Interpretation of Unbiasedness:

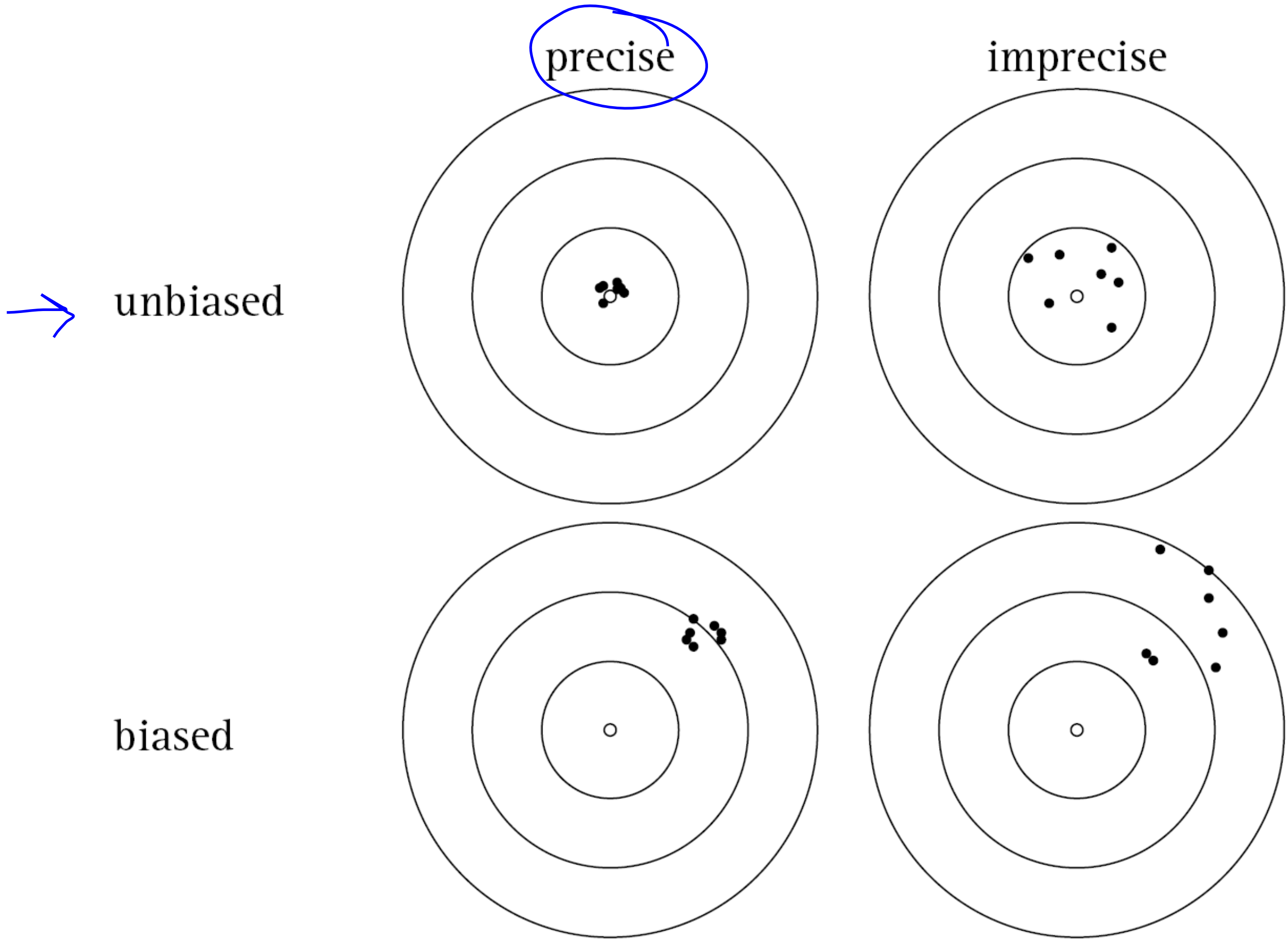


Figure 1: Top left: Unbiased estimator with small variance; Top right: Unbiased estimator with large variance; Bottom left: Biased with small variance; Bottom right: Biased with large variance

We have stated that mean squared error (MSE) can be decomposed into a variance and bias term:

$$\mathrm{MSE}_\theta(T_n) = \mathbb{E}\big[(T_n - \theta)^2\big] = \mathrm{Var}_\theta(T_n) + (b_n(\theta))^2$$

where $b_n(\theta) = \mathbb{E}(T_n) - \theta$. With this formula and the figure above it makes sense to look for an unbiased estimator with the smallest possible variance in the class of unbiased estimators.

Corollary for *i.i.d.* case

Proof:

Since $X_1, X_2, \ldots, X_n$ are *i.i.d.*

$$I_X(\theta) = \quad \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{x};\theta)\right)^2\right] = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^{n} f(x_i;\theta)\right)^2\right]$$

$$(a+b)^2 = a^2+b^2+2ab$$

$$= \mathbb{E}_\theta\left[\left(\sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)^2\right]$$

$$= \sum_{i=1}^{n}\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)^2\right] \; nI_{X_1}(\theta)$$

$$+ \sum_{i\neq j}\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)\left(\frac{\partial}{\partial\theta}\log f(x_j;\theta)\right)\right]$$

Continuing with the second term for $i \neq j$:

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_i;\theta)\right)\right]\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_j;\theta)\right)\right] = 0\times 0 = 0 \qquad \text{by independence.}$$

Therefore,

$$I_{\mathbf{X}}(\theta) = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(\mathbf{x};\theta)\right)^2\right] = n\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_1;\theta)\right)^2\right] = nI_{X_1}(\theta)$$

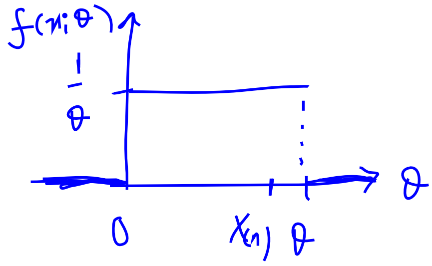
which gives the result. □

Note: this result tells us that the Fisher-information number gets bigger has the sample size increases and we have more information about θ, and therefore we have a smaller bound on the variance of the best unbiased estimator.

Example 4.30

This example illustrates that when the condition (*) is violated, we could have estimators which are unbiased and have a variance less than the Cramer Rao lower bound.

Let $X_1, X_2, \ldots, X_n$ be *i.i.d.* Uniform$(0, \theta)$ random variables so that the support of the density depends on θ and the condition (*) is violated. Here



$$f(x, \theta) = \frac{1}{\theta} I_{(0,\theta)}(x)$$

and

$X_{(n)} = \max(X_1, \ldots, X_n)$

$$F(x;\theta) = \begin{cases} 0 & \text{if} \quad x \le 0 \\ \frac{x}{\theta} & \text{if} \quad 0 < x < \theta \\ 1 & \text{if} \quad x \ge \theta \end{cases}$$

Then

$$\begin{aligned} F_{X_{(n)}}(y) = P(X_{(n)} \le y) &= P(X_1 \le y, X_2 \le y, \ldots, X_n \le y) = P(X_1 < y) \cdot P(X_2 < y) \cdots P(X_n < y) \\ &= \left[P(X_1 \le y)\right]^n \\ &= \left(\frac{y}{\theta}\right)^n \end{aligned}$$

Therefore,

$$f_{X_{(n)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_{X_{(n)}}(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{y}{\theta}\right)^n = \frac{ny^{n-1}}{\theta^n}, \quad 0 < y < \theta$$

and zero elsewhere.

Consider

$$\mathbb{E}(X_{(n)}) = \int_0^\theta y \cdot \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+1} \frac{y^{n+1}}{\theta^n}\Big|_0^\theta = \frac{n}{n+1}\theta \neq \theta$$

which implies that $X_{(n)}$ is biased as an estimator for θ. However,

$$T = \frac{n+1}{n} X_{(n)},$$

$\frac{1+1}{1} = 2$ $\qquad$ $\frac{101}{10} = 1.1$

turns out be unbiased for estimating θ since

$$\mathbb{E}_\theta(T) = \frac{n+1}{n} \cdot \mathbb{E}_\theta(X_{(n)}) = \frac{\cancel{n+1}}{\cancel{n}} \cdot \frac{\cancel{n}}{\cancel{n+1}}\theta = \theta$$

$n \to \infty \quad = 1$

Now

$T = \frac{n+1}{n} X_{(n)}$

$$\begin{aligned} \mathrm{Var}(T) &= \mathbb{E}(T^2) - \left[\mathbb{E}(T)\right]^2 \\ &= \left(\frac{n+1}{n}\right)^2 \int_0^\theta y^2 \cdot \frac{ny^{n-1}}{\theta^n} \mathrm{d}y - \theta^2 \\ &= \left(\frac{n+1}{n}\right)^2 \left[\frac{n}{n+2} \frac{y^{n+2}}{\theta^n}\right]_0^\theta - \theta^2 \\ &= \frac{(n+1)^2}{n(n+2)}\theta^2 - \theta^2 \times \frac{n(n+2)}{n(n+2)} \end{aligned}$$

$$= \theta^2 \frac{\cancel{n^2} + \cancel{2n} + 1 - \cancel{n^2} - \cancel{2n}}{n(n+2)} = \frac{\theta^2}{n(n+2)}$$

After a bit more algebra we get:

$$\text{Var}(T) = \frac{\theta^2}{n(n+2)}$$

However, for $f(x;\theta) = \frac{1}{\theta}$ for $0 < x < \theta$ we have $\log f(x;\theta) = -\log\theta$,

$$\frac{\partial \log f(x;\theta)}{\partial \theta} = -\frac{1}{\theta}$$

Therefore, the information in the sample is:

$$I_{\mathbf{X}_1}(\theta) \overset{\text{def}}{=} \mathbb{E}_\theta\left[\left(\frac{\partial \log f(x;\theta)}{\partial\theta}\right)^2\right] = \mathbb{E}_\theta\left[\left(-\frac{1}{\theta}\right)^2\right] = \frac{1}{\theta^2}.$$

By recklessly applying the Cramer Rao lower bound we obtain:

$$\text{CRLB} = \frac{\theta^2}{n} \qquad \frac{(\tau'(\theta))^2}{n \cdot I_{X_1}(\theta)} = \frac{1^2}{n \cdot 1/\theta^2} = \frac{\theta^2}{n}$$

However as we now see

$$\text{Var}(T) = \underline{\frac{\theta^2}{n(n+2)}} < \frac{\theta^2}{n}. := \text{CRLB}$$

Reason: the contradiction occurs as the condition (*) was violated in this example since the density depends on the support of the parameter θ.

Proof of the form $V(X,\theta) = k_n(\theta)(W(X) - \tau(\theta))$:

For the Cramer Rao lower bound to be attainable we require the score to be in the form

$$V(X,\theta) = k_n(\theta)(W(X) - \tau(\theta))$$

To see why, we note that the Cramer-Rao inequality in (6) can be written as

$$\left[\mathrm{Cov}_\theta\Big(W(X), \frac{\partial}{\partial\theta}\log L(x,\theta)\Big)\right]^2 \leq \mathrm{Var}_\theta(W(X))\mathrm{Var}_\theta\Big(\frac{\partial}{\partial\theta}\log L(x,\theta)\Big)$$

Then recall that $\mathbb{E}_\theta(W(X)) = \tau(\theta)$ and $\mathbb{E}_\theta\left(\frac{\partial}{\partial\theta}\log L(X,\theta)\right) = \mathbb{E}_\theta(V(X,\theta)) = 0$.

We have equality when (if and only if) $W(X) - \tau(\theta)$ is proportional to $\frac{\partial}{\partial\theta}\log L(X,\theta) - 0$. That is,

$$k_n(\theta)\big[W(X) - \tau(\theta)\big] = \frac{\partial}{\partial\theta}\log L(X,\theta) - 0 = V(X,\theta)$$

and the form is complete. $\square$

Theorem 4.14

$Cov(V, W) = Cov(K_n(\theta)(W - \tau), W)$

$= Cov(K_n(\theta)W, W) + 0$

$= K_n(\theta) Cov(W, W)$

$= K_n(\theta) Var(W)$

Proof:

For attainability we have

$$V(X,\theta) = k_n(\theta)\big[W(X) - \tau(\theta)\big]$$

Then we have:

$$\mathrm{Cov}^2(V(X,\theta), W(X)) = (k_n(\theta))^2(\mathrm{Var}(W))^2$$

By rearranging (7) knowing that we have attainability gives:

$$\mathrm{Cov}^2(W,V) = \left(\frac{\partial}{\partial\theta}\tau(\theta)\right)^2 = \mathrm{Var}(W)I_{\mathbf{X}}(\theta)$$

Since all variances are not equal to zero it means that $k_n(\theta) \neq 0$ for any θ.

However, the MLE for $\widehat{\tau(\theta)}_{\text{mle}} = \tau(\hat{\theta}_{\text{mle}})$ and for the MLE $\hat{\theta}_{\text{mle}}$ we have:

$$\begin{aligned} 0 = V(X, \hat{\theta}_{\text{mle}}) &= k_n(\hat{\theta}_{\text{mle}})(W - \tau(\hat{\theta}_{\text{mle}})) \\ &= \underbrace{k_n(\hat{\theta}_{\text{mle}})}_{\neq 0}\underbrace{(W - \widehat{\tau(\theta)}_{\text{mle}})}_{=0} \end{aligned}$$

Since we know that $k_n(\hat{\theta}_{\text{mle}}) \neq 0$ it must hold that:

$$W(X) = \widehat{\tau(\theta)}_{\text{mle}}.$$

$\square$

Note: $\widehat{\tau(\theta)}_{\text{mle}} = \tau(\hat{\theta}_{\text{mle}})$ is the transformation invariance property of the MLE. It is not true in general for the UMVUE discussed next. You can see this with the Poisson example

$$\hat{\theta}_{\text{umvue}} = \bar{X}$$

$\tau(\theta) = e^{-\theta}$

but

$$\widehat{\tau(\theta)}_{\text{umvue}} = \left(1 - \frac{1}{n}\right)^{n\bar{X}} \neq e^{-\bar{X}} = \tau(\hat{\theta}_{\text{umvue}}).$$

$\hat{\tau}(T) = E(W|T)$

W - unbiased

T - sufficient

Theorem 4.15 (Rao-Blackwell Theorem)

Proof:

(i) First,

$$\mathbb{E}_\theta\big[\hat{\tau}(T)\big] = \mathbb{E}_\theta\big[\mathbb{E}(W|T)\big] = \mathbb{E}_\theta\big[W\big] = \tau(\theta)$$

holds by applying the "iterative" property of expected values. Hence, $\hat{\tau}(T)$ is an unbiased estimator for $\tau(\theta)$.

(ii) Second, we will now show that "always" $\boxed{\mathrm{Var}(Y|X) \leq \mathrm{Var}(Y)}$. That is, the variance is never increased after conditioning. Let $a(X) = \mathbb{E}(Y|X)$. Then

$$\begin{aligned}
\mathrm{Var}(Y) &= \mathbb{E}\big[(Y - \mathbb{E}(Y))^2\big] \\
&= \mathbb{E}\big[(Y - a(X) + a(X) - \mathbb{E}(Y))^2\big] \\
&= \mathbb{E}\big[(Y - a(X))^2\big] + \mathbb{E}\big[(a(X) - \mathbb{E}(Y))^2\big] + \underbrace{\cancel{2\mathbb{E}\big[(Y - a(X))(a(X) - \mathbb{E}(Y))\big]}}_{=0}
\end{aligned}$$

By considering the latter term:

$$\begin{aligned}
\mathbb{E}\big[(Y - a(X))(a(X) - \mathbb{E}(Y))\big] &= \mathbb{E}_X\Big[\mathbb{E}\big[(Y - a(X))(a(X) - \mathbb{E}(Y))|X\big]\Big] \\
&= \mathbb{E}_X\Big[(a(X) - \mathbb{E}(Y))\mathbb{E}(Y - a(X)|X)\Big] \\
&= \mathbb{E}_X\Big[(a(X) - \mathbb{E}(Y))(\mathbb{E}(Y|X) - a(X))\Big] \\
&= \mathbb{E}_X\Big[(a(X) - \mathbb{E}(Y))\underbrace{(a(X) - a(X))}_{=0}\Big] \\
&= 0
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Var}(Y) &= \underbrace{\mathbb{E}\big[(Y - a(X))^2\big]}_{\geq 0} + \mathbb{E}\big[(a(X) - \mathbb{E}(Y))^2\big] \\
&\geq \mathbb{E}\big[(a(X) - \mathbb{E}(Y))^2\big] \\
&= \mathbb{E}\big[(a(X) - \mathbb{E}(a(X)))^2\big] \\
&= \mathrm{Var}(a(X))
\end{aligned}$$

$E(Y) = E[E(Y|X)]$

$= E[a(X)]$

and we have shown that

$$\mathrm{Var}(Y) \geq \mathrm{Var}(Y|X)$$

$\square$

Theorem 4.16

Proof:

Let W be an UMVUE for $\tau(\theta)$. Suppose W' is another UMVUE for $\tau(\theta)$ and consider the estimator

$$W^* = \frac{1}{2}(W + W')$$

Then

$$\mathbb{E}(W^*) = \frac{1}{2}(\mathbb{E}(W) + \mathbb{E}(W')) = \frac{1}{2}(\tau(\theta) + \tau(\theta)) = \tau(\theta)$$

and

Apply Cauchy-Schwartz 'ly

$\mathrm{Cov}(W, W') \leq [\mathrm{Var}W \, \mathrm{Var}W']$

W - UMVUE

$$\begin{aligned}
\mathrm{Var}(W^*) &= \mathrm{Var}\left(\frac{1}{2}W + \frac{1}{2}W'\right) \\
&= \frac{1}{4}\mathrm{Var}(W) + \frac{1}{4}\mathrm{Var}(W') + \frac{1}{2}\mathrm{Cov}(W, W') \\
&\leq \frac{1}{4}\mathrm{Var}(W) + \frac{1}{4}\mathrm{Var}(W') + \frac{1}{2}\left[\mathrm{Var}(W)\mathrm{Var}(W')\right]^{1/2} \\
&= \mathrm{Var}(W) \qquad (*)
\end{aligned}$$

where we have applied the Cauchy-Schwartz ineqlity and the fact that $\mathrm{Var}(W) = \mathrm{Var}(W')$.

However, if the above inequality is strict then the UMVUE of W is contradicted - cannot happen!

Hence, we must have equality for all θ. Since the inequality is an application of Cauchy-Schwartz, we can have equality only if

$$W' = a(\theta)W + b(\theta)$$

$a(\theta) \,\mathrm{Var}\, W = \mathrm{Var}\, W$

$a(\theta) = 1$

Now:

$$\mathrm{Cov}(W, W') = \mathrm{Cov}(W, a(\theta)W + b(\theta)) = \mathrm{Cov}(W, a(\theta)W) = a(\theta)\mathrm{Var}(W)$$

but $\mathrm{Cov}(W, W') = \mathrm{Var}(W)$ since we have equality in (*). Hence $a(\theta) = 1$ and since $\mathbb{E}_\theta(W') = \tau(\theta)$ we must also have $b(\theta) = 0$ and hence $W = W'$ showing that W is unique. □

$E W' = E W + b(\theta) = \tau(\theta) + b(\theta) = \tau(\theta)$

$\Rightarrow b(\theta) = 0$

$W = W'$

Theorem 4.17

Proof:

Suppose that W is an UMVUE for $\tau(\theta)$ that is uncorrelated with all unbiased estimators of zero. Let W' be any other unbiased estimator for $\tau(\theta)$ i.e. $\mathbb{E}(W') = \tau(\theta)$ holds.

$\Longleftarrow$: Now consider $W' = W + (W' - W)$ and calculate

$$\mathrm{Var}_\theta(W') = \mathrm{Var}_\theta(W) + \mathrm{Var}_\theta(W' - W) + 2\mathrm{Cov}_\theta(W, W' - W)$$

But $W' - W$ is an unbiased estimator of zero since $\mathbb{E}(W - W') = 0$ and since W is uncorrelated with all unbiased estimators of zero we have $\mathrm{Cov}_\theta(W, W' - W) = 0$. This leads us to:

≥ 0

$$\mathrm{Var}_\theta(W') = \mathrm{Var}_\theta(W) + \mathrm{Var}_\theta(W' - W) \geq \mathrm{Var}_\theta(W)$$

since $\mathrm{Var}_\theta(W - W') \geq 0$. Hence, any W' which is unbiased fo $\tau(\theta)$ cannot beat W in terms of variance. This implies W is already the UMVUE of $\tau(\theta)$.

$\Longrightarrow$: If there is a $U(X)$ with $\mathbb{E}_\theta(U) = 0$ for all θ which is correlated with W (i.e. $\mathrm{Cov}_\theta(W, U) \neq 0$) then W cannot be the UMVUE of $\tau(\theta)$. Indeed, consider $W + aU$ where a is allowed to be varied. Then:

$E[aU] = aE(U) = 0$

$$\mathbb{E}_\theta(W + aU) = \mathbb{E}_\theta(W) = \tau(\theta)$$

which means it is in the class of unbiased estimators of $\tau(\theta)$. Now consider, choosing optimal values for a^* for which

$$\mathrm{Var}_\theta(W + a^*U) < \mathrm{Var}_\theta(W)$$

will hold, so that W would not be an UMVUE for $\tau(\theta)$. Now the quadratic function:

$$g(a) = a^2\mathrm{Var}_\theta(U) + 2a\mathrm{Cov}(U, W) + \mathrm{Var}_\theta(W)$$

will have a minumum at

$g'(a) = 2a\mathrm{Var}(U) + 2\mathrm{Cov}(U,W) = 0$

$$\boxed{a^* = -\frac{\mathrm{Cov}(U, W)}{\mathrm{Var}_\theta(U)} \neq 0}$$

by setting $g'(a) = 0$.

Hence, $W + a^*U$ will be unbiased for $\tau(\theta)$ but its variance will be truly less than $\mathrm{Var}_\theta(W)$ and hence W cannot be the UMVUE since $W + a^*U$ would be better.

This a^* will only be zero (and then W will be the UMVUE) when

$$\mathrm{Cov}_\theta(U, W) = 0$$

That is, the UMVUE of W will be uncorrelated with all unbiased estimators of zero. $\square$

Theorem 4.18

$\mathbb{E}W = \tau(\theta)$

Proof:

Let W be an unbiased estimator of $\tau(\theta)$. Then the Rao-Blackwell Theorem gives:

$$\hat{\tau}(T) = \mathbb{E}(W|T)$$

$\mathbb{E}\tilde{W} = \tau(\theta)$

is such that $\mathrm{Var}(\hat{\tau}(T)) \leq \mathrm{Var}(W)$ for all θ. Let $\tilde{W}$ be any other unbiased estimator and $\tilde{\tau}(T) = \mathbb{E}(\tilde{W}|T)$. Then

$$\mathbb{E}_\theta\big[\underbrace{\hat{\tau}(T) - \tilde{\tau}(T)}_{g(t)}\big] = 0$$

for all θ, and by completeness of T, it follows that

$$P_\theta(\hat{\tau}(T) = \tilde{\tau}(T)) = 1$$

for all θ. Hence $\hat{\tau}(T)$ is the unique UMVUE. □

Application of Lehmann-Scheffe:

$\mathbb{E}(X^2|X) = X^2$

1. Find a complete and sufficient statistic T
2. If we can find an unbiased estimator $V(T)$ then we've found the UMVUE since $\mathbb{E}(V(T)|T) = V(T)$.
3. Otherwise, find an unbiased estimator $W(X)$ can then compute $\hat{\tau}(T) = \mathbb{E}(W|T)$.