Tatra Mt. Math. Publ. **17** (1998), 133–139

TATRA MOUNTAINS
Mathematical Publications

BAYESIAN AND OTHER BIASED-COIN DESIGNS FOR SEQUENTIAL CLINICAL TRIALS

Anthony C. Atkinson

ABSTRACT. A relationship is established between the Bayesian biased-coin design suggested by Ball *et al.* (1993) and the randomised optimum design suggested by Atkinson (1982). Properties of the Bayesian design are found by simulation.

1. Introduction

Patients arrive sequentially and are to be given one of t treatments. In order to avoid the suspicion of conscious or unconscious bias, treatments should be allocated at random. But, because it is not known when the trial will terminate, there needs to be a balance of the number of patients receiving each treatment. If there are just two treatments and no measurements of prognostic factors on the patients, random allocation chooses either treatment with probability one half. To enforce balance the under-represented treatment should be assigned with probability one. A compromise between these two schemes is the biased-coin design of Efron [1971] in which the under-represented treatment is assigned with a probability between 0.5 and one. Efron chose 2/3, a value which he shows has good properties and which is also the simplest ratio of integers in the range one half to one. To extend these ideas to t treatments and prognostic factors requires modelling the response.

Optimum design theory for an appropriate linear model is developed in §2 and a randomized form of D_A-optimality is introduced. The Bayesian biased-coin design of Ball, Smith, and Verdinelli (1993) is described in §3 and the relationship between the two schemes is established. Simulation is used to augment theoretical results on the loss in information caused by the randomisation. More background and an introduction to the large mathematical statistical literature on the allocation of treatments in sequential clinical trials is given by Atkinson (1999b).

2000 Mathematics Subject Classification: Primary: 92C50; Secondary: 62K05.
Keywords: balance, bias, biased-coin design, D_A-optimum design, loss, randomization.

2. Biased-Coin D_A-optimum designs

Each patient receives one treatment which is allocated in the knowledge of a vector of prognostic factors. In many trials the response will be approximately normally distributed, perhaps after transformation: one example would be the logarithm of survival time. It is then appropriate to use the regression model

$$E(Y) = G\omega = E\alpha + Z\gamma\,, \tag{1}$$

for the responses of the first n patients. In (1) E is the $n \times t$ matrix of indicator variables for the treatments with one non-zero entry per row, and Z is the $n \times (q-1)$ matrix of prognostic factors, including interactions and other terms, if required. The γ are nuisance parameters. However interest is only in contrasts between the α, not in the mean level of response. For example, if there are only two treatments, experiments should be designed to estimate $\alpha_1 - \alpha_2$ with minimum variance. With more than two treatments a set of $t-1$ contrasts orthogonal to the mean is required, for example

$$L^T = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 1 & 0 & -1 & \dots & 0 \\ & & \vdots & & \\ 1 & 0 & 0 & \dots & -1 \end{pmatrix}. \tag{2}$$

Because the γ in (1) are nuisance parameters, the contrasts in (2) need augmenting by a matrix of zeroes

$$A^T = (L^T \;\; 0) \tag{3}$$

to reflect the required interest in the parameters. In all there are q nuisance parameters.

One rule for allocating treatments uses a randomised version of the iterative construction of D_A-optimum designs minimizing $|A^T(G^TG)^{-1}A|$. In the iterative construction of optimum designs the $(n+1)$st patient receives the treatment for which the variance after n trials,

$$d_A(j,n,z_{n+1}) = g_{n+1}^T(G_n^TG_n)^{-1}A\{A^T(G_n^TG_n)^{-1}A\}^{-1}A^T(G_n^TG_n)^{-1}g_{n+1}\,, \tag{4}$$

is maximum. In (4) g_{n+1} is the $(n+1)$st row of G_{n+1}. The design depends not only on the previous allocations and the matrix of prognostic factors Z_n, contained in G_n, but also on z_{n+1}, the vector of prognostic factors for the new patient. To emphasize these dependencies the variances have been written $d_A(j,n,z_{n+1})$, $j = 1,\dots,t$. In the iterative construction all treatments will asymptotically be allocated equally often and the variances will tend to equality.

To provide a randomised form of this iterative construction, treatment j is allocated with probability

$$\pi_A(j\,|\,z_{n+1}) = \frac{d_A(j,n,z_{n+1})}{\sum_{j=1}^{t} d_A(j,n,z_{n+1})}\,. \tag{5}$$

In (5) the variances $d_A(\cdot)$ could be replaced by any monotone function $\psi\{d_A(\cdot)\}$. For example Ball, Smith, and Verdinelli (1993) described a procedure with a Bayesian biased coin. In the next section it is shown that their procedure is a member of this more general class.

This sequential scheme does not depend on the outcome of the treatment, that is the observations y, unlike the adaptive designs reviewed in Rosenberger (1996).

3. A Bayesian biased-coin

To include both variance and bias in a single criterion Ball, Smith, and Verdinelli (1993) suggest that the probabilities of treatment selection $\pi_B(j\,|\,z_{n+1})$ be chosen to maximize the utility

$$U = \sum_{j=1}^{t} \pi_B(j\,|\,z_{n+1})\phi(M_{j,n+1}) - \gamma\left\{\sum_{j=1}^{t} \pi_B(j\,|\,z_{n+1}) \log \pi_B(j\,|\,z_{n+1})\right\}, \tag{6}$$

where

$$M_{j,n+1} = G^T_{j,n+1} G_{j,n+1}\,,$$

is the information matrix after $n+1$ patients, if treatment j is allocated to the $(n+1)$st patient. In (6) when $\gamma = 0$ the sequential allocation is non-random, the next treatment being chosen to maximize $\phi(M_{j,n+1})$. For D_A-optimality

$$\phi(M_{j,n+1}) = -\log |A^T(G^T_{j,n+1}G_{j,n+1})^{-1}A| \tag{7}$$

so that the sequential construction of the optimum designs is recovered. As $\gamma \to \infty$, the procedure tends towards equal allocation through maximization of $\left\{\sum \pi_B(\cdot)\log \pi_B(\cdot)\right\}$. These two values of γ thus provide procedures which respectively maximize balance and minimize selection bias. It is however not clear what are the properties of the procedure for any intermediate value of γ. However the criterion can be informatively re-written in terms of variances.

Differentiation of (6) shows that the optimum treatment allocation probabilities are given by

$$\pi_B(j\,|\,z_{n+1}) = \exp\left\{\phi(M_{j,n+1})/\gamma\right\} \Big/ \sum_{k=1}^{t} \exp\left\{\phi(M_{k,n+1})/\gamma\right\}, \tag{8}$$

with $\phi(\cdot)$ given by (7). A simpler form of this allocation probability is found by using the variances of §2. The relationship comes from the extension of the standard matrix result

$$|G_{j,n+1}^T G_{j,n+1}| = \{1 + d(j,n,z_{n+1})\}|G_n^T G_n|, \tag{9}$$

and yields

$$\pi_B(j\,|\,z_{n+1}) = \frac{\{1 + d_A(j,n,z_{n+1})\}^{1/\gamma}}{\sum_{k=1}^{t} \{1 + d_A(k,n,z_{n+1})\}^{1/\gamma}}. \tag{10}$$

The Bayesian criterion is thus one generalization of (5) with $\psi(u) = (1+u)^{1/\gamma}$.

4. Loss

To compare designs we use the increase in the variance of the estimated parameters which is caused by the slight imbalances resulting from randomisation.

The D_A-optimum design for the model (1) requires minimization of

$$|A^T(G^TG)^{-1}A| = |L^T\{E^TE - E^TZ(Z^TZ)^{-1}Z^TE\}^{-1}L|. \tag{11}$$

For the optimum design, in which an equal number of patients is allocated to each treatment, and there is balance over all prognostic factors,

$$|A^T(G^TG)^{-1}A| = t^t/(n^{t-1}).$$

The efficiency of any other design is then

$$\mathcal{E}_n = \left(\frac{t^t/(n^{t-1})}{|A^T(G^TG)^{-1}A|}\right)^{1/(t-1)},$$

so that the effective number of trials is $n\mathcal{E}_n$. The loss is then

$$\mathcal{L}_n = n(1 - \mathcal{E}_n). \tag{12}$$

This important measure expresses the loss of information due to imbalance. If the design is exactly balanced, $\mathcal{L}_n$ is zero. Otherwise the loss of information is expressed in terms of number of patients. For the randomized designs studied here $\mathcal{L}_n$ is a random variable. The results of Smith [1984a] and of Smith [1984b], derived using martingale theory, provide a basis for calculating asymptotic values $\mathcal{L}_\infty$ for some schemes, but not for the Bayesian one of Section 3. Burman [1996] focused attention on $\mathcal{L}_n$. He also gave explicit expressions for $\mathcal{L}_\infty$ for two schemes. For completely random allocation the expected value of $\mathcal{L}_\infty = q$, the number of nuisance parameters. This result is derived by Cox [1982] by elementary arguments, but goes back much further in the literature on randomisation and experimental design. It is for example implicit in Cox [1957]. Burman [1996, p. 39] uses Smith's results to show that the asymptotic loss for randomised D_A-optimality (5) has the surprising value of $q/5$. Other rules give other functions of q for $\mathcal{L}_\infty$. To study the properties of $\mathcal{L}_n$ for small samples, Burman [1996] used simulation, as is done here for the Bayesian criterion. But first some analytical progress is possible.

The variance $d_A(\cdot)$ in (5) and (10) is not normalized. Initially, when $n = t$, the design cannot be balanced and the loss is high. Then, as n increases, especially for small values of γ in (10), the Bayesian rule is almost that of the iterative construction of the design so that the loss rapidly decreases. Finally, for larger values of n the loss increases again with n. Comparison of (5) with (10) is informative about the asymptotic behaviour of the Bayesian procedure. Because the variance $d_A(\cdot)$ has not been normalized it will have the value $(t-1)/n$ for a balanced design. In (5) multiplication of the variances by n will leave the criterion unchanged and will recover the original formulation of Atkinson [1982] in which the probabilities converge to values which give an asymptotic loss of $q/5$. But the Bayesian criterion (10) would be affected by this multiplication. The probabilities in (10) converge to the value $1/t$, leading to random allocation and a loss of q as $n \to \infty$.

Simulations support this interpretation. Fig. 1 is the average loss for 10,000 simulations when $q = 5$ and $t = 3$. The values of γ were chosen to give a wide distribution of values of $\mathcal{L}_{200}$. As the theoretical considerations above suggested, the Bayesian procedure gives a loss which initially decreases. But, as n increases, the allocation becomes increasingly random and the loss begins to rise towards q. For given values of the variances $d_A(\cdot)$, decreasing the value of γ makes the allocation probabilities less uniform and so decreases the loss. For $\gamma = 0.03$ $\mathcal{L}_{200} = 2.539$.

Simulations for $q = 10$ and, again $t = 3$, are in Fig. 2. Not only are the values of γ the same in the two plots, but the values of the losses for $q = 10$ are very close to twice those for $q = 5$, suggesting some invariance. Now for $\gamma = 0.03$, $\mathcal{L}_{200} = 4.978$. In fact it is hard to tell apart the right hand halves of the two figures if the losses for $q = 5$ are doubled to bring them to the same magnitude as those for $q = 10$.

The bias associated with these rules is evaluated by simulation in Atkinson [1999a], which also includes plots of losses for the Bayesian rules when $t = 2$. These plots are very similar to those of Figures 1 and 2.

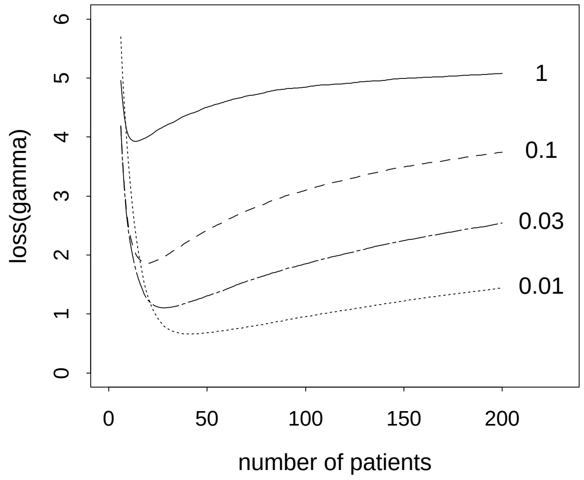

FIGURE 1. Loss $\mathcal{L}_n$ for four Bayesian strategies for sequential allocation of treatments with $t = 3$ and $q = 5$. Means of $1,000$ simulations.

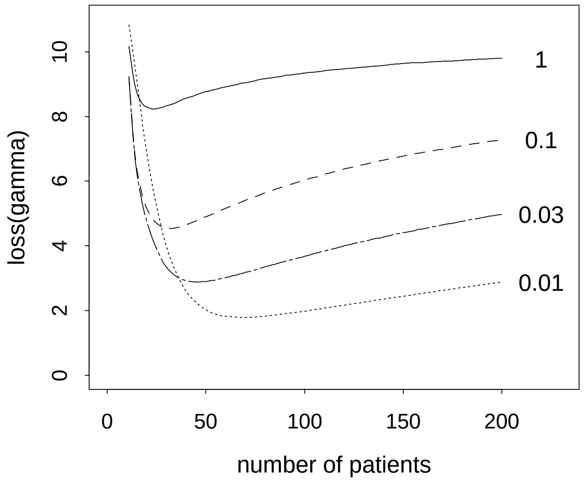

FIGURE 2. Loss $\mathcal{L}_n$ for four Bayesian strategies for sequential allocation of treatments with $t = 3$ and $q = 10$. Means of $1,000$ simulations.

REFERENCES

[1] ATKINSON, A. C.: *Optimum biased coin designs for sequential clinical trials with prognostic factors*, Biometrika **69** (1982), 61–67.

[2] ATKINSON, A. C.: *The comparison of designs for sequential clinical trials with covariate information*, (under revision).

[3] ATKINSON, A. C.: *Optimum biased-coin designs for sequential treatment allocation with covariate information*, Statist. Medicine **18** (1999), 1741–1752.

[4] BALL, F. G.—SMITH, A. F. M.—VERDINELLI, I.: *Biased coin designs with a Bayesian bias*, J. Statist. Plann. Inference **34** (1993), 403–421.

[5] BURMAN, C. F.: *On Sequential Treatment Allocations in Clinical Trials*, Department of Mathematics, Göteborg, 1996.

[6] COX, D. R.: *The use of a concomitant variable in selecting an experimental design*, Biometrika **44** (1957), 150–158.

[7] COX, D. R.: *Randomization and concomitant variables in the design of experiments*, in: Statistics and Probability: (G. Kallianpur, J. K. Krishnaiah and J. K. Ghosh, eds.), Essays in Honor of C. R. Rao, North-Holland, Amsterdam, 1982, pp. 197–202.

[8] EFRON, B.: *Forcing a sequential experiment to be balanced*, Biometrika **58** (1971), 403–417.

[9] ROSENBERGER, W. F.: *New directions in adaptive designs*, Statist. Sci. **11** (1996), 137–149.

[10] SMITH, R. L.: *Properties of biased coin designs in sequential clinical trials*, Ann. Statist. **12** (1984), 1018—1034.

[11] SMITH, R. L.: *Sequential treatment allocation using biased coin designs*, J. Roy. Statist. Soc., Ser. B **46** (1984), 519–543.

Received March 25, 1998

Department of Statistics
London School of Economics
and Political Science
Houghton Street
UK–London WC2A 2AE
GREAT BRITAIN
E-mail: a.c.atkinson@lse.ac.uk