CAUSAL INFERENCE Part III / Michaelmas 2020
Example Sheet 1 (of 3)

In the questions below on randomised experiments, we assume the treatment A is binary.

1. (Exercise 2.5) Consider a stratified randomised experiment with m groups. Suppose group j has n_j units, among which n_{1j} receive the treatment. What is the treatment assignment mechanism of this experiment?

2. (Exercise 2.17) Consider a completely randomised experiment in which the treatment assignments are sampled without replacement. Suppose n_1 out of the n units receive treatment. Let $\hat{\beta}$ be the difference-in-means estimator of $\mathrm{SATE} = \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - Y_i(0)$. Show that

$$\mathrm{Var}\left(\hat{\beta} \mid \boldsymbol{Y}(0), \boldsymbol{Y}(1)\right) = \frac{1}{n_0} S_0^2 + \frac{1}{n_1} S_1^2 - \frac{S_{01}^2}{n}, \tag{1}$$

where $n_0 = n - n_1$, $S_a^2 = \sum_{i=1}^{n}(Y_i(a) - \bar{Y}(a))^2/(n-1)$, $\bar{Y}(a) = \sum_{i=1}^{n} Y_i(a)/n$ for $a = 0, 1$, and $S_{01}^2 = \sum_{i=1}^{n}(Y_i(1) - Y_i(0) - \mathrm{SATE})^2/(n-1)$. Then construct an unbiased estimator of S_a^2.
Hint: Let $Y_i^*(a) = Y_i(a) - \bar{Y}(a), a = 0, 1$. Show that

$$\mathrm{Var}\left(\hat{\beta} \mid \boldsymbol{Y}(0), \boldsymbol{Y}(1)\right) = \mathbb{E}\Big[\Big(\sum_{i=1}^{n} \frac{A_i}{n_1} Y_i^*(1) - \frac{1 - A_i}{n_0} Y_i^*(0)\Big)^2\Big].$$

Then expand the sum of squares and use $\mathbb{E}[A_i A_{i'}] = \frac{n_1}{n}\frac{n_1 - 1}{n-1}, i \neq i'$ and $\sum_{i=1}^{n} Y_i^*(a) = 0$.

3. (Exercise 2.21) Consider two randomisation tests for the sharp null hypothesis $H_0 : Y_i(1) - Y_i(0) = \beta,\ i = 1, \ldots, n$. Let $\mathcal{F} = (\boldsymbol{X}_{[n]}, \boldsymbol{Y}_{[n]}(0), \boldsymbol{Y}_{[n]}(1))$. The first test uses the distribution of $T_1(\boldsymbol{A}_{[n]}, \boldsymbol{X}_{[n]}, \boldsymbol{Y}_{[n]}(0))$ given $\mathcal{F}$, while the second test uses the distribution of $T_2(\boldsymbol{A}_{[n]}, \boldsymbol{X}_{[n]}, \boldsymbol{Y}_{[n]}(\boldsymbol{A}_{[n]}))$ given $\mathcal{F}$. Show that the two tests are equivalent by constructing a one-to-one mapping between the functions T_1 and T_2.

4. (Exercise 2.24) Suppose $(\boldsymbol{X}_i, A_i, Y_i)$ are iid, $\mathbb{P}(A_i \mid \boldsymbol{X}_i) = \pi$, $\mathbb{E}[\boldsymbol{X}_i] = 0$. Let $(\boldsymbol{X}, A, Y)$ be a generic random vector from the same distribution. Define

$$\begin{aligned}
(\alpha_1, \beta_1) &= \operatorname*{arg\,min}_{\alpha,\beta} \mathbb{E}[(Y - \alpha - \beta A)^2], \\
(\alpha_2, \beta_2, \boldsymbol{\gamma}_2) &= \operatorname*{arg\,min}_{(\alpha,\beta,\boldsymbol{\gamma})} \mathbb{E}[(Y - \alpha - \beta A - \boldsymbol{\gamma}^T \boldsymbol{X})^2], \\
(\alpha_3, \beta_3, \boldsymbol{\gamma}_3, \boldsymbol{\delta}_3) &= \operatorname*{arg\,min}_{(\alpha\beta,\boldsymbol{\gamma},\boldsymbol{\delta})} \mathbb{E}[(Y - \alpha - \beta A - \boldsymbol{\gamma}^T \boldsymbol{X} - A \cdot (\boldsymbol{\delta}^T \boldsymbol{X}))^2].
\end{aligned}$$

Express $\boldsymbol{\gamma}_2$, $\boldsymbol{\gamma}_3$, and $\boldsymbol{\delta}_3$ in terms of in terms of the distribution of $(\boldsymbol{X}, A, Y)$ and in terms of the distribution of $(\boldsymbol{X}, A, Y(0), Y(1))$ (assuming SUTVA and the randomisation assumption).

5. (Exercises 2.29, 2.31) Consider the same setting as above, and let $\hat{\beta}_m$ be the least squares estimator of β in the mth regression. Let $\epsilon_1, \epsilon_2, \epsilon_3$ be the error terms in the three regression models:

$$\epsilon_m = Y - \alpha_m - \beta_m A - \boldsymbol{\gamma}_m^T \boldsymbol{X} - A(\boldsymbol{\delta}_m^T \boldsymbol{X}),\ m = 1, 2, 3.$$

Here we are using the convention $\boldsymbol{\gamma}_1 = 0$ and $\boldsymbol{\delta}_2 = \boldsymbol{\delta}_3 = \boldsymbol{0}$. Use the M-estimation theory for linear regression to show that, as $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_m - \beta) \xrightarrow{d} \mathrm{N}(0, V_m),\ \text{where } V_m = \frac{\mathbb{E}[(A - \pi)^2 \epsilon_m^2]}{\pi^2 (1 - \pi)^2},\ m = 1, 2, 3.$$

Under what conditions do we have $V_1 = V_2 = V_3$? Show that $V_2 \leq V_1$ is not always true.

6. (Exercises 3.5, 3.6) Show that a directed graph is acyclic if and only if the vertices can be relabelled in a way that the edges are monotone in the label (this is called a *topological ordering*). In other words, there exists a permutation $(k_1, \ldots, k_p)$ of $(1, \ldots, p)$ such that $(i, j) \in E$ implies $k_i < k_j$. The use the topological ordering to show that for any $J \subset [p]$, there exists $i \notin J$ such that all the descendants of i in a DAG $\mathcal{G}$ are in J.

7. (Exercise 3.10) Suppose the vertices in a DAG $\mathcal{G}$ is labelled according to a topological ordering and the random variables $\boldsymbol{X}_{[p]}$ are generated from a linear SEM according to $\mathcal{G}$. What property does the matrix of path coefficients have? Use this property to show that $\mathrm{Cov}(\boldsymbol{X}_{[p]})$ is positive definite.

8. (Exercise 3.15) Suppose $\boldsymbol{X}_{[p]}$ satisfies the linear SEM according to a DAG $\mathcal{G}$. Modify the path analysis equation for the covariance of X_i and X_j so that it is still true when the random variables are not standardised.

9. (Exercise 3.17) In each of the two cases below, give an example of a DAG and a linear SEM such that X is not an descendant of A or Y in the graph, $\mathrm{CE}(A, Y) = 0$ but the coefficient of A in a linear regression of Y on A and X is not equal to 0:

   (a) There is no d-connected path between A and Y;

   (b) X is on every d-connected path between A and Y.

10. There was an error in the previous version of this question. In this question, you will learn how to invert the randomisation test to obtain point estimators and confidence intervals for the treatment effect. Throughout the question we assume $H_0 : Y_i(1) - Y_i(0) = \beta,\ i = 1, \ldots, n$ is true but β is unknown. Consider a test statistic $T(\boldsymbol{A}_{[n]}, \boldsymbol{X}_{[n]}, \boldsymbol{Y}_{[n]})$; to simplify the notation, we will suppress the subscript $[n]$ in this question. Suppose the randomisation p-value is computed using the first approach in the lectures:

$$P_1 = P_1(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}) = \mathbb{P}^*\Big(T(\boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{Y}(0)) \leq T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y} - \beta\boldsymbol{A}) \mid \boldsymbol{X}, \boldsymbol{Y}(0)\Big),$$

where $\boldsymbol{A}^*$ is an independent copy of $\boldsymbol{A}$ and $\mathbb{P}^*$ means that the probability is with respect to the distribution of $\boldsymbol{A}^*$.
The Hodges-Lehmann estimator is given by the value of β such that the observed test statistic is equal to its expectation

$$T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y} - \beta\boldsymbol{A}) = \mathbb{E}^*\big[T(\boldsymbol{A}^*, \boldsymbol{X}, \boldsymbol{Y}(0)) \mid \boldsymbol{X}, \boldsymbol{Y}(0)\big].$$

For many test statistics, the right hand side does not depend on $\boldsymbol{Y}(0)$ (let's call it E). However, the solution to the above equation may not always exist. In that case, we may define

$$\hat{\beta}_{\mathrm{HL}} = \frac{\inf\{\beta \mid T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y} - \beta\boldsymbol{A}) < E\} + \sup\{\beta \mid T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y} - \beta\boldsymbol{A}) > E\}}{2}$$

if $T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y} - \beta\boldsymbol{A})$ is decreasing in β.

   (a) Consider a completely randomised experiments where n_1 out of n treated units are drawn by sampling without replacement. Suppose the test statistic is the difference-in-means estimator:

$$T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}) = \frac{1}{n_1}\sum_{i=1}^{n} A_i Y_i - \frac{1}{n - n_1}\sum_{i=1}^{n}(1 - A_i) Y_i.$$

   Show that the Hodges-Lehmann estimator corresponding to this randomisation test is also the difference-in-means estimator.

   (b) In a pairwise randomised experiment, let $1 \leq X_i \leq m = n/2$ to denote the pair which unit i is assigned to. Let D_j be the treated-minus-control difference in the jth pair

$$D_j = \sum_{i=1}^{n} I(X_i = j) \cdot (2A_i - 1) Y_i,\ j = 1, \ldots, m.$$

   The *sign statistic* is given by

$$T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}) = \sum_{j=1}^{m} \mathrm{sgn}(D_j),$$

   where sgn is the sign function

$$\mathrm{sgn}(x) = \begin{cases} -1, & \text{if } x < 0, \\ 0, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases}$$

Show that the Hodges-Lehmann estimator for this test is the sample median

$$\hat{\beta}_{\mathrm{HL}} = \begin{cases} D_{(\frac{m+1}{2})}, & \text{if } m \text{ is odd,} \\ \frac{1}{2}\{D_{(\frac{m}{2})} + D_{(\frac{m}{2}+1)}\}, & \text{if } m \text{ is even,} \end{cases}$$

where $D_{(j)}$ denotes the jth order statistic of $D_1, \ldots, D_m$.

(c) Consider the interval estimator $\mathrm{CI}_\alpha = \{\beta \mid P_2 \geq \alpha\}$ where $P_2 = P_2(\beta)$ is the randomisation p-value defined in the lectures. Under SUTVA and H_0, show that CI_α is a $(1-\alpha)$-confidence interval for β (i.e., it covers β with probability at least $1-\alpha$).

(d) Derive CI_α for the sign test in the setting in (b).