Statistics and Probability Letters 119 (2016) 11–20

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Covariate adjustment in randomization-based causal inference for 2^K factorial designs

Jiannan Lu

Analysis and Experimentation, Microsoft Corporation, One Microsoft Way, Redmond, WA 98052, USA

ARTICLE INFO

Article history:
Received 6 February 2016
Received in revised form 12 July 2016
Accepted 12 July 2016
Available online 19 July 2016

Keywords:
Potential outcome
Variance reduction
Finite-population asymptotics

ABSTRACT

We develop finite-population asymptotic theory for covariate adjustment in randomization-based causal inference for 2^K factorial designs. In particular, we confirm that both the unadjusted and the covariate-adjusted estimators of the factorial effects are asymptotically unbiased and normal, and the latter is more precise than the former.

1. Introduction

Randomization is often considered the gold standard for causal inference (Rubin, 2008). A well-established methodology to conduct causal inference is the potential outcomes framework (Neyman, 1923; Rubin, 1974), which defines the causal effect of a binary treatment factor as the comparison between the potential outcomes under treatment and control. In the presence of multiple binary treatment factors, we can evaluate them simultaneously under the 2^K factorial design framework (Fisher, 1935; Yates, 1937). Several researchers (e.g., Kempthrone, 1952, 1955; Wilk and Kempthrone, 1956; Bailey, 1981, 1991; Dasgupta et al., 2015) advocated conducting randomization-based causal inference for 2^K factorial designs, which has several advantages over the widely-used regression-based inference. For example, randomization-based inference is applicable to the finite-population setting, and therefore may be more reasonable in practice (e.g., Miller, 2006; Lu et al., 2015). For more discussion on the comparison and reconciliation of randomization-based and regression-based inferences for 2^K factorial designs, see Lu (2016).

In randomization-based causal inference, covariate adjustment (Cochran, 1977) is a variance reduction technique widely used by researchers (e.g., Deng et al., 2013; Miratrix et al., 2013). In an illuminating paper, Lin (2013) demonstrated the advantages of performing covariate adjustment for randomized treatment–control studies (i.e., 2^1 factorial designs). However, to our best knowledge, for 2^K factorial designs which are of great importance from both theoretical and practical perspectives, similar discussions appear to be absent; it is unclear whether covariate adjustment is beneficial for 2^K factorial designs, and if so, how to quantify said benefit. In this paper we answer this question, by extending the discussions in Lin (2013) and illustrating the advantages of performing covariate adjustment in 2^K factorial designs. To be specific, we derive the closed-form expressions for the asymptotic precisions of the unadjusted and covariate-adjusted estimators, and thus accurately measure the precision gained by covariate adjustment.

The paper proceeds as follows. Section 2 reviews randomization-based inference for 2^K factorial designs. Section 3 introduces the covariate-adjusted estimator for 2^K factorial designs. Section 4 derives the asymptotic precisions of the unadjusted and covariate-adjusted estimators. Section 5 concludes and discusses possible future directions.

E-mail address: jiannl@microsoft.com.

http://dx.doi.org/10.1016/j.spl.2016.07.010

2. Randomization inference for 2^K factorial designs

In this section, we review the randomization-based inference framework for 2^K factorial designs (Dasgupta et al., 2015; Lu, 2016). For consistency we adopt the notations in Lu (2016).

2.1. 2^K factorial designs

2^K factorial designs consist of K distinct treatment factors, each of which has two levels coded as -1 and 1. To simplify future notations we let $J = 2^K$. To define 2^K factorial designs, we rely on a $J \times J$ orthogonal matrix $\boldsymbol{H} = (\boldsymbol{h}_0, \ldots, \boldsymbol{h}_{J-1})$, which is often referred to as the model matrix (Wu and Hamada, 2009). We construct the model matrix in the following recursive way (Espinosa et al., 2016; Lu, 2016):

1. Let $\boldsymbol{h}_0 = \boldsymbol{1}_J$;
2. For $k = 1, \ldots, K$, construct $\boldsymbol{h}_k$ by letting its first 2^{K-k} entries be -1, the next 2^{K-k} entries be 1, and repeating 2^{k-1} times;
3. If $K \geq 2$, order all subsets of $\{1, \ldots, K\}$ with at least two elements, first by cardinality and then lexicography. For $k = 1, \ldots, J-1-K$, let σ_k be the kth subset and $\boldsymbol{h}_{K+k} = \prod_{l \in \sigma_k} \boldsymbol{h}_l$, where "$\prod$" stands for entry-wise product.

The jth row of the sub-matrix $\tilde{\boldsymbol{H}} = (\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K)$ is the jth treatment combination $\boldsymbol{z}_j$. To further illustrate the construction of the model matrix, we adopt the example in Lu (2016).

Example 1. Let $K = 2$. By following the above recursive procedure, we obtain $\boldsymbol{h}_0 = \boldsymbol{1}$, $\boldsymbol{h}_1 = (-1, -1, 1, 1)'$, $\boldsymbol{h}_2 = (-1, 1, -1, 1)'$, and $\boldsymbol{h}_3 = (1, -1, -1, 1)'$. Consequently, for 2^2 factorial designs the model matrix is:

$$\boldsymbol{H} = \begin{array}{c} \begin{matrix} \boldsymbol{h}_0 & \boldsymbol{h}_1 & \boldsymbol{h}_2 & \boldsymbol{h}_3 \end{matrix} \\ \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{array}.$$

The four treatment combinations are $\boldsymbol{z}_1 = (-1, -1)$, $\boldsymbol{z}_2 = (-1, 1)$, $\boldsymbol{z}_3 = (1, -1)$ and $\boldsymbol{z}_4 = (1, 1)$.

2.2. Randomization-based inference

We allow $N \geq 2J$ experimental units in the design. To describe the randomization-based inference framework, we follow a three-step procedure.

First, under the Stable Unit Treatment Value Assumption (Rubin, 1980) that for $j = 1, \ldots, J$ there is only one version of the treatment combination $\boldsymbol{z}_j$, and no interference among the experimental units, let $Y_i(\boldsymbol{z}_j)$ be the potential outcome of unit i under treatment combination $\boldsymbol{z}_j$, and $\bar{Y}(\boldsymbol{z}_j) = N^{-1} \sum_{i=1}^{N} Y_i(\boldsymbol{z}_j)$ be the average potential outcome across all the experimental units. Let $\boldsymbol{Y}_i = \{Y_i(\boldsymbol{z}_1), \ldots, Y_i(\boldsymbol{z}_J)\}'$ and $\bar{\boldsymbol{Y}} = \{\bar{Y}(\boldsymbol{z}_1), \ldots, \bar{Y}(\boldsymbol{z}_J)\}'$.

Next, we randomly assign $n_j \geq 2$ units to treatment combination $\boldsymbol{z}_j$. Let

$$W_i(\boldsymbol{z}_j) = \begin{cases} 1, & \text{if unit } i \text{ is assigned treatment } \boldsymbol{z}_j, \\ 0, & \text{otherwise,} \end{cases}$$

and let $Y_i^{\text{obs}} = \sum_{j=1}^{J} W_i(\boldsymbol{z}_j) Y_i(\boldsymbol{z}_j)$ be the observed outcome for unit i, and therefore the average observed outcome across all experimental units that are assigned to treatment combination $\boldsymbol{z}_j$ is $\bar{Y}^{\text{obs}}(\boldsymbol{z}_j) = n_j^{-1} \sum_{i=1}^{N} W_i(\boldsymbol{z}_j) Y_i(\boldsymbol{z}_j)$. Furthermore, we let $\bar{\boldsymbol{Y}}^{\text{obs}} = \{\bar{Y}^{\text{obs}}(\boldsymbol{z}_1), \ldots, \bar{Y}^{\text{obs}}(\boldsymbol{z}_J)\}'$.

Finally, we define the factorial effects as

$$\tau(l) = \frac{1}{2^{K-1}} \boldsymbol{h}_l' \bar{\boldsymbol{Y}} \quad (l = 1, \ldots, J-1),$$

and their randomization-based estimators as

$$\hat{\tau}_{\text{rb}}(l) = \frac{1}{2^{K-1}} \boldsymbol{h}_l' \bar{\boldsymbol{Y}}^{\text{obs}} \quad (l = 1, \ldots, J-1). \tag{1}$$

Its randomness is solely from the treatment assignment $W_i(\boldsymbol{z}_j)$'s.

3. Covariate adjustment in 2^K factorial designs

The idea behind the randomization-based estimator is estimating the average potential outcome $\bar{Y}(\boldsymbol{z}_j)$ by its corresponding average observed outcome $\bar{Y}^{\text{obs}}(\boldsymbol{z}_j)$. However, as shown in Cochran (1977) and later mentioned in Lin (2013), utilizing the pre-treatment covariates can potentially improve the precision of $\bar{Y}^{\text{obs}}(\boldsymbol{z}_j)$, and consequently that