UPPSALA UNIVERSITET

Working Paper 2019:2

Department of Statistics

Asymptotic inference for optimal re-randomization designs

Mårten Schultzberg and Per Johansson

Working Paper 2019:2
April 2019; September 2019 (revision)
Department of Statistics
Uppsala University
Box 513
SE-751 20 UPPSALA
SWEDEN

Title: Asymptotic inference for optimal re-randomization designs
Author: Mårten Schultzberg and Per Johansson
E-mail: marten.schultzberg@statistik.uu.se

Asymptotic Inference for Optimal Rerandomization Designs

Mårten Schultzberg and Per Johansson

September 27, 2019

Abstract

Recently an experimental design strategy, called rerandomization, has been proposed as an alternative or complement to traditional blocked designs. The idea of rerandomization is to remove, from consideration, those allocations with large imbalances in observed covariates according to a balance criterion, and then randomize within the set of acceptable allocations. This paper clarifies the concept of an 'optimal' rerandomization design for inferences using the mean-difference sample estimator, $\widehat{SATE}$. Based on the Mahalanobis distance criterion for balancing the covariates, we show that standard asymptotic inference to the population, from which the units in the sample are randomly drawn, is possible using only the set of best, or 'optimal', allocations.

Keywords: Rerandomization, Optimal design, Asymptotic inference

1 Introduction

In randomized experiments, the treatment assignment is unconfounded (Rubin, 1978; Imbens and Rubin, 2015) or equivalently strongly ignorable (Rosenbaum and Rubin, 1983), which permits valid inference to a large collection of well-defined estimands. For this reason, among others, randomized experiments are seen as the gold standard for causal inference. However, in a single realized randomized allocation, the imbalance in both observed and unobserved covariates can be substantial. This can lead to poor precision and low efficiency in the inference. To reduce these potential imbalances, blocking (also called stratification) on observed covariates has been used in experimental design, especially with a few discrete covariates. Recently, the alternative strategy of rerandomization has been suggested, reducing the risk of randomly drawing a bad allocation even with a large set of essentially continuous covariates (Morgan and Rubin, 2012). The idea is to remove from considerations allocations with large imbalances in the observed covariates based on a covariate-balance criterion, and then randomize within the remaining set of allocations with better balanced covariates.

With the goal of estimating treatment effects within the observed sample, such as the sample average treatment effect (SATE), Morgan and Rubin (2012) suggested using the Mahalanobis distance between the covariate-mean vectors of assigned treated and controls as the covariate-balance criterion for removing unbalanced allocations. Kallus (2018) proposed algorithms for finding 'optimal' designs for the estimation of the population average treatment effect (PATE). The optimality is defined as minimizing the maximum conditional sample variance of the mean difference estimator, $\widehat{SATE}$, under a structural relationship between potential outcomes (under treatment and control) and the covariates in the population. However, Kallus (2018) does not derive the sampling distribution of the $\widehat{SATE}$ under this design, and instead proposes that inferences should be based on a bootstrap procedure.

This paper clarifies the concept of an 'optimal' design for *inferences* to the PATE based on the $\widehat{SATE}$. Under complete randomization, the asymptotic properties of this estimator are derived by taking expectations over all possible samples and all possible allocations of the sampled units to treatment and control in the experiment. The $\widehat{SATE}$ is unbiased for PATE, asymptotically normal with a known variance over random sampling and complete randomized treatment allocation. Using blocking only to restrict allocations, standard asymptotic inference still applies within each block making the properties of the $\widehat{SATE}$ straightforward to derive. However, removing allocations with large covariate imbalances affects the properties of this estimator in a more intricate way.

The discussion of 'optimal' inference to the PATE is restricted to the strategy suggested in Morgan and Rubin (2012) and studied further by Li et al. (2018). Here, the operating characteristics, i.e., the expectation, variance, and asymptotic distribution of the $\widehat{SATE}$ are known; the 'optimal' designs suggested by Kallus (2018) includes rerandomization based on the Mahalanobis distance as a special case.

There are two results that we report here: First, using the Mahalanobis-distance criterion when the cardinality of the set of allocations fulfilling the covariance-balance criterion (to be defined below) is close to its minimum, the large sample sampling distribution of the $\widehat{SATE}$ is normally distributed with known variance. Inference to SATE, however, may be degenerate in this case because the resulting distribution may have zero variance. Second, the large sample asymptotic distribution of $\widehat{SATE}$ under Mahalanobis-based rerandomization is usually well approximated by a normal distribution also when a larger 'near optimal' set is used, and the difference in efficiency compared to using the 'optimal' set is typically very small (Morgan and Rubin, 2012), which means that, using a slightly larger 'near optimal' set, admits non-degenerate inference to both SATE and PATE without substantially decreasing efficiency of estimation to PATE.

The next section discusses rerandomization using the Mahalanobis metric. Section 3 provides the main results concerning asymptotic inference and Section 4 concludes.

2 Rerandomization based on the Mahalanobis criterion

In line with Morgan and Rubin (2012), consider a trial with n units, with n_1 assigned to treatment and n_0 assigned to control. Let $W_i = 1$ or $W_i = 0$ if unit i is assigned treatment or control, respectively, and define $\mathbf{W} = (W_1, ..., W_n)'$. Furthermore let $\mathbf{X}$ be the $n \times K$ matrix of fixed covariates in the sample ($\mathbf{x}_i, i = 1, ..., n$) with covariance matrix $cov(\mathbf{X})$. In a balanced experiment, i.e., $n_1 = n_0 = n/2$, there are $\binom{n}{n_1} = n_A$ possible treatment allocation (assignment) vectors, thus $\mathbf{W}^j$, $j = 1, ..., n_A$ and $\mathbb{W} = (\mathbf{W}^1, ..., \mathbf{W}^{n_A})$ the complete set of allocations.

The Mahalanobis distance for allocation j is

$$M(\mathbf{W}^j, \mathbf{X}) = \frac{N}{4} \widehat{\tau}_X^{j\,\prime} cov(\mathbf{X})^{-1} \widehat{\tau}_X^j, \; j = 1, ..., n_A$$

where

$$\widehat{\tau}_X^j = \frac{1}{n_1} \sum_{i=1}^{n_1} W_i^j \mathbf{x}_i' - \frac{1}{n_0} \sum_{i=1}^{n_0} (1 - W_i^j) \mathbf{x}_i', = \overline{\mathbf{X}}_T^j - \overline{\mathbf{X}}_C^j, \; j = 1, ..., n_A.$$

Following Li et al. (2018), denote $\widehat{\tau}_X$ the estimator of the covariate mean difference vector over complete randomization. Morgan and Rubin (2012) suggested accepting the treatment assignment vector $\mathbf{W}^j$ only when

$$M(\mathbf{W}^j, \mathbf{X}) \leq a,$$

where a is a positive constant.

Due to the asymptotic normality of difference in means implied by the Central Limit Theorem, $M(\mathbf{W}^j, \mathbf{X})$ follows a χ^2_K distribution asymptotically. This means that a quantile of this distribution can be used as an allocation inclusion/exclusion criterion. Let $\mathrm{P}(\chi^2_K \leq a) = p_a$, then to randomize within the set of the 0.01% best balanced allocations implies setting a so that $p_a = 0.0001$.

The minimum number of allocations in the set of allocations with the smallest Mahalanobis distance (i.e. $M(\mathbf{W}^j, \mathbf{X}) \simeq 0$ or $p_a \simeq 0$) in an experiment with $n_1 = n_0$ is two.[1] Because the Mahalanobis distance of an allocation ($\mathbf{W}^j$) is always exactly the same as for $(1 - \mathbf{W}^j)$, i.e. a mirror allocation. This minimal set, containing only the allocations with the smallest Mahalanobis distance across all possible allocations, contains the optimal set in terms of covariate balance. Since, assuming at least one continuous covariate, there are always $\binom{N}{N/2}/2$ unique Mahalanobis values over the set of all $\binom{N}{N/2}$ allocations. This means that the large sample rerandomization criterion that gives the minimum set can be written as $a^{min} := a : p_a^{min} = \frac{1}{\frac{\binom{N}{N/2}}{2}} = \frac{2}{\binom{N}{N/2}}$. We refer to this inclusion criterion as the 'best allocation inclusion criteria' (BAIC) and the set of allocations fulfilling BAIC is denoted $\mathcal{M}_{BAIC}$. However, with ties in the minimum of the Mahalanobis distance $\mathcal{M}_{BAIC}$ is larger than two [2]. If $\mathcal{M}_{BAIC}$ is large enough, non-degenerate inference to the SATE is possible (Johansson et al., 2018). However, because the cardinality of $\mathcal{M}_{BAIC}$ is not restricted by design, BAIC does not generally allow for non-degenerate inference to SATE. If there are too few allowed allocations, probabilistic inference to SATE is not helpful (Johansson et al., 2018).

3 Asymptotic Theory

Consider an experiment where the n units are evenly divided into two groups: $n_1 = n_2 = n/2$. Let $Y_i(w)$ denote the potential outcome when unit i is exposed to w. The sample average treatment effect is defiend as

$$\tau = \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - Y_i(0).$$

With random sampling from a fixed population of size N or from a superpopulation, τ is also the population average treatement effect (PATE). Define the sample means $\overline{Y(w)} = \frac{1}{n}\sum_{i=1}^{n} Y_i(w)$, $w = 0, 1$, then the sample

[1] In an experiment with $n_1 \neq n_0$ the minimum set could consist of just one allocation. See Morgan and Rubin (2012, p. 9) for an example with $n = 3$ and $n_1 \equiv 2$.

[2] This could happen with discrete data.

variance and the variance of the treatment effect are equal to

$$S^2_{Y(w)} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(w) - \overline{Y(w)})^2, w = 0, 1,$$

and

$$\begin{aligned} S_{Y(1)Y(0)} &= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i(1) - Y_i(0) - (\overline{Y(1)} - \overline{Y(0)})^2 \\ &= \frac{1}{n-1}\sum_{i=1}^{n}(\tau_i - \tau)^2 = S^2_\tau \end{aligned}$$

Define the sample means of treate and controls

$$\overline{Y}_1 = \frac{1}{n_1}\sum_{i:W_i=1}^{n_1} Y_i \text{ and } \overline{Y}_0 = \frac{1}{n_0}\sum_{i:W_i=0}^{n_0} Y_i.$$

Under **SUTVA** (no interference between individuals and the same treatment, Rubin 1980) the observed Y_i is equal to $Y(W_i)$, then the SATE estimator

$$\widehat{SATE} = \overline{Y}_1 - \overline{Y}_0,$$

is an unbiased estimator of τ (Neyman, 1923).

In Li and Ding (2017), the authors give a complete and comprehensive review of the conditions for CLT based inference in the causal setting, including new less stringent conditions for several settings. Using CLTs where the n units in the sample is embedded into an infinite sequence of finite populations with increasing sizes the authors show that under complete randomization

$$P\lim \sqrt{n}(\widehat{SATE} - \tau) \xrightarrow{d} N(0, V_\tau),$$

where

$$V_\tau = \frac{n}{n_1}S^2_{Y(1)} + \frac{n}{n_0}S^2_{Y(0)} - S^2_\tau.$$

Under the superpopulation assumption and inference to the PATE the third term vanish (as treated and controls are sampled independently).

Using the same conditions as in Li and Ding (2017), Li et al. (2018) showed that the asymptotic distribution of $\widehat{SATE}$ after randomly choosing an allocation from the set $\mathcal{A}_a = \{\mathbb{W}|M(\mathbf{W}^j, \mathbf{X}) \leq a\}$ is

$$P\lim \sqrt{n}(\widehat{SATE} - \tau)|\hat{\tau}_X \xrightarrow{d} \sqrt{V_{\tau\tau}}Q, \tag{1}$$

where $Q = \sqrt{(1-R^2)}\varepsilon_0 + \sqrt{R^2}L_{K,a}$; ε_0 is a standard normal variable (for Y in the space orthogonal to the covariates), $L_{K,a}$ is the projection of Y into the space of covariates and is thus affected by the rerandomization, and

$$R^2 = \frac{\frac{n}{n_1}s^2_{Y(1)|\mathbf{X}} + \frac{n}{n_0}s^2_{Y(0)|\mathbf{X}} - \mathbf{s}^2_{\tau|\mathbf{X}}}{V_{\tau\tau}},$$

where $S^2_{Y(w)|\mathbf{X}}$ and $S^2_{\tau|\mathbf{X}}$ are the variances of the linear projection of $\mathbf{Y}(w)$ and τ on $\mathbf{X}$, respectively. Under homogeneous treatment effects, i.e., $S^2_\tau = 0$ and $S^2_{\tau|\mathbf{X}} = 0$, it follows that $R^2 = s^2_{Y(0)|\mathbf{X}}/S^2_{Y(0)}$, that is, the R^2 in a regression of $\mathbf{Y}(0)$ on $\mathbf{X}$. Li et al. (2018) show that under this condition, R^2 can be consistently estimated using the linear projection of the outcomes of the treated and control units values on $\mathbf{X}$. The second part of Q has the following distribution

$$L_{K,a} \sim \chi_{K,a} S \sqrt{\beta_K},$$

where $\chi_{K,a} = \chi^2_K | \chi^2_K \leq a$, S a random variable taking values ± 1 with probability 1/2, and $\beta_K \sim \beta(1/2, (K-1)/2)$ degenerating to a point mass at 1 when $K = 1$.

3.1 Asymptotic theory for the optimal design

We now investigate the properties of Q as a approaches 0, i.e., when the set of included allocations is restricted more and more such that the cardinality of the set of allowed allocations goes towards its minimum.

From the assumption that the n units are randomly sampled from the population, the asymptotic results when $n \to \infty$ in Li et al. (2018) can be used to derive the asymptotic distribution in the situation when a approaches 0. In this situation, it is only the second term of Equation 1, $\sqrt{R^2}L_{K,a}$, that depends on a, where $\sqrt{R^2}$ is a constant that can be estimated consistently. $L_{K,a}$ is symmetric and unimodal around zero, with variance

$$\mathrm{Var}(L_{K,a}) = P(\chi^2_{K+2,a} < a)/P(\chi^2_{K,a} < a). \tag{2}$$

For fixed K, $\mathrm{Var}(L_{K,a}) \to 0$ as $a \to 0$ (Li et al., 2018). That is, the $L_{K,a}$ distribution will converge to point mass at zero when a, or equivalently p_a, goes to zero. This implies that for large n using the minimum criterion a^{min},

$$P\lim \sqrt{n}(\widehat{SATE} - \tau)|\widehat{\tau}_X \xrightarrow{d} \sqrt{V_\tau}\left(\sqrt{(1-R^2)}\varepsilon_0\right). \tag{3}$$

The intuition of the result is that randomizing the treatment assignment within the set of allocations containing only the very best allocations will, in large samples, result in a realized treatment allocation with a Mahalanobis distance close to zero, which means that essentially all variation in $\widehat{SATE}$ that is explained by group differences in $\mathbf{X}$ is removed, that is, all the variance in the linear projection of $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ on $\mathbf{X}$ is eliminated.

Interestingly, because the allocations in $\mathcal{M}_{BAIC}$ are allocation from a random sample of units from the population, the $\widehat{SATE}$ has has stochastic variation across repeated sampling, i.e., a sampling distribution[3], and therefore non-degenerate inference to PATE is usually possible also using BAIC.

Under Theorem 1 in (Li et al., 2018, pp. 8) and using Equation 3, the test statistic is thus

$$\frac{\sqrt{n}(\widehat{SATE}-\tau)|\sqrt{n}\hat{\tau}_x \in \mathcal{M}_{BAIC}}{\sqrt{V_\tau}\sqrt{1-R^2}} \sim N(0,1), \tag{4}$$

and all the well established asymptotic results for the standard normal distribution apply. The sampling distribution is well defined for $R^2 < 1$. In the unlikely situation that $R^2 = 1$, i.e., that the variation in $\mathbf{X}$ explains *all* variation in both $\mathbf{Y}(0)$ and $\mathbf{Y}(1)$ and *all* variation in $\mathbf{X}$ is removed by rerandomization, the samplings distribution is degenerated to a point mass at zero.

To summarize, Table 1 displays the sampling distribution of the $\widehat{SATE}$ under Mahalanobis-based rerandomization for different regions of p_a. Under complete randomization ($p_a = 1$), the distribution is normal, for p_a in the interval $(0, 1)$, the sampling distribution is Q, and for BAIC ($p_a \to 0$) the sampling distribution is again normal but with a scaled variance. These results have the important implication that the large sample

Table 1: The sampling distribution of the $\widehat{SATE}$ under under Mahalanobis-based rerandomization for different values of p_a.

	Complete rand.	Rerendomization	$\text{Rerandomization}_{\text{BAIC}}$
Reran. crit	$p_a = 1$	$0 < p_a < 1$	$p_a \to 0$
$Var(\sqrt{n}(\hat{\tau}-\tau))$	$V_{\tau\tau}$	$V_{\tau\tau}(1-(1-\nu_{K,a})R^2)$	$V_{\tau\tau}(1-R^2)$
$\sqrt{n}(\hat{\tau}-\tau) \sim$	$N(0, V_{\tau\tau})$	$Q(V_{\tau\tau}, R^2, K, a)$	$N(0, V_{\tau\tau}(1-R^2))$

sampling distribution under Mahalanobis-based rerandomization with sufficiently small p_a is simply normal

[3]With one exception described below.

with mean zero but with variance $V_\tau(1-R^2)$ instead of V_τ as under complete randomization.

3.2 Choosing p_a

This section presents a suggestion for choosing p_a in the design phase such that the (scaled) standard inferences can be used in the analysis.

The variance of Q is equal to

$$(1-R^2)Var(\epsilon_0)+R^2Var(L_{K,a})=(1-R^2)+R^2Var(L_{K,a})$$

where the equality comes from $\epsilon_0 \sim N(0,1)$. Because the asymptotic variance of $L_{K,a}$ is known (cf. Equation 2), this allows us to calculate the relative importance of the second term for a specific R^2. The variance ratio (VR) of the second term to the overall variance of the estimator under Mahalanobis based rerandomization given R^2 and a, equals

$$VR=\frac{R^2Var(L_{K,a})}{R^2Var(L_{K,a})+(1-R^2)}=\frac{P(\chi^2_{K+2,a}<a)/P(\chi^2_{K,a}<a)}{P(\chi^2_{K+2,a}<a)/P(\chi^2_{K,a}<a)+(\frac{1}{R^2}-1)}. \tag{5}$$

To illustrate, consider a setting with $R^2=0.5$, four covariates, and a small inclusion criterion, e.g. $a=10^{-4}$ which corresponds to $p_a\approx 1.25\times10^{-9}$. It follows that

$$\begin{aligned}VR&=\frac{Var(L_{4,10^{-4}})}{Var(L_{4,10^{-4}})+(1/0.5-1)}\\&=\frac{\frac{P(\chi^2_6\leq10^{-4})}{P(\chi^2_4\leq10^{-4})}}{\frac{P(\chi^2_6\leq10^{-4}))}{P(\chi^2_4\leq10^{-4})}+1}\\&\approx1.66\times10^{-5},\end{aligned} \tag{6}$$

which means that around 2 1000th of a percent of the variance of Q arises from the variance of the second term. Clearly, including or excluding this term has no practical importance for inferences in this case. For a large sample, e.g, $n=100$, there would still be around 1.26×10^{20} ($=1.25\times10^{-9}\times\binom{100}{50}$) allocations fulfilling this rerandomization criterion. Thus choosing a to be 10^{-4} would also enable inference to the units of the sample while simultaneously allowing standard methods to make inference to the units of the population.

In real experiments, R^2 is not known in the design phase, which complicates this type of analysis. However, the VR can be calculated for various hypothetical values of R^2 that are larger than the expected empirical R^2 to create an upper bound for the VR. Figure 1 displays the VR as a function of a for

$n = 100, K = 5$, and $R^2 = (0.05, 0.1, ..., 0.90, 0.95)$. It is clear that, in this case, even if R^2 is large and the sample size is only moderately large, there are still enough allocations fulfilling the criterion. For example, if a is set to 10^{-5}, there are still 1.69×10^{15} allocations fulfilling the criterion. That this small sample size allows for the simplification even for very large R^2 illustrates that these results often can be applied in typical experimental settings.

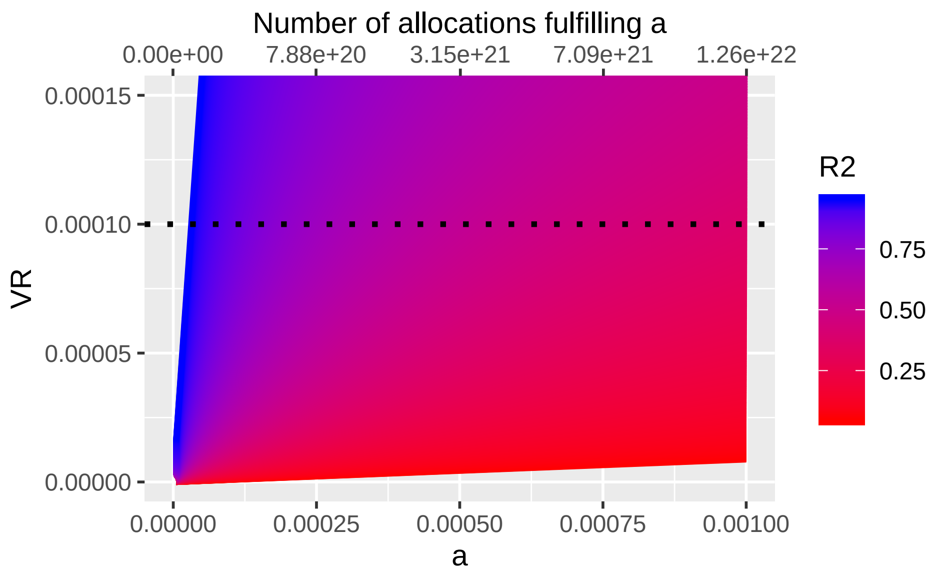

Figure 1: The variance ratio as a function of the Mahalanobis inclusion criterion a, for $K = 4$, $n = 100$ and different values of R^2. The values on the top of the abscissa are the number of allocations satisfying the criterion and the dotted line indicates the VR which implies that variance in the sampling distribution due to the second term in Equation 1 is smaller than 0.01%.

Based on a limited set of simulations, there seem to be no problem with inference, using the normal approximation, if the ratio of the variance (VR) in the sampling distribution due to the second term in Equation 1 is smaller than 0.01%. In the example above, using $VR < 0.01\%$ would allow for the (scaled) standard asymptotic inferences for all $R^2 < 0.95$. These simple calculations can be performed for any experiment to quickly evaluate the possibilities of using the simplified results presented in this paper, or, to select a such that, for the largest plausible R^2, the VR is smaller than the rule of thumb.

4 Discussion

The asymptotic sampling distribution for the difference in mean estimator, $\widehat{SATE}$, for inferences to the PATE under Mahalanobis-based rerandomization design is investigated, specifically, how the sampling distribution is affected by letting the Mahalanobis-based rerandomization inclusion criterion approach the 'optimal' design.

The $\widehat{SATE}$ under complete randomization is asymptotically normal. However, removing allocations associated with large covariates-differences between treated and controls, as with 'rerandomization designs', affects the properties of this estimator in ways that depend on the balance criterion used to discard unbalanced allocations. Morgan and Rubin (2012) and Li et al. (2018) used the well known properties of the affinely invariant Mahalanobis distance to derive properties of the $\widehat{SATE}$ after randomizing in the set of allocations with a Mahalanobis distance smaller than a specified inclusion criterion.

The $\widehat{SATE}$ has a non-degenerate sampling distribution for repeated sampling inference to the population. However, deterministically choosing an 'optimal' design can lead to a degenerate sampling distribution for inference to the units in the experiment. Based on the results in Li et al. (2018), we show that the asymptotic sampling distribution under Mahalanobis-based rerandomization simplifies to a normal distribution when the inclusion criterion is small. When the sample size is moderately large, there will be large number of allocations that fulfills even very restrictive inclusion criteria, which enables inference to the PATE or the sample estimand (SATE) using standard results (Johansson et al., 2018).

For these reasons, when using Mahalanobis-based rerandomization, it can be advisable to set the inclusion criterion for admissible allocations slightly smaller than suggested in Li et al. (2018) so that standard asymptotic inference can be used for PATE, and slightly larger than the minimum so that inference is possible to both SATE and PATE. To this end we suggest a simple-to-use rule of thumb for when the simplified asymptotics can be used that makes it possible to choose the Mahalanobis criterion accordingly.

5 Acknowledgements

The authors are grateful to Donald B. Rubin for feedback on previous versions of this paper

References

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences.* Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.

Johansson, P., Rubin, D. B., and Schultzberg, M. (2018). Design of Randomized Controlled Trials: Inference and Optimality.

Kallus, N. (2018). Optimal a priori balance in the design of controlled experiments. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 80(1):85–112.

Li, X. and Ding, P. (2017). General Forms of Finite Population Central Limit Theorems with Applications to Causal Inference. *Journal of the American Statistical Association*, 112(520):1759–1769.

Li, X., Ding, P., and Rubin, D. B. (2018). Asymptotic theory of rerandomization in treatment-control experiments. *Proceedings of the National Academy of Sciences*, 115(37):9157 – 9162.

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *Annals of Statistics*, 40(2):1263–1282.

Neyman, J. (1923). On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. *Annals of Agricultural Sciences.*

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rubin, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6(1):34–58.

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. *Journal of the American Statistical Association*, 75(371):591–593.