Development Engineering 1 (2016) 12–25

Contents lists available at ScienceDirect

Development Engineering

journal homepage: www.elsevier.com/locate/deveng

The pursuit of balance in sequential randomized trials

Raymond P. Guiteras [a,*], David I. Levine [b], Thomas H. Polley [c]

[a] Department of Economics, 3114 Tydings Hall, University of Maryland, College Park, MD 20142, USA
[b] U.C. Berkeley Haas, USA
[c] Duke University, USA

ARTICLE INFO

Article history:
Received 7 August 2015
Received in revised form
6 October 2015
Accepted 7 November 2015
Available online 2 December 2015

JEL classification:
C93
O12

Keywords:
Stratification
Sequential randomization
Design of Experiments

ABSTRACT

In many randomized trials, subjects enter the sample sequentially. Because the covariates for all units are not known in advance, standard methods of stratification do not apply. We describe and assess the method of D_A-optimal sequential allocation (Atkinson, 1982) for balancing stratification covariates across treatment arms. We provide simulation evidence that the method can provide substantial improvements in precision over commonly employed alternatives. We also describe our experience implementing the method in a field trial of a clean water and handwashing intervention in Dhaka, Bangladesh, the first time the method has been used. We provide advice and software for future researchers.

1. Introduction

Randomized-controlled trials (RCTs) are an increasingly important tool for policy evaluation and estimation of economic parameters. However, they are expensive, and efficient use of limited resources (funding, inputs from implementation partners, and researchers' time) requires that they be designed carefully. In an important contribution, Bruhn and McKenzie (2009) reviewed stratification methods that were common in economics RCTs at the time, and showed that large gains in precision were available by adopting more sophisticated stratification methods from the clinical trials literature. These stratification methods require researchers to obtain stratification covariates from all subjects prior to randomization. However, this is not always feasible. In clinical trials, subjects are often allocated to treatment as they arrive. In field trials, operational constraints may prevent defining and surveying the full sample frame in advance. In such situations, subjects must be assigned *sequentially*, with the researcher only learning the value of the stratification variables for that subject's at the time of enrollment and assignment.[1]

In this paper, we propose the use of D_A-*optimal sequential allocation* (Atkinson, 1982) to improve balance and power when subjects are enrolled sequentially. The D_A-optimal method minimizes imbalance given the constraint of not knowing covariate values in advance. We describe the method and its properties, and provide an algorithm for its implementation. We conduct a set of simulations, based on Bruhn and McKenzie (2009), and show that the D_A-optimal method offers clear benefits relative to commonly used sequential alternatives. In fact, surprisingly, optimal sequential designs are comparably well-balanced to stratifications performed with full knowledge of covariates in advance. In spite of these practical advantages, the method had not, to our knowledge and according to three survey articles, ever been employed in the field.[2] We describe our experience implementing the method in a water treatment and hygiene intervention in Dhaka, Bangladesh (Guiteras et al., 2015), and offer practical advice on its implementation under field conditions. Implementation was feasible with standard software (Stata), and produced an allocation that was well-balanced both on the stratification variables chosen *ex ante* and, *ex post*, on other important variables that were not included in the stratification.

* Corresponding author.
E-mail address: guiteras@econ.umd.edu (R.P. Guiteras).

[1] Examples of sequential randomization in economics include Beaman and Magruder (2012), which randomized without stratification, and Bronchetti et al. (2013), which stratified using the block randomization method we describe in Section 5.1.2.

[2] See McEntegart (2003), Table 1 in Taves (2010), and Ciolino et al. (2011). Confirmed by personal communication with J. Cicolino, Northwestern University, January 17, 2014.

http://dx.doi.org/10.1016/j.deveng.2015.11.001

2. Theory

Our exposition follows Atkinson (2002), with some changes in notation. First, we lay out the model and notation. Second, we develop the theory for the traditional situation of a fixed population of N subjects, for whom covariates X have been collected in advance. Third, we introduce sequential designs using a simplified case where the researcher is concerned with the precision of all estimated parameters, both treatment effects and nuisance parameters (coefficients on stratification variables). Finally, we adapt the sequential design to the standard situation where only precisely estimated treatment effects are of interest.

2.1. Model and notation

Suppose the researcher is conducting an individual-level trial with J treatments, including the control treatment. We first consider a linear model with homogeneous treatment effects and i.i.d. errors. In Section 3, we discuss extensions, including heteroscedasticity, nonlinear models, and cluster designs. The model for unit i is

$$y_i = d_i'\alpha + x_i'\beta + \varepsilon_i = w_i'\theta + \varepsilon_i, \tag{1}$$

where d_i is a $J \times 1$ vector of indicator variables assigning unit i to a single treatment (i.e., exactly one element of d_i is equal to one), x_i is a $K \times 1$ vector of covariates, and ε_i is an error term. Without loss of generality, we order the treatments with the control condition first. Let $d_i(j)$ indicate assignment to the jth treatment; that is, $d_i(1) = (1\ 0 \cdots 0)'$, $d_i(2) = (0\ 1\ 0 \ldots 0)'$, etc. We are interested in estimating *contrasts* between the elements of α; that is, $\alpha_1 - \alpha_2$, $\alpha_1 - \alpha_3$, etc. The control group mean is a nuisance parameter,[3] as are the K elements of β (the coefficients on the covariates), so we have $K+1$ nuisance parameters and $J-1$ parameters of interest.[4]

2.2. Optimal designs with baseline covariates

First, consider a population of N subjects, for whom the researcher has obtained baseline covariates X prior to randomization. The population regression model is given by

$$E[Y] = D\alpha + X\beta = W\theta, \tag{2}$$

where D is the $N \times J$ matrix assigning all subjects to treatment (i.e., $D = (d_1 \cdots d_n)'$). X is the $N \times K$ matrix of covariates, and α and β are as before. Given the covariates X, our goal is to choose D to minimize the variance of our estimated treatment effect. As a simple example, with one treatment plus a control condition, $J=2$, we are interested in the *contrast* $\alpha_1 - \alpha_2$ and wish to minimize $V[\hat{\alpha}_1 - \hat{\alpha}_2]$.

A useful matrix to create contrasts is

$$L'_{(J-1)\times J} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & & -1 \end{bmatrix}.$$

Now we can create a vector of contrasts by premultiplying α by L':

$$L'\alpha = \begin{bmatrix} \alpha_1 - \alpha_2 \\ \vdots \\ \alpha_1 - \alpha_J \end{bmatrix}.$$

To annihilate the nuisance parameters, we augment L' with a $(J-1) \times K$ matrix of zeros, and define

$$A' = [L'\ \ 0].$$

The variance of $\hat{\alpha}$ is proportional to square root of the determinant of the generalized variance[5]:

$$\left|A'(W'W)^{-1}A\right| = \left|L'\left\{D'D - D'X(X'X)^{-1}X'D\right\}^{-1}L\right|. \tag{3}$$

This quantity is minimized when $D'X = 0$; that is, when the treatment assignment is *orthogonal* to the covariates, which is to say that the treatments are *balanced* across the covariates. When $D'X = 0$, the generalized variance simplifies, and the determinant is

$$\left|A'(W'W)^{-1}A\right| = \left|L'(D'D)^{-1}L\right| = J^J/N^{J-1}$$

This minimum possible value is the standard against any other treatment assignment D. Note that this value is increasing in J and decreasing in N, which matches our intuition that the variance will increase with the number of treatments and decrease with the number of observations.

The relative efficiency of a design D is the ratio of the determinant of the generalized variance to this minimal value:

$$\mathcal{E} = \left\{\frac{J^J/N^{J-1}}{A'(W'W)^{-1}A}\right\}^{1/(J-1)},$$

where $1/(J-1)$ is a scale factor. A smaller denominator $|A'(W'W)^{-1}A|$ leads to higher $\mathcal{E}$, implying a more efficient design. Note that $\mathcal{E} = 1$ for an exactly balanced design. A useful representation is the *loss*

$$\mathcal{L} = N(1 - \mathcal{E}),$$

which is expressed as the effective loss of observations relative to an optimal design. That is, a non-optimal design D with N units is as precise as an optimal design with $N(1-\mathcal{E})$ fewer units. For an exactly balanced design, $\mathcal{L} = 0$.

Although not the focus of this paper, this framework can be used for near-optimal randomization in cases where a researcher can collect baseline data prior to randomization. Specifically, create a large number S of random allocations $\{D^1, \ldots, D^s, \ldots, D^S\}$ and choose the allocation D^s with lowest associated loss.[6] Kasy (2013) considers a more general Bayesian framework, and provides a search algorithm to find an optimal allocation.[7]

2.3. Sequential D-optimality

To extend to sequential randomized trials, we first consider the simple case where all elements of $\theta = (\alpha', \beta')'$ are of interest. Our goal is to minimize the variance of $\hat{\theta}$. The variance of $\hat{\theta}$ is proportional to the inverse of the design matrix $(W'W)^{-1}$, so we want to minimize $(W'W)^{-1}$ or, equivalently, maximize $|W'W|$, which will give us a *D-optimum design*.

Suppose the first n units have been allocated, with the resulting

[3] We are not interested in α_1 *per se*, but a precise estimate $\hat{\alpha}_1$ is necessary to estimate contrasts precisely.

[4] A more familiar setup for economics readers would include an intercept term as a covariate, so $\alpha_{\sim 1}$ would have $J-1$ elements (corresponding to the $J-1$ treatment conditions excluding the control) and the augmented covariate vector $(1, x')'$ would have $K+1$ elements including the intercept. This turns out to be less convenient for some of the matrix algebra below.

[5] Recall that $W = [D\ X]$, so $W'W = \left[\begin{smallmatrix} D' \\ X' \end{smallmatrix}\right][D\ X] = \left[\begin{smallmatrix} D'D & D'X \\ X'D & X'X \end{smallmatrix}\right]$. Then use results on inverses of partitioned matrices and use the zero block of the matrix A to zero out several terms.

[6] To conduct randomization inference, rather than choose the allocation with minimum $\mathcal{L}$, the researcher can instead specify an acceptable maximum $\overline{\mathcal{L}}$, retain $R+1$ draws with loss less than $\overline{\mathcal{L}}$, select one of these $R+1$ at random, and retain the remaining R for randomization inference. Code is available from the authors on request.

[7] This optimal allocation is unique if any element of x is continuous, and may be unique (in finite samples) even for discrete x with a large number of treatments and covariate cells. See also Bertsimas et al. (2015).