36-705 Intermediate Statistics Homework #8
Solutions

November 27, 2017

Problem 1 (15 points)

For each $i$, $Y_i^{obs} \in \{0, 1\}$. We define $W_i = 1$, when $i \in T$ and $W_i = 0$, when $i \in C$. Since individuals are randomly assigned to receive treatment and control, $W \perp (Y(1), Y(0))$. Hence, for $i \in T$,

$$\mathbb{E}[Y_i^{obs}] = \mathbb{E}[Y(1)|W = 1] = \mathbb{E}[Y(1)],$$

and for $i \in C$,

$$\mathbb{E}[Y_i^{obs}] = \mathbb{E}[Y(0)|W = 0] = \mathbb{E}[Y(0)].$$

Therefore,

$$\mathbb{E}\left[\frac{2}{n} \sum_{i \in T} Y_i^{obs}\right] = \mathbb{E}[Y(1)], \quad \mathbb{E}\left[\frac{2}{n} \sum_{i \in C} Y_i^{obs}\right] = \mathbb{E}[Y(0)].$$

Now since $Y_i^{obs} \in \{0, 1\}$, by Hoeffding’s inequality we have that,

$$\mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in T} Y_i^{obs} - \mathbb{E}[Y(1)]\right| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{4}\right),$$

and

$$\mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in C} Y_i^{obs} - \mathbb{E}[Y(0)]\right| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{4}\right).$$

Now notice that,

$$\mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in T} Y_i^{obs} - \frac{2}{n} \sum_{i \in C} Y_i^{obs} - \tau\right| \geq t\right) = \mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in T} Y_i^{obs} - \frac{2}{n} \sum_{i \in C} Y_i^{obs} - \mathbb{E}[Y(1)] + \mathbb{E}[Y(0)]\right| \geq t\right)$$

$$\leq \mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in T} Y_i^{obs} - \mathbb{E}[Y(1)]\right| + \left|\frac{2}{n} \sum_{i \in C} Y_i^{obs} - \mathbb{E}[Y(0)]\right| \geq t\right)$$

$$\leq \mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in T} Y_i^{obs} - \mathbb{E}[Y(1)]\right| \geq \frac{t}{2}\right) + \mathbb{P}\left(\left|\frac{2}{n} \sum_{i \in C} Y_i^{obs} - \mathbb{E}[Y(0)]\right| \geq \frac{t}{2}\right)$$

$$\leq 4 \exp\left(-\frac{nt^2}{4}\right).$$

Setting $\delta = 4 \exp\left(-\frac{nt^2}{4}\right)$, we get $t = \sqrt{\frac{4}{n} \log\left(\frac{4}{\delta}\right)}$. Therefore, with probability at least $1 - \delta$,

$$|\hat{\tau} - \tau| = \left|\frac{2}{n} \sum_{i \in T} Y_i^{obs} - \frac{2}{n} \sum_{i \in C} Y_i^{obs} - \tau\right| \leq \sqrt{\frac{4}{n} \log\left(\frac{4}{\delta}\right)}.$$

Hoeffding is a reasonable tool here because $Y_i^{obs}$ are independent bounded random variables and we are looking for a bound of a sample mean from it’s population mean.
Problem 2 (20 points)

In this case we observe \((X_i, W_i, Y_i^{obs})\) for every \(i\) and have \(W \perp (Y(1), Y(0))|X\). Also we assume that the propensity score, \(\pi(x) = \mathbb{P}(W = 1|X = x)\) are known and that
\[
\pi_{\text{min}} \leq \pi(x) \leq \pi_{\text{max}}.
\]

Now \(\tau\) is estimated by
\[
\hat{\tau}_2 = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{Y_i^{obs} W_i}{\pi(X_i)} - \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)} \right].
\]

Notice that,
\[
\mathbb{E} \left[ \frac{Y_i^{obs} W_i}{\pi(X_i)} \right] = \mathbb{E} \left[ \frac{Y^{obs} W}{\pi(x)} \right] \quad \text{and} \quad \mathbb{E} \left[ \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)} \right] = \mathbb{E} \left[ \frac{Y^{obs} (1 - W)}{1 - \pi(x)} \right].
\]

We also have from the Lecture notes 26-6 that when \(W \perp (Y(1), Y(0))|X\),
\[
\tau = \mathbb{E} \left[ \frac{Y^{obs} W}{\pi(x)} \right] - \mathbb{E} \left[ \frac{Y^{obs} (1 - W)}{1 - \pi(x)} \right].
\]

Now to use the Hoeffding’s inequality again, we notice that for every \(i\),
\[
0 \leq \frac{Y_i^{obs} W_i}{\pi(X_i)} \leq \frac{1}{\pi_{\text{min}}} \quad \text{and} \quad 0 \leq \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)} \leq \frac{1}{1 - \pi_{\text{max}}}.
\]

Therefore using Hoeffding’s inequality we have that,
\[
\mathbb{P}\left( \left| \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i^{obs} W_i}{\pi(X_i)} - \mathbb{E} \left[ \frac{Y^{obs} W}{\pi(x)} \right] \right| \geq \frac{t}{2} \right) \leq 2 \exp \left( - \frac{nt^2 \pi_{\text{min}}^2}{2} \right),
\]

and
\[
\mathbb{P}\left( \left| \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)} - \mathbb{E} \left[ \frac{Y^{obs} (1 - W)}{1 - \pi(x)} \right] \right| \geq \frac{t}{2} \right) \leq 2 \exp \left( - \frac{nt^2 (1 - \pi_{\text{max}})^2}{2} \right).
\]

If we assign \(A = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i^{obs} W_i}{\pi(X_i)} - \mathbb{E} \left[ \frac{Y^{obs} W}{\pi(x)} \right] \) and \(B = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)} - \mathbb{E} \left[ \frac{Y^{obs} (1 - W)}{1 - \pi(x)} \right] \), then
\[
\mathbb{P}(|\hat{\tau}_2 - \tau| \geq t) \leq \mathbb{P}(|A| + |B| \geq t) \\
\leq \mathbb{P}(|A| \geq \frac{t}{2}) + \mathbb{P}(|B| \geq \frac{t}{2}) \\
\leq 2 \exp \left( - \frac{nt^2 \pi_{\text{min}}^2}{2} \right) + 2 \exp \left( - \frac{nt^2 (1 - \pi_{\text{max}})^2}{2} \right) \\
\leq 4 \exp \left( - \frac{nt^2 \min\{\pi_{\text{min}}^2, (1 - \pi_{\text{max}})^2\}}{2} \right).
\]

Setting \(\delta = 4 \exp \left( - \frac{nt^2 \min\{\pi_{\text{min}}^2, (1 - \pi_{\text{max}})^2\}}{2} \right)\), we get \(t = \sqrt{\frac{2}{n \min\{\pi_{\text{min}}^2, (1 - \pi_{\text{max}})^2\}} \log \left( \frac{4}{\delta} \right)}\). Therefore, with probability at least \(1 - \delta\),
\[
|\hat{\tau}_2 - \tau| \leq \sqrt{\frac{2}{n \min\{\pi_{\text{min}}^2, (1 - \pi_{\text{max}})^2\}} \log \left( \frac{4}{\delta} \right)}.
\]

Note that we needed \(\pi_{\text{min}}\) and \(\pi_{\text{max}}\) to bound \(\frac{Y_i^{obs} W_i}{\pi(X_i)} - \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)}\). Once we assumed that, we had that \(\frac{Y_i^{obs} W_i}{\pi(X_i)} - \frac{Y_i^{obs} (1 - W_i)}{1 - \pi(X_i)}\) were independent bounded random variables whose mean was \(\tau\). So again Hoeffding was a reasonable tool here.
Problem 3 (15 points)

We want to find a bound for $|\hat{\tau}_3 - \tau|$, but first we find a bound for $|\hat{\tau} - \tau|$ that holds with probability $1 - \delta$, where

$$\hat{\tau}_3 = \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i)) \quad \text{and} \quad \hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} (\mu_1(X_i) - \mu_0(X_i)).$$

Now for every $i$,

$$\mathbb{E} [\mu_1(X_i) - \mu_0(X_i)] = \mathbb{E}_X [\mu_1(X) - \mu_0(X)] = \tau,$$

and if we use the information that $0 \leq Y^{obs} \leq 1$, we have that $-1 \leq \mu_1(X_i) - \mu_0(X_i) \leq 1$. Then by Hoeffding’s,

$$\mathbb{P} (|\hat{\tau} - \tau| \geq t) \leq 2 \exp \left( - \frac{2t^2}{4} \right) = 2 \exp \left( - \frac{t^2}{2} \right).$$

Therefore with probability at least $1 - \delta$, we have that

$$|\hat{\tau} - \tau| \leq \sqrt{\frac{2}{n} \log \left( \frac{2}{\delta} \right)}.$$

Hence, with probability at least $1 - \delta$, we have that

$$|\hat{\tau}_3 - \tau| \leq |\hat{\tau}_3 - \hat{\tau}| + |\hat{\tau} - \tau|$$

$$\leq \left| \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_1(X_i) - \mu_1(X_i)) \right| + \left| \frac{1}{n} \sum_{i=1}^{n} (\hat{\mu}_0(X_i) - \mu_0(X_i)) \right| + |\hat{\tau} - \tau|$$

$$\leq \sup_x |\hat{\mu}_1(x) - \mu_1(x)| + \sup_x |\hat{\mu}_0(x) - \mu_0(x)| + |\hat{\tau} - \tau|$$

$$\leq 2C \left( \frac{\log n}{n} \right)^\frac{\beta}{2+\alpha} + \sqrt{\frac{2}{n} \log \left( \frac{2}{\delta} \right)}.$$

Note that now along with the $O \left( \frac{1}{\sqrt{n}} \right)$ term we have an additional $O \left( \frac{\log n}{n} \right)^\frac{\beta}{2+\alpha}$ term, which makes the rate of convergence slower than in either a trial or when the propensity scores are known.

Alternate bound: Without using the information that the data is binary, we could instead notice that we can bound $\mu_1(X_i) - \mu_0(X_i)$ by,

$$|\mu_1(X_i) - \mu_0(X_i)| \leq |\hat{\mu}_1(X_i) - \mu_1(X_i)| + |\hat{\mu}_0(X_i) - \mu_0(X_i)| + \hat{\mu}_1(X_i) + \hat{\mu}_0(X_i)$$

$$\leq \sup_x |\hat{\mu}_1(x) - \mu_1(x)| + \sup_x |\hat{\mu}_0(x) - \mu_0(x)| + 2\mu_{\max}$$

$$\leq 2C \left( \frac{\log n}{n} \right)^\frac{\beta}{2+\alpha} + 2\mu_{\max}$$

$$\leq 2 \left( C \left( \frac{\log n}{n} \right)^\frac{\beta}{2+\alpha} + \mu_{\max} \right).$$
Hence by using Hoeffding’s inequality we get,

$$\mathbb{P}(|\tilde{\tau} - \tau| \geq t) \leq 2 \exp\left(-\frac{2nt^2}{16\left(C\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+\gamma}} + \mu_{\text{max}}\right)^2}\right).$$

Therefore with probability at least $1 - \delta$, we have that

$$|\tilde{\tau} - \tau| \leq 4\left(C\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+\gamma}} + \mu_{\text{max}}\right)\sqrt{\frac{1}{2n}\log\left(\frac{2}{\delta}\right)}.$$

Then we get that with probability at least $1 - \delta$,

$$|\tilde{\tau}_3 - \tau| \leq 2C\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+\gamma}} + 4\left(C\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+\gamma}} + \mu_{\text{max}}\right)\sqrt{\frac{1}{2n}\log\left(\frac{2}{\delta}\right)}.$$

**Problem 4 (20 points)**

From problem 2 in this homework, we know that with probability at least $1 - \delta$,

$$|\tilde{\tau}_2 - \tau| \leq \sqrt{\frac{2}{n \min\{\pi_{min}^2, (1 - \pi_{max})^2\}} \log\left(\frac{4}{\delta}\right)}.$$

Now notice that,

$$|\tilde{\tau}_4 - \tilde{\tau}_2| \leq \frac{1}{n} \sum_{i=1}^{n} \left| Y_i^{\text{obs}} W_i \left( \frac{1}{\hat{\pi}(X_i)} - \frac{1}{\pi(X_i)} \right) - Y_i^{\text{obs}} \left( 1 - W_i \right) \left( \frac{1}{1 - \hat{\pi}(X_i)} - \frac{1}{1 - \pi(X_i)} \right) \right|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} \left| \frac{1}{\hat{\pi}(X_i)} - \frac{1}{\pi(X_i)} \right| + \left| \frac{1}{1 - \hat{\pi}(X_i)} - \frac{1}{1 - \pi(X_i)} \right|$$

$$\leq \sup_x |\hat{\pi}(x) - \pi(x)| \frac{1}{\pi_{min} \pi_{min}} + \sup_x |\hat{\pi}(x) - \pi(x)| \frac{1}{(1 - \pi_{max}) (1 - \pi_{max})}$$

$$\leq C\left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\beta+\gamma}} \left[ \frac{1}{\pi_{min} \pi_{min}} + \frac{1}{(1 - \pi_{max}) (1 - \pi_{max})} \right].$$

Therefore, with probability at least $1 - \delta$,

$$|\tilde{\tau}_4 - \tau| \leq |\tilde{\tau}_4 - \tilde{\tau}_2| + |\tilde{\tau}_2 - \tau|$$

$$\leq C\left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\beta+\gamma}} \left[ \frac{1}{\pi_{min} \pi_{min}} + \frac{1}{(1 - \pi_{max}) (1 - \pi_{max})} \right] + \sqrt{\frac{2}{n \min\{\pi_{min}^2, (1 - \pi_{max})^2\}} \log\left(\frac{4}{\delta}\right)}.$$

In problem 4 the rate of convergence involved a $O\left(\left(\frac{\log n}{n}\right)^{\frac{\beta}{2\beta+\gamma}}\right)$ term where as the rate of convergence now involves a $O\left(\left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\beta+\gamma}}\right)$ term. So depending on if $\beta > \gamma$, the previous rate is better than the current rate.
Problem 5 (30 points)

We have that

\[ \hat{\tau}_5 = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \frac{W_i}{\pi(X_i)} \right) (Y_{i}^{\text{obs}} - \hat{\mu}_1(X_i)) - \left( \frac{1-W_i}{1-\pi(X_i)} \right) (Y_{i}^{\text{obs}} - \hat{\mu}_0(X_i)) \right] . \]

We now show that this estimator can be re-written as:

\[ \hat{\tau}_5 = \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\mu}_1(X_i) + \frac{W_i}{\pi(X_i)} (Y_{i}^{\text{obs}} - \hat{\mu}_1(X_i)) \right] - \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\mu}_0(X_i) + \frac{1-W_i}{1-\pi(X_i)} (Y_{i}^{\text{obs}} - \hat{\mu}_0(X_i)) \right] \]

Again, splitting the last two terms gives us,

\[ \hat{\tau}_5 = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \frac{W_i}{\pi(X_i)} \right) Y_{i}^{\text{obs}} - \hat{\mu}_1(X_i) \right] + \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i) - \frac{1-W_i}{1-\pi(X_i)} \hat{\mu}_0(X_i) \]

Again, splitting the last two terms gives us,

\[ \hat{\tau}_5 = \frac{1}{n} \sum_{i=1}^{n} \left[ \left( \frac{W_i}{\pi(X_i)} \right) Y_{i}^{\text{obs}} - \hat{\mu}_1(X_i) \right] + \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i) - \frac{1-W_i}{1-\pi(X_i)} \hat{\mu}_0(X_i) \]

Therefore we can write,

\[ |\hat{\tau}_5 - \tau| \leq |\hat{\tau}_2 - \tau| + \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\mu}_1(X_i) - \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i) \right] + \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\mu}_0(X_i) - \frac{1-W_i}{1-\pi(X_i)} \hat{\mu}_0(X_i) \right] \]

Let us set,

\[ A = \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\mu}_1(X_i) - \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i) \right] , \quad B = \frac{1}{n} \sum_{i=1}^{n} \left[ \hat{\mu}_0(X_i) - \frac{1-W_i}{1-\pi(X_i)} \hat{\mu}_0(X_i) \right] , \]
Also, similarly in order to find a bound for $|A|$, note that

$$
\mathbb{E}[A] = \mathbb{E}
\left[
\hat{\mu}_1(X_i) - \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i)
\right] = \mathbb{E}
\left[
\hat{\mu}_1(X) - \frac{W}{\pi(X)} \hat{\mu}_1(X)
\right]_X
$$

$$
= \mathbb{E}
\left[
\hat{\mu}_1(X) - \frac{\mathbb{E}[W|X]}{\pi(X)} \hat{\mu}_1(X)
\right]
$$

$$
= \mathbb{E}
\left[
\hat{\mu}_1(X) - \frac{\pi(X)}{\pi(X)} \hat{\mu}_1(X)
\right] = 0.
$$

Also,

$$
\hat{\mu}_1(X_i) - \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i) = \hat{\mu}_1(X_i) \left[ 1 - \frac{W_i}{\pi(X_i)} \right] \leq \hat{\mu}_{\text{max}},
$$

and

$$
\hat{\mu}_1(X_i) - \frac{W_i}{\pi(X_i)} \hat{\mu}_1(X_i) = \hat{\mu}_1(X_i) \left[ 1 - \frac{W_i}{\pi(X_i)} \right] \geq \hat{\mu}_{\text{min}} \left[ 1 - \frac{1}{\pi_{\text{min}}} \right].
$$

Therefore, using Hoeffding’s inequality,

$$
\mathbb{P}
\left|
\mathbb{E}[A] \right| \geq \frac{t}{5}
\right) \leq 2 \exp\left(-\frac{2nt^2\pi_{\text{min}}^2}{25\hat{\mu}_{\text{max}}^2}\right).
$$

Similarly in order to find a bound for $|B|$, note that similar to the previous case,

$$
\mathbb{E}[B] = \mathbb{E}
\left[
\hat{\mu}_0(X_i) - \frac{1-W_i}{1-\pi(X_i)} \hat{\mu}_0(X_i)
\right] = 0.
$$

Also,

$$
\hat{\mu}_{\text{max}} \left[ 1 - \frac{1}{1-\pi_{\text{max}}} \right] \leq \hat{\mu}_0(X_i) \left[ 1 - \frac{1-W_i}{1-\pi(X_i)} \right] \leq \hat{\mu}_{\text{max}}.
$$

Therefore, using Hoeffding’s inequality,

$$
\mathbb{P}
\left|
\mathbb{E}[B] \right| \geq \frac{t}{5}
\right) \leq 2 \exp\left(-\frac{2nt^2(1-\pi_{\text{max}})^2}{25\hat{\mu}_{\text{max}}^2}\right).
$$
In order to find a bound for $|D|$, note that

$$E[D] = E \left[ W_i (Y_i^{\text{obs}} - \mu_1(X_i)) \left( \frac{1}{\pi(X_i)} - \frac{1}{\pi(X_i)} \right) \right] = E \left[ E \left[ W_i (Y_i^{\text{obs}} - \mu_1(X_i)) \left( \frac{1}{\pi(X_i)} - \frac{1}{\pi(X_i)} \right) \right| X_i \right] = E[0] = 0,$$

and we can find the bound for every $i$ as:

$$\left| W_i (Y_i^{\text{obs}} - \mu_1(X_i)) \left( \frac{1}{\pi(X_i)} - \frac{1}{\pi(X_i)} \right) \right| \leq \frac{1}{\hat{\pi}_{\min} \pi_{\min}}.$$

Therefore by Hoeffding’s inequality,

$$P \left( |D| \geq \frac{t}{5} \right) \leq 2 \exp \left( -\frac{2nt^2 \hat{\pi}_{\min}^2 \pi_{\min}^2}{100} \right).$$

Similarly following the arguments for finding an inequality for $|D|$ we can find a similar inequality for $|E|$ as,

$$P \left( |E| \geq \frac{t}{5} \right) \leq 2 \exp \left( -\frac{2nt^2 (1 - \hat{\pi}_{\max})^2 (1 - \pi_{\max})^2}{100} \right).$$

Now following the derivations of problem 2,

$$P \left( |\hat{\tau} - \tau| \geq \frac{t}{5} \right) \leq 4 \exp \left( -\frac{2nt^2 \min \{\hat{\pi}_{\min}^2, (1 - \pi_{\max})^2\}}{100} \right).$$

Hence,

$$P \left( |\hat{\tau} - \tau| + |A| + |B| + |D| + |E| \geq t \right) \leq P \left( |\hat{\tau} - \tau| \geq \frac{t}{5} \right) + P \left( |A| \geq \frac{t}{5} \right) + P \left( |B| \geq \frac{t}{5} \right) + P \left( |D| \geq \frac{t}{5} \right) + P \left( |E| \geq \frac{t}{5} \right) \leq 4 \exp \left( -\frac{2nt^2 \min \{\hat{\pi}_{\min}^2, (1 - \pi_{\max})^2\}}{100} \right) + 2 \exp \left( -\frac{2nt^2 \pi_{\min}^2}{25 \hat{\pi}_{\max}^2} \right) + 2 \exp \left( -\frac{2nt^2 (1 - \pi_{\max})^2}{25 \hat{\pi}_{\max}^2} \right) + 2 \exp \left( -\frac{2nt^2 (1 - \hat{\pi}_{\max})^2}{25 \pi_{\min}^2} \right) \leq 12 \exp \left( -\frac{2nt^2 \min \{\hat{\pi}_{\min}^2, (1 - \pi_{\max})^2\}}{100} \right).$$

Since, $Y_i^{\text{obs}} \leq 1, 25 \hat{\pi}_{\max}^2 \leq 100$. Also $\pi_{\min}^2 \geq \hat{\pi}_{\min}^2 \pi_{\min}^2$ and $(1 - \pi_{\max})^2 \geq (1 - \hat{\pi}_{\max})^2 (1 - \pi_{\max})^2$. Hence,

$$P \left( |\hat{\tau} - \tau| + |A| + |B| + |D| + |E| \geq t \right) \leq 12 \exp \left( -\frac{2nt^2 \min \{\hat{\pi}_{\min}^2, \pi_{\min}^2, (1 - \pi_{\max})^2\}}{100} \right).$$

Therefore with probability at least $1 - \delta$,

$$|\hat{\tau} - \tau| + |A| + |B| + |D| + |E| \leq \sqrt{n \min \{\hat{\pi}_{\min}^2, \pi_{\min}^2, (1 - \pi_{\max})^2\}} \log \left( \frac{50}{\delta} \right).$$
Now we bound $|F|$ and $|G|$. Let us first look at $|F|$.

$$|F| = \left| \frac{1}{n} \sum_{i=1}^{n} W_i (\mu_1(X_i) - \hat{\mu}_1(X_i)) \left( \frac{1}{\hat{\pi}(X_i)} - \frac{1}{\pi(X_i)} \right) \right|$$

$$\leq \frac{1}{n} \sum_{i=1}^{n} |W_i (\mu_1(X_i) - \hat{\mu}_1(X_i))| \left| \frac{1}{\hat{\pi}(X_i)} - \frac{1}{\pi(X_i)} \right| .$$

Now, for every $i$,

$$|W_i (\mu_1(X_i) - \hat{\mu}_1(X_i))| \leq |\mu_1(X_i) - \hat{\mu}_1(X_i)|$$

$$\leq C \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} .$$

Therefore,

$$|F| \leq C \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \sup_x |\pi(X_i) - \hat{\pi}(X_i)|$$

$$\leq C \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \left( \log n \right)^{\frac{\gamma}{\gamma + d}} \frac{1}{\hat{\pi}_\min \pi_\min} .$$

Similarly we can get a bound for $|G|$ as,

$$|G| = \left| \frac{1}{n} \sum_{i=1}^{n} (1 - W_i) (Y_i^\text{obs} - \hat{\mu}_0(X_i)) \left( \frac{1}{1 - \hat{\pi}(X_i)} - \frac{1}{1 - \pi(X_i)} \right) \right|$$

$$\leq C \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \sup_x |\pi(X_i) - \hat{\pi}(X_i)|$$

$$\leq C \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \left( \log n \right)^{\frac{\gamma}{\gamma + d}} \frac{1}{(1 - \hat{\pi}_\max)(1 - \pi_\max)} .$$

Therefore with probability at least $1 - \delta$, we have

$$|\hat{\tau}_5 - \tau| \leq |\hat{\tau}_2 - \tau| + |A| + |B| + |D| + |E| + |F| + |G|$$

$$\leq \sqrt{\frac{50}{n \min \left\{ \hat{\pi}_\min \pi_\min, (1 - \hat{\pi}_\max)^2 (1 - \pi_\max)^2 \right\} \log \left( \frac{12}{\delta} \right)}}$$

$$+ C \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \left( \log n \right)^{\frac{\gamma}{\gamma + d}} \left[ \frac{1}{\hat{\pi}_\min \pi_\min} + \frac{1}{(1 - \hat{\pi}_\max)(1 - \pi_\max)} \right] .$$

This estimator is doubly robust because the estimator is consistent if either your estimate for the regression functions or of the propensity score is consistent (i.e. if either $\beta > 0$ or $\gamma > 0$). Furthermore, when both estimators are consistent (i.e. both $\beta > 0$ and $\gamma > 0$) the estimator dominates both $\hat{\tau}_3$ and $\hat{\tau}_4$ because,

$$\left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \left( \log n \right)^{\frac{\gamma}{\gamma + d}} \leq \left( \frac{\log n}{n} \right)^{\frac{\beta}{2\beta + d}} \text{ and } \left( \frac{\log n}{n} \right)^{\frac{\gamma}{\gamma + d}} .$$

So the rate of convergence is better than both the previous rates of convergence.
Alternate Solution: As seen before we can write \(|\hat{\tau}_2 - \tau|\) as:

\[
|\hat{\tau}_2 - \tau| \leq |\hat{\tau}_2 - \tau| + |A - B + D - E| + |F| + |G|
\]

where \(A, B, C, D, E, F\) and \(G\) are as defined before. Also notice in the previous solution that,

\[
\]

Therefore if we can bound \(A - B + D - E\) then we can find the Hoeffding’s bound for them together. Putting the bounds we found earlier together, we get

\[
A - B + D - E \geq \hat{\mu}_{\max} \left[1 - \frac{1}{\pi_{\min}}\right] - \hat{\mu}_{\max} - \frac{1}{\pi_{\min} \pi_{\min}} - \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})}
\]

\[
= -\frac{1}{\pi_{\min}} - \frac{1}{\pi_{\min} \pi_{\min}} - \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})} = a \text{ (say)},
\]

\[
A - B + D - E \leq \hat{\mu}_{\max} - \hat{\mu}_{\max} \left[1 - \frac{1}{1 - \pi_{\max}}\right] + \frac{1}{\pi_{\min} \pi_{\min}} + \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})}
\]

\[
= \frac{1}{1 - \pi_{\max}} + \frac{1}{\pi_{\min} \pi_{\min}} + \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})} = b \text{ (say)}.
\]

Then,

\[
b - a = \frac{1}{1 - \pi_{\max}} + \frac{1}{\pi_{\min} \pi_{\min}} + \frac{1}{3} \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})} + \frac{1}{\pi_{\min} \pi_{\min}} + \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})}
\]

\[
\leq \frac{1}{\pi_{\min} \pi_{\min}} + \frac{1}{3} \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})} \leq \frac{1}{6} \min \{\hat{\pi}_{\min} \pi_{\min}, (1 - \hat{\pi}_{\max})(1 - \pi_{\max})\}
\]

Therefore, using Hoeffding’s inequality,

\[
P\left(|A - B + D - E| \geq \frac{t}{2}\right) \leq 2 \exp \left(-\frac{2nt^2 \min \{\hat{\pi}_{\min}^2, (1 - \hat{\pi}_{\max})^2\}}{36 \times 4}\right).
\]

Now following the derivations of problem 2,

\[
P\left(|\hat{\tau}_2 - \tau| \geq \frac{t}{2}\right) \leq 4 \exp \left(-\frac{2nt^2 \min \{\hat{\pi}_{\min}^2 (1 - \pi_{\max})^2\}}{16}\right).
\]

Hence,

\[
P\left(|\hat{\tau}_2 - \tau| + |A - B + D - E| \geq t\right) \leq P\left(|\hat{\tau}_2 - \tau| \geq \frac{t}{2}\right) + P\left(|A - B + D - E| \geq \frac{t}{2}\right)
\]

\[
\leq 4 \exp \left(-\frac{2nt^2 \min \{\hat{\pi}_{\min}^2 (1 - \pi_{\max})^2\}}{100}\right)
\]

\[
+ 2 \exp \left(-\frac{2nt^2 \min \{\hat{\pi}_{\min}^2, (1 - \hat{\pi}_{\max})^2 (1 - \pi_{\max})^2\}}{36 \times 4}\right)
\]

\[
\leq 6 \exp \left(-\frac{nt^2 \min \{\hat{\pi}_{\min}^2, (1 - \hat{\pi}_{\max})^2 (1 - \pi_{\max})^2\}}{72}\right).
\]
Therefore with probability at least $1 - \delta$,

$$|\hat{\tau}_2 - \tau| + |A - B + D - E| \leq \sqrt{\frac{72}{n \min\{\hat{\pi}_2\hat{\pi}_2, (1 - \hat{\pi}_{\max})^2 (1 - \pi_{\max})^2\}} \log \left(\frac{6}{\delta}\right)}.$$

Adding with this the bounds for $|F|$ and $|G|$ as derived in the previous solution, we get with probability at least $1 - \delta$,

$$|\hat{\tau}_5 - \tau| \leq |\hat{\tau}_2 - \tau| + |A - B + D - E| + |F| + |G| \leq \sqrt{\frac{72}{n \min\{\hat{\pi}_2\hat{\pi}_2, (1 - \hat{\pi}_{\max})^2 (1 - \pi_{\max})^2\}} \log \left(\frac{6}{\delta}\right)}$$

$$+ C^2 \left(\frac{\log n}{n}\right)^{\frac{\beta}{2\gamma + 2}} \left(\frac{\log n}{n}\right)^{\frac{\gamma}{2\gamma + 2}} \left[\frac{1}{\hat{\pi}_{\min} \pi_{\min}} + \frac{1}{(1 - \hat{\pi}_{\max})(1 - \pi_{\max})}\right].$$