Causal Inference from 2^k factorial designs using the potential outcomes model

Tirthankar Dasgupta
Harvard University
(joint work with Donald B. Rubin)
DEMA 2011

Motivating example: The New York School Study

- Assessment of effects of five new interventions on approximately 400 New York schools.
- Response: Annual review report scores.
- Appropriate setting for using a 2^5 factorial design (or possibly a fractional factorial).
- For each factor, 1: treatment, 0: control.

Factorial effects and their estimation

- Treatment combination: a binary k-tuple $\mathbf{x} = (x_1, \ldots, x_k)$, where $x_j \in \{0, 1\}$ for $j = 1, \ldots, k$.
- Response model: $y(\mathbf{x}) = \tau(\mathbf{x}) + \epsilon(\mathbf{x})$.
- Factorial effects: the $2^k - 1$ mutually orthogonal linear contrasts $\sum_{\mathbf{x}} l(\mathbf{x})\tau(\mathbf{x})$'s where $l(\mathbf{x}) = \pm 1/2^{k-1}$.
- An unbiased estimator: $\sum_{\mathbf{x}} l(\mathbf{x})\bar{y}(\mathbf{x})$.

Defining factorial effects: some key questions

- What are factorial effects in terms of the finite population of schools?
  - Average causal effects? Why average?
  - Effects of additivity of treatment effects on the inference?
  - Inference for a finite population?

Complexities in experimentation

- Bailey (2008, Ch. 7) : experiments with people and animals.

Complexities in experimentation

- Bailey (2008, Ch. 7) : experiments with people and animals.
- Experimental units (schools) may not accept randomly assigned treatment combinations.

Complexities in experimentation

- Bailey (2008, Ch. 7) : experiments with people and animals.
- Experimental units (schools) may not accept randomly assigned treatment combinations.
- Infeasible combinations (e.g, 00 in a 2^2 design).

Complexities in experimentation

- Bailey (2008, Ch. 7) : experiments with people and animals.
- Experimental units (schools) may not accept randomly assigned treatment combinations.
- Infeasible combinations (e.g, 00 in a 2^2 design).
- Semi-observational and observational studies with a factorial structure.

Potential outcomes model: treatment-control studies

Experimental Unit	Potential response under treatment(1)	control(0)	Causal Effect
1	$Y_1(1)$	$Y_1(0)$	$Y_1(1) - Y_1(0)$
...	...	...	...
N	$Y_N(1)$	$Y_N(0)$	$Y_N(1) - Y_N(0)$
Average	$\bar{Y}(1)$	$\bar{Y}(0)$	$\bar{Y}(1) - \bar{Y}(0)$

Potential outcomes model: treatment-control studies

- A population of units.

Potential outcomes model: treatment-control studies

- A population of units.
- Potential outcomes $\mathbf{Y}(1), \mathbf{Y}(0)$.

Potential outcomes model: treatment-control studies

- A population of units.
- Potential outcomes $\mathbf{Y}(1), \mathbf{Y}(0)$.
- Unit-level causal effects $Y_i(1) - Y_i(0)$.

Potential outcomes model: treatment-control studies

- A population of units.
- Potential outcomes $\mathbf{Y}(1), \mathbf{Y}(0)$.
- Unit-level causal effects $Y_i(1) - Y_i(0)$.
- Causal estimand (Population-level causal effect) $\tau(\mathbf{Y}(1), \mathbf{Y}(0))$.

Potential outcomes model: treatment-control studies

- A population of units.
- Potential outcomes $\mathbf{Y}(1), \mathbf{Y}(0)$.
- Unit-level causal effects $Y_i(1) - Y_i(0)$.
- Causal estimand (Population-level causal effect) $\tau(\mathbf{Y}(1), \mathbf{Y}(0))$.
- Assignment mechanism $\mathbf{W}$ (known for randomized experiments).

Potential outcomes model: treatment-control studies

- A population of units.
- Potential outcomes $\mathbf{Y}(1), \mathbf{Y}(0)$.
- Unit-level causal effects $Y_i(1) - Y_i(0)$.
- Causal estimand (Population-level causal effect) $\tau(\mathbf{Y}(1), \mathbf{Y}(0))$.
- Assignment mechanism $\mathbf{W}$ (known for randomized experiments).
- Observed outcomes $Y_i^{obs} = W_i Y_i(1) + (1 - W_i) Y_i(0)$ (half of the potential outcomes are observed).

Potential outcomes model: treatment-control studies

- A population of units.
- Potential outcomes $\mathbf{Y}(1), \mathbf{Y}(0)$.
- Unit-level causal effects $Y_i(1) - Y_i(0)$.
- Causal estimand (Population-level causal effect) $\tau(\mathbf{Y}(1), \mathbf{Y}(0))$.
- Assignment mechanism $\mathbf{W}$ (known for randomized experiments).
- Observed outcomes $Y_i^{obs} = W_i Y_i(1) + (1 - W_i) Y_i(0)$ (half of the potential outcomes are observed).
- Fundamental problem of causal inference: inference of the missing data and hence the causal estimand.

Potential outcomes model: A historical perspective

- *If we say, "This boy has grown tall because he has been well fed," we are not merely tracing out the cause and effect in the individual instance; we are suggesting that he might quite probably have been worse fed, and that in this case he would have been shorter* (Fisher, 1918).
- First formal notation by Neyman (1923).
- Causal inference from randomized experiments by Kempthorne (1952), Cox (1958).
- Formalized and extended by Rubin (1974) for causal inference from randomized experiments and observational studies.

Causal inference from 2^k factorial designs

- $N = 4r$ experimental units available for study.
- A typical treatment combination $\mathbf{x} = (x_1, \ldots, x_k)$, where $x_j \in \{0, 1\}$.
- Potential outcomes $Y_i(\mathbf{x})$ (2^k for each unit).
- Assumption: SUTVA (Rubin 1980).

Definition of unit-level factorial effects

For a non-null vector $(b_1, \ldots, b_k)$ with $b_j \in \{0, 1\}$, $j = 1, \ldots, k$, the factorial effect $\theta = 1^{b_1} \ldots k^{b_k}$ for the ith experimental unit, $i = 1, \ldots, N$ is:

$$\theta_i = \frac{1}{2^{k-1}} \Big(\sum_{\mathbf{x} \in V_1(\theta)} Y_i(\mathbf{x}) - \sum_{\mathbf{x} \in V_0(\theta)} Y_i(\mathbf{x}) \Big),$$

where $\{V_1(\theta), V_0(\theta)\}$ represent a disjoint partition of the set of all treatment combinations, such that

$$V_1(\theta) = \{\mathbf{x} : \sum_{j=1}^{k} b_j x_j = 1\}, V_0(\theta) = \{\mathbf{x} : \sum_{j=1}^{k} b_j x_j = 0\},$$

where additions are over the set of integers $\{0, 1\}$ *modulo* 2.

Estimands or population-level factorial effects for a finite population

- Can be any summary of the unit-level effects $\theta_1, \dots, \theta_N$.
- Most typically the average:

$$\theta^{FS} = \frac{1}{N} \sum_{i=1}^{N} \theta_i,$$

Extending the definition to a super population

- Allow the population, from which the $2^k r$ experimental units are sampled, to grow infinitely large,
- Factorial effect for super population:

$$\theta^{SP} = \lim_{N \to \infty} \theta^{FS} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \theta_i$$

Assignment variables

- Assignment variables:

$$W_i(\mathbf{x}) = \begin{cases} 1, & \text{if } i\text{th unit assigned to } \mathbf{x} \\ 0, & \text{otherwise} \end{cases}$$

- Observed outcome corresponding to the ith experimental unit:

$$Y_i^{\text{obs}} = \sum_{\mathbf{x}} W_i(\mathbf{x}) Y_i(\mathbf{x}).$$

Unbiased estimation of factorial effects

- A natural estimator of θ^{FS}:

$$\hat{\theta} = \frac{1}{2^{k-1}}\Big(\sum_{\mathbf{x} \in V_1(\theta)} \bar{Y}^{obs}(\mathbf{x}) - \sum_{\mathbf{x} \in V_0(\theta)} \bar{Y}^{obs}(\mathbf{x})\Big),$$

where

$$\bar{Y}^{obs}(\mathbf{x}) = \frac{1}{r} \sum_{i=1}^{N} W_i(\mathbf{x}) Y_i(\mathbf{x}),$$

represents the average of the r observations corresponding to treatment combination $\mathbf{x}$.

Neymanian inference: sampling distribution of estimated factorial effects

Theorem 1: For a completely randomized treatment assignment mechanism, the variance of $\hat{\theta}$ is

$$\frac{1}{2^{2(k-1)}r} \sum_{\mathbf{x}} S^2(\mathbf{x}) - \frac{1}{N} S^2(\theta),$$

where

$$S^2(\mathbf{x}) = \frac{1}{N-1} \sum_{i=1}^{N} \left(Y_i(\mathbf{x}) - \bar{Y}(\mathbf{x}) \right)^2,$$

$$S^2(\theta) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\theta_i - \theta^{FS} \right)^2.$$

Sampling distribution of estimated factorial effects (contd.)

Theorem 2: For a completely randomized treatment assignment mechanism, the covariance between the estimated factorial effects $\hat{\theta}_1$ and $\hat{\theta}_2$ is

$$\frac{1}{2^{2(k-1)}r}\Big[\sum_{\mathbf{x}\in V_0(\theta_1)\bigcap V_1(\theta_2)} S^2(\mathbf{x}) - \sum_{\mathbf{x}\in V_0(\theta_1)\bigcap V_1(\theta_2)} S^2(\mathbf{x})$$
$$- \sum_{\mathbf{x}\in V_1(\theta_1)\bigcap V_0(\theta_2)} S^2(\mathbf{x}) + \sum_{\mathbf{x}\in V_1(\theta_1)\bigcap V_1(\theta_2)} S^2(\mathbf{x})\Big] - \frac{1}{N} S^2(\theta_1, \theta_2),$$

where

$$S^2(\mathbf{x}) = \frac{1}{N-1}\sum_{i=1}^{N}\Big(Y_i(\mathbf{x}) - \bar{Y}(\mathbf{x})\Big)^2,$$
$$S^2(\theta_1, \theta_2) = \frac{1}{N-1}\sum_{i=1}^{N}\Big[\theta_{1i} - \theta_1^{FS}\Big]\Big[\theta_{2i} - \theta_2^{FS}\Big].$$

If additivity of treatment effects hold

- Additivity of all treatment effects:
  $Y_i(\mathbf{x}) - Y_i(\mathbf{x}^*) = c(\mathbf{x}, \mathbf{x}^*)$ for all i and for all pairs $\mathbf{x}, \mathbf{x}^*$.
- $S^2(\mathbf{x}) = S^2 \ \forall \mathbf{x}$, $S^2(\theta) = 0 \ \forall \theta$, $S^2(\theta_1, \theta_2) = 0 \ \forall \theta_1 \neq \theta_2$.
- Variance of each estimated factorial effect is the same.
- Covariance between each pair of estimated factorial effects is zero.
- Consequently, $\mathbb{COV}(\hat{\boldsymbol{\theta}}) = \frac{4}{2^k} S^2 \mathbf{I}_{2^k}$.

If additivity is violated

- Estimators are not uncorrelated.
- Estimators of variance will be conservative.
- Covariance needs to be considered for testing of a group of effects
- Covariance can either be overestimated or underestimated.

Fisherian inference: Randomization tests

- Randomization tests (Fisher's sharp null hypothesis).
- Permits use of any reasonable test statistic without the knowledge of its (asymptotic) distribution.
- Impute missing potential outcomes under the sharp null and compute the randomization distribution of estimated factorial effects.
- Used for unreplicated factorial designs (Permutation tests) by Loughin and Noble (1997), studied by Kimel, Benjamini and Steinberg (2008) - found inferior to competing methods (e.g., Lenth's test).

Testing Fisher's sharp null hypothesis for a 2^2 design

Experimental Unit	Treatment combination $\mathbf{x} = (x_1, x_2)$ (0,0)	(0,1)	(1,0)	(1,1)
1	?	?	$Y_1^{obs} = Y_1(10)$	?
2	?	$Y_2^{obs} = Y_2(01)$	?	?
3	?	?	?	$Y_3^{obs} = Y_3(11)$
...	...	...	...	...
...	...	...	...	...
N	$Y_N^{obs} = Y_N(00)$	?	?	?

Testing Fisher's sharp null hypothesis (contd.)

- Impute missing potential outcomes under the sharp null:

Experimental Unit	Treatment combination $\mathbf{x} = (x_1, x_2)$ (0,0)	(0,1)	(1,0)	(1,1)
1	Y_1^{obs}	Y_1^{obs}	Y_1^{obs}	Y_1^{obs}
2	Y_2^{obs}	Y_2^{obs}	Y_2^{obs}	Y_2^{obs}
3	Y_3^{obs}	Y_3^{obs}	Y_3^{obs}	Y_3^{obs}
...	...	...	...	...
...	...	...	...	...
N	Y_N^{obs}	Y_N^{obs}	Y_N^{obs}	Y_N^{obs}

- Resample from each column under the specified randomization scheme.
- Problems with the adaptive version.

A Bayesian framework for inference

- "The Fisher randomization test gives the posterior predictive distribution of the mean treatment difference under a model of no treatment effect and fixed units with fixed responses" (Rubin 1984).

Steps in Bayesian inference

1. Based on Rubin (1978).

Steps in Bayesian inference

1. Based on Rubin (1978).
2. Specify a probabilistic model for $f(\mathbf{Y}|\Theta)$ and suitable prior distributions for hyperparameters Θ.

Steps in Bayesian inference

1. Based on Rubin (1978).
2. Specify a probabilistic model for $f(\mathbf{Y}|\Theta)$ and suitable prior distributions for hyperparameters Θ.
3. Obtain the conditional posterior $f(\Theta|\mathbf{Y}^{obs}, \mathbf{W})$.

Steps in Bayesian inference

1. Based on Rubin (1978).
2. Specify a probabilistic model for $f(\mathbf{Y}|\Theta)$ and suitable prior distributions for hyperparameters Θ.
3. Obtain the conditional posterior $f(\Theta|\mathbf{Y}^{obs}, \mathbf{W})$.
4. Obtain $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \Theta, \mathbf{W})$.

Steps in Bayesian inference

1. Based on Rubin (1978).
2. Specify a probabilistic model for $f(\mathbf{Y}|\Theta)$ and suitable prior distributions for hyperparameters Θ.
3. Obtain the conditional posterior $f(\Theta|\mathbf{Y}^{obs}, \mathbf{W})$.
4. Obtain $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \Theta, \mathbf{W})$.
5. Using (2) and (3), obtain $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W})$ (Imputation model).

Steps in Bayesian inference

1. Based on Rubin (1978).
2. Specify a probabilistic model for $f(\mathbf{Y}|\Theta)$ and suitable prior distributions for hyperparameters Θ.
3. Obtain the conditional posterior $f(\Theta|\mathbf{Y}^{obs}, \mathbf{W})$.
4. Obtain $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \Theta, \mathbf{W})$.
5. Using (2) and (3), obtain $f(\mathbf{Y}^{mis}|\mathbf{Y}^{obs}, \mathbf{W})$ (Imputation model).
6. Obtain $f(\theta|\mathbf{Y}^{obs}, \mathbf{W})$.

A Naive model

- Potential outcomes model: $Y_i | \Theta \sim N_m(\boldsymbol{\tau}, \sigma^2 \mathbf{I}_m)$, $\boldsymbol{\tau} \sim N_m(\boldsymbol{\tau}_0, \sigma_\tau^2 I_m)$, σ^2 known, $m = 2^k$, completely randomized assignment model.
- For any factorial effect θ,

$$\begin{aligned} \mathbb{E}(\theta | \mathbf{Y}^{obs}, \mathbf{W}) &= \frac{1}{2^{k-1}} \Big[\sum_{\mathbf{x} \in V_1(\theta)} \big(\frac{r}{N} + \alpha(\mathbf{x}) \frac{N-r}{N} \big) \bar{Y}^{obs}(\mathbf{x}) \\ &\quad - \sum_{\mathbf{x} \in V_0(\theta)} \big(\frac{r}{N} + \alpha(\mathbf{x}) \frac{N-r}{N} \big) \bar{Y}^{obs}(\mathbf{x}) \Big], \end{aligned}$$

where

$$\alpha(\mathbf{x}) = \frac{\tau_0(\mathbf{x})/\sigma_\tau^2 + r/\sigma^2}{1/\sigma_\tau^2 + r/\sigma^2}.$$

- As $\sigma_\tau^2 \longrightarrow \infty$,

$$\mathbb{E}(\theta | \mathbf{Y}^{obs}, \mathbf{W}) \longrightarrow \frac{1}{2^{k-1}} \Big(\sum_{\mathbf{x} \in V_1(\theta)} \bar{Y}^{obs}(\mathbf{x}) - \sum_{\mathbf{x} \in V_0(\theta)} \bar{Y}^{obs}(\mathbf{x}) \Big)$$

Conclusions

- A framework for causal inference from 2^k factorial experiments using the potential outcomes framework is suggested.

Conclusions

- A framework for causal inference from 2^k factorial experiments using the potential outcomes framework is suggested.
- Focus on social, biomedical, educational experiments
  - Enables inference of any causal estimand defined as a function of potential outcomes
  - Can handle semi and fully observational studies having a factorial structure

Conclusions

- A framework for causal inference from 2^k factorial experiments using the potential outcomes framework is suggested.
- Focus on social, biomedical, educational experiments
  - Enables inference of any causal estimand defined as a function of potential outcomes
  - Can handle semi and fully observational studies having a factorial structure
- Potential applications in industrial experimentation?
  - Complex randomization structures
  - Large unit-to-unit variability (e.g., substrates in nanotechnology)
  - Non-normal response

The missing data

- Observed and missing data are linear transformations of vectors of potential outcomes:

$$Y_i^{obs} = \mathbf{W}_i'\mathbf{Y}_i,$$

$$\mathbf{Y}_i^{mis} = \mathbf{W}_i^{mis}\mathbf{Y}_i,$$

- Example: For a 2^2 design, if $W_i(00) = 1$,

$$Y_i^{obs} = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \mathbf{Y}_i,$$

$$Y_i^{mis} = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \mathbf{Y}_i$$