0.7716	0.8192
2.5	0.5954	0.6667	0.7188	0.7876	0.8300	0.8695
5	0.6319	0.7127	0.7653	0.8288	0.8654	0.8982
∞	0.6667	0.75	0.8	0.8571	0.8889	0.9167

Weights w^* on treatment 1 for D-optimum designs as a function of τ and k: first and last lines, limiting weights as $\tau \to 0$ and 1. For all k, the weights are 0.5 for τ = 1. For fixed k, reading down each column gives a symmetrical sigmoid curve which becomes increasingly steep around $\tau = 1$ as k increases. The values of τ in the lower half of the table are the reciprocals of those in the upper half, so that, working out from $\tau = 1$, the weights for τ and $1 - \tau$ sum to one.

Nuisance Parameters

- Using optimum design theory when only the α are of interest, the γ beng nuisance parameters.

$$y_i = \beta^T f(x_i) + \sigma_\delta \epsilon_i = \alpha_1 \delta_i + \alpha_2 (1 - \delta_i) + \sum_{j=1}^{k} \gamma_j x_{ij} + \sigma_{\delta_i} \epsilon_i.$$

- Generalized D- or D_A-optimality is useful when only some linear combinations of the parameters are of interest. A particular case is D_S-optimality, when the linear combinations pick out subsets of parameters.
- Let A be a $p \times s$ matrix of known constants, with $s < p$. Then the D_A-optimum design minimizes the generalized variance $|A^T M^{-1}(\xi) A|$. For $s = 1$ we obtain the $p \times 1$ vector a and the scalar variance $a^T M^{-1}(\xi) a$ is minimized.

Neyman Allocation Again

- With the vector $a = (1 \quad -1 \quad 0 \quad \ldots \quad 0)^T$, the design minimizes Var$(\hat{\alpha}_1 - \hat{\alpha}_2)$.
- From the information matrix the design therefore minimizes

$$a^T M^{-1}(\xi) a = 1/w + \tau/(1-w).$$

Differentiation and setting the derivative to zero yields

$$w^* = 1/(\sqrt{\tau} + 1) = \sigma_1/(\sigma_1 + \sigma_2),$$

the Neyman solution, since $\tau = \sigma_2^2/\sigma_1^2$.

- For the optimum design for the two treatment parameters individually

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

when the quantity to be minimized is

$$|A^T M^{-1}(\xi) A| = \begin{vmatrix} 1/w & 0 \\ 0 & \tau/1-w \end{vmatrix} = \frac{\tau}{w(1-w)},$$

so that $w^* = 1/2$, regardless of the value of τ. (See Fedorov).

Estimation of τ

- D-optimum designs for homoskedastic regression do not depend on the value of σ^2 (although power will).
- The heteroskedastic designs here do depend on the value of τ.
- The value of τ can be estimated from a sequential experiment.
  - For each n estimate σ_j^2 from patients who have received treatment j using unweighted least squares. (So obtain j sets of estimates of the parameters α and γ).
  - Hence obtain an estimate of τ.
  - Use this estimate to construct Σ and so obtain one set of parameter estimates and then in the sequential design.
- Is this the best procedure? Could iterate the estimation of τ.