The Randomization Principle in Causal Inference: A Modern Look at Some Old Ideas

Qingyuan Zhao

Statistical Laboratory, University of Cambridge

2 December, 2022 @ Imperial College, London

The randomization principle in causal inference

We should use randomization in

1. The **design** of an **experiment**.
2. The **analysis** of an **experiment**.

We should mimic randomization in

3. The **design** of an **observational study**.
4. The **analysis** of an **observational study**.

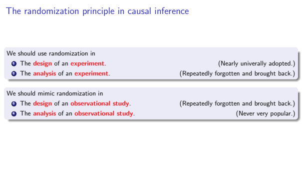
The randomization principle in causal inference

We should use randomization in

1. The **design** of an **experiment**. (Nearly univerally adopted.)
2. The **analysis** of an **experiment**. (Repeatedly forgotten and brought back.)

We should mimic randomization in

3. The **design** of an **observational study**. (Repeatedly forgotten and brought back.)
4. The **analysis** of an **observational study**. (Never very popular.)



This is partly due to a lack of precise description and understanding of the randomization principle. This talk will try to use modern tools in causal inference to better understand randomization and will have two parts.

Outline

Outline

Fisher and randomization

- Randomization is R A Fisher's first principle of experimental design It has profoundly changed how modern science is being done.
- *Statistical Methods for Research Workers* (1925) $\rightarrow$ Fisher (1926)[1] $\rightarrow$ *Design of Experiments* (1935).

[1] Ronald Aylmer Fisher (1926). "The Arrangement of Field Experiments". In: *Journal of the Ministry of Agriculture* 33, pp. 503–513. DOI: 10.23637/ROTHAMSTED.8V61Q.

Fisher and randomization

- Randomization is R A Fisher's first principle of experimental design It has profoundly changed how modern science is being done.
- *Statistical Methods for Research Workers* (1925) $\rightarrow$ Fisher (1926)[1] $\rightarrow$ *Design of Experiments* (1935).
- Nowadays we take this idea for granted. But this was not the case even decades after *DOE*.
- For example, W S Gosset ("Student") repeatedly disagreed with Fisher.

  I do not expect to convince you but I do not agree with your controlled randomness. You would want a large lunatic asylum for the operators who are apt to make mistakes enough even at present.

  (Gosset proofreading SMRW, 1924)

[1] Ronald Aylmer Fisher (1926). "The Arrangement of Field Experiments". In: *Journal of the Ministry of Agriculture* 33, pp. 503–513. DOI: 10.23637/ROTHAMSTED.8V61Q.

Randomization before and after Fisher

- Peirce and Jastrow (1884)[2] is believed to be the first randomized experiment.[3]
- Richet (1880s): Can we deteck weak powers of telepathy?
- Coover (starting from 1912): Randomized controlled experiments.
- Bradford Hill argued forcefully (in the 1940s) for randomized clinical trials.

[2] Charles Sanders Peirce and Joseph Jastrow (1884). "On small differences in sensation". In: *Memoirs of the National Academy of Sciences* 3.

[3] Ian Hacking (1988). "Telepathy: Origins of Randomization in Experimental Design". In: *Isis* 79.3, pp. 427–451. ISSN: 1545-6994. DOI: 10.1086/354775; Stephen M. Stigler (1978). "Mathematical Statistics in the Early States". In: *The Annals of Statistics* 6.2, pp. 239–265. DOI: 10.1214/aos/1176344123.

Randomization before and after Fisher

- Peirce and Jastrow (1884)[2] is believed to be the first randomized experiment.[3]
- Richet (1880s): Can we deteck weak powers of telepathy?
- Coover (starting from 1912): Randomized controlled experiments.
- Bradford Hill argued forcefully (in the 1940s) for randomized clinical trials.
- The psychologists and Hill emphasized on how randomization eliminates personal idiosyncracies and confounding bias.
- Fisher surely knew this point by his heart:

  Randomisation properly carried out ... relieves the experimenter from the anxiety of considering and estimating the magnitude of ***the innumerable causes by which his data may be disturbed****.* *(DOE, p. 44).*

[2] Charles Sanders Peirce and Joseph Jastrow (1884). "On small differences in sensation". In: *Memoirs of the National Academy of Sciences* 3.

[3] Ian Hacking (1988). "Telepathy: Origins of Randomization in Experimental Design". In: *Isis* 79.3, pp. 427–451. ISSN: 1545-6994. DOI: 10.1086/354775; Stephen M. Stigler (1978). "Mathematical Statistics in the Early States". In: *The Annals of Statistics* 6.2, pp. 239–265. DOI: 10.1214/aos/1176344123.

Randomization before and after Fisher

- Peirce and Jastrow (1884)[2] is believed to be the first randomized experiment.[3]
- Richet (1880s): Can we deteck weak powers of telepathy?
- Coover (starting from 1912): Randomized controlled experiments.
- Bradford Hill argued forcefully (in the 1940s) for randomized clinical trials.
- The psychologists and Hill emphasized on how randomization eliminates personal idiosyncracies and confounding bias.
- Fisher surely knew this point by his heart:

  Randomisation properly carried out ... relieves the experimenter from the anxiety of considering and estimating the magnitude of **the innumerable causes by which his data may be disturbed**. *(DOE, p. 44).*

[2] Charles Sanders Peirce and Joseph Jastrow (1884). "On small differences in sensation". In: *Memoirs of the National Academy of Sciences* 3.

[3] Ian Hacking (1988). "Telepathy: Origins of Randomization in Experimental Design". In: *Isis* 79.3, pp. 427–451. ISSN: 1545-6994. DOI: 10.1086/354775; Stephen M. Stigler (1978). "Mathematical Statistics in the Early States". In: *The Annals of Statistics* 6.2, pp. 239–265. DOI: 10.1214/aos/1176344123.

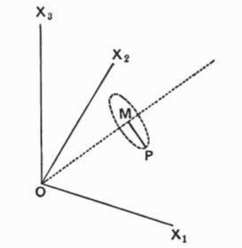
1. Peirce and Jastrow: To test whether there is a threshold in our sensation of pressure, experimental subjects first experienced a weight of 1kg and then a second weight either slightly heavier or slightly lighter than the first, which was determined by well shuffled decks of cards.
2. Richet used a long sequence of trials in which an "agent" drew a playing card at random and concentrated upon it for a short time, after which a "reagent" guessed the suit of the card.
3. Coover not only randomized the card, but also whether the trial would be regular or control (in which the agent did not look at the card at all).
4. Hacking's conclusion: Fisher was well aware of psychophysics research, but Fisherian randomization involves a very different level of sophistication.

Outline

Fisher's geometric intuition

- Fisher repeatedly used geometric insights, starting from his proof of Gosset's conjectured t-distribution.

  This result, although arrived at by empirical methods, was established almost beyond reasonable doubt...[but] the form **establishes itself instantly when the distribution of the sample is viewed geometrically***.* *(Fisher 1915)*[4]



- His argument involved representing n observations as a point P in the n-dimensional space: "For, given $\bar{x}$ and μ_1, P must lie on a sphere in $n-1$ dimensions."
- But the actual derivation of the t-distribution is much more involved than what Fisher indicated.

[4] Ronald Aylmer Fisher (1915). "Frequency distribution of the values of the correlation coeffients in samples from an indefinitely large population". In: *Biometrika* 10.4, pp. 507–521. DOI: 10.1093/biomet/10.4.507.

Randomization and analysis of variance

> *One way of making sure that a valid estimate of error will be obtained is to* **arrange the plots deliberately at random** *. . . ; in such a case an estimate of error, derived in the usual way from the variations of sets of plots treated alike, may be applied to test the significance of the observed difference between the averages of plots treated differently.* *(Fisher 1926)*[5]

> *His confidence in the result, however, depended on the geometric representation that was by then second nature to him. . . . he could see that* **randomization would produce a symmetry in that pattern rather like that produced by a kaleidoscope***, and which approximated the required spherical symmetry available, in particular, from standard normal theory assumptions.* *(Box 1980)*[6]

- Again, the math is not straightforward. The formal connection was not established until 1950s by requiring "additivity" (homogeneous treatment effect).[7]

---

[5] Ronald Aylmer Fisher (1926). "The Arrangement of Field Experiments". In: *Journal of the Ministry of Agriculture* 33, pp. 503–513. DOI: 10.23637/ROTHAMSTED.8V61Q.

[6] Joan Fisher Box (1980). "R. A. Fisher and the Design of Experiments, 1922-1926". In: *The American Statistician* 34.1, pp. 1–7. DOI: 10.1080/00031305.1980.10482701.

[7] Oscar Kempthorne (1955). "The Randomization Theory of Experimental Inference". In: *Journal of the American Statistical Association* 50.271, pp. 946–967. DOI: 10.2307/2281178.

Randomization and analysis of variance

> *One way of making sure that a valid estimate of error will be obtained is to* **arrange the plots deliberately at random** *...; in such a case an estimate of error, derived in the usual way from the variations of sets of plots treated alike, may be applied to test the significance of the observed difference between the averages of plots treated differently.* (Fisher 1926)[5]

> *His confidence in the result, however, depended on the geometric representation that was by then second nature to him. ...he could see that* **randomization would produce a symmetry in that pattern rather like that produced by a kaleidoscope**, *and which approximated the required spherical symmetry available, in particular, from standard normal theory assumptions.* (Box 1980)[6]

- Again, the math is not straightforward. The formal connection was not established until 1950s by requiring "additivity" (homogeneous treatment effect).[7]

[5] Ronald Aylmer Fisher (1926). "The Arrangement of Field Experiments". In: *Journal of the Ministry of Agriculture* 33, pp. 503–513. DOI: 10.23637/ROTHAMSTED.8V61Q.

[6] Joan Fisher Box (1980). "R. A. Fisher and the Design of Experiments, 1922-1926". In: *The American Statistician* 34.1, pp. 1–7. DOI: 10.1080/00031305.1980.10482701.

[7] Oscar Kempthorne (1955). "The Randomization Theory of Experimental Inference". In: *Journal of the American Statistical Association* 50.271, pp. 946–967. DOI: 10.2307/2281178.

1. Fisher emphasized the importance of randomization in quantifying statistical error.
2. The main result is that the randomization distribution of the F-statistic is approximately the F-distribution under Fisher's sharp null.

Randomization test

- Fisher initially suggested in *DOE* that randomization test can be used to substitute the *t*-test when normality is not true.

  In these discussions it seems to have escaped recognition that the physical act of randomisation, ..., affords the means, ..., of examining the wider hypothesis in which no normality of distribution is implied. *(DOE, p. 45)*

- Pitman (1937) seems to be the first who realized the full potential of randomization tests.[8]
- However, this is also confused with related concepts/terms, especially *permutation tests*.
- But the semantics are clearly different:
  - **Randomization** tests emphasize on the basis of inference (probabilistic).
  - **Permutation** tests emphasize on the computational algorithm (non-probabilistic).

[8] Patrick Onghena (2017). "Randomization Tests or Permutation Tests? A Historical and Terminological Clarification". In: *Randomization, Masking, and Allocation Concealment*, 209–228. DOI: 10.1201/9781315305110-14; E. J. G. Pitman (1937). "Significance Tests Which May Be Applied To Samples From Any Populations". In: *Supplement to the Journal of the Royal Statistical Society* 4.1, pp. 119–130. DOI: 10.2307/2984124.

What is randomization test?

Below: a precise description of conditional randomization tests that is a folklore among a small group of causal inference researchers.

What is randomization test?

Below: a precise description of conditional randomization tests that is a folklore among a small group of causal inference researchers.

Setup

- N units, treatment $\boldsymbol{Z} \in \mathcal{Z}$ is randomized.
- Potential outcomes $\boldsymbol{Y}(\boldsymbol{z}) = (Y_1(\boldsymbol{z}), \dots, Y_N(\boldsymbol{z}))$; Consistency: $\boldsymbol{Y} = (Y_1, \dots, Y_N) = Y(\boldsymbol{Z})$.
- P.O. schedule $\boldsymbol{W} = (\boldsymbol{Y}(\boldsymbol{z}) : \boldsymbol{z} \in \mathcal{Z}) \in \mathcal{W}$.

What is randomization test?

Below: a precise description of conditional randomization tests that is a folklore among a small group of causal inference researchers.

Setup

- N units, treatment $\boldsymbol{Z} \in \mathcal{Z}$ is randomized.
- Potential outcomes $\boldsymbol{Y}(\boldsymbol{z}) = (Y_1(\boldsymbol{z}), \dots, Y_N(\boldsymbol{z}))$; Consistency: $\boldsymbol{Y} = (Y_1, \dots, Y_N) = Y(\boldsymbol{Z})$.
- P.O. schedule $\boldsymbol{W} = (\boldsymbol{Y}(\boldsymbol{z}) : \boldsymbol{z} \in \mathcal{Z}) \in \mathcal{W}$.

Assumption (Randomization)

$\boldsymbol{Z} \perp\!\!\!\perp \boldsymbol{W}$ and the density function $\pi(\cdot)$ of $\boldsymbol{Z}$ is **known** *and positive everywhere.*

- Remark: We will condition on observed covariates $\boldsymbol{X}$.

Null hypothesis

A typical sharp null hypothesis assumes that certain potential outcomes are equal or related.

- Example 1: no interference $H_0 : Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}^*)$ whenever $z_i = z_i^*$;
- Example 2: constant treatment effect τ (on top of no interference) $H_0 : Y_i(1) - Y_i(0) = \tau$.

Null hypothesis

A typical sharp null hypothesis assumes that certain potential outcomes are equal or related.

- Example 1: no interference $H_0 : Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}^*)$ whenever $z_i = z_i^*$;
- Example 2: constant treatment effect τ (on top of no interference) $H_0 : Y_i(1) - Y_i(0) = \tau$.

Definition

A sharp null hypothesis H defines an **imputability mapping**

$$\begin{aligned} \mathcal{H} :\ & \mathcal{Z} \times \mathcal{Z} \to 2^{[N]}, \\ & (\boldsymbol{z}, \boldsymbol{z}^*) \mapsto \mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*), \end{aligned}$$

where $\mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*)$ is the largest subset of $[N] = \{1, \ldots, N\}$ such that $\boldsymbol{Y}_{\mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*)}(\boldsymbol{z}^*)$ is imputable from $\boldsymbol{Y}(\boldsymbol{z})$ under H.

Null hypothesis

A typical sharp null hypothesis assumes that certain potential outcomes are equal or related.

- Example 1: no interference $H_0 : Y_i(\boldsymbol{z}) = Y_i(\boldsymbol{z}^*)$ whenever $z_i = z_i^*$;
- Example 2: constant treatment effect τ (on top of no interference) $H_0 : Y_i(1) - Y_i(0) = \tau$.

Definition

A sharp null hypothesis H defines an **imputability mapping**

$$\begin{aligned} \mathcal{H} :\ & \mathcal{Z} \times \mathcal{Z} \to 2^{[N]}, \\ & (\boldsymbol{z}, \boldsymbol{z}^*) \mapsto \mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*), \end{aligned}$$

where $\mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*)$ is the largest subset of $[N] = \{1, \ldots, N\}$ such that $\boldsymbol{Y}_{\mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*)}(\boldsymbol{z}^*)$ is imputable from $\boldsymbol{Y}(\boldsymbol{z})$ under H.

Fully sharp means that $\mathcal{H}(\boldsymbol{z}, \boldsymbol{z}^*) \equiv [N]$. Otherwise **partially sharp**.

Conditional randomization tests (CRT)

- It is sometimes useful to not use the full randomness in $\boldsymbol{Z}$. Consider any function $g : \mathcal{Z} \to [M]$ and a collection of test statistics: $T_j : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$, $j \in [M]$.
- The *p-value* of the CRT is given by

$$P(\boldsymbol{Z}, \boldsymbol{W}) = \mathbb{P}\left\{ T_{g(\boldsymbol{Z})}(\boldsymbol{Z}', \boldsymbol{W}) \leq T_{g(\boldsymbol{Z})}(\boldsymbol{Z}, \boldsymbol{W}) \mid g(\boldsymbol{Z}') = g(\boldsymbol{Z}), \boldsymbol{Z}, \boldsymbol{W} \right\}.$$

where $\boldsymbol{Z}^*$ is an independent copy of $\boldsymbol{Z}$ given $\boldsymbol{W}$.

Conditional randomization tests (CRT)

- It is sometimes useful to not use the full randomness in $\boldsymbol{Z}$. Consider any function $g : \mathcal{Z} \to [M]$ and a collection of test statistics: $T_j : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$, $j \in [M]$.
- The *p-value* of the CRT is given by

$$P(\boldsymbol{Z}, \boldsymbol{W}) = \mathbb{P}\left\{ T_{g(\boldsymbol{Z})}(\boldsymbol{Z}', \boldsymbol{W}) \le T_{g(\boldsymbol{Z})}(\boldsymbol{Z}, \boldsymbol{W}) \mid g(\boldsymbol{Z}') = g(\boldsymbol{Z}), \boldsymbol{Z}, \boldsymbol{W} \right\}.$$

  where $\boldsymbol{Z}^*$ is an independent copy of $\boldsymbol{Z}$ given $\boldsymbol{W}$.
- Validity: This test always satisfies

$$\mathbb{P}\left\{P(\boldsymbol{Z}, \boldsymbol{W}) \le \alpha \mid g(\boldsymbol{Z}), \boldsymbol{W}\right\} \le \alpha, \ \forall \alpha \in [0,1], \boldsymbol{z} \in \mathcal{Z}.$$

- Computability: Suppose Assumption 1 is satisfied and the test statistics are imputable (in the sense that $T_{g(\boldsymbol{z})}(\boldsymbol{z}', \boldsymbol{W})$ only depends on $\boldsymbol{W}$ through $\boldsymbol{Y}_{\mathcal{H}(\boldsymbol{z},\boldsymbol{z}')}(\boldsymbol{z}')$ for all $\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}$). Then $P(\boldsymbol{Z}, \boldsymbol{W})$ only depends on $\boldsymbol{Z}$ and $\boldsymbol{Y}$.

Conditional randomization tests (CRT)

- It is sometimes useful to not use the full randomness in $\boldsymbol{Z}$. Consider any function $g : \mathcal{Z} \to [M]$ and a collection of test statistics: $T_j : \mathcal{Z} \times \mathcal{W} \to \mathbb{R}$, $j \in [M]$.
- The *p-value* of the CRT is given by

$$P(\boldsymbol{Z}, \boldsymbol{W}) = \mathbb{P}\left\{ T_{g(\boldsymbol{Z})}(\boldsymbol{Z}', \boldsymbol{W}) \le T_{g(\boldsymbol{Z})}(\boldsymbol{Z}, \boldsymbol{W}) \mid g(\boldsymbol{Z}') = g(\boldsymbol{Z}), \boldsymbol{Z}, \boldsymbol{W} \right\}.$$

  where $\boldsymbol{Z}^*$ is an independent copy of $\boldsymbol{Z}$ given $\boldsymbol{W}$.
- Validity: This test always satisfies

$$\mathbb{P}\left\{ P(\boldsymbol{Z}, \boldsymbol{W}) \le \alpha \mid g(\boldsymbol{Z}), \boldsymbol{W} \right\} \le \alpha, \ \forall \alpha \in [0,1], \boldsymbol{z} \in \mathcal{Z}.$$

- Computability: Suppose Assumption 1 is satisfied and the test statistics are imputable (in the sense that $T_{g(\boldsymbol{z})}(\boldsymbol{z}', \boldsymbol{W})$ only depends on $\boldsymbol{W}$ through $\boldsymbol{Y}_{\mathcal{H}(\boldsymbol{z},\boldsymbol{z}')}(\boldsymbol{z}')$ for all $\boldsymbol{z}, \boldsymbol{z}' \in \mathcal{Z}$). Then $P(\boldsymbol{Z}, \boldsymbol{W})$ only depends on $\boldsymbol{Z}$ and $\boldsymbol{Y}$.

1. Without randomization (Assumption 1), the distribution of $\boldsymbol{Z}^* \mid \boldsymbol{W} \overset{d}{=} \boldsymbol{Z} \mid \boldsymbol{W}$ is unknown.
2. Randomization guarantees validity, but the test is not always computable.

Example: Fisher's exact test for 2×2 contingency tables

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Fisher observed that the null probability of observing $(N_{00}, N_{01}, N_{10}, N_{11})$ **given the marginal totals** is given by the hypergeometric distribution. An exact test can then be immediately derived.

Example: Fisher's exact test for 2×2 contingency tables

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Fisher observed that the null probability of observing $(N_{00}, N_{01}, N_{10}, N_{11})$ **given the marginal totals** is given by the hypergeometric distribution. An exact test can then be immediately derived.

- This is a **unconditional randomization** test if the randomization fixes $N_{0\cdot}$ and $N_{1\cdot}$ (as in the famous tea-tasting example).

Example: Fisher's exact test for 2×2 contingency tables

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Fisher observed that the null probability of observing $(N_{00}, N_{01}, N_{10}, N_{11})$ **given the marginal totals** is given by the hypergeometric distribution. An exact test can then be immediately derived.

- This is a **unconditional randomization** test if the randomization fixes $N_{0\cdot}$ and $N_{1\cdot}$ (as in the famous tea-tasting example).
- This is a **conditional randomization** test if the treatments are assigned by Bernoulli trials.

Example: Fisher's exact test for 2×2 contingency tables

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Fisher observed that the null probability of observing $(N_{00}, N_{01}, N_{10}, N_{11})$ **given the marginal totals** is given by the hypergeometric distribution. An exact test can then be immediately derived.

- This is a **unconditional randomization** test if the randomization fixes $N_{0\cdot}$ and $N_{1\cdot}$ (as in the famous tea-tasting example).
- This is a **conditional randomization** test if the treatments are assigned by Bernoulli trials.
- This is a **conditional quasi-randomization** test in the "two Binomials" setup: $N_{00} \sim \mathrm{Bin}(N_{0\cdot}, \pi_0)$, $N_{10} \sim \mathrm{Bin}(N_{1\cdot}, \pi_1)$, and the null hypothesis is $H_0 : \pi_0 = \pi_1$.

Example: Fisher's exact test for 2×2 contingency tables

		Outcome Y 0	1	Total
Treatment A	0	N_{00}	N_{01}	$N_{0\cdot}$
	1	N_{10}	N_{11}	$N_{1\cdot}$
	Total	$N_{\cdot 0}$	$N_{\cdot 1}$	N

Fisher observed that the null probability of observing $(N_{00}, N_{01}, N_{10}, N_{11})$ **given the marginal totals** is given by the hypergeometric distribution. An exact test can then be immediately derived.

- This is a **unconditional randomization** test if the randomization fixes $N_{0\cdot}$ and $N_{1\cdot}$ (as in the famous tea-tasting example).
- This is a **conditional randomization** test if the treatments are assigned by Bernoulli trials.
- This is a **conditional quasi-randomization** test in the "two Binomials" setup: $N_{00} \sim \text{Bin}(N_{0\cdot}, \pi_0)$, $N_{10} \sim \text{Bin}(N_{1\cdot}, \pi_1)$, and the null hypothesis is $H_0 : \pi_0 = \pi_1$.
- This is always a permutation test, although Monte Carlo approximation is not needed.

Example: Evidence factors

- Consider $\mathcal{Z} = \{\text{non-smoking } (0), \text{light smoking } (1), \text{heavy smoking } (2)\}^n$; Y is lung cancer.
- To test the hypothesis $H : Y(0) = Y(1) = Y(2)$, we may use a randomization test that compares non-smokers with smokers.
- To test the hypothesis $H : Y(1) = Y(2)$, we may use a conditional randomization test that compares light smokers with heavy smokers; this amounts to conditioning on $\boldsymbol{g}(\boldsymbol{Z}) = (1_{\{Z_i=0\}})_{i=1}^n$.

[9] Paul R. Rosenbaum (Nov. 2017). "The General Structure of Evidence Factors in Observational Studies". In: *Statistical Science* 32.4, pp. 514–530. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/17-STS621.
[10] Yao Zhang and Qingyuan Zhao (2021). "Multiple Conditional Randomization Tests". In: arXiv: 2104.10618 [math.ST].

Example: Evidence factors

- Consider $\mathcal{Z} = \{\text{non-smoking (0)}, \text{light smoking (1)}, \text{heavy smoking (2)}\}^n$; Y is lung cancer.
- To test the hypothesis $H: Y(0) = Y(1) = Y(2)$, we may use a randomization test that compares non-smokers with smokers.
- To test the hypothesis $H: Y(1) = Y(2)$, we may use a conditional randomization test that compares light smokers with heavy smokers; this amounts to conditioning on $\boldsymbol{g}(\boldsymbol{Z}) = (1_{\{Z_i=0\}})_{i=1}^n$.
- Rosenbaum (2017)[9] confirmed the intuition that two tests should be "independent" by exploiting the knit product of permutation groups.
- A more general viewpoint: **sequential** conditional randomization tests $\Longrightarrow$ a much simpler proof by the law of iterated expectation.[10]

[9] Paul R. Rosenbaum (Nov. 2017). "The General Structure of Evidence Factors in Observational Studies". In: *Statistical Science* 32.4, pp. 514–530. ISSN: 0883-4237, 2168-8745. DOI: 10.1214/17-STS621.

[10] Yao Zhang and Qingyuan Zhao (2021). "Multiple Conditional Randomization Tests". In: arXiv: 2104.10618 [math.ST].

See our papers[11] for ...

- A discussion on the terminology.
- Different views of conditioning: on a function of $\boldsymbol{Z}$, on a partition of $\mathcal{Z}$, or on a sub σ-algebra.
- A discussion on post-randomization.
- A review of methods to construct computable tests in the causal interference literature.
- More examples: Permutation tests for treatment effect; tests for (conditional) independence; conformal inference.
- General conditions on when multiple conditional randomization tests are (nearly) independent.
- Applications to the stepped wedge trial design.

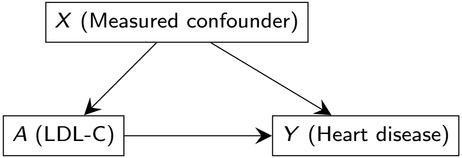
[11] Yao Zhang and Qingyuan Zhao (2021). "Multiple Conditional Randomization Tests". In: arXiv: 2104.10618 [math.ST]; Yao Zhang and Qingyuan Zhao (2022). "What Is a Randomization Test?" In: arXiv: 2203.10980 [stat.ME].

Outline

I will not talk about matching.

No unmeasured confounders/ignorability/exchangeability

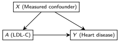
- It is typically assumed that all confounders X are measured, so that the observational study mimics a randomized experiment.



- It is often assumed that observations are drawn i.i.d. from this graph. Modern theory for causal graphical models interprets this as $A \perp\!\!\!\perp Y(a) \mid X$.
- But the role of randomization is obscure.
- For this reason, natural experiments are usually thought to be more credible.

No unmeasured confounders/ignorability/exchangeability

- It is typically assumed that all confounders X are measured, so that the observational study mimics a randomized experiment.

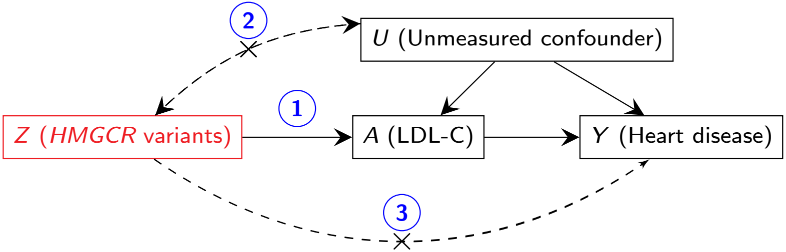


- It is often assumed that observations are drawn i.i.d. from this graph. Modern theory for causal graphical models interprets this as $A \perp Y(a) \mid X$.
- But the role of randomization is obscure.
- For this reason, natural experiments are usually thought to be more credible.

All randomness in A that cannot be explained by X is assumed to be randomized. This is very different from the active randomization in experiments.

Mendelian randomization

- Mendelian randomization tries to use randomness in genetic inheritance to aid causal inference.
- The most popular view is that genetic variants are used as **instrumental variables**.



- Modern causal graphical theory says this means that $Z \perp\!\!\!\perp (A(z), Y(z,a))$ and $Y(z,a) = Y(a)$.
- But the role of randomization is still not entirely clear.

Pre-history of Mendelian randomization

- Wright (1923), in a defence of his method of path coefficients, argues that the validity of this method "rests on the validity of the premises, i.e., on the evidence for Mendelian heridity", and the "universality" of Mendelian laws justifies ascribing a causal interpretation to his findings.[12]

[12] Sewall Wright (1923). "The Theory of Path Coefficients: A Reply to Niles's Criticism". In: *Genetics* 8.3, pp. 239–255. DOI: 10.1093/genetics/8.3.239.

[13] George Davey Smith and Shah Ebrahim (2003). "'Mendelian Randomization': Can Genetic Epidemiology Contribute To Understanding Environmental Determinants of Disease?" In: *International Journal of Epidemiology* 32.1, pp. 1–22. DOI: 10.1093/ije/dyg070.

Pre-history of Mendelian randomization

- Wright (1923), in a defence of his method of path coefficients, argues that the validity of this method "rests on the validity of the premises, i.e., on the evidence for Mendelian heridity", and the "universality" of Mendelian laws justifies ascribing a causal interpretation to his findings.[12]
- Fisher must have also known this by heart. Below are quotes from his 1951 Bateson lecture.

And here I may mention a connection between our two subjects which seems not to be altogether accidental, namely that the **"factorial" method of experimentation** *. . . derives its structure, and its name, from the simultaneous inheritance of* **Mendelian factors***.*

Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. **The different genotypes possible from the same mating have been beautifully randomised by the meiotic process.**

[12] Sewall Wright (1923). "The Theory of Path Coefficients: A Reply to Niles's Criticism". In: *Genetics* 8.3, pp. 239–255. DOI: 10.1093/genetics/8.3.239.

[13] George Davey Smith and Shah Ebrahim (2003). "'Mendelian Randomization': Can Genetic Epidemiology Contribute To Understanding Environmental Determinants of Disease?" In: *International Journal of Epidemiology* 32.1, pp. 1–22. DOI: 10.1093/ije/dyg070.

Pre-history of Mendelian randomization

- Wright (1923), in a defence of his method of path coefficients, argues that the validity of this method "rests on the validity of the premises, i.e., on the evidence for Mendelian heridity", and the "universality" of Mendelian laws justifies ascribing a causal interpretation to his findings.[12]
- Fisher must have also known this by heart. Below are quotes from his 1951 Bateson lecture.

  And here I may mention a connection between our two subjects which seems not to be altogether accidental, namely that the **"factorial" method of experimentation** *. . . derives its structure, and its name, from the simultaneous inheritance of* **Mendelian factors***.*

  Genetics is indeed in a peculiarly favoured condition in that Providence has shielded the geneticist from many of the difficulties of a reliably controlled comparison. **The different genotypes possible from the same mating have been beautifully randomised by the meiotic process.**

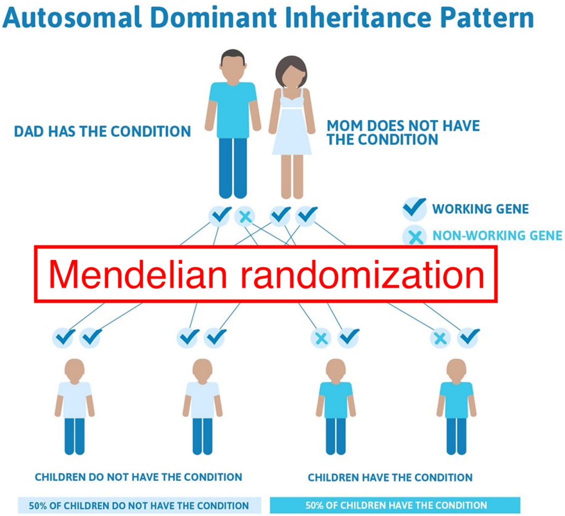
- Independent proposals appeared in 1970s-90s before Davey Smith and Ebrahim (2003) brought the idea to the front stage.[13]

---

[12] Sewall Wright (1923). "The Theory of Path Coefficients: A Reply to Niles's Criticism". In: *Genetics* 8.3, pp. 239–255. DOI: 10.1093/genetics/8.3.239.

[13] George Davey Smith and Shah Ebrahim (2003). "'Mendelian Randomization': Can Genetic Epidemiology Contribute To Understanding Environmental Determinants of Disease?" In: *International Journal of Epidemiology* 32.1, pp. 1–22. DOI: 10.1093/ije/dyg070.

Genetic inheritance as a natural experiment



Genetic trio studies

Data: Genotypes and phenotypes of mother, father, and offspring.

- $M/F/Z$: mother/father/offspring.
- Superscript f/m: Haplotypes inherited from father/mother.
- So $M_j^f \in \{0,1\}$ is mother's haplotype at locus j inherited from her father.
- No superscript means genotypes: $Z_j = Z_j^f + Z_j^m \in \{0,1,2\}$.

[14] R S Spielman, R E McGinnis, and W J Ewens (Mar. 1993). "Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM).". In: *American Journal of Human Genetics* 52.3, pp. 506–516. ISSN: 0002-9297.

[15] Stephen Bates et al. (Sept. 2020). "Causal Inference in Genetic Trio Studies". In: *Proceedings of the National Academy of Sciences* 117.39, pp. 24117–24126. DOI: 10.1073/pnas.2007743117.

[16] J B S Haldane (1919). "The combination of linkage values and the calculation of distance between the loci fo linked factors.". In: *Journal of Genetics* 8, pp. 299–309.

Genetic trio studies

Data: Genotypes and phenotypes of mother, father, and offspring.

- $M/F/Z$: mother/father/offspring.
- Superscript f/m: Haplotypes inherited from father/mother.
- So $M_j^f \in \{0,1\}$ is mother's haplotype at locus j inherited from her father.
- No superscript means genotypes: $Z_j = Z_j^f + Z_j^m \in \{0,1,2\}$.

- Spielman, McGinnis, and Ewens (1993)[14]: Conditional on parental haplotypes.
- Bates et al. (2020)[15]: Use existing meiosis models to obtain $\boldsymbol{Z} \mid \boldsymbol{M}^m, \boldsymbol{M}^f, \boldsymbol{F}^m, \boldsymbol{F}^f$.
- Haldane (1919)[16]: Ancestry indicator $\boldsymbol{U}$ roughly follows a Poisson process.

[14] R S Spielman, R E McGinnis, and W J Ewens (Mar. 1993). "Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM).". In: *American Journal of Human Genetics* 52.3, pp. 506–516. ISSN: 0002-9297.

[15] Stephen Bates et al. (Sept. 2020). "Causal Inference in Genetic Trio Studies". In: *Proceedings of the National Academy of Sciences* 117.39, pp. 24117–24126. DOI: 10.1073/pnas.2007743117.

[16] J B S Haldane (1919). "The combination of linkage values and the calculation of distance between the loci fo linked factors.". In: *Journal of Genetics* 8, pp. 299–309.

Illustration of within-family Mendelian randomization

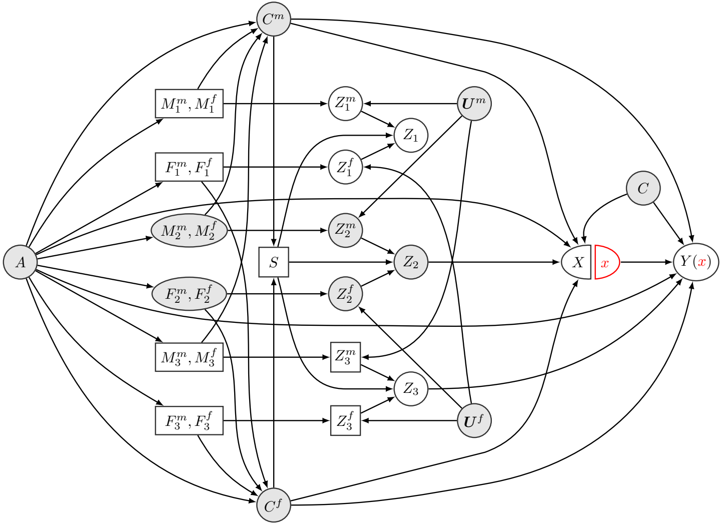


Illustration of within-family Mendelian randomization

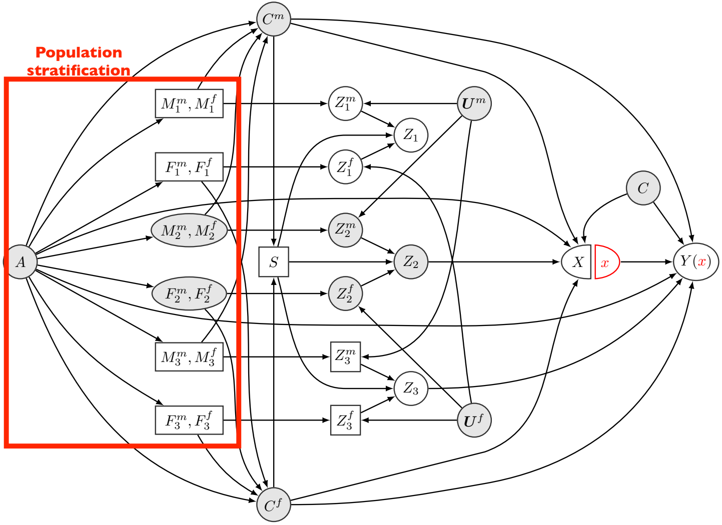


Illustration of within-family Mendelian randomization

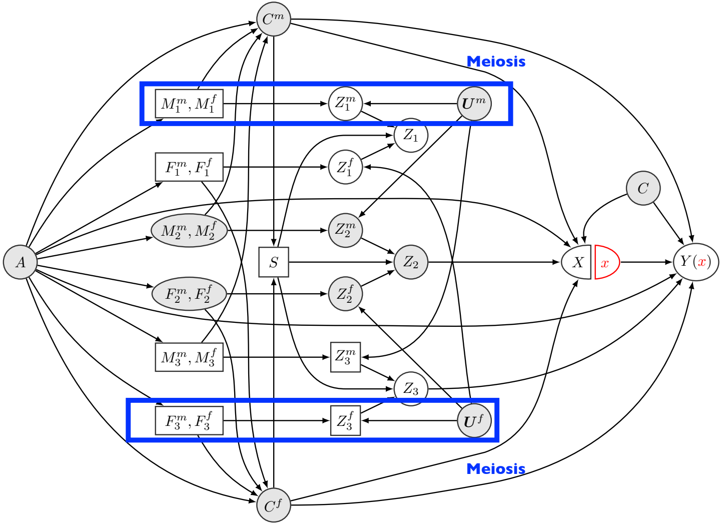


Illustration of within-family Mendelian randomization

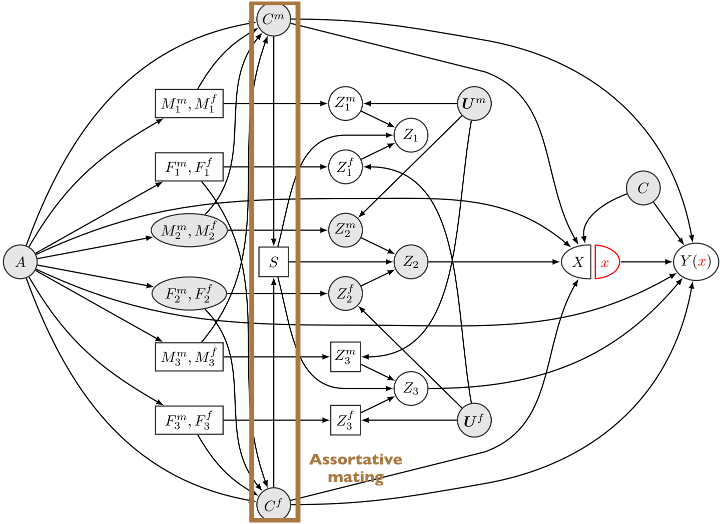


Illustration of within-family Mendelian randomization

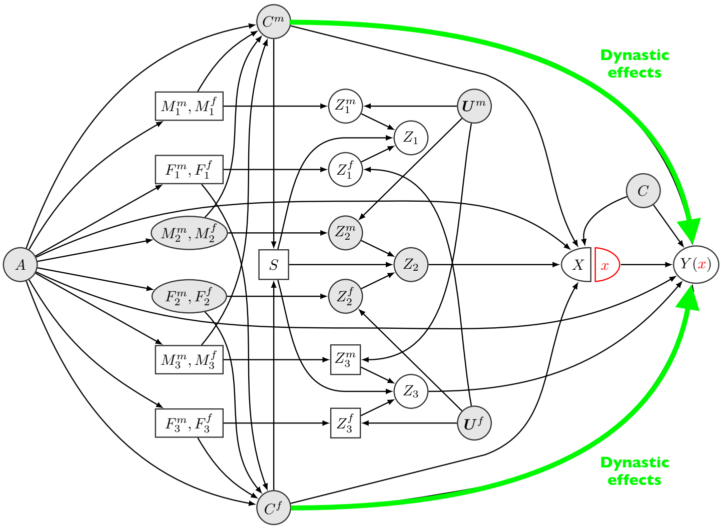


Illustration of within-family Mendelian randomization

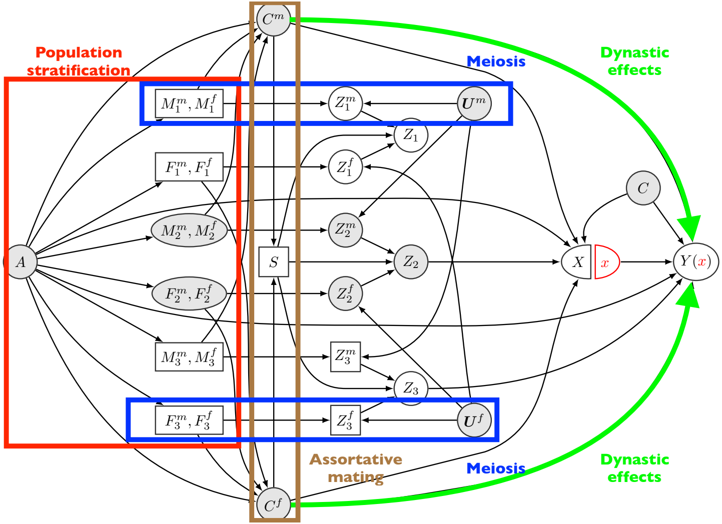
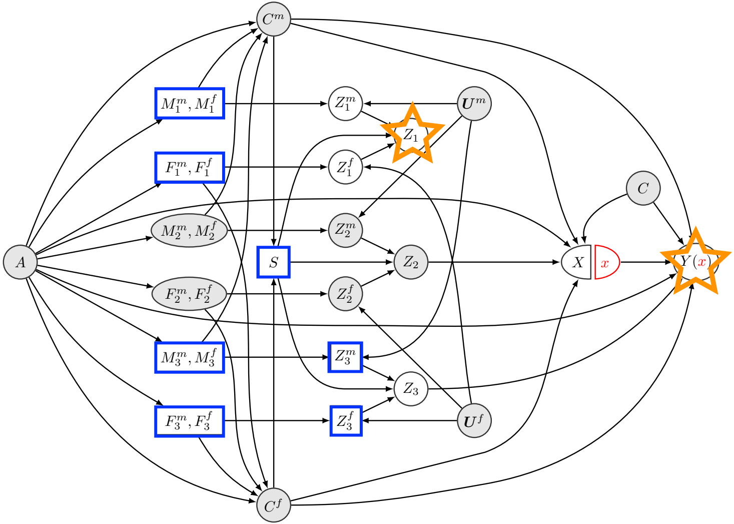
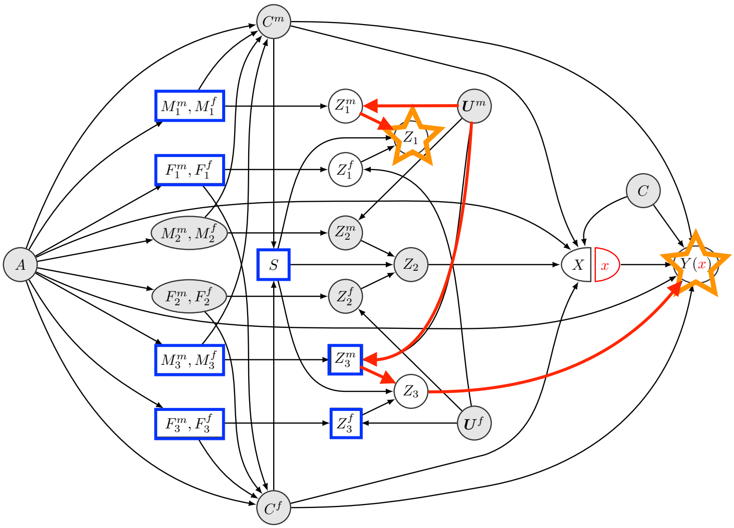


Illustration of within-family Mendelian randomization



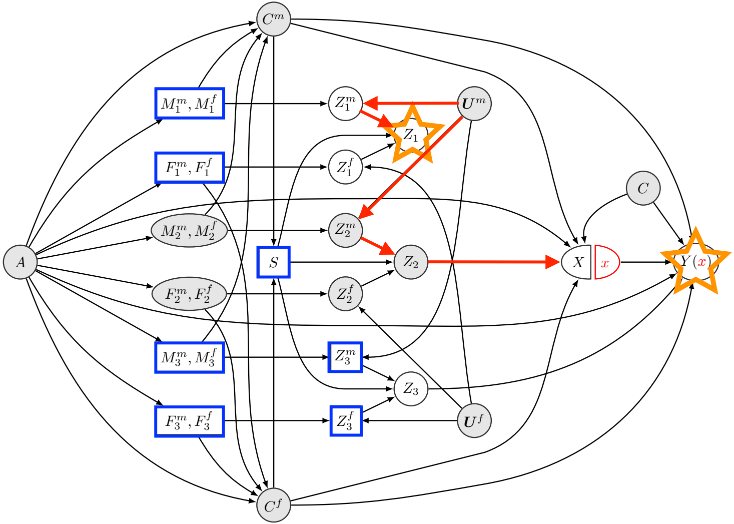
d-separation: $Z_1 \perp\!\!\!\perp Y(x) \mid (\boldsymbol{M}_1^{mf}, \boldsymbol{F}_1^{mf}, \boldsymbol{M}_3^{mf}, \boldsymbol{F}_3^{mf}, \boldsymbol{Z}_3^{mf})$.

Illustration of within-family Mendelian randomization



d-separation: $Z_1 \perp\!\!\!\perp Y(x) \mid (\boldsymbol{M}_1^{mf}, \boldsymbol{F}_1^{mf}, \boldsymbol{M}_3^{mf}, \boldsymbol{F}_3^{mf}, \boldsymbol{Z}_3^{mf})$.

Illustration of within-family Mendelian randomization



d-separation: $Z_1 \perp\!\!\!\perp Y(x) \mid (\boldsymbol{M}_1^{mf}, \boldsymbol{F}_1^{mf}, \boldsymbol{M}_3^{mf}, \boldsymbol{F}_3^{mf}, \boldsymbol{Z}_3^{mf})$.

See our paper[18] for ...

- A detailed account of the history of MR.
- A detailed explanation of the different components of this graph.
- A discussion various bias-inducing paths and sufficient adjustment sets.
- An "almost exact" randomization test, following previous ideas.[17]
- Simplification under Haldane's Poisson process model with no mutation.
- Combining techniques from multiple hypothesis testing (especially for partial conjunction nulls).
- "Proof-of-concept" examples.

[17] Hyunseung Kang, Laura Peck, and Luke Keele (2018). "Inference for Instrumental Variables: A Randomization Inference Approach". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 181.4, pp. 1231–1254. ISSN: 1467-985X. DOI: 10.1111/rssa.12353; Paul R. Rosenbaum (1996). "Identification of Causal Effects Using Instrumental Variables: Comment". In: *Journal of the American Statistical Association* 91.434, pp. 465–468. ISSN: 0162-1459. DOI: 10.2307/2291633.

[18] Matthew J Tudball, George Davey Smith, and Qingyuan Zhao (2022). "Almost Exact Mendelian Randomization". In: arXiv: 2208.14035 [stat.ME].

Closing remarks

- Two dominating principles in causal inference:
  - **Randomization**: design, blocking/matching, randomization test, exactness.
  - **Identification**: graphs, do- and potential outcomes calculus, i.i.d. sampling, semiparametric inference.
- We can gain a much better understanding about randomization by using tools developed primarily for identification. I believe there is also much to gain in the other direction.

[19] Ronald Aylmer Fisher (1922). "On the Mathematical Foundations of Theoretical Statistics". In: *Philosophical Transactions of the Royal Society of London. Series A* 222.594-604, pp. 309–368. DOI: 10.1098/rsta.1922.0009.

Closing remarks

- Two dominating principles in causal inference:
  - **Randomization**: design, blocking/matching, randomization test, exactness.
  - **Identification**: graphs, do- and potential outcomes calculus, i.i.d. sampling, semiparametric inference.
- We can gain a much better understanding about randomization by using tools developed primarily for identification. I believe there is also much to gain in the other direction.
- Can we further close this gap? Here is a sobering remark, again from Fisher.

  The postulate of randomness thus resolves itself into the question, 'Of what population is this a random sample?' which must frequently be asked by every practical statistician.

  (*Fisher 1922*)[19]

[19] Ronald Aylmer Fisher (1922). "On the Mathematical Foundations of Theoretical Statistics". In: *Philosophical Transactions of the Royal Society of London. Series A* 222.594-604, pp. 309–368. DOI: 10.1098/rsta.1922.0009.