7 Appendix

7.1 Bias and Variance: Proof of Theorem 1

Proof. We provide full computations for the bias $\mathbb{E}\left[\hat{v}_\tau - V_\tau\right]$ and variance $\mathbb{E}[(\hat{v}_\tau - \mathbb{E}[\hat{v}_\tau])^2]$

We compute the expectation of the estimator at one data point, $\frac{1}{h}K\left(\frac{\tau(x_i)-t_i}{h}\right)\frac{y_i}{Q_i}$, omitting the $\frac{1}{n}$ term. By linearity of expectation:

$$\mathbb{E}\left[\frac{1}{h}K\left(\frac{\tau(x_i)-t_i}{h}\right)\frac{y_i}{Q_i}\right] = \mathbb{E}\left[\sum_{i=1}^{n}\frac{1}{nh}K\left(\frac{\tau(x_i)-t_i}{h}\right)\frac{y_i}{Q_i}\right]$$

The analysis follows the structure of standard bias and variance calculations for kernel density estimation [17]. We can express the conditional expectation of the kernel estimator via the integral convolution of the kernel and the conditional density. Note that by the symmetric property of kernel functions, $K\left(\frac{\tau(x_i)-t_i}{h}\right) = K\left(\frac{t_i-\tau(x_i)}{h}\right)$: we use them interchangeably. By iterated expectation and the definition of conditional expectation:

$$\begin{aligned}\mathbb{E}\left[\frac{1}{h}K\left(\frac{t_i-\tau(x_i)}{h}\right)\frac{y_i}{Q_i}\right] &= \mathbb{E}\left[\mathbb{E}\left[\frac{1}{h}K\left(\frac{t_i-\tau(x_i)}{h}\right)\frac{y_i}{Q_i}\Big| x_i\right]\right]\\ &= \mathbb{E}\left[\int \frac{y_i}{f_{T|X}(y_i \mid t')h}K\left(\frac{t'-\tau(x_i)}{h}\right)f_{Y,T|X}(y_i, t' \mid x_i)dt'\ dy\right]\end{aligned}$$

By a change of variables, let $u = \frac{t'-\tau(x_i)}{h}$. Then $t' = hu + \tau(x_i)$ and $dt' = \frac{du}{h}$. Changing variables in the integral corresponds to computing a local expansion of the conditional outcome density $f_{Y|T,x}$ around $\tau(x_i)$.

$$\mathbb{E}\left[\int y_i K(u)\frac{f_{Y,T|X}(y_i, \tau(i)+hu \mid x_i)}{f_{T|X}(y_i \mid \tau(x_i)+hu)}du\ dy\right] = \mathbb{E}\left[\int y_i K(u) f_{Y|T,X}(y_i \mid hu+\tau(x_i), x_i)du\ dy\right] \tag{1}$$

We use the definition of Bayes' rule and conditional densities, $f_{Y,T|X} = f_{Y|T,X}f_{T|X}$, with the definition of Q as $f_{T|X}(t_i, x_i)$, to transform the conditional density from the conditional density to the target density, $f_{Y|T,X}$.

Consider a 2nd order Taylor expansion of $f_{Y|T,X}$ around $T = \tau(x_i)$:

$$\begin{aligned}&f_{Y|T,X}(y_i \mid \tau(x_i)+hu, x_i)\\ &\approx f_{Y|T,X}(y_i \mid \tau(x_i), x_i) + hu\left(\frac{\partial}{\partial T}f_{Y|T,X}(y_i \mid \tau(x_i))\right) + \frac{(hu)^2}{2}\frac{\partial^2}{\partial T^2}f_{Y|T,X}(y_i \mid \tau(x_i), x_i) + o(h^2)\end{aligned}$$

Then we can compute the conditional expectation by integrating the approximation to the density term by term, where $\kappa_j(K) = \int u^j K(u)du$ denotes the jth kernel moment. The second order term describes the bias.

$$\begin{aligned}\mathbb{E}\left[\frac{1}{h}K\left(\frac{\tau(x_i)-t_i}{h}\right)\frac{y_i}{Q_i}\right] &= \mathbb{E}\left[\int y_i K(u) f_{Y|T,x}(y_i, \tau(x_i)+hu)du\ dy\right] \text{ by eq.1}\\ &= \mathbb{E}\left[\int y_i K(u) f_{Y|T,x}(y_i, \tau(x_i))du\ dy\right] + \mathbb{E}\left[\int y_i K(u) hu \frac{\partial}{\partial T}f_{Y|T,x}(y_i, \tau(x_i))du\ dy\right]\\ &+ \mathbb{E}\left[\int y_i K(u)(hu)^2\frac{\partial^2}{\partial T^2}f_{Y|T,x}(y_i, \tau(x_i))du\ dy\right] + \mathbb{E}\left[\int \frac{y_i}{n}o(h^2)K(u)du\ dy\right]\end{aligned}$$

For a symmetric kernel, the odd-order moments integrate to 0.

$$\begin{aligned}&= \mathbb{E}\left[\int y_i f_{Y|T,x}(y_i \mid \tau(x_i))dy\right] + \mathbb{E}\left[\int \frac{1}{2}y_i\frac{\partial^2}{\partial T^2}f_{Y|T,x}(y_i, \tau(x_i))h^2\kappa_2(K)dy\right]\\ &+ o(h^2)\int y_i dy \text{ since } \int K(u)du = 1\\ &= \mathbb{E}[Y(\tau(x_i))] + \kappa_2(K)\mathbb{E}\left[\int \frac{y_i h^2}{2}\frac{\partial^2}{\partial T^2}f_{Y|T,x}(y_i, \tau(x_i))dy\right] + o(h^2)\int y_i dy\end{aligned}$$

For the bias to vanish asymptotically, we require that $h^2 \to 0$, assuming that outcomes y_i are bounded, and that the second derivative of the conditional density of y given T, X is bounded.

We also consider the multivariate case. We assume the kernel function for the vector $\mathbf{u}$ is a product kernel, in the sense that it is the product of univariate kernels: $K(\mathbf{u}) = \prod K(u_i)$, each with bandwidth h_i. The multidimensional change of variables takes the form $t = \tau(x_i) + H\mathbf{u}$. Then by the multivariate Taylor expansion of the conditional density $f_{Y|T,X}$, the bias is:

$$\text{Bias}(\tau) = \frac{\kappa_2(K)}{2} \sum_{j=1}^{d} \frac{\partial^2}{\partial T_j^2} f_{Y|T,X}(y_i \mid \tau(x_i), x_i) h_j^2 + o(\sum_{j=1}^{d} h_j^2)$$

Calculations for Variance: A similar analysis follows for considering the variance; since we assume data (X_i, Y_i) are i.i.d. it suffices to consider the varaince of one term of the estimator.

$$Var\left[\sum_i^n \frac{1}{nh}\frac{y_i}{Q} K\left(\frac{t_i - \tau(x_i)}{h}\right)\right] = \frac{1}{nh^2} Var\left[\frac{y_i}{Q} K\left(\frac{t_i - \tau(x_i)}{h}\right)\right]$$
$$= \frac{1}{n}\left(\mathbb{E}\left[\left(\frac{y_i}{hQ} K\left(\frac{t_i - \tau(x_i)}{h}\right)\right)^2\right] - \mathbb{E}\left[\frac{y_i}{hQ} K\left(\frac{t_i - \tau(x_i)}{h}\right)\right]^2\right)$$

We will rewrite the squared expectation $\left(\mathbb{E}\left[\frac{y_i}{h} K\left(\frac{t_i - \tau(x_i)}{h}\right)\right]\right)^2$ using the bias analysis as approximately $\left(\mathbb{E}[Y(\tau(X))]h^2 \mathbb{E}\left[\int \frac{y_i}{2} \frac{\partial}{\partial T^2} f_{Y|T,x}(y_i \mid \tau(x_i), x_i)\kappa_2(K) dy\right] + o(h^2) \int y_i dy\right)^2$ for bounded outcomes y_i.

We compute $\mathbb{E}\left[\left(\frac{y_i}{h} K\left(\frac{T - \tau(x_i)}{h}\right)\right)^2\right]$ by analyzing a Taylor expansion of the conditional density after a change of variables:

$$\mathbb{E}\left[\left(\frac{y_i}{hQ_i} K\left(\frac{t_i - \tau(x_i)}{h}\right)\right)^2\right] = \mathbb{E}\left[\int\int_{-\infty}^{\infty} \frac{y_i^2}{h} K\left(\frac{t' - \tau(x_i)}{h}\right)^2 \frac{f_{Y,T}(y_i, t')}{f_{T|X}(\tau(x_i) + hu \mid x_i)^2} dt' dy\right]$$
$$= \mathbb{E}\left[\int\int_{-\infty}^{\infty} \frac{1}{h} \frac{y_i^2 K(u)^2 f_{Y|T,x_i}(y, \tau(x_i) + hu)}{f_{T|X}(\tau(x_i) + hu \mid x_i)} du dy\right] \text{ by Bayes' rule}$$

For convenience, we denote the quotient $\frac{f_{Y|T,x_i}(y, \tau(x_i) + hu, x_i)}{f_{T|X}(\tau(x_i) + hu|x_i)} = g_y(\tau(x_i) + hu)$. We expand this function around the argument T. In the asymptotic perspective, we omit the exact expressions of the derivatives g', g'', but we will require that the treatment density is nonzero for $\tau(x_i)$ (a standard overlap condition required for counterfactual policy evaluation). Then consider each term of the expansion in turn:

$$\mathbb{E}\left[\int\int_{-\infty}^{\infty} y_i^2 K(u)^2 \left(g(\tau(x_i)) + g'(\tau(x_i))(hu) + (hu)^2 \frac{g''(\tau(x_i))}{2}\right) du dy\right]$$

The first term is equivalent to

$$\frac{R(K)}{h} \mathbb{E}\left[\int \frac{y_i^2 f_{Y|T,X}(y_i \mid \tau(x_i), x_i) dy}{f_{T|X}(\tau(x_i) \mid x_i)}\right] = \frac{R(K)}{h} \mathbb{E}\left[\frac{\mathbb{E}[Y_i^2 \mid \tau(x_i), X]}{f_{T|X}(\tau(x_i) \mid x_i)}\right]$$

The second term and third terms are equivalently would require integrating $\int u^2 K(u) du$ by parts, which requires specification on the structure of the kernel. Since the integration of the terms evaluate to constants, under the assumption that $h \to 0$, the integral of the second and third terms of the expansion is $O(h) = o(\frac{1}{h})$.

So

$$\frac{1}{h}\mathbb{E}\left[\left(\frac{y_i}{Q} K\left(\frac{t_i - \tau(x_i)}{h^2}\right)\right)^2\right] \approx \frac{R(K)}{h} \mathbb{E}\left[\frac{\mathbb{E}[Y^2 \mid \tau(X), X]}{f_{T|X}(\tau(X) \mid X)}\right] + o(\frac{1}{h})$$

$R(k)$ is the 'roughness' term, where $R(k) = \int K^2(u) du$.

Combining these results:

$$\begin{aligned}
&\frac{1}{nh^2} Var\left[\frac{y_i}{Q} K\left(\frac{T-\tau(x_i)}{h}\right)\right] \\
&\approx \frac{R(K)}{nh}\mathbb{E}\left[\frac{\mathbb{E}[Y^2 \mid \tau(X), X]}{f_{T|X}(\tau(X)\mid X)}\right] + o\left(\frac{1}{nh}\right) - \frac{1}{n}\left(\mathbb{E}[Y(\tau(X))] + Bias\right)^2 \\
&\approx \frac{R(K)}{nh}\mathbb{E}\left[\frac{\mathbb{E}[Y^2 \mid \tau(X), X]}{f_{T|X}(\tau(X)\mid X)}\right] + \frac{O(h^4)}{n} + o\left(\frac{1}{nh}\right)
\end{aligned}$$

We discuss the multivariate case: again, we use the analysis of bias for the squared expectation term. For $\mathbb{E}\left[\left(\frac{y_i}{\prod_j^d h_j} K\left(\frac{T-\tau(x_i)}{\prod_j^d h_j}\right)\right)^2\right]$, we repeat the same analysis and use the product kernel form to decompose $K(\vec{u})^2 = \prod_{j=1}^d K(u_j)^2$. Then

$$\begin{aligned}
&\mathbb{E}\left[\int\int_{-\infty}^{\infty} y_i^2 \prod_{j=1}^d K(u_j)^2 \left(g(\tau(x_i)) + \sum_{j=1}^d \frac{\partial}{\partial u_i} g(\tau(x_i)) h_i u_i + \sum_{j=1}^d (h_i u_i)^2 \frac{\partial^2}{\partial u_i^2} g(\tau(x_i))\right) d\mathbf{u} dy\right] \\
&= \frac{R(K)^d}{\prod_j^d h_j}\mathbb{E}\left[\frac{\mathbb{E}[Y_i^2 \mid \tau(x_i), X]}{f_{T|X}(\tau(x_i)\mid x_i)}\right] + O\left(\prod_j^d h_j\right)
\end{aligned}$$

□

7.2 Analysis of Mean Squared Error: Proof of Theorem 2

Proof. Proof of Theorem 2: We analyze the MSE and characterize the optimal bandwidth which minimizes the MSE.

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{nh}\sum_{i=1}^n K\left(\frac{\tau(x_i)-t_i}{h}\right)\frac{y_i}{Q_i} - \mathbb{E}[Y(\tau(X))]\right)^2\right] = (\text{Bias}_\tau)^2 + \text{Var}_\tau \\
&= (\text{Bias}_\tau)^2 + \frac{R(K)}{nh}\mathbb{E}\left[\frac{\mathbb{E}[Y_i^2 \mid \tau(X), X]}{f_{T|X}(\tau(x_i)+hu\mid x_i)}\right] + o\left(\frac{1}{nh}\right) - \frac{1}{n}\left(\mathbb{E}[Y(\tau(X))] + \text{Bias}_\tau\right)^2 \\
&= \frac{R(K)}{nh}\mathbb{E}\left[\frac{\mathbb{E}[Y_i^2 \mid \tau(X), X]}{f_{T|X}(\tau(x_i)+hu\mid x_i)}\right] + o\left(\frac{1}{nh}\right) + O\left(\frac{h^4}{n}\right) + \frac{o(h^4)}{n} + O\left(\frac{1}{n}\right)
\end{aligned}$$

For notational convenience, we denote the following constant factors:

$C_1 = \mathbb{E}\left[\int \frac{y_i}{2}\frac{\partial}{\partial T^2} f_{Y|T,x}(y_i, \tau(x_i))\kappa_2(K)dy\right]$, $C_2 = \int y_i dy$, $C_3 = R(K)\mathbb{E}\left[\frac{1}{Q}\mathbb{E}[Y_i^2 \mid \tau(x_i), X]\right]$

If we want to optimize the bias-variance tradeoff of the asymptotic mean squared error, we choose the optimal bandwidth h such that neither term dominates the other.

$$=(Bias)^2 + Variance \approx h^4\left(C_1^2 - O\left(\frac{1}{n}\right)\right) + \frac{1}{nh}C_3 + o(h^4)(2C_1C_2 + C_2^2)$$

Optimizing the leading terms of the asymptotic MSE with respect to the bandwidth h:

$$\frac{d}{dh}(MSE) = 4C_1^2 h^3 - \frac{C_3}{nh^2} = 0$$

$$h^* = \left(\frac{C_3}{4C_1^2 n}\right)^{\frac{1}{5}} = \left(\frac{R(K)\mathbb{E}\left[\frac{\mathbb{E}[Y^2|\tau(X),X]}{f_{T|X}(\tau(X),X)}\right]}{4(\mathbb{E}\left[\int \frac{y_i}{2}\frac{\partial}{\partial T^2} f_{Y|T,x}(y_i,\tau(x_i))\kappa_2(K)dy\right])^2 n}\right)^{\frac{1}{5}} = O\left(n^{-\frac{1}{5}}\right)$$

The order of the optimal bandwidth is $O(\frac{1}{n^5})$; however in general, it will rely on the true density which is unknown a priori. □

7.3 Analysis of consistency (Proof of Theorem 3)

Proof. The convergence analysis of Theorem 2 provides rate on the convergence of the estimator: at the optimal bandwidth $h = O\left(n^{-\frac{1}{5}}\right)$, and therefore the asymptotic terms $O\left(\frac{h^4}{n}\right)$ and $O\left(\frac{1}{nh}\right) = O(n^{-\frac{4}{5}})$ control the rate of convergence. Since the MSE decays at the rate $O(n^{-\frac{4}{5}})$, the estimator converges in probability at the rate $n^{-\frac{2}{5}}$. This somewhat slower rate of convergence is standard in kernel density estimation.

For weak and strong uniform consistency results, we cite the results of [20] which hold if the density is uniformly continuous with bounded variation, and under additional moment conditions. In addition to absolute integrability of the kernel function, we require additionally that $\int \|xlog(x)\|x\|^{\frac{1}{2}}\|dK(x) < \infty$. These conditions are satisfied by most kernel functions, including the Gaussian kernel.

Then, the empirical estimator converges almost surely. Theorem B of [20] provides exact rates of the convergence as well: $((nh)^{-1}log(\frac{1}{h}))^{\frac{1}{2}} \xrightarrow{a.s.} 0$. With an optimal bandwidth of $h = O(n)^{-\frac{1}{5}}$, this corresponds to a $n^{\frac{2}{5}}\sqrt{log(n)}$ rate for strong uniform consistency. □

7.4 Analysis of consistency for normalized estimator

First we show that the normalization term converges to 1 in probability: $\hat{w}_n = \mathbb{E}\left[\sum \frac{1}{nhQ_i} K\left(\frac{\tau(x_i)-t}{h}\right)\right] \to 1$.

$$\mathbb{E}\left[\sum \frac{1}{nhQ_i} K\left(\frac{\tau(x_i)-t}{h}\right)\right] = \int \frac{1}{nhQ_i} K\left(\frac{\tau(x_i)-t}{h}\right) dtdy$$

$$= \int \frac{f_{Y|T,X}(y \mid t,x) + o(h^2)}{f_{T|X}(t,x)} dtdy = \int f_{Y,T|X}(y,t \mid x) dtdy + o(h^2) = 1 + o(h^2) \to 1$$

Then, by Slutsky's theorem, since each term in the quotient converges in probability and $\hat{w}_n \to_p 1$ which is a constant, $\hat{v}_\tau^{\text{norm}} \to V_\tau$.

7.5 Consistency of Policy Optimization

Our generalization bound follows from direct application of the upper bound of Theorem 8 of [1], the previous analysis characterizing the expectation of the estimator as $V_\tau = \hat{v}_\tau + \text{Bias}(\tau)$, and the fact that $V_{\tau^*} \leq V_{\hat{\tau}}$:

$$V_{\tau^*} - \text{Bias}(\tau^*) = \text{Risk}(\tau^*) \leq \text{Risk}(\hat{\tau}_n) \leq V_{\hat{\tau}_n} - \text{Bias}(\hat{\tau}_n)$$

In our setting, since the cost function is assumed bounded with constant $\bar{M}_c$, there is an additional multiplicative factor on the concentration term $\sqrt{\frac{8ln(2/\delta)}{n}}$.

The kernel function is bounded by definition and Lipschitz continuous depending on its specific functional form; we assume the Lipschitz constant of $K(u)$ is L_K, which is multiplied by $\frac{1}{h^2}$ in the empirical estimator. By Assumption 6, the inverse propensity weights are bounded $\frac{1}{Q_i} \in [1, \bar{M}_Q]$, $\forall x, T_i$.

The result follows from the Rademacher generalization bound [1, Thm. 8] concentrating $\mathbb{E}\hat{v}_\tau$ near $\mathbb{E}\hat{v}_\tau$, Thm. 1 relating V_τ to $\mathbb{E}\hat{v}_\tau$, and the Rademacher comparison lemma [14, Thm. 4.12].

7.6 Gradient of estimator with Epanechnikov kernel for optimization

For concreteness, we compute the gradient of our estimator for use in optimization algorithms for the Epanechnikov kernel. Note that the Epanechnikov kernel $K(u) = \frac{3}{4}(1-u^2)\mathbb{I}\{|u| \leq 1\}$is nondifferentiable at the boundary where $u = 1$; we assume this happens with probability 0 on real data. The optimization problem is nonsmooth and nonconvex.

The gradient of the un-normalized linear policy estimator is:

$$\nabla \hat{v}_\tau = \sum_{i=1}^{n} -x_i \frac{2(\beta^T x_i - t_i)}{h} \frac{y_i}{Q_i} \mathbb{I}\{|\frac{\beta^T x_i - t_i}{h}| \leq 1\}$$

We can compute the gradient directly of the normalized estimator via the quotient rule: for simplicity, denote the sum of probability ratios as $\text{norm}(Q)$, and denote the indicator function for each data point $\mathbb{I}\{|\frac{\beta^T x_i - t_i}{h}| \leq 1\} = \mathbb{I}(\beta^T x_i)$.

$$\nabla \hat{v}_\tau^{\text{norm}} = \frac{-3}{2h} \frac{\left(\text{norm}(Q)\left(\sum_i^n \frac{x_i y_i}{Q_i} \frac{\beta^T x_i - t_i}{h} \mathbb{I}(\beta^T x_i)\right) - \hat{v}_\tau \left(\sum_{i=1}^n \frac{x_i}{Q_i} \frac{\beta^T x_i - t_i}{h} \mathbb{I}(\beta^T x_i)\right)\right)}{\text{norm}(Q)^2}$$

7.7 Optimizing with the Triangular kernel

In the paper we reference the Epanechnikov kernel for concreteness and because it enjoys favorable properties such as statistical efficiency, as well as smoothness and concavity on its support. We remark that using the triangular kernel $K(u) = 1 - |u|, |u| \leq 1$ can allow us to rewrite the policy optimization problem (**??**) as a difference of convex programs, a special structural form of nonconvex optimization that can admit faster optimization by the Concave-Convex procedure (which alternatingly majorizes the concave portion). The global convergence of such a procedure remains an open question but it works well in practice [13]. The advantage of such a formulation is that vs. using numerical procedures for nonconvex optimization (BFGS) which have limited support for constraints, we can use modeling languages for convex solvers with richer support for constraints.

Denote $\ell^h(a, u) = \max(a, u)$ as the hinge loss. Then the expression for the triangular kernel can be expressed as $K(u) = \ell^h(-1, u) - 2\ell^h(0, u) + l^h(1, u)$. Since the hinge loss is a convex function, we can write the objective as a difference of convex functions depending on the sign of y_i, which is data. The program remains convex if we optimize over a convex policy space.