This contains solutions for Homework 10. Please note that we have included several additional comments and approaches to the problems to give you better insight.

**Problem 1.** Let \((X_1, Y_1, Z_1), \ldots, (X_n, Y_n, Z_n) \sim P.\) Assume that \(X_i \in \{0, 1\}\) and let \(\theta = \mathbb{E}[Y_1] - \mathbb{E}[Y_0]\) denote the causal effect. Assume there are no unmeasured confounders, that is, \(X \perp (Y_0, Y_1) \mid Z.\) Let
\[
\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(0, Z_i)
\]
where \(\hat{\mu}(x, z)\) is an estimate of \(\mu(x, z)\). Suppose that
\[
\sup_{x, z} |\hat{\mu}(x, z) - \mu(x, z)| \overset{P}{\to} 0.
\]
Show that \(\hat{\theta} \overset{P}{\to} \theta.\)

**Solution 1.** By the definition of converge in probability, it’s sufficient to show
\[
\forall \varepsilon > 0, \lim_{n \to \infty} \mathbb{P}\left( |\hat{\theta} - \theta| \leq \varepsilon \right) = 1
\]
Given
\[
\sup_{x, z} |\hat{\mu}(x, z) - \mu(x, z)| \overset{P}{\to} 0
\]
That is,
\[
\forall \varepsilon' > 0, \lim_{n \to \infty} \mathbb{P}\left( \sup_{x, z} |\hat{\mu}(x, z) - \mu(x, z)| \leq \varepsilon' \right) = 1
\]
(1)

Notice
\[
|\hat{\theta} - \theta| = \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(1, z) p(z) dz - \int \mu(0, z) p(z) dz \right|
\]
\[
\leq \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(1, Z_i) - \int \mu(1, z) p(z) dz \right| + \left| \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}(0, Z_i) - \int \mu(0, z) p(z) dz \right|
\]
\[
\leq \int \left| \frac{1}{n} \sum_{i=1}^{n} \mu(1, z) - \hat{\mu}(1, Z_i) \right| p(z) dz + \int \left| \frac{1}{n} \sum_{i=1}^{n} \mu(0, z) - \hat{\mu}(0, Z_i) \right| p(z) dz
\]
\[
\leq 2 \int \sup_{x, z} |\hat{\mu}(x, z) - \mu(x, z)| p(z) dz
\]
\[
= 2 \sup_{x, z} |\hat{\mu}(x, z) - \mu(x, z)|
\]
Thus

\[
\sup_{x,z} |\hat{\mu}(x,z) - \mu(x,z)| \leq \frac{\epsilon}{2} \Rightarrow |\hat{\theta} - \theta| \leq \epsilon
\]

which implies

\[
P\left( \sup_{x,z} |\hat{\mu}(x,z) - \mu(x,z)| \leq \frac{\epsilon}{2} \right) \leq P\left( |\hat{\theta} - \theta| \leq \epsilon \right)
\]

By (1), let \( \epsilon' = \frac{\epsilon}{2} \),

\[
\lim_{n \to \infty} P\left( \sup_{x,z} |\hat{\mu}(x,z) - \mu(x,z)| \leq \frac{\epsilon}{2} \right) = 1
\]

Therefore

\[
\lim_{n \to \infty} P\left( |\hat{\theta} - \theta| \leq \epsilon \right) = 1
\]
**Problem 2.** Let \((X,Y)\) be such that \(P(Y = 1) = P(Y = 0) = 1/2\), \(X|Y = 0 \sim \text{Uniform}(-3,1)\) and \(X|Y = 1 \sim \text{Uniform}(-1,3)\).

(a) Find \(m(x) = P(Y = 1|X = x)\)

(b) Find the Bayes classification rule \(h_*(x)\)

(c) Find the risk of the Bayes classifier \(P(Y \neq h_*(X))\)

**Solution 2.** Let \(p_0(x) \sim \text{Unif}(-3,1)\) and \(p_1(x) \sim \text{Unif}(-1,3)\).

(a) Proof.

\[
m(x) = \frac{P(Y = 1|X = x)}{P(X = x)}
= \frac{P(X = x|Y = 1) P(Y = 1)}{P(X = x)}
= \frac{P(X = x|Y = 1) P(Y = 1)}{P(X = x|Y = 1) P(Y = 1) + P(X = x|Y = 0) P(Y = 0) + P(X = x|Y = 1) P(Y = 1)}
= \frac{\frac{1}{2} p_{X|Y}(x|Y = 1)}{\frac{1}{2} p_{X|Y}(x|Y = 0) + \frac{1}{2} p_{X|Y}(x|Y = 1) + \frac{1}{2} p_{X|Y}(x|Y = 1)}
= \frac{p_1(x)}{p_0(x) + p_1(x)}
\]

\[
= \begin{cases} 
  \frac{1}{2} & \text{if } x \in (-1, 1] \\
  1 & \text{if } x \in (1, 3) \\
  0 & \text{otherwise}
\end{cases}
\]

(b) In this case we note that the Bayes classification rule is:

\[
h_*(x) = 1 \left( m(x) \geq \frac{1}{2} \right)
= 1(x \in (-1, 3))
\]
(c) We can calculate the risk of the Bayes classifier as follows:

\[ R(h_*) = \mathbb{P}_{X,Y}(Y \neq h_*(X)) \]
\[ = \mathbb{P}_{X,Y}(Y \neq h_*(X)|Y = 0)\mathbb{P}(Y = 0) + \mathbb{P}_{X,Y}(Y \neq h_*(X)|Y = 1)\mathbb{P}(Y = 1) \]
\[ = \mathbb{P}_{X,Y}(h_*(X) = 1|Y = 0)\mathbb{P}(Y = 0) + \mathbb{P}_{X,Y}(h_*(X) = 0|Y = 1)\mathbb{P}(Y = 1) \]
\[ = \mathbb{P}_{X,Y}(X \in (-1,3)|Y = 0)\mathbb{P}(Y = 0) + \mathbb{P}_{X,Y}(X \notin (-1,3)|Y = 1)\mathbb{P}(Y = 1) \]
\[ = \frac{1}{2} \int_{-1}^{3} p_0(x) dx + \frac{1}{2}(1 - \int_{-1}^{3} p_1(x) dx) \]
\[ = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2}(1 - 1) \]
\[ = \frac{1}{4} \]
Problem 3. Let $\mathcal{A}$ consist of all sets of the form $A = [a, b] \cup [c, d]$ where $a \leq b \leq c \leq d$. Find the VC dimension of $\mathcal{A}$.

Solution 3. The VC dimension of $\mathcal{A}$ is 4.

Let $F = \{x_1, x_2, x_3, x_4\}$ be an increasingly ordered set. $\mathcal{A}$ can pick up all its subsets.

1. $\emptyset$: Let $a > x_4$, $A \cap F = \emptyset$.
2. 1-element subsets: Let $a = b = x_i$, $c > x_4$. Then $A \cap F = \{x_i\}$.
3. 2-element subsets: Let $a = b = x_i$, $c = d = x_j$. Then $A \cap F = \{x_i, x_j\}$.
4. 3-element subsets:
   \[
   \begin{cases}
   \{x_1, x_2, x_3\} : & a = x_1, b = x_2, c > x_4. \\
   \{x_1, x_2, x_4\} : & a = x_1, b = x_2, c = d = x_4. \\
   \{x_1, x_3, x_4\} : & a = b = x_1, c = x_3, d = x_4. \\
   \{x_2, x_3, x_4\} : & a = x_2, b = x_4, c > x_4.
   \end{cases}
   \]
5. 4-element subsets: Let $a = x_1, b = x_4$.

However, if $|F| = 5$, $F$ cannot be shattered by $\mathcal{A}$. Let $F = \{x_1, x_2, x_3, x_4, x_5\}$ be an increasingly ordered set. There is no $A \in \mathcal{A}$ such that
\[
A \cap F = \{x_1, x_3, x_5\}
\]

Thus $d(\mathcal{A}) = 4$. 


Problem 4. Suppose that \((Y, X)\) are random variables where \(Y \in \{0, 1\}\) and \(X \in \mathbb{R}\). Suppose that

\[
X|Y = 0 \sim \text{Normal}(0,1)
\]

and that

\[
X|Y = 1 \sim \text{Normal}(2,1).
\]

Further suppose that \(P(Y = 0) = P(Y = 1) = 1/2\).

(a) Find \(m(x) = P(Y = 1|X = x)\).

(b) Let \(A = A_1 \cup A_2\) where \(A_1 = \{(\infty, a) : a \in \mathbb{R}\}\) and \(A_2 = \{(b, \infty) : b \in \mathbb{R}\}\). Find the VC dimension of \(A\).

(c) Let \(H = \{h_A : A \in \mathcal{A}\}\) where \(h_A(x) = 1\) if \(x \in A\) and \(h_A(x) = 0\) if \(x \notin A\). Show that the Bayes rule \(h_*\) is in \(H\).

Solution 4.(a)

\[
P(Y = 1|X = x) = \frac{P(Y = 1, X = x)}{P(X = x)} = \frac{P(X|Y = 1)P(Y = 1)}{P(X|Y = 0)P(Y = 0) + P(X|Y = 1)P(Y = 1)}
\]

\[=
\frac{\frac{1}{2} \exp\left\{-\frac{(x-2)^2}{2}\right\}}{\frac{1}{2} \exp\left\{-\frac{(x-2)^2}{2}\right\} + \frac{1}{2} \exp\left\{-\frac{x^2}{2}\right\}}
\]

\[= \frac{1}{1 + \exp\{4(1-x)\}}
\]

(b) The VC dimension of \(A\) is 2.

Let \(F = \{x_1, x_2\}\) be an increasingly ordered set. \(A\) can pick up all its subsets. Let \(A = (\infty, a) \cup (b, \infty)\).

(a) \(\emptyset\): Let \(a < x_1, b > x_2\), \(A \cap F = \emptyset\).

(b) 1-element subsets:

\[
\begin{cases}
\{x_1\} : & x_1 < a < x_2, b > x_2. \\
\{x_2\} & : a < x_1, x_1 < b < x_2.
\end{cases}
\]

(c) 2-element subsets: Let \(a > x_2\).

However, if \(|F| = 3\), \(F\) cannot be shattered by \(A\). Let \(F = \{x_1, x_2, x_3\}\) be an increasingly ordered set. There is no \(A \in \mathcal{A}\) such that

\[
A \cap F = \{x_2\}
\]

Thus \(d(A) = 2\).

(c) The Bayes rule is

\[
h_*(x) = I(m(x) \geq \frac{1}{2})
\]
Since \( m(x) = \frac{1}{1 + \exp\left(\frac{x}{4(1-x)}\right)} \) is an increasing function of \( x \), we have

\[
\{ x : m(x) \geq \frac{1}{2} \} = \{ x : x \geq m^{-1}(\frac{1}{2}) \} = [1, \infty] \in \mathcal{A}
\]

Thus

\[
h_*(x) = h_{[1,\infty]} \in \mathcal{H}
\]
Problem 5. Let $\hat{m}_h(x) = \sum_i Y_i w_i(x)$ be the kernel regression estimator.  
Prove that 
\[
\frac{1}{n} \sum_i (Y_i - \hat{m}_{h(-i)}(X_i))^2 = \frac{1}{n} \sum_i \left( \frac{Y_i - \hat{m}_h(X_i)}{1 - S_{ii}} \right)^2
\]
where $S_{ij} = w_j(X_j)$.

Solution 5. Since, 
\[
\hat{m}_{h(-i)}(x) = \frac{\sum_{j \neq i} Y_j K_h(x - X_j)}{\sum_{j \neq i} K_h(x - X_j)} = \frac{\sum_{j} Y_j K_h(x - X_j) - Y_i K_h(x - X_i)}{\sum_{j} K_h(x - X_j)} 
\]
\[
= \frac{\sum_{j} Y_j K_h(x - X_j) - Y_i K_h(x - X_i)}{\sum_{j} K_h(x - X_j)} 
\]
\[
= \frac{\sum_{j} Y_j w_j(x) - Y_i w_i(x)}{1 - w_i(x)} 
\]
\[
= \frac{\hat{m}_h(x) - Y_i w_i(x)}{1 - w_i(x)} 
\]

Hence, 
\[
\frac{1}{n} \sum_i (Y_i - \hat{m}_{h(-i)}(X_i))^2 = \frac{1}{n} \sum_i \left( \frac{\hat{m}_h(X_i) - Y_i w_i(X_i)}{1 - w_i(X_i)} \right)^2 
\]
\[
= \frac{1}{n} \sum_i \left( \frac{Y_i - S_{ii}}{1 - S_{ii}} - \frac{\hat{m}_h(X_i) - Y_i S_{ii}}{1 - S_{ii}} \right)^2 
\]
\[
= \frac{1}{n} \sum_i \left( \frac{Y_i - \hat{m}_h(X_i)}{1 - S_{ii}} \right)^2 
\]