CAUSAL INFERENCE Part III / Michaelmas 2023
Example Sheet 1 (of 3)

All questions below on randomized experiments assume the Neyman-Rubin causal model (which entails consistency of potential outcomes and identity exposure mapping) and exogeneity of randomization. Additional terminology and notation can be found in `http://www.statslab.cam.ac.uk/~qz280/teaching/causal-2023/slides.pdf`.

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

2. Let T be a real-valued random variable and F be its cumulative distribution function: $F(t) = \mathbb{P}(T \leq t)$. Show that $F(T)$ stochastically dominates the uniform distribution over $[0,1]$ in the sense that
$$\mathbb{P}(F(T) \leq \alpha) \leq \alpha \quad \text{for all} \quad 0 \leq \alpha \leq 1.$$
Further, show that this inequality becomes an equality when F is continuous. (*Hint*: Consider the quantile function: $F^{-}(\alpha) = \inf\{t : F(t) \geq \alpha\}$.)

3. In this question we assume $\mathcal{A} = \{0,1\}$ and the treatment is assigned by sampling without replacement, that is, n_1 out of the n units receive the treatment at random. Show that the variance of the difference-in-means estimator
$$\hat{\beta} = \frac{1}{n_1}\sum_{i=1}^{n} A_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i) Y_i$$
in its randomzation distribution given the potential outcomes schedule $\boldsymbol{W}$ is given by
$$\text{Var}\left(\hat{\beta} \mid \boldsymbol{W}\right) = \frac{1}{n_0} S(0)^2 + \frac{1}{n_1} S(1)^2 - \frac{1}{n} S(0,1)^2. \tag{1}$$

4. Consider the repeated sampling setting: $\mathcal{A} = \{0,1\}$, $(X_i, A_i, (Y_i(a))_{a\in\mathcal{A}})$ are i.i.d, and let (X, A, Y) be a generic random vector from the same distribution. Further, suppose $A \perp\!\!\!\perp X$ and $\mathbb{E}(X) = 0$. Define
$$\begin{aligned}
(\alpha_1, \beta_1) &= \arg\min_{\alpha,\beta} \mathbb{E}[(Y - \alpha - \beta A)^2], \\
(\alpha_2, \beta_2, \gamma_2) &= \arg\min_{(\alpha,\beta,\gamma)} \mathbb{E}[(Y - \alpha - \beta A - \gamma X)^2], \\
(\alpha_3, \beta_3, \gamma_3, \delta_3) &= \arg\min_{(\alpha,\beta,\gamma,\delta)} \mathbb{E}[(Y - \alpha - \beta A - \gamma X - A \cdot (\delta X))^2].
\end{aligned}$$
Express γ_2, γ_3, and δ_3 in terms of the distribution of (X, A, Y). Then express them in terms of the distribution of $(X, A, Y(0), Y(1))$.

5. In the same setting as above, let $\hat{\beta}_m$, $m = 1,2,3$, be the least squares estimator of β in the mth regression problem. It has been shown in the lectures that
$$\sqrt{n}(\hat{\beta}_m - \beta) \xrightarrow{d} \text{N}(0, V_m), \text{ where } V_m = \frac{\mathbb{E}[(A-\pi)^2 \epsilon_m^2]}{\pi^2(1-\pi)^2}, \; m = 1,2,3,$$
where ϵ_m is the residual in the corresponding regression problem. When do we have $V_1 = V_2 = V_3$? Give a counterexample that shows $V_2 \leq V_1$ is not always true.

6. Consider the independence testing problem in 2×2 contigency tables with counts $(N_{00}, N_{01}, N_{10}, N_{11})$. Suppose the row margins $N_{0\cdot} = N_{00} + N_{01}$ and $N_{1\cdot} = N_{10} + N_{11}$ are fixed, and $N_{01} \sim \text{Bin}(N_{0\cdot}, \pi_0)$ is independent of $N_{11} \sim \text{Bin}(N_{0\cdot}, \pi_1)$. Show that Fisher's exact test is still valid for testing the null hypothesis $H_0 : \pi_0 = \pi_1$ by deriving the conditional distribution of N_{11} given the column margins $N_{\cdot 0}$ and $N_{\cdot 1}$. Compare this with Fisher's exact test in the randomization model discussed in lectures.

7. When the treatment is binary (i.e. $\mathcal{A} = \{0, 1\}$), a popular choice of randomization test statistic is the signed rank:

$$T(\boldsymbol{A}, \boldsymbol{W}) = \sum_{i=1}^{n} A_i \psi\Big(\frac{r_i}{n}\Big), \text{ where } r_i = \text{rank of } |Y_i(0)| \text{ among } |Y_1(0)|, \dots, |Y_n(0)|, \qquad (2)$$

and $\psi : (0, 1) \to \mathbb{R}$ is a transformation of the normalized rank. For example, the identity transformation $\psi(r) = r$ corresponds to Wilcoxon's rank sum statistic.

(a) Show that, if $\boldsymbol{A}$ is randomized by sampling without replacement, the signed rank statistic $T(\boldsymbol{A}, \boldsymbol{W})$ is *distribution-free* in the sense that its randomization distribution does not depend on $\boldsymbol{W}$.

(b) With the choice $\psi(r) = r$, use your results in part (a) to derive an asymptotic z-test of Fisher's sharp null hypothesis: $H_0 : Y_i(0) = Y_i(1)$ for all i. This test should compare T with quantiles of the standard normal distribution after scaling.

8. Continuing from Question 7, we will consider how to invert the randomization test to obtain point estimators and confidence intervals for the treatment effect. We assume the treatment effect is a constant β, i.e. $Y_i(1) - Y_i(0) = \beta,\ i = 1, \dots, n$, but β is unknown. Consider the test statistic $T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{W}) = T(\boldsymbol{A}, \boldsymbol{X}, \boldsymbol{Y}(0))$ where $\boldsymbol{Y}(0) = (Y_1(0), \dots, Y_n(0)) = \boldsymbol{Y} - \beta\boldsymbol{A}$.

The Hodges-Lehmann estimator $\hat{\beta}_{\text{HL}}$ 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 (let's call it E) does not depend on $\boldsymbol{Y}(0)$. However, the solution to the above equation may not always exist. In that case, we define

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

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

(a) Suppose treatment is assigned by sampling without replacement (that is, n_1 out of the n units receive the treatment at random), and 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 corresponding Hodges-Lehmann estimator is also the difference-in-means estimator.

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

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

The *sign statistic* is given by

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

where sgn is the sign function

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

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

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

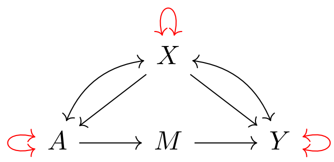
where $D_{(j)}$ denotes the jth order statistic of $D_1, \dots, D_m$.

(c) Consider the set estimator $\mathcal{C}_\alpha = \{\beta \mid P \geq \alpha\}$ (usually an interval) where $P = P(\beta)$ is the randomization p-value for the null hypothesis $H_0 : Y_i(1) - Y_i(0) = \beta, \forall i$. Show that $\mathcal{C}_\alpha$ is a confidence set for β with confidence level at least $(1-\alpha)$. Explain how the distribution-free property in Question 7 may be useful to compute $\mathcal{C}_\alpha$.

9. Show that a directed mixed graph $\mathcal{G} = (\mathcal{V} = [p], \mathcal{D}, \mathcal{B})$ 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 $(k_1, \dots, k_p)$ of $(1, \dots, p)$ such that $(i, j) \in \mathcal{D}$ implies $k_i < k_j$. Then use this topological ordering to show that for any strict subset $J \subset [p]$, there exists $i \notin J$ such that all the descendants of i in $\mathcal{G}$ are in J.

10. How would you represent a randomized experiment with a generic exposure mapping using a graph? Your graph should contain the following vertices: pre-treatment covariates $X_1, \dots, X_n$, treatment Z, exposures $A_1, \dots, A_n$, and outcomes $Y_1, \dots, Y_n$.

11. Consider a linear SEM corresponding to the following acyclic directed mixed graph (ADMG). Use both the trek rule and Wright's path analysis to find $\mathrm{Cov}(A, Y)$ and verify that the two expressions are the same.



12. Consider the partition of a random vector $\boldsymbol{V} = (V_1, V_2, \boldsymbol{V}_3)$, and denote its covariance and inverse covariance matrix as (assuming the covariance matrix is indeed invertible)

$$\mathrm{Cov}(\boldsymbol{V}) = \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \boldsymbol{\Sigma}_{13} \\ \Sigma_{21} & \Sigma_{22} & \boldsymbol{\Sigma}_{23} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{32} & \boldsymbol{\Sigma}_{33} \end{pmatrix}, \quad \boldsymbol{\Sigma}^{-1} = \boldsymbol{\Omega} = \begin{pmatrix} \Omega_{11} & \Omega_{12} & \boldsymbol{\Omega}_{13} \\ \Omega_{21} & \Omega_{22} & \boldsymbol{\Omega}_{23} \\ \boldsymbol{\Omega}_{31} & \boldsymbol{\Omega}_{32} & \boldsymbol{\Omega}_{33} \end{pmatrix}.$$

The *partial correlation* of V_1 and V_2 given $\boldsymbol{V}_3$ is defined as

$$\mathrm{Cor}(V_1, V_2 \mid \boldsymbol{V}_3) = \mathrm{Cor}(V_1 - \boldsymbol{\Sigma}_{13}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{V}_3, V_2 - \boldsymbol{\Sigma}_{23}\boldsymbol{\Sigma}_{33}^{-1}\boldsymbol{V}_3).$$

Show that $\mathrm{Cor}(V_1, V_2 \mid \boldsymbol{V}_3) = -\Omega_{12}/\sqrt{\Omega_{11}\Omega_{22}}$. [*Hint:* You may find the block matrix inversion formula given in the lectures useful.]