Chapter 4

Donald B. Rubin

Yau Mathematical Sciences Center, Tsinghua University

September 2, 2021

Introduction

Four specific examples of classical randomized assignment mechanisms are discussed

1. Bernoulli trials,
2. completely randomized experiments,
3. stratified randomized experiments (randomized blocks),
4. paired randomized experiments,

which all satisfy the four criteria necessary for assignment mechanisms to be classified as classical randomized experiments.

That is, the assignment mechanism: (*i*) is *individualistic*; (*ii*) is *probabilistic*, (*iii*) is *unconfounded* and (*iv*) has a known functional form that is controlled by the researcher.

Introduction

The key difference between the four types is in the set of assignment vectors $\mathbf{W}$ (the N–dimensional vector with elements $W_i \in \{0, 1\}$) with positive probability.

Let the set of all possible values be denoted by $\mathbb{W} = \{0, 1\}^N$, with cardinality 2^N, and let the subset of values for $\mathbf{W}$ with positive probability be denoted by $\mathbb{W}^+$.

In the first example of randomized experiments, Bernoulli trials, each of the 2^N possible vectors $\mathbf{W}$ defining the treatment assignments of the full population of size N has positive probability (such trials put positive probability on assignments even in which all units receive the same treatment).

The remaining three types of classical randomized experiments impose increasingly restrictive sets of conditions on the set $\mathbb{W}^+$ of values of $\mathbf{W}$ with positive probability. If imposed judiciously, these restrictions can lead to more precise inferences.

Notation

From chapter 3 we know that the assignment mechanism in a classical randomized experiment can be written as

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = c \cdot \prod_{i=1}^{N} e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i},$$

for $\mathbf{W} \in \mathbb{W}^+$, and zero elsewhere.

Here $\mathbb{W}^+ \subset \mathbb{W}$ is the subset of the set of possible values for $\mathbf{W}$ with positive probability. The constant c ensures that the probabilities add to unity:

$$c = \left(\sum_{\mathbf{W} \in \mathbb{W}^+} \prod_{i=1}^{N} e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i} \right)^{-1}.$$

Bernoulli Trials

The simplest Bernoulli experiment tosses a fair coin for each unit.

Because the coin is fair, the unit level probabilities and the propensity scores are all 0.5. Because the tosses are independent we get

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = 0.5^N, \tag{1}$$

for all vectors $\mathbf{W}$. Here $\mathbb{W}^+ = \{0,1\}^N = \mathbb{W}$.

Slightly more generally with the propensity score $q \in (0,1)$, equation (1) becomes

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = q^{N_t} \cdot (1-q)^{N_c}, \tag{2}$$

Here, the probabilities of the different $\mathbf{W}$ vectors depend solely on the number of treated and control units, but still $\mathbb{W}^+ = \{0,1\}^N$.

Letting $q \in (0.5, 1)$, may be attractive when trying to induce people with a serious disease to enroll in a blinded placebo controlled experiment of a promising new drug.

Bernoulli Trials

A generalization of Bernoulli trials allows the unit probabilities to vary with the unit's covariate values. This situation can occur, for example, when certain types of patients are thought to do better on one treatment than another (based e.g., age, sex, race).

Thus, each unit has a special coin tossed, with probability, $e(X_i)$ to be treated; consequently,

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i} \right]. \tag{3}$$

Here again $\mathbb{W}^+ = \mathbb{W}$.

Our formal definition of a Bernoulli trial requires that assignments to treatment are independent across all units in the population:

Bernoulli Trials

Definition

(BERNOULLI TRIALS)
A Bernoulli trial is a classical randomized experiment with an assignment mechanism such that the assignments for all units are independent.

Bernoulli Trials

PROOF: If assignment to treatment is independent across all observations in the population, then the probability of observing a specific assignment vector $\mathbf{W}$, $\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))$, will simply equal the product of each unit's probability of assignment:

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1))^{W_i} \cdot (1 - p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)))^{1-W_i} \right].$$

Combined with the fact that $p_i(\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = e(X_i)$ for all i, implied by the fact that a Bernoulli trial is a classical randomized experiment, it follows that the normalizing constant is $c = 1$, and that the general form of the assignment mechanism for this type of randomized experiment is

$$\Pr(\mathbf{W}|\mathbf{X}, \mathbf{Y}(0), \mathbf{Y}(1)) = \prod_{i=1}^{N} \left[e(X_i)^{W_i} \cdot (1 - e(X_i))^{1-W_i} \right],$$

as in equation (3). $\square$

Completely Randomized Experiments

The simplest completely randomized experiment takes an even number of units and partitions them at random in two groups (e.g. by putting labels on the N unit in an urn), with exactly one half of the sample receiving the active treatment and the remaining units receiving the control treatment.

The assignment mechanism is;

$$\Pr(\mathbf{W}|\mathbf{X},\mathbf{Y}(0),\mathbf{Y}(1)) = \begin{cases} \begin{pmatrix} N \\ N_t \end{pmatrix}^{-1} & \text{if } \sum_{i=1}^{N} W_i = N_t, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

In this design, the propensity scores are equal for all units, namely N_t/N. The only requirement is that N_t is fixed in advance and $0 < N_t < N$.

Bernoulli Trials

Definition

(Completely Randomized Experiment)
A completely randomized experiment is a classical randomized experiment with an assignment mechanism satisfying

$$\mathbb{W}^{+} = \left\{ \mathbf{W} \in \mathbb{W} \middle| \sum_{i=1}^{N} W_i = N_t \right\},$$

for some preset positive $N_t \in \{1, 2, \ldots, N-1\}$.

The number of possible assignment vectors, $\mathbf{W}$, is under this design $A = \binom{N}{N_t}$.

As all assignment vectors in this set are equally likely, the probability for any one is equal to A^{-1}. Thus, in completely randomized experiments, the assignment mechanism is given by equation (4).

Bernoulli Trials

Completely randomized experiments are not without drawbacks. Consider, for example, a study with $N = 20$ units, ten men and ten women, where the potential treatment and control outcomes are *a priori* thought to vary substantially by sex. Then, although a completely randomized design with $N_t = 10$ would ensure that ten units get treated, there is the possibility that all ten of them are men (or women).

Stratified Randomized Experiments

The population of units is first partitioned into *blocks* or *strata* so that the units within each block are similar with respect to some (functions of) covariates thought to be predictive of potential outcomes. Then, within each block, we conduct a completely randomized experiment, with assignments independent across blocks.

The simplest randomized block experiment involves two blocks, say males and females, where independent completely randomized experiments are conducted for each group.

Thus, the assignment mechanism is the product of expression (4) for males and females, respectively.

Stratified Randomized Experiments

In general, more strata can be used. Let $B_i \in \{1, \ldots, J\}$ indicate the block of the i-th unit, with $B_i = B(X_i)$ a function of the pretreatment variables X_i, with a total of J blocks, and let $B_i(j)$ be the binary indicator for the event $B_i = j$.

Then the assignment mechanism is the product of J versions of expression (4), each version having N_j and N_{jt} indexed by the J distinct values of $B_i \in \{1, \ldots, J\}$.

The unit level probabilities are common for all units within a block, but can vary across blocks.

The main reason for generally preferring randomized blocks designs to completely randomized designs is that the former designs control balance in the covariates used to define blocks in treatment and control groups.

Stratified Randomized Experiments

Definition

(Stratified Randomized Experiment)
A stratified randomized experiment with J blocks is a classical randomized experiment with an assignment mechanism satisfying

$$\mathbb{W}^{+} = \left\{ \mathbf{W} \in \mathbb{W} \,\middle|\, \sum_{i:B_i=j}^{N} W_i = N_t(j), \text{for } j = 1, 2, \ldots, J \right\},$$

for some preset $N_t(j)$ such that $N_j > N_t(j) > 0$, for $j = 1, \ldots, J$.

In this setting, the unit-level assignment probability, or equivalently in our situation with a classical randomized experiment, the propensity score, $e(X_i)$, is equal to $N_t(j)/N(j)$ for all units with $B_i = j$. As this representation makes explicit, this probability can vary with the stratum indicator.

Stratified Randomized Experiments

Often, however, the unit-level assignment probabilities are identical across the strata so that $e(x) = q$ for all x.

The equality of sample size for treatment and control group is achieved within each stratum, whereas in complete randomization this is ensured only across the full sample.

If the covariate defining B_i corresponds to substantive information about the units, in the sense that B_i is predictive of the potential outcomes,$(Y_i(0), Y_i(1))$, randomizing within the strata will lead to more precise inferences.

Even if there is no predictive power of the blocking indicator B_i, stratification does not reduce actual precision if defined appropriately.

Paired Randomized Experiments

Definition

(Paired Randomized Experiment)
A paired randomized experiment is a stratified randomized experiment with $N(j) = 2$ and $N_t(j) = 1$ for $j = 1, \ldots, N/2$, and

$$\mathbb{W}^+ = \left\{ \mathbf{W} \in \mathbb{W} \;\middle|\; \sum_{i:B_i=j}^{N} W_i = 1, \text{for } j = 1, 2, \ldots, N/2 \right\}.$$

Example: An experiment of an education program where we have access to pretest scores at the design stage. The top two students on the pretest form the first pair, the next two form the next pair, and so forth.

Discussion

The way in which these four designs differ is in the set of values allowed for the vector of treatment indicators, $\mathbf{W}$, that is, in the definition of the $\mathbb{W}^{+}$.

As an illustration, let N be even and let the single variable X_i take on $N/2$ different values, with the number of units with $X_i = x$ equal to 2 for all $x \in \{1, \ldots, N/2\}$. Furthermore let $e(x) = 1/2$ for all x.

Paired design: Makes direct use of the fact that the number of units with $X_i = x$ is 2 for all $x = 1, \ldots, N/2$. Thus $B_i = X_i$.

Block design: Two blocks, $B_i = 1$ if $X_i \leq N/4$ and $B_i = 2$ if $X_i > N/4$. That is the number of units assigned to the 'treatment' and 'controls' within each block is $N/4$.

Discussion

Table 4.1: Number of Possible Non-Zero Values for the Assignment Vector by Design and Sample Size

Type of Experiment and Design	# of Possible Assignments Cardinality of $\mathbb{W}^+$	Number of Units (N) 4	8	16	32
Bernoulli trial	2^N	16	256	65,536	4.2×10^9
Completely Randomized Experiment	$\binom{N}{N/2}$	6	70	12,870	0.6×10^9
Stratified Randomized Experiment	$\binom{N/2}{N/4}^2$	4	36	4,900	0.2×10^9
Paired Randomized Experiment	$2^{N/2}$	4	16	256	65,536

Discussion

In this particular sequence of designs with fixed N, the number of distinct values of the assignment vector with positive probability, that is, the cardinality of the set $\mathbb{W}^+$, gradually decreases. The argument for choosing successively more restrictive designs is to eliminate "unhelpful" assignment vectors that are *a priori* unlikely to lead to precise causal inferences.

Imposing the first restriction — from Bernoulli trials to completely randomized experiments — is obvious because we need both treated and control unites.

The further restrictions, to stratified and paired randomized experiments, have similar advantages, when the grouping into strata or pairs is based on covariates that are related to the potential outcomes.

Discussion

In an extreme case, if the pre-treatment variable, X_i, upon which the stratification or pairing is based, perfectly predicts both potential outcomes, there will be no uncertainty remaining regarding the treatment effect across the N units or within the subgroups defined by the covariate.

On the other hand, if the blocks or pairs are formed in a way unrelated to the potential outcomes (e.g., by randomly drawing units to assign block labels B_i) the precision of estimators for treatment effects in stratified or paired randomized experiments is no greater than that for the corresponding estimators under completely randomized experiments.