Re-randomization in Clinical Trials

Tuo Wang (rotation with Prof. Menggang Yu)

1. Introduction

In many clinical trials, the ultimate goal is to estimate the treatment effect. Balanced randomization, which eliminate the effect of confounders, is the key to get correct causal effect. In a clinical trial with a large sample size, balanced randomization can be easily achieved. However, in a clinical trial with a smaller sample size, such as trials that need cluster randomization, a randomization can make treatment group and control group with significantly unbalanced covariates. In practice, one can only include a small amount of stratification in cluster randomization, leaving many other equally or more important covariates stay unbalanced. However, we can also rerandomize and check the balance of covariates until a certain criterion satisfied. Morgan and Rubin (2012) proposed a general procedure for rerandomization. First, collect the covariate data of all the participants and specify a balance criterion determining when a randomization is acceptable. Then, randomize the subjects and check the balance criterion. If the criterion is satisfied then initiate the trial with this randomization. Otherwise, randomize the subjects again. However, because of rerandomization, normal statistical inference procedures don't work. Instead, one can analyze the data by using a randomization test. The most important part of this general procedure is to decide the balance criterion. Morgan and Rubin (2012) proposed a criterion using Mahalanobis distance. Xu and Kalbfleisch (2010) proposed a rerandomization design, the balance match weighted (BMW) design, using propensity score matching. In section 2, we will breifly introduce these two methods and we propose an revised version of BMW design by substituting the propensity score matching with multivariate caliper matching using the Mahalanobis distance. In section 3, we conducted simulations to compare the BMW design with the revised version BMW design.

2. Methods

2.1 Rerandomization using Mahalanobis distance. Let $\boldsymbol{X}$ be the $n \times p$ covariate matrix representing n observations with p covariates. Let $\widehat{\Sigma}$ be the sample covariance matrix of $\boldsymbol{X}$. Then the estimated Mahalanobis distance between subject i and j, who has covariates $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ respectively, is defined as:

$$d(\boldsymbol{X}_i, \boldsymbol{X}_j) = (\boldsymbol{X}_i - \boldsymbol{X}_j)^T \widehat{\Sigma}^{-1} (\boldsymbol{X}_i - \boldsymbol{X}_j)$$

Let $\boldsymbol{A} = (A_1 \ldots, A_n)^T$ be a n-dimensional treatment assignment vector. $A_i = 0$ if subject i is in the control group and $A_i = 1$ if subject i in the treatment group. Assume the proportion of treatment and control groups are fixed before randomization. Let the proportion of treatment group be $p_T = \frac{1}{n}\sum_{i=1}^{n} A_i$. Let $\bar{\boldsymbol{X}}_T$ and $\bar{\boldsymbol{X}}_C$ be the covariate means in treatment and control groups respectively. The Mahalanobis distance can be calculated by:

$$\begin{aligned} M &= (\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_C)^T [\mathrm{cov}(\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_C)]^{-1} (\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_C) \\ &= n p_T (1 - p_T) (\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_C)^T \widehat{\Sigma}^{-1} (\bar{\boldsymbol{X}}_T - \bar{\boldsymbol{X}}_C) \end{aligned}$$

A randomization is considered as acceptable when M is smaller than a certain number m. m can be pre-specified by first specifying the proportion of acceptable randomizations. If $M \geq m$, then we can rerandomize the data and calculate M again with the new randomized groups.

2.2 Rerandomization using propensity score matching (PCM). In an observational study, treated and control subjects may differ in terms of covariates, so direct comparison of the outcomes may be biased. The propensity score matching can be used to control for confounders (Rosenbaum and Rubin, 1984). Following the notations in section 2.1, the propensity score is defined as $e(\boldsymbol{X}_i) = \Pr(A_i = 1|\boldsymbol{X}_i)$. Subjects with similar propensity score are paired together. In practice, the true propensity score is always unknown but can be estimated using logistic regression. Xu and Kalbfleisch (2010) proposed rerandomization design called the balance match weighted (BMW) design, which using optimal full matching on propensity score. The BMW design has following steps:

1. Pre-specified the ratio of controls to treatments that is to be permitted within a matched set. Specifically, let the ratio stays between $\frac{1}{k}$ and k;
2. Randomize half of the subjects to treatment group and the other to control group to get sets T and C;
3. Fit logistic regression and compute the estimated propensity scores and create $|T| \times |C|$ matrix of estimated propensity score distances;
4. Apply full matching with constrain k and record the total distance Δ_k;
5. Repeat steps 2 and 4 M times and choose the randomized sample with the minimum total distance.

Obviously, the large M is the smaller the total distance we can get. However, large M will increase the complexity in randomization test for calculating statistical inference. Based on the simulation in Xu and Kalbfleisch (2010), the authors recommended using $M \in [10, 20]$ and $k = 2$. The optimal full matching can be conducted uing `Optmatch` package in R developed by Hansen (2004).

2.2 Rerandomization using multivariate caliper matching (MCM). There are two major drawbacks about propensity score matching. First, two individuals with the same propensity score may differ in important ways. Second, propensity score matching is a single covariate matching, which ignore the interaction between different covariates. This motivates us to use multivariate caliper matching in the BMW design. The multivariate caliper matching is quite similar to the propensity score matching except two aspects. First, instead of using propensity score as distance, it use Mahalanobis distance. Second it penalize large distance. The idea is that two individuals can be close on the propensity score to a degree, once this degree is achieved, covariates $\boldsymbol{X}$ may affect the distance. Define ω as the caliper width. With ω, if two individuals have propensity scores that differ more tahn ω, we will add a penalty to the Mahalanobis distance between subject i and subject j. Explicitly,

$$d_{new}(\mathbf{x_i}, \mathbf{x_j}) = \begin{cases} d(\mathbf{x_i}, \mathbf{x_j}) + p \times \left|\text{logit}(e(\mathbf{x_i})) - \text{logit}(e(\mathbf{x_j}))\right|, & \text{if } \left|\text{logit}(e(\mathbf{x_i})) - \text{logit}(e(\mathbf{x_j}))\right| \geq w \\ d(\mathbf{x_i}, \mathbf{x_j}), & \text{if } \left|\text{logit}(e(\mathbf{x_i})) - \text{logit}(e(\mathbf{x_j}))\right| < w \end{cases}$$

In practice people use $p = 1000$ and $\omega = 0.5 \times \text{sd}(e(\boldsymbol{X}))$. Therefore the BMW design with multivariate caliper matching has the following steps:

1. Pre-specified the number k defined in section 2.1 and pre-specified p and ω
2. Randomize half of the subjects to treatment group and the other to control group to get sets T and C;
3. Fit logistic regression and compute the estimated propensity scores and create $|T| \times |C|$ matrix of Mahalanobis distances with caliper;
4. Apply full matching with constrain k and record the total distance Δ_k;
5. Repeat steps 2 and 4 M times and choose the randomized sample with the minimum total distance.

3. Simulation Studies

We used a similar simulation setting as described in Xu and Kalbfleisch (2010). Suppose we have n subjects, for each subject , we generate p covariates $X_1, \cdots, X_p$. Let A_i be the treatment assignment for subject i. The response was calculated from

$$Y_i = \beta A_i + \sum_{j=1}^{p} \gamma_j X_{ij} + \epsilon_i,$$

where the noise $\epsilon_i, i = 1, \ldots, n$ are i.i.d. samples from $\mathcal{N}(0, 1)$. In the simulations we considered the following set-up:

- consider sample size $n = 30$ and number of covariates $p = 8$
- True treatment effect $\beta = 0.7$
- consider two different set-up of the true confounding effects. (1) $\gamma_j = 1, j = 1, \ldots, 8$. (2)$\gamma_1 = \gamma_2 = 10$ $\gamma_j = 1, j = 3, \ldots, 8$
- consider three distributions of the covariates.

  (1). $X_1, X_2, \ldots, X_8 \overset{i.i.d}{\sim}$ Bernoulli(0.5),

  (2). $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right)$, $X_3, X_4, X_5, X_6, X_7, X_8 \overset{i.i.d}{\sim}$ Bernoulli(0.5),

  (3). $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 & 1.5 \\ 1.5 & 3 \end{pmatrix} \right)$, $X_3, X_4, X_5, X_6, X_7, X_8 \overset{i.i.d}{\sim}$ Bernoulli(0.5)

For this design, the treatment effect estimator based on the unstratified pooled sample is $\hat{\beta}_{pool} = \bar{Y}_T - \bar{Y}_C$ and the MSE is:

$$MSE(\hat{\beta}_{pool}|T, C, X) = \left\{ \sum_{j=1}^{r} \gamma_j (\bar{X}_{jT} - \bar{X}_{jT}) \right\}^2 + \frac{2\sigma^2}{n}$$

We compare the MSE of the treatment effect estimator for the BMW design with the multivariate caliper matching with the completely randomized design (CR) and the BMW design with the propensity score matching. We simulated 500 replicates for each scenario and computed the mean of the MSE and mean MSE percent reduction. The results are summarized in table 1.

Table 1: Simulation results. The numbers in the parentheses are the reduction of MSE.

γ	M	MSE (CR)	MSE (BMW1)	MSE (BMW2)
$X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8 \overset{i.i.d}{\sim} Bernoulli(0.5)$				
(1.0,1.0,1.0,1.0, 1.0,1.0,1.0,1.0)	10	0.335	0.179 (45.46%)	0.188 (43.83%)
(10.0,10.0,1.0,1.0, 1.0,1.0,1.0,1.0)	10	6.54	3.06 (53.16%)	3.50 (46.55%)
$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right)$, $X_3, X_4, X_5, X_6, X_7, X_8 \overset{i.i.d}{\sim} Bernoulli(0.5)$				
(1.0,1.0,1.0,1.0, 1.0,1.0,1.0,1.0)	10	0.739	0.344 (53.45%)	0.331 (55.17%)
(10.0,10.0,1.0,1.0, 1.0,1.0,1.0,1.0)	10	44.204	20.17 (54.36%)	20.28 (54.12%)
$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim MVN\left(\begin{pmatrix} 5 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 & 1.5 \\ 1.5 & 3 \end{pmatrix} \right)$, $X_3, X_4, X_5, X_6, X_7, X_8 \overset{i.i.d}{\sim} Bernoulli(0.5)$				
(1.0,1.0,1.0,1.0, 1.0,1.0,1.0,1.0)	10	1.52	0.81 (46.39%)	0.77 (48.82%)
(10.0,10.0,1.0,1.0, 1.0,1.0,1.0,1.0)	10	128.02	54.02 (57.8%)	52.01 (59.3%)

CR: Completely randomized design, BMW1: BMW design with the propensity score matching, BMW2: BMW design with multivariate caliper matching.

4. Discussion

Simulation shows that when there is no correlation between covariates, BMW design with pcm has lower MSE than with mcm. When there is correlation between covariates, BCM design with mcm has lower MSE than pcm, though the difference is not prominent. In general this BMW design propose by Xu and Kalbfleisch (2010) provide a useful rerandomization procedure for small trials.

For future work. First, we should explore more simulation setting with the covariates generated in a more correlated and complicated way and apply this procedure to real data. Second, this BMW design, no matter with pcm or mcm, ignore the tiers of the covariates and treat every covariate equally important. However, in practice, we may know some information about the covariates before the a trial start. For example, in a trail with 10 observed covariates, suppose we know the first 5 covariates are more important than the last 5 covariates. Then, ideally, we would want to balance the first 5 important covariates first. Morgan and Rubin (2017) proposed a general framework of rerandomization to balance tiers of covariates, but there are not many publications about rerandomization using matching procedure on balancing tiers of covariates. Matching procedure always reduce high-dimensional data into a scalar such as propoensity score and Mahalanobis distance. Propensity score is a special case of balancing score (Rosenbaum and Rubin, 1984) and balancing score can be any dimension. Can we first reduce the dimension of the covariates, get a balancing score, which has at least two dimensions, and match the subjects based on balancing score. In this case we can include tiers of covariates easily in the rerandomization procedure.

References

[1] Hansen, B. B. (2004). Full matching in an observational study of coach- ing for the SAT. *Journal of the American Statistical Association* 99, 609–618.

[2] Morgan, K. L., and Rubin, D. B. (2012). Rerandomization to Improve Co- variate Balance in Experiments. *The Annals of Statistics*, 40, 1263–1282.

[3] Morgan, K. L., and Rubin, D. B. (2016). Rerandomization to Balance Tiers of Covariates. *Journal of the American Statistical Association*, 110, 1412–1421.

[4] Rosenbaum, P. R. and Rubin, D. B. (1984). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.

[5] XU, Z. and KALBFLEISCH, J. D. (2010). Propensity score matching in randomized clinical trials. *Biometrics* 66 813–823.