Homework 11
36-705
Due: Thursday November 12th by 3pm.

1. Suppose we randomize n people to treatment or control. The observed data are $(A_1, Y_1), \ldots, (A_n, Y_n)$ where A_i is binary ($A_i = 0$ means control and $A_i = 1$ means treatment). We have $P(A_i = 0) = P(A_i = 1) = 1/2$. We want to estimate the average treatment effect:

$$\tau = \mathbb{E}[Y(1) - Y(0)]$$

where $Y(0)$ and $Y(1)$ are the potential outcomes (counterfactuals). Let

$$\widehat{\tau}_1 = \frac{1}{n_1} \sum_{i:A_i=1} Y_i - \frac{1}{n_0} \sum_{i:A_i=0} Y_i.$$

Note that n_1 and n_0 are random.

(a) Find the limiting distribution of $\sqrt{n}(\widehat{\tau} - \tau)$.

(b) Construct an asymptotic $1 - \alpha$ confidence interval for τ.

2. Suppose we have an observational study with data $(X_1, A_1, Y_1), \ldots, (X_n, A_n, Y_n)$ where A_i is a binary treatment.

$$A \perp\!\!\!\perp (Y(1), Y(0)) | X.$$

(a) Suppose that the propensity scores $\pi(X) = \mathbb{P}(A = 1|X)$ is known. Also assumme that

$$0 < \delta \leq P(A = 1|X) \leq 1 - \delta < 1$$

for some δ. A natural estimator then is the Horvitz-Thompson/Inverse-Propensity Score estimator:

$$\widehat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{Y_i A_i}{\pi(X_i)} - \frac{Y_i(1 - A_i)}{1 - \pi(X_i)} \right].$$

Find the limiting distribution of $\sqrt{n}(\widehat{\tau} - \tau)$.

(b) Construct an asymptotic $1 - \alpha$ confidence interval for τ.

(c) Suppose we don't know $\pi(x)$ but we have an estimate $\widehat{\pi}$ and that

$$\sup_x |\widehat{\pi}(x) - \pi(x)| \leq \epsilon_n$$

where $\epsilon_n \to 0$. Let

$$\widehat{\tau} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{Y_i A_i}{\widehat{\pi}(X_i)} - \frac{Y_i(1 - A_i)}{1 - \widehat{\pi}(X_i)} \right].$$

Show that $\widehat{\tau}^2 \xrightarrow{P} \tau$.

3. Suppose we have an observational study with data $(X_1, A_1, Y_1), \ldots, (X_n, A_n, Y_n)$ where $A_i \in \mathbb{R}$. Assume that

$$A \perp\!\!\!\perp (Y(a) : \ a \in \mathbb{R}) | X.$$

In this case, we showed that

$$\psi(a) \equiv \mathbb{E}[Y(a)] = \int \mu(x, a) p(x) dx$$

where $\mu(x, a) = \mathbb{E}[Y|X = x, A = a]$.

(a) Supposse that $\mu(x,a)$ is know. Let

$$\widehat{\psi}(a) = \frac{1}{n}\sum_i \widehat{\mu}(X_i, a).$$

Find the limiting distribution of $\sqrt{n}(\widehat{\psi}(a) - \psi(a))$.

(b) Suppose that $\mu(x,a)$ is unknown but that $p(a|x)$ is known. Fix $\epsilon > 0$ and define

$$\widehat{\psi}_\epsilon(a) = \frac{1}{2n\epsilon}\sum_i \frac{Y_i\ I(|A_i - a| < \epsilon)}{p(A_i|X_i)}$$

where $I(|A_i - a| < \epsilon) = 1$ if $|A_i - a| < \epsilon$ and $I(|A_i - a| < \epsilon) = 0$ otherwise.

Find the mean of $\widehat{\psi}_\epsilon(a)$. Let's denote the mean by $\psi_\epsilon(a)$. Show that $\psi_\epsilon(a) \to \psi(a)$ as $\epsilon \to 0$ (assuming $\mu(x,a)$ is sufficiently smooth).