Discussion on "Causal Inference in Genetic Trio Studies"

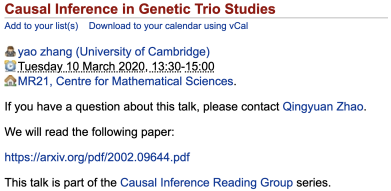
Qingyuan Zhao

Statistical Laboratory, University of Cambridge

January 26, 2021 @ Online Causal Inference Seminar

Congratulations to Stephen, Matteo, Chiara, and Emmanuel!

- Beautiful paper & presentation.
- Our reading group read arXiv:v1 two weeks after posted!



Causal Inference in Genetic Trio Studies

Add to your list(s) Download to your calendar using vCal

yao zhang (University of Cambridge)
Tuesday 10 March 2020, 13:30-15:00
MR21, Centre for Mathematical Sciences.

If you have a question about this talk, please contact Qingyuan Zhao.

We will read the following paper:

https://arxiv.org/pdf/2002.09644.pdf

This talk is part of the Causal Inference Reading Group series.

- Gained a lot of inspirations. Honored to discuss it.

This discussion

- Some historical context for DTT.
- Explain the basic ideas using causal DAGs.
- Some questions.

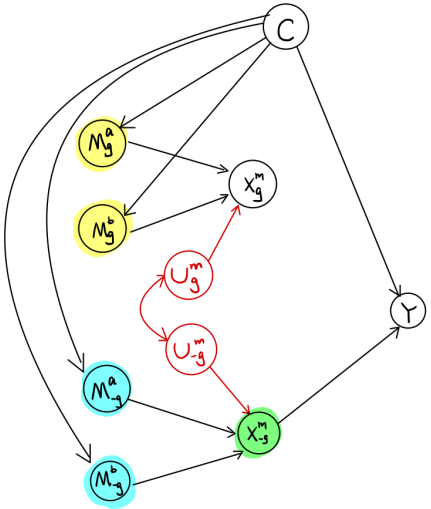
Mendelian randomization

- DTT is an instance of Mendelian randomization, although right now MR usually refers to using genetic variation as instrumental variables.

Historical context for DTT and causal inference

- Sewall Wright used selective inbreeding of guinea pigs to investigate the causes of colour variation. When criticized, he defended the method by **"the universality of Mendelian inheritance under sexual reproduction"**.
- Sir Ronald Fisher's 1925 *Statistical Method for Research Workers*: Randomization as the **"reasoned basis for inference"**.
- In Fisher's 1951 Bateson Lecture: **"the 'factorial' method of experimentation ... derives its structure and its name, from the simultaneous inheritance of Mendelian factors."**
- Same lecture: **"The different genotypes possible from the same mating have been beautifully randomised by the meiotic process. A more perfect control of conditions is scarcely possible."**
- A great talk by George Davey Smith (on YouTube): Mendelian randomization—where did it come from and where is it going?

What is the DTT?



U^m: "Mendelian Randomization"

Assortment / Recombination

Haldane's HMM: Distribution $X^m \mid M^a, M^b$

From the DAG: Ignorability

$$X_g^m \perp\!\!\!\perp Y(x_g^m = 0) \mid X_{-g}^m, M_g, M_{-g}$$

0. DTT: Randomization test.
1. GWAS: conditions on nothing.
2. TDT: only conditions on M_g
3. Multiple linear regression / vanilla knockoff: only conditions on X_{-g}^m

Remark Important to condition on M_{-g}.

Some questions

- Currently advertised as a way to confirm GWAS findings. Power/Type I error versus GWAS in realistic scenarios? When will we start to prefer DTT?
- Tradeoff between power and localization: couldn't find a power calculator.
- Meiosis model: Haldane's HMM assumes crossover is a Poisson process. In reality, there was evidence for positive crossover interference (with crossovers more evenly spaced than would be expected with random placement). A better model is the Gamma renewal process.[1] How much difference will this make for the DTT?

[1] Otto, S. P. & Payseur, B. A. Crossover Interference: Shedding Light on the Evolution of Recombination. *Annual Review of Genetics* **53,** 19–44 (2019).