Supplementary Material

Proofs

Proof of Theorem 1. Let us write SATT as

$$\text{SATT} = \tfrac{1}{n_1}\sum_{i\in\mathcal{T}_1} Y_i - \tfrac{1}{n_1}\sum_{i\in\mathcal{T}_1} Y_i(0).$$

It is then clear that SATT differs from $\hat{\tau}_W$ only in the second term, that is,

$$\begin{aligned}\hat{\tau} - \text{SATT} &= \tfrac{1}{n_1}\sum_{i\in\mathcal{T}_1} Y_i(0) - \sum_{i\in\mathcal{T}_0} W_i Y_i(0)\\ &= \sum_{i=1}^n (-1)^{T_i+1} W_i Y_i(0)\\ &= \sum_{i=1}^n (-1)^{T_i+1} W_i f_0(X_i) + \sum_{i=1}^n (-1)^{T_i+1} W_i \epsilon_i,\end{aligned}$$

where we recognize the last term as E_W. For each term of E_W we have

$$\begin{aligned}&\mathbb{E}[(-1)^{T_i+1} W_i \epsilon_i \big| X, T]\\ &\qquad= (-1)^{T_i+1} W_i \left(\mathbb{E}\left[Y_i(0)|X,T\right] - f_0(X_i)\right)\\ &\qquad= (-1)^{T_i+1} W_i \left(\mathbb{E}\left[Y_i(0)|X\right] - f_0(X_i)\right) = 0,\end{aligned}$$

where the first equality is by definition of ϵ_i and the fact that $W_i = W_i(X,T)$ and the second is by Assumption 1. □

Proof of Theorem 2. Let D be the distance matrix $D_{ii'} = \delta(X_i, X_{i'})$. For this choice of $(\mathcal{F}, \|\cdot\|)$, by linear optimization duality we get

$$\begin{aligned}\mathfrak{E}(W;\mathcal{F}) &= \tfrac{1}{n_1}\sup_{v_i - v_{i'} \le D_{ii'} \forall i,i'} \left(\sum_{i\in\mathcal{T}_1} v_i - \sum_{i\in\mathcal{T}_0} n_1 W_i v_i\right)\\ &= \tfrac{1}{n_1}\min_S \quad \sum_{i,i'} D_{ii'} S_{ii'}\\ &\qquad \text{s.t.} \quad S \in \mathbb{R}_+^{n\times n}\\ &\qquad\qquad \sum_{i'=1}^n (S_{ii'} - S_{i'i}) = 1 \quad \forall i \in \mathcal{T}_1\\ &\qquad\qquad \sum_{i'=1}^n (S_{ii'} - S_{i'i}) = -n_1 W_i \quad \forall i \in \mathcal{T}_0.\end{aligned}$$

This describes a min-cost network flow problem with sources $\mathcal{T}_1$ with inputs 1, sinks $\mathcal{T}_0$ with outputs W_i, edges between every two nodes with costs $D_{ii'}$ and without capacities. Consider any source $i \in \mathcal{T}_1$ and any sink $i' \in \mathcal{T}_0$ and any path $i, i_1, \dots, i_m, i'$. By the triangle inequality, $D_{ii'} \le D_{ii_1} + D_{i_1 i_2} + \cdots + D_{i_m i'}$. Therefore, as there are no capacities, it is always preferable to send the flow from the sources to the sinks along the direct edges from $\mathcal{T}_1$ to $\mathcal{T}_0$. That is, we can eliminate all other edges and write

$$\begin{aligned}\mathfrak{E}(W;\mathcal{F}) = \tfrac{1}{n_1}\min_S \quad & \sum_{i\in\mathcal{T}_1,\, i'\in\mathcal{T}_0} D_{ii'} S_{ii'}\\ \text{s.t.} \quad & S \in \mathbb{R}_+^{\mathcal{T}_1\times\mathcal{T}_0}\\ & \sum_{i'\in\mathcal{T}_0} S_{ii'} = 1 \quad \forall i \in \mathcal{T}_1\\ & \sum_{i\in\mathcal{T}_1} S_{ii'} = n_1 W_i \quad \forall i' \in \mathcal{T}_0.\end{aligned}$$

In the case of with replacement and $\mathcal{W}_0 = \mathcal{W}_0^{\text{probability}}$, using the transformation $W_i' = n_1 W_i$, we get

$$\begin{aligned}&\min_{W\in\mathcal{W}} \mathfrak{E}(W;\mathcal{F})\\ &= \tfrac{1}{n_1}\min_{S,W'} \quad \sum_{i\in\mathcal{T}_1,\, i'\in\mathcal{T}_0} D_{ii'} S_{ii'}\\ &\qquad \text{s.t.} \quad S \in \mathbb{R}_+^{\mathcal{T}_1\times\mathcal{T}_0}\\ &\qquad\qquad W_i' \in \mathbb{R}_+^{\mathcal{T}_0}\\ &\qquad\qquad \sum_{i\in\mathcal{T}_0} W_i' = n_1\\ &\qquad\qquad \sum_{i'\in\mathcal{T}_0} S_{ii'} = 1 \quad \forall i \in \mathcal{T}_1\\ &\qquad\qquad \sum_{i\in\mathcal{T}_1} S_{ii'} - W_i' = 0 \quad \forall i' \in \mathcal{T}_0.\end{aligned}$$

This describes a min-cost netwrok flow problem with sources $\mathcal{T}_1$ with inputs 1; nodes $\mathcal{T}_0$ with 0 exogenous flow; one sink with output n_1; edges from each $i \in \mathcal{T}_1$ to each $i' \in \mathcal{T}_0$ with flow variable $S_{ii'}$, cost $D_{ii'}$, and without capacity; and edges from each $i \in \mathcal{T}_0$ to the sink with flow variable W_i' and without cost or capacity. Because all data is integer, the optimal solution of $W' = n_1 W$ is integer [1]. Hence, since $\mathcal{W}_0^{n_1\text{-multisubset}} \subseteq \mathbb{Z}/n_1$, the solution is the same when we restrict to $\mathcal{W}_0 = \mathcal{W}_0^{n_1\text{-multisubset}}$. This solution (in terms of W') is equal to sending the whole input 1 from each source in $\mathcal{T}_1$ to the node in $\mathcal{T}_0$ with smallest distance and from there routing this flow to the sink, which corresponds exactly to one-to-one matching with replacement.

In the case of no replacement and $\mathcal{W}_0 = \mathcal{W}_0^{n_1^{-1}\text{-bounded}}$, using the transformation $W_i' = n_1 W_i$, we get

$$\begin{aligned}&\min_{W\in\mathcal{W}} \mathfrak{E}(W;\mathcal{F})\\ &= \tfrac{1}{n_1}\min_{S,W'} \quad \sum_{i\in\mathcal{T}_1,\, i'\in\mathcal{T}_0} D_{ii'} S_{ii'}\\ &\qquad \text{s.t.} \quad S \in \mathbb{R}_+^{\mathcal{T}_1\times\mathcal{T}_0}\\ &\qquad\qquad W_i' \in \mathbb{R}_+^{\mathcal{T}_0}\\ &\qquad\qquad \sum_{i\in\mathcal{T}_0} W_i' = n_1\\ &\qquad\qquad W_i' \le 1 \quad \forall i \in \mathcal{T}_0\\ &\qquad\qquad \sum_{i'\in\mathcal{T}_0} S_{ii'} = 1 \quad \forall i \in \mathcal{T}_1\\ &\qquad\qquad \sum_{i\in\mathcal{T}_1} S_{ii'} - W_i' = 0 \quad \forall i' \in \mathcal{T}_0.\end{aligned}$$

This describes the same min-cost netwrok flow problem except that the edges from each $i \in \mathcal{T}_0$ to the sink have a capacity of 1. Because all data is integer, the optimal solution of S and $W' = n_1 W$ is integer [1]. Hence, since $\mathcal{W}_0^{n_1\text{-subset}} \subseteq \mathbb{Z}/n_1$, the solution is the same when we restrict to $\mathcal{W}_0 = \mathcal{W}_0^{n_1\text{-subset}}$. The optimal $S_{ii'}$ is integer and so, by $\sum_{i'\in\mathcal{T}_0} S_{ii'} = 1$, for each $i \in \mathcal{T}_1$ there is exactly one $i' \in \mathcal{T}_0$ with $S_{ii'} = 1$ and all others are zero. $S_{ii'} = 1$ denotes matching i with i'. The optimal W_i' is integral and so, by $W_i' \le 1$, $W_i' \in \{0,1\}$. Hence, for each $i \in \mathcal{T}_0$,

$\sum_{i' \in \mathcal{T}_1} S_{ii'} \in \{0,1\}$ so we only use node i at most once. The cost of S is exactly the sum of pairwise distances in the match. Hence, the optimal solution corresponds exactly to one-to-one matching without replacement. □

Proof of Corollary 3. Apply Theorem 2 with the metric

$$\delta'(x,x') = \mathbb{I}_{[x \neq x']} \max\{\delta(x,x'),\, \delta_0\}. \quad \square$$

Proof of Theorem 4. This choice of space leads to

$$\mathfrak{E}(W;\mathcal{F}) = \sum_{j=1}^{M} \left| \frac{1}{n_1} \sum_{i \in \mathcal{T}_1} \mathbb{I}_{[C(X_i)=j]} - \sum_{i \in \mathcal{T}_0} W_i \mathbb{I}_{[C(X_i)=j]} \right|.$$

That is, the worst-case f assigns ± 1 to each partition in order to make the difference of values in that partition be nonnegative. Then clearly the optimal choice of $W \in \mathbb{R}^{\mathcal{T}_0}$ is to make each of these absolute values equal zero. This happens exactly when, for each $i \in \mathcal{T}_0$,

$$\begin{aligned} W_i &= \tfrac{1}{n_1} \tfrac{|i' \in \mathcal{T}_1 : C(X_{i'}) = C(X_i)|}{|i' \in \mathcal{T}_0 : C(X_{i'}) = C(X_i)|} \\ &= \tfrac{1}{n_1} \tfrac{\text{num treatment subjects in same partition as } i}{\text{num control subjects in same partition as } i}, \end{aligned}$$

where $0/0 = 0$ and we never encounter dividing a positive integer by 0 due to the no-extrapolation assumption. Because the weight is nonnegative, the solution is unchanged when restricting to nonnegative weights. □

Proof of Theorem 5. By duality of norms,

$$\begin{aligned} \mathfrak{E}(W;\mathcal{F}) &= \sup_{\beta^T V \beta \le 1} \beta^T \left(\frac{1}{n_1} \sum_{i \in \mathcal{T}_1} X_i - \sum_{i \in \mathcal{T}_0} W_i X_i \right) \\ &= M_V(W). \end{aligned}$$

The optimal W minimizes this discrepancy over subsamples from control with the allowable size. □

Proof of Theorem 6. We have

$$\begin{aligned} &\mathfrak{E}^2(W;\mathcal{F}) = \max_{\|f\| \le 1} \left(\textstyle\sum_{i=1}^n (-1)^{T_i+1} W_i f(X_i) \right)^2 \\ &= \left\langle \textstyle\sum_{i=1}^n (-1)^{T_i+1} W_i \mathcal{K}(X_i,\cdot), \sum_{i=1}^n (-1)^{T_i+1} W_i \mathcal{K}(X_i,\cdot) \right\rangle \\ &= \textstyle\sum_{i,j=1}^n (-1)^{T_i+T_j} W_i W_j K_{ij}, \end{aligned}$$

which when written in block form gives rise to the result. □

Proof of Theorem 7. First we show $\mathfrak{E}_{\min}(\mathcal{F}) \to 0$ a.s. by showing that we can construct a feasible $\tilde{W}$ such that $\mathfrak{E}(\tilde{W};\mathcal{F}) \to 0$ a.s. Let $p(x) = \mathbb{P}(T=1|X=x)$. By Assumption 1, $0 < p(X) < 1$ a.s. So there exists $\alpha > 0$ such that $q(x) = \alpha p(x)/(1-p(x))$ is a.s. in $(0,1)$. For each i, let $\tilde{W}_i' \in \{0,1\}$ be Bernoulli with probability $q(X_i)$. Then we have that $X_i \big| T=0, \tilde{W}_i' = 1$ is distributed as $X_i \big| T=1$. Let $n_0' = \sum_{j \in \mathcal{T}_0} \tilde{W}_j'$ and note that $n_0' \ge \underline{n}_0'$ eventually a.s. For each $i \in \mathcal{T}_0$, set $\tilde{W}_i = \tilde{W}_i'/n_0'$. Let $\zeta(f) = \mathbb{E}[f(X_1) \mid T=1]$, $\xi_i(f) = (T_i + \tilde{W}_i')(f(X_i) - \zeta(f))$. Let $A_0 = \frac{1}{n_0}\sum_{i \in \mathcal{T}_0} \xi_i$ and $A_1 = \frac{1}{n_1}\sum_{i \in \mathcal{T}_1} \xi_i$. Adding and subtracting ζ, we see $\mathcal{E}(\tilde{W};f) = A_1(f) - (n_0/n_0')A_0(f)$. By construction of $\tilde{W}_i'$, we see that $\mathbb{E}[\xi_i] = 0$ (i.e., Bochner integral). By (5), $\|\xi\|_*$ has (a) second or (b) first moment. By (1), each ξ_i is independent. Therefore, by [2] for (5)(a) (since B-convexity of $\mathcal{F}$ implies B-convexity of $\mathcal{F}^*$; see [20]) or by [6] for (5)(b), a law of large numbers holds yielding, a.s., $\|A_0\|_* \to 0$ and $\|A_1\|_* \to 0$. Since $(n_0/n_0') \to \alpha \mathbb{E}[p(X_1)] < \infty$ a.s., we have that $\|\mathcal{E}(\tilde{W};\cdot)\|_* \to 0$ a.s. By (3) $\tilde{W}$ is feasible, so, a.s.

$$\mathfrak{E}_{\min}(\mathcal{F}) = \mathfrak{E}(W;\mathcal{F}) \le \mathfrak{E}(\tilde{W};\mathcal{F}) = \|\mathcal{E}(\tilde{W};\cdot)\|_* \to 0.$$

Fix $\epsilon > 0$. By (4), theres is a $g_0 \in \mathcal{F}$ such that $\sup_x |f_0(x) - g_0(x)| \le \epsilon/2$. Hence,

$$\begin{aligned} |\mathcal{E}(W;f_0)| &\le |\mathcal{E}(W;g_0)| + 2 \sup_{i=1,\dots,n} |f_0(X_i) - g_0(X_i)| \\ &\le \|g_0\|\, \mathfrak{E}(W;\mathcal{F}) + \epsilon = \|g_0\|\, \mathfrak{E}_{\min} + \epsilon \to \epsilon. \end{aligned}$$

Since true for any $\epsilon > 0$, $|\mathcal{E}(W;f_0)| \to 0$ a.s. □