A. Supplemental Material

These appendices cover all the proofs for the paper. We begin with a high-level analysis outline (Appendix B) followed by more detailed proofs (Appendices C–G). An additional example model (Weibull regression) satisfying Assumptions A1–A4 is presented in Appendix I. Guarantees for nearest neighbor and kernel variants of the Nelson-Aalen estimator are in Appendix J. Additional information on experimental results is in Appendix K.

Before presenting the proof of the kernel survival estimator result, we present an intermediate result for what we call the fixed-radius NN survival estimator.

Fixed-radius NN survival estimator. We find all training subjects with feature vectors at most a user-specified distance $h > 0$ from x. Let $\mathcal{N}_{\text{NN}(h)}(x) \subseteq [n]$ denote their indices. Then the fixed-radius NN estimator is $\widehat{S}^{\text{NN}(h)}(t|x) := \widehat{S}^{\text{KM}}(t|\mathcal{N}_{\text{NN}(h)}(x))$.

This estimator is a special case of the kernel survival estimator with kernel $K(s) = \mathbb{1}\{s \le 1\}$. However, because this estimator weights all neighbors found within radius h equally, we can actually derive a stronger guarantee than for the kernel estimator.

Theorem A.1 (Fixed-radius NN pointwise guarantees)**.** *Under Assumptions A1–A4, let $\varepsilon \in (0,1)$ be a user-specified error tolerance. Suppose that the threshold distance satisfies $h \in (0, h^*]$ with $h^* = (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$, and the number of training data satisfies $n \ge \frac{144}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,h})}$. For any $x \in supp(\mathbb{P}_X)$,*

$$\begin{aligned}&\mathbb{P}\Big(\sup_{t\in[0,\tau]} |\widehat{S}^{\text{NN}(h)}(t|x) - S(t|x)| > \varepsilon\Big)\\ &\le \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\theta}{16}\Big) + \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})}{8}\Big)\\ &\quad+2\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^4}{1296}\Big)+\frac{8}{\varepsilon}\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^2}{324}\Big).\end{aligned} \tag{7}$$

Moreover, if there exist constants $p_{\min} > 0$, $d > 0$, and $r^* > 0$ such that $\mathbb{P}_X(\mathcal{B}_{x,r}) \ge p_{\min} r^d$ for all $r \in (0, r^*]$, then using the numbers $c_2 = \Theta\big(\frac{\theta^{(4\alpha+d)/(5\alpha+2d)}}{\Lambda^{d/(5\alpha+2d)}}\big)$ and $c_3 = \Theta\big(\frac{\Lambda^{d/(2\alpha+d)}}{\theta^{(4\alpha+d)/(2\alpha+d)}}\big)$ as in Corollary 3.1, letting $c_1' := (\frac{\theta c_3}{18\Lambda})^{1/\alpha} = \Theta(\frac{1}{(\theta\Lambda)^{2/(2\alpha+d)}})$, and choosing threshold

$$h_n := c_1'\Big(\frac{\log(c_2 n)}{n}\Big)^{\frac{1}{2\alpha+d}},$$

we have, with probability 1,

$$\limsup_{n\to\infty}\left\{\frac{\sup_{t\in[0,\tau]}|\widehat{S}^{\text{NN}(h_n)}(t|x) - S(t|x)|}{c_3\big(\frac{\log(c_2 n)}{n}\big)^{\alpha/(2\alpha+d)}}\right\} < 1.$$

Bound (7) matches that of the k-NN estimator (bound (2)) with k replaced by $\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h})$, and every instance of h^* in the k-NN bound replaced by user-specified threshold h, which we ask to be at most h^*. The main change is that we now directly control how close training subjects must be to x to be declared as neighbors, but we lose control over how many of them there are. The second term in bound (7) is the penalty for not having at least $\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h})$ neighbors.

The technical core of the paper resides in the analysis of the k-NN survival estimator (proofs of Theorems 3.1 and Corollary 3.1). Our proofs for the analogous fixed-radius NN and kernel estimator guarantees primarily focus on aspects that differ from the k-NN case.

B. Analysis Outline

We outline the proof strategy for establishing the nonasymptotic k-NN estimator result (Theorem 3.1). The fixed-radius NN and kernel analyses are similar. We denote $d_{\mathcal{I}}^{+}(t)$ to be the number of training subjects in $\mathcal{I} \subseteq [n]$ who survive beyond time t, i.e., $d_{\mathcal{I}}^{+}(t) := \sum_{j\in\mathcal{I}} \mathbb{1}\{Y_j > t\}$.

As with the analysis of the Kaplan-Meier estimator by Földes & Rejtö (1981), we decompose the log of the k-NN estimate $\widehat{S}^{k\text{-NN}}(t|x)$ into three terms with the help of a Taylor expansion. By Assumption A2, two deaths happen at the same time with probability 0, so

$$\widehat{S}^{k\text{-NN}}(t|x) = \prod_{i\in\mathcal{N}_{k\text{-NN}}(x)} \Big(\frac{d^{+}_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i)}{d^{+}_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i)+1}\Big)^{\delta_i\mathbb{1}\{Y_i\le t\}}.$$

Taking the log of both sides, and noting that for any positive real number z, we have $\log(1+z) = \sum_{\ell=1}^{\infty}\frac{1}{\ell}(\frac{z}{z+1})^{\ell}$, we get

$$\begin{aligned}&\log\widehat{S}^{k\text{-NN}}(t|x)\\ &= -\sum_{i\in\mathcal{N}_{k\text{-NN}}(x)}\delta_i\mathbb{1}\{Y_i\le t\}\log\Big(1+\frac{1}{d^{+}_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i)}\Big)\\ &= -\sum_{i\in\mathcal{N}_{k\text{-NN}}(x)}\delta_i\mathbb{1}\{Y_i\le t\}\sum_{\ell=1}^{\infty}\frac{1}{\ell(d^{+}_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i)+1)^{\ell}}.\\ &= U_1(t|x)+U_2(t|x)+U_3(t|x),\end{aligned} \tag{8}$$

where

$$U_1(t|x) = \frac{1}{k}\sum_{i\in\mathcal{N}_{k\text{-NN}}(x)} -\frac{\delta_i\mathbb{1}\{Y_i\le t\}}{S_{\text{Y}}(Y_i|x)},$$

$$U_2(t|x) = -\sum_{i\in\mathcal{N}_{k\text{-NN}}(x)}\frac{\delta_i\mathbb{1}\{Y_i\le t\}}{d^{+}_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i)+1} - U_1(t|x),$$

$$U_3(t|x) = -\sum_{i\in\mathcal{N}_{k\text{-NN}}(x)}\delta_i\mathbb{1}\{Y_i\le t\}\sum_{\ell=2}^{\infty}\frac{1}{\ell(d^{+}_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i)+1)^{\ell}}.$$

For large enough k and n, it turns out that $U_1(t|x)$ converges to $\log S(t|x)$ while $U_2(t|x)$ (first-order Taylor approximation error) and $U_3(t|x)$ (sum of higher-order Taylor series terms) both go to 0.

The first term $U_1(t|x)$ corresponds to a k-NN regression estimate that averages the "label" variable $\xi_i := -\frac{\delta_i \mathbb{1}\{Y_i \le t\}}{S_Y(Y_i|x)}$ across the k nearest neighbors. Note that the label variable ξ_i perfectly knows the observed time Y's tail distribution $S_Y(\cdot|x)$. Provided that the k nearest neighbors have feature vectors within distance h^* of x, then it turns out that $\mathbb{E}[\xi_i] \approx \log S(t|x)$. Thus, having the nearest neighbors close to x aims to control the bias of the k-NN regression estimator $U_1(t|x)$.

To control the variance of regression estimator $U_1(t|x)$, i.e., for the k labels being averaged to be close to its expectation, intuitively we want k to be sufficiently large. However, how fast the average label converges to its expectation depends on whether the label variables ξ_i's are correlated. The joint distribution of these k label variables ξ_i's is not straightforward to analyze. To circumvent this issue, we use a key proof technique by Chaudhuri & Dasgupta (2014). Specifically, let $\widetilde{X}$ denote the feature vector of the $(k+1)$-st nearest neighbor of x. Then conditioned on $\widetilde{X}$, the k nearest neighbors' feature vectors appear as i.i.d. samples from $\mathbb{P}_X$ restricted to the open ball $\mathcal{B}^o_{x,\rho(x,\widetilde{X})}$. Thus, upon conditioning on $\widetilde{X}$, regression estimate $U_1(t|x)$ indeed becomes the average of k label variables ξ_i's that appear i.i.d., so Hoeffding's inequality tells us how fast their average converges to their expectation.

Since the regression estimate $U_1(t|x)$ assumes perfect knowledge of the distribution of the Y_i's (encoded in the tail probability $S_Y(\cdot|x)$), unsurprisingly the first-order Taylor approximation error $U_2(t|x)$ is about how well we can estimate $S_Y(\cdot|x)$. In particular, it turns out that $|U_2(t|x)|$ can be upper-bounded by how close the empirical distribution of the k nearest neighbors' Y values is to the CDF $1 - S_Y(\cdot|x)$. Thus, the problem boils down to one of CDF estimation, for which there is once again a bias-variance sort of decomposition. The bias term is controlled by making sure that the k nearest neighbors' feature vectors are within distance h^* of x. To control the variance, once again, we apply Chaudhuri and Dasgupta's proof technique of conditioning on the $(k+1)$-st nearest neighbor's feature vector $\widetilde{X}$. By doing this conditioning, the k nearest neighbors' observed times Y_i's become i.i.d., so the DKW inequality can be applied to bound the empirical distribution's deviation from its expectation.

In analyzing both $U_2(t|x)$ and $U_3(t|x)$, we remark that a key ingredient needed for our proof is that among the k nearest neighbors, the number of them that survive beyond time τ (which is precisely $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau)$) is sufficiently large. In the equations for $U_2(t|x)$ and $U_3(t|x)$, note that $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(Y_i) \ge d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau)$ whenever $Y_i \le \tau$. Thus by making $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau)$ large, the denominator terms of $U_2(t|x)$ and $U_3(t|x)$ are becoming big. This shrinks $|U_3(t|x)|$ to 0, and only partially helps in controlling $|U_2(t|x)|$, with the CDF estimation discussion above fully bringing $|U_2(t|x)|$ to 0.

Relating to the Nelson-Aalen estimator. When there are no ties in survival and censoring times, the Nelson-Aalen estimator is given by

$$\widehat{H}^{\text{NA}}(t) := \sum_{i=1}^{n} \frac{\delta_i \mathbb{1}\{Y_i \le t\}}{d^+_{[n]}(Y_i) + 1}.$$

Note that the first term in the definition of $U_2(t|x)$ is precisely a negated version of a k-NN variant of the Nelson-Aalen estimator! By showing that $U_1(t|x) + U_2(t|x)$ converges to $\log S(t|x)$, we can readily establish a nonasymptotic error bound for a k-NN Nelson-Aalen-based estimator for $H(t|x) := -\log S(t|x)$. We state guarantees for k-NN and kernel Nelson-Aalen-based estimators in Appendix J.

C. Proof of Theorem 3.1

To keep the exposition of the overall proof strategy clear, we defer proofs of supporting lemmas to the end of this section (in Appendices C.1–C.8). Much of the high-level proof structure is based on the nonasymptotic analysis of the Kaplan-Meier estimator by Földes & Rejtö (1981). In addition to making changes to incorporate nearest neighbor analysis, we also make some technical changes to Földes and Rejtö's proof, which we mention in Appendix C.9.

Following our analysis outline of Section B, we denote $\widetilde{X}$ to be the feature vector of the $(k+1)$-st nearest neighbor to x. We will be using this variable throughout this section.

As we discussed after the presentation of Theorem 3.1, there are four key bad events. We now precisely state what these bad events are. For each bad event, we also show how to control its probability to be arbitrarily small. After presenting these probability bounds, we explain why none of these bad events happening implies that $|\widehat{S}^{k\text{-NN}}(t|x) - S(t|x)| \le \varepsilon/3$. This factor of $1/3$ is important in the argument by Földes & Rejtö (1981) that translates an error guarantee for a fixed $t \in [0, \tau]$ to one that holds simultaneously across all $t \in [0, \tau]$, i.e., $\sup_{t\in[0,\tau]} |\widehat{S}^{k\text{-NN}}(t|x) - S(t|x)| \le \varepsilon$.

The first bad event is that not enough of the k nearest neighbors survive beyond the time horizon τ. Note that our convergence arguments for $U_2(t|x)$ and $U_3(t|x)$ later require that $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau) > k\theta/2$. Thus, our first bad event is

$$\mathcal{E}^{k\text{-NN}}_{\text{bad}\,\tau}(x) := \{d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau) \le k\theta/2\}.$$

We control $\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\text{bad}\,\tau}(x))$ to be arbitrarily small by having the number of nearest neighbors k be sufficiently large, which in turn requires the number of training data $n \ge k$ to be sufficiently large.

Lemma C.1. *Under Assumptions A1–A3, let $x \in supp(\mathbb{P}_X)$*

and $\varepsilon \in (0,1)$. *We have*

$$\mathbb{P}\big(\mathcal{E}^{k\text{-NN}}_{\textit{bad }\tau}(x)\big) \le \exp\Big(-\frac{k\theta}{8}\Big).$$

Next, for the terms $U_1(t|x)$ to converge to $\log S(t|x)$ and $U_2(t|x)$ to 0, we ask that the k nearest neighbors found for x be within a critical distance h^* that will depend on Hölder continuity constants of Assumption A4. This leads us to the next bad event:

$$\mathcal{E}^{k\text{-NN}}_{\text{far neighbors}}(x) := \{\rho(x,\widetilde{X}) \ge h^*\}.$$

Of course, if the $(k+1)$-st nearest neighbor is less than distance h^* away from x, then so are the k nearest neighbors. We control $\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\text{far neighbors}}(x))$ to be arbtrarily small by making the number of training subjects n sufficiently large. By sampling more training data, the k nearest neighbors found for x will gradually get closer to x.

Lemma C.2 (Chaudhuri & Dasgupta 2014, Lemma 9)**.** *Under Assumption A1, if* $k \le \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h^*})$*, then*

$$\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\textit{far neighbors}}(x)) \le \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h^*})}{8}\Big).$$

This lemma holds for any choice of distance $h^* > 0$ although for our analysis, we will choose $h^* = (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$. This particular choice of h^* is explained later on in Lemmas C.6 and C.7.

To get to our next bad event, we first relate $U_2(t|x)$ to a CDF estimate. Specifically, the function

$$\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x) := \frac{d^+_{\mathcal{N}_{k\text{-NN}}(x)}(s)}{k} = \frac{1}{k}\sum_{i\in\mathcal{N}_{k\text{-NN}}(x)} \mathbb{1}\{Y_i > s\}$$

is one minus an empirical distribution function. The next lemma bounds $|U_2(t|x)|$ in terms of $\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}$.

Lemma C.3. *Under Assumptions A1–A3, let* $x \in$ *supp*$(\mathbb{P}_X)$ *and* $t \in [0,\tau]$. *When event* $\mathcal{E}^{k\text{-NN}}_{\textit{bad }\tau}(x)$ *does not happen,*

$$\begin{aligned}&|U_2(t|x)|\\ &\quad\le \frac{2}{k\theta^2} + \frac{2}{\theta^2}\sup_{s\in[0,\tau]}|S_{\mathrm{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x)|\widetilde{X}]|\\ &\qquad+\frac{2}{\theta^2}\sup_{s\ge 0}|\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x)|\widetilde{X}]|.\end{aligned} \tag{9}$$

The third bad event corresponds to the empirical distribution function being too far from its expectation:

$$\begin{aligned}&\mathcal{E}^{k\text{-NN}}_{\text{bad EDF}}(x)\\ &\quad:= \Big\{\sup_{s\ge 0}|\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x)|\widetilde{X}]| > \frac{\varepsilon\theta^2}{36}\Big\},\end{aligned}$$

where importantly the expectation is, as with handling $U_1(t|x)$, a function of the $(k+1)$-st nearest neighbor $\widetilde{X}$. We control $\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\text{bad EDF}}(x))$ to be arbitrarily small by making the number of nearest neighbors k sufficiently large. The rate of convergence for the empirical distribution function is given by the DKW inequality.

Lemma C.4. *Under Assumptions A1–A3, for any* $x \in$ *supp*$(\mathbb{P}_X)$*,*

$$\mathbb{P}\big(\mathcal{E}^{k\text{-NN}}_{\textit{bad EDF}}(x)\big) \le 2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big).$$

The last bad event is that $U_1(t|x)$ is not close to its expectation $\mathbb{E}[U_1(t|x)|\widetilde{X}]$:

$$\mathcal{E}^{k\text{-NN}}_{\text{bad }U_1}(t,x) := \{|U_1(t|x) - \mathbb{E}[U_1(t|x)|\widetilde{X}]| \ge \varepsilon/18\}.$$

We control $\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\text{bad }U_1}(t,x))$ to be small by making the number of nearest neighbors k is sufficiently large.

Lemma C.5. *Under Assumptions A1–A3, let* $x \in$ *supp*$(\mathbb{P}_X)$ *and* $t \in [0,\tau]$. *Then*

$$\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\textit{bad }U_1}(t,x)) \le 2\exp\Big(-\frac{k\varepsilon^2\theta^2}{162}\Big).$$

At this point, we have collected all four main bad events. When none of these bad events happen, then starting from equation (8), applying the triangle inequality a few times, and using inequality (9), we get

$$\begin{aligned}&|\log\widehat{S}^{k\text{-NN}}(t|x) - \log S(t|x)|\\ &= |U_1(t|x) - \log S(t|x) + U_2(t|x) + U_3(t|x)|\\ &\le |U_1(t|x) - \log S(t|x)| + |U_2(t|x)| + |U_3(t|x)|\\ &\le |U_1(t|x) - \mathbb{E}[U_1(t|x)|\widetilde{X}]|\\ &\quad + |\mathbb{E}[U_1(t|x)|\widetilde{X}] - \log S(t|x)| + |U_2(t|x)| + |U_3(t|x)|\\ &\le |U_1(t|x) - \mathbb{E}[U_1(t|x)|\widetilde{X}]|\\ &\quad + |\mathbb{E}[U_1(t|x)|\widetilde{X}] - \log S(t|x)|\\ &\quad + \frac{2}{k\theta^2} + \frac{2}{\theta^2}\sup_{s\in[0,\tau]}|S_{\mathrm{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x)|\widetilde{X}]|\\ &\quad + \frac{2}{\theta^2}\sup_{s\ge 0}|\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x)|\widetilde{X}]|\\ &\quad + |U_3(t|x)|.\end{aligned} \tag{10}$$

We show that the RHS is at most $\varepsilon/3$ by ensuring that each of its six terms is at most $\varepsilon/18$. The 1st and 5th terms are at most $\varepsilon/18$ since bad events $\mathcal{E}^{k\text{-NN}}_{\text{bad }U_1}(t,x)$ and $\mathcal{E}^{k\text{-NN}}_{\text{bad EDF}}(x)$ do not happen. The 3rd term is at most $\varepsilon/18$ by recalling that the theorem assumes $k \ge \frac{72}{\varepsilon\theta^2}$, so $\frac{2}{k\theta^2} \le \frac{2}{(\frac{72}{\varepsilon\theta^2})\theta^2} = \frac{\varepsilon}{36} < \frac{\varepsilon}{18}$.

The 2nd, 4th, and 6th RHS terms of inequality (10) remain to be bounded. We tackle these in the next three lemmas. Note that these lemmas are deterministic. The first two lemmas ask that the k nearest neighbors be sufficiently close to x and make use of Hölder continuity; these lemmas explain why critical distance $h^* = (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$ and why Λ is defined the way it is.

Lemma C.6. *Under Assumptions A1–A4* (*this lemma uses Hölder continuity of $S_{\mathrm{C}}(t|\cdot)f_{\mathrm{T}}(t|\cdot)$*), *let $x \in supp(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. If bad event $\mathcal{E}^{k\text{-NN}}_{far\ neighbors}(x)$ does not happen, and $h^* \le [\frac{\varepsilon\theta}{18(\lambda_{\mathrm{T}}\tau+(f^*_{\mathrm{T}}\lambda_{\mathrm{C}}\tau^2)/2)}]^{1/\alpha}$, then*

$$|\mathbb{E}[U_1(t|x)|\widetilde{X}] - \log S(t|x)| \le \frac{\varepsilon}{18}.$$

Lemma C.7. *Under Assumptions A1–A4* (*this lemma uses Hölder continuity of $S_{\mathrm{Y}}(t|\cdot)$*), *let $x \in supp(\mathbb{P}_X)$ and $\varepsilon \in (0,1)$. If bad event $\mathcal{E}^{k\text{-NN}}_{far\ neighbors}(x)$ does not happen, and $h^* \le [\frac{\varepsilon\theta^2}{36(\lambda_{\mathrm{T}}+\lambda_{\mathrm{C}})\tau}]^{1/\alpha}$, then*

$$\frac{2}{\theta^2}\sup_{s\in[0,\tau]}|S_{\mathrm{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\mathrm{Y}}(s|x)|\widetilde{X}]| \le \frac{\varepsilon}{18}.$$

Lemma C.8. *Under Assumptions A1–A3, let $x \in supp(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. If bad event $\mathcal{E}^{k\text{-NN}}_{bad\ \tau}(x)$ does not happen, and $k \ge \frac{72}{\varepsilon\theta^2}$, then $|U_3(t|x)| \le \varepsilon/18$.*

Putting together the pieces so far, provided that all the bad events do not happen, then we have bounded all six RHS terms of inequality (10) by $\varepsilon/18$:

$$|\log \widehat{S}^{k\text{-NN}}(t|x) - \log S(t|x)| \le 6\cdot\frac{\varepsilon}{18} = \frac{\varepsilon}{3}.$$

For any $a,b \in (0,1]$, we have $|a-b| \le |\log a - \log b|$, so the above inequality implies that we also have

$$|\widehat{S}^{k\text{-NN}}(t|x) - S(t|x)| \le \frac{\varepsilon}{3}.$$

To establish Theorem 3.1, we need to guarantee that $\sup_{t\in[0,\tau]}|\widehat{S}^{k\text{-NN}}(t|x) - S(t|x)| \le \varepsilon$. A sufficient condition that accomplishes this task is to ask that $|\widehat{S}^{k\text{-NN}}(t|x) - S(t|x)| \le \varepsilon/3$ for a finite collection of times t within the interval $[0,\tau]$. Specifically, we partition the interval $[0,\tau]$ into $L(\varepsilon)$ pieces such that $0 = \eta_0 < \eta_1 < \cdots < \eta_{L(\varepsilon)} = \tau$, where:

- $S(\eta_{j-1}|x) - S(\eta_j|x) \le \varepsilon/3$ for $j = 1,\dots,L(\varepsilon)$,
- $L(\varepsilon) \le 4/\varepsilon$.

We can always produce a partition satisfying the above conditions because the most S can change from 0 to τ is by a value of 1 (S is one minus a CDF and is continuous). In this worst case scenario of S changing by 1, by placing the points η_j's at times where S drops by exactly $\varepsilon/3$ in value (except across the last piece $[\eta_{L(\varepsilon)-1}, \eta_{L(\varepsilon)}]$, where S could drop by less than $\varepsilon/3$), then $L(\varepsilon) = \lceil \frac{1}{\varepsilon/3} \rceil = \lceil 3/\varepsilon \rceil \le 4/\varepsilon$ where the last inequality holds for $\varepsilon \in (0,1]$. When S changes by less than 1, $L(\varepsilon)$ could be smaller.

We shall ask that $|\widehat{S}(\eta_j|x) - S(\eta_j|x)| \le \varepsilon/3$ for each $j = 1,2,\dots,L(\varepsilon)$. Note that $\widehat{S}(\cdot|x)$ is piecewise constant and monotonically decreasing. Moreover, $\widehat{S}(0|x) = S(0|x) = 1$ (the probability of a death happening at $t=0$ is 0). Thus, by having $\widehat{S}(\cdot|x)$ differ from $S(\cdot|x)$ by at most $\varepsilon/3$ at each η_j for $j = 1,\dots,L(\varepsilon)$, we are guaranteed that $|\widehat{S}(t|x) - S(t|x)| \le \varepsilon$ for any time $t \in [0,\tau]$. In summary, here are all the bad events of interest:

- $\mathcal{E}^{k\text{-NN}}_{\text{bad }\tau}(x)$
- $\mathcal{E}^{k\text{-NN}}_{\text{far neighbors}}(x)$
- $\mathcal{E}^{k\text{-NN}}_{\text{bad EDF}}(x)$
- $\mathcal{E}^{k\text{-NN}}_{\text{bad }U_1}(t,x)$ for $t = \eta_1,\eta_2,\dots,\eta_{L(\varepsilon)}$

The lemmas require $\frac{72}{\varepsilon\theta^2} \le k \le \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h^*})$, and $h^* \le [\min\{\frac{\varepsilon\theta}{18(\lambda_{\mathrm{T}}\tau+(f^*_{\mathrm{T}}\lambda_{\mathrm{C}}\tau^2)/2)}, \frac{\varepsilon\theta^2}{36(\lambda_{\mathrm{T}}+\lambda_{\mathrm{C}})\tau}\}]^{1/\alpha}$. Union bounding over all the bad events,

$$\begin{aligned}&\mathbb{P}(\text{at least one bad event happens})\\ &\quad\le \mathbb{P}\big(\mathcal{E}^{k\text{-NN}}_{\text{bad }\tau}(x)\big) + \mathbb{P}\big(\mathcal{E}^{k\text{-NN}}_{\text{far neighbors}}(x)\big)\\ &\qquad + \mathbb{P}\big(\mathcal{E}^{k\text{-NN}}_{\text{bad EDF}}(x)\big) + \sum_{\ell=1}^{L(\varepsilon)}\mathbb{P}\big(\mathcal{E}^{k\text{-NN}}_{\text{bad }U_1}(\eta_\ell,x)\big)\\ &\quad\le \exp\Big(-\frac{k\theta}{8}\Big) + \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h^*})}{8}\Big)\\ &\qquad + 2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big) + \frac{8}{\varepsilon}\exp\Big(-\frac{k\varepsilon^2\theta^2}{162}\Big). \qquad \square\end{aligned}$$

C.1. Proof of Lemma C.1

The key idea is that regardless of where each nearest neighbor $x' \in \mathcal{N}_{k\text{-NN}}(x)$ lands in feature space $\mathcal{X}$, the probability that its observed time (the corresponding Y variable) exceeds τ is $S_{\mathrm{Y}}(\tau|x') \ge \theta$ (Assumption A3). This means that $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau)$ stochastically dominates a Binomial(k,θ) random variable. Hence,

$$\begin{aligned}\mathbb{P}(\mathcal{E}^{k\text{-NN}}_{\text{bad }\tau}(x)) &= \mathbb{P}\Big(d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau) \le \frac{k\theta}{2}\Big)\\ &\le \mathbb{P}\Big(\text{Binomial}(k,\theta) \le \frac{k\theta}{2}\Big)\\ &\le \exp\Big(-\frac{1}{2\theta}\cdot\frac{(k\theta-\frac{k\theta}{2})^2}{k}\Big)\\ &= \exp\Big(-\frac{k\theta}{8}\Big),\end{aligned}$$

where the second inequality uses a Chernoff bound for the binomial distribution. $\square$

C.2. Proof of Lemma C.2

This proof is by Chaudhuri & Dasgupta (2014, Lemma 9). Let $x \in \text{supp}(\mathbb{P}_X)$ and $h^* > 0$. Let $\widetilde{X}$ denote the $(k+1)$-st nearest neighbor of x, and $N_{x,h^*} \sim$

Binomial$(n, \mathbb{P}_X(\mathcal{B}_{x,h^*}))$ denote the number of training data that land within distance h^* of x. Note that $\rho(x, \widetilde{X}) \geq h^*$ implies that $N_{x,h^*} \leq k$. Therefore, with the help of a Chernoff bound for the binomial distribution (with the assumption $1 \leq k \leq \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h^*})$),

$$\begin{aligned}
&\mathbb{P}(\rho(x,\widetilde{X}) \geq h^*) \\
&\quad \leq \mathbb{P}(N_{x,h^*} \leq k) \\
&\quad \leq \exp\Big(-\frac{(n\mathbb{P}_X(\mathcal{B}_{x,h^*}) - k)^2}{2n\mathbb{P}_X(\mathcal{B}_{x,h^*})}\Big) \\
&\quad \leq \exp\Big(-\frac{(n\mathbb{P}_X(\mathcal{B}_{x,h^*}) - \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h^*}))^2}{2n\mathbb{P}_X(\mathcal{B}_{x,h^*})}\Big) \\
&\quad = \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h^*})}{8}\Big). \qquad \square
\end{aligned}$$

C.3. Proof of Lemma C.3

We abbreviate the set of k nearest training subjects $\mathcal{N}_{k\text{-NN}}(x)$ as the set $\mathcal{I}$. We frequently use the fact that the function $d_{\mathcal{I}}^+$ monotonically decreases. Provided that bad event $\mathcal{E}_{\text{bad }\tau}^{k\text{-NN}}(x)$ does not happen, then we have $d_{\mathcal{I}}^+(t) > k\theta/2$ for all $t \in [0,\tau]$. Then

$$\begin{aligned}
&|U_2(t|x)| \\
&= \Big|\frac{1}{k}\sum_{i\in\mathcal{I}} \delta_i \mathbb{1}\{Y_i \leq t\}\Big[\frac{k}{d_{\mathcal{I}}^+(Y_i)+1} - \frac{1}{S_{\text{Y}}(Y_i|x)}\Big]\Big| \\
&\leq \frac{1}{k}\sum_{i\in\mathcal{I}} \delta_i \mathbb{1}\{Y_i \leq t\}\Big|\frac{k}{d_{\mathcal{I}}^+(Y_i)+1} - \frac{1}{S_{\text{Y}}(Y_i|x)}\Big| \\
&\leq \frac{1}{k}\sum_{i\in\mathcal{I}} \delta_i \mathbb{1}\{Y_i \leq t\} \sup_{s\in[0,\tau]}\Big|\frac{k}{d_{\mathcal{I}}^+(s)+1} - \frac{1}{S_{\text{Y}}(s|x)}\Big| \\
&\leq \sup_{s\in[0,\tau]}\Big|\frac{k}{d_{\mathcal{I}}^+(s)+1} - \frac{1}{S_{\text{Y}}(s|x)}\Big| \\
&= \sup_{s\in[0,\tau]}\Big|\frac{kS_{\text{Y}}(s|x) - d_{\mathcal{I}}^+(s) - 1}{(d_{\mathcal{I}}^+(s)+1)S_{\text{Y}}(s|x)}\Big| \\
&\leq \frac{k}{(d_{\mathcal{I}}^+(\tau)+1)S_{\text{Y}}(\tau|x)} \sup_{s\in[0,\tau]}\Big|S_{\text{Y}}(s|x) - \frac{d_{\mathcal{I}}^+(s)}{k} - \frac{1}{k}\Big| \\
&\leq \frac{k}{d_{\mathcal{I}}^+(\tau)\theta} \sup_{s\in[0,\tau]}\Big|S_{\text{Y}}(s|x) - \frac{d_{\mathcal{I}}^+(s)}{k} - \frac{1}{k}\Big| \\
&< \frac{2}{k\theta}\cdot\frac{k}{\theta} \sup_{s\in[0,\tau]}\Big|S_{\text{Y}}(s|x) - \frac{d_{\mathcal{I}}^+(s)}{k} - \frac{1}{k}\Big| \\
&= \frac{2}{\theta^2} \sup_{s\in[0,\tau]}\Big|S_{\text{Y}}(s|x) - \frac{d_{\mathcal{I}}^+(s)}{k} - \frac{1}{k}\Big| \\
&\leq \frac{2}{\theta^2}\Big(\frac{1}{k} + \sup_{s\in[0,\tau]}\Big|S_{\text{Y}}(s|x) - \frac{d_{\mathcal{I}}^+(s)}{k}\Big|\Big).
\end{aligned}$$

Using abbreviation $\widehat{S}_{\text{Y}}(s) := \widehat{S}_{\text{Y}}^{k\text{-NN}}(s|x) = d_{\mathcal{I}}^+(s)/k$,

$$\begin{aligned}
&\sup_{s\in[0,\tau]} |S_{\text{Y}}(s|x) - \widehat{S}_{\text{Y}}(s)| \\
&\leq \sup_{s\in[0,\tau]} |S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}_{\text{Y}}(s)|\widetilde{X}]| + \sup_{s\geq 0} |\widehat{S}_{\text{Y}}(s) - \mathbb{E}[\widehat{S}_{\text{Y}}(s)|\widetilde{X}]|.
\end{aligned}$$

Putting together the two inequalities above,

$$\begin{aligned}
|U_2(t|x)| \leq \frac{2}{k\theta^2} + \frac{2}{\theta^2}\sup_{s\in[0,\tau]} |S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}_{\text{Y}}(s)|\widetilde{X}]| \\
+ \frac{2}{\theta^2}\sup_{s\geq 0} |\widehat{S}_{\text{Y}}(s) - \mathbb{E}[\widehat{S}_{\text{Y}}(s)|\widetilde{X}]|. \qquad \square
\end{aligned}$$

C.4. Proof of Lemma C.4

This proof technique is from Chaudhuri & Dasgupta (2014, Lemma 10), modified to handle the survival analysis setup. The randomness can be described as follows:

1. Sample a feature vector $\widetilde{X} \in \mathcal{X}$ from the marginal distribution of the $(k+1)$-st nearest neighbor of x.
2. Sample k feature vectors i.i.d. from $\mathbb{P}_X$ conditioned on landing in the ball $\mathcal{B}^o_{x,\rho(x,\widetilde{X})}$.
3. Sample $n-k-1$ feature vectors i.i.d. from $\mathbb{P}_X$ conditioned on landing in $\mathcal{X} \setminus \mathcal{B}^o_{x,\rho(x,\widetilde{X})}$.
4. Randomly permute the n feature vectors sampled.
5. For each feature vector X_i, sample its corresponding observed time Y_i and censoring indicator δ_i.

As a technical remark, the above description of randomness requires Assumption A1 to hold in addition to using randomized tie breaking when finding the k nearest neighbors. Moreover, to incorporate this tie breaking into the theory, the definition of the open ball needs to be changed slightly, upon which the proof strategy still carries through. For details, see Section 2.7 in the Appendix of Chaudhuri & Dasgupta (2014).

The points sampled in step 2 are precisely the k nearest neighbors of x. Thus, using the Y_i variables corresponding specifically to the feature vectors generated in step 2 (let's call these k variables $Y_{(1)},\dots,Y_{(k)}$), construct the function $\Psi_s(\widetilde{X}) := \frac{1}{k}\sum_{\ell=1}^k \mathbb{1}\{Y_{(\ell)} > s\}$.

Note that $\widehat{S}_{\text{Y}}^{k\text{-NN}}(s|x) = \Psi_s(\widetilde{X})$, and after conditioning on $\widetilde{X}$, empirical distribution function $1 - \Psi_s(\widetilde{X})$ is constructed from i.i.d. samples from the CDF

$$1 - \mathbb{E}[\Psi(s) \mid \widetilde{X}] = 1 - \underbrace{\mathbb{P}(Y > s \mid X \in \mathcal{B}^o_{x,\rho(x,\widetilde{X})})}_{:=\overline{\Psi}(\widetilde{X})}.$$

Letting $\mathbb{P}_{\widetilde{X}}$ refer to the marginal distribution of $\widetilde{X}$ (from step 1 of the procedure above), then by the DKW inequality,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{s\ge 0}|\widehat{S}_{\mathrm{Y}}^{k\text{-NN}}(s|x)-\mathbb{E}[\widehat{S}_{\mathrm{Y}}^{k\text{-NN}}(s|x)|\widetilde{X}]|>\frac{\varepsilon\theta^2}{36}\Big)\\
&=\mathbb{P}\Big(\sup_{s\ge 0}|\Psi_s(\widetilde{X})-\overline{\Psi}(\widetilde{X})|>\frac{\varepsilon\theta^2}{36}\Big)\\
&=\int_{\mathcal{X}}\mathbb{P}\Big(\sup_{s\ge 0}|\Psi_s(\widetilde{X})-\overline{\Psi}(\widetilde{X})|>\frac{\varepsilon\theta^2}{36}\,\Big|\,\widetilde{X}=\widetilde{x}\Big)d\mathbb{P}_{\widetilde{X}}(\widetilde{x})\\
&\le\int_{\mathcal{X}}2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big)d\mathbb{P}_{\widetilde{X}}(\widetilde{x})\\
&=2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big).
\end{aligned}$$

□

C.5. Proof of Lemma C.5

Again, we use the proof technique by Chaudhuri & Dasgupta (2014, Lemma 10), slightly modified. The randomness can be described as follows:

1. Sample a feature vector $\widetilde{X}\in\mathcal{X}$ from the marginal distribution of the $(k+1)$-st nearest neighbor of x.
2. Sample k feature vectors i.i.d. from $\mathbb{P}_X$ conditioned on landing in the ball $\mathcal{B}^o_{x,\rho(x,\widetilde{X})}$.
3. Sample $n-k-1$ feature vectors i.i.d. from $\mathbb{P}_X$ conditioned on landing in $\mathcal{X}\setminus\mathcal{B}^o_{x,\rho(x,\widetilde{X})}$.
4. Randomly permute the n feature vectors sampled.
5. For each feature vector X_i, sample its corresponding observed time Y_i and censoring indicator δ_i.
6. Let $\xi_i=-\frac{\delta_i\mathbb{1}\{Y_i\le t\}}{S_{\mathrm{Y}}(Y_i|x)}$ for each i.

The points sampled in step 2 are the k nearest neighbors of x. In particular, $U_1(t|x)$ is the average of k terms that become i.i.d. after we condition on the $(k+1)$-st nearest neighbor $\widetilde{X}$:

$$U_1(t|x)=\frac{1}{k}\sum_{\ell=1}^{k}\xi_\ell(\widetilde{X}),$$

where $\xi_\ell(\widetilde{X})$ is the ξ_i variable corresponding to one of the feature vectors drawn in step 2 (which depends on $\widetilde{X}$). Each $\xi_\ell(\widetilde{X})$ has expectation

$$\begin{aligned}
\overline{\xi}(\widetilde{X})&:=\mathbb{E}_{Y,\delta}\Big[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\mathrm{Y}}(Y|x)}\,\Big|\,X\in\mathcal{B}^o_{x,\rho(x,\widetilde{X})}\Big]\\
&=\mathbb{E}[U_1(t|x)|\widetilde{X}],
\end{aligned}\tag{11}$$

which is a function of random variable $\widetilde{X}$. Moreover, each $\xi_\ell(\widetilde{X})$ is bounded in $[-\frac{1}{S_{\mathrm{Y}}(t|x)},0]$.

Letting $\mathbb{P}_{\widetilde{X}}$ refer to the marginal distribution of the $(k+1)$-st nearest neighbor (from step 1 of the procedure above), then by Hoeffding's inequality, $S_{\mathrm{Y}}(\cdot|x)$ monotonically decreasing, and Assumption A3,

$$\begin{aligned}
&\mathbb{P}\Big(|U_1(t|x)-\mathbb{E}[U_1(t|x)|\widetilde{X}]|\ge\frac{\varepsilon}{12}\Big)\\
&=\mathbb{P}\Big(\Big|\frac{1}{k}\sum_{\ell=1}^{k}\xi_\ell(\widetilde{X})-\overline{\xi}(\widetilde{X})\Big|\ge\frac{\varepsilon}{12}\Big)\\
&=\int_{\mathcal{X}}\mathbb{P}\Big(\Big|\frac{1}{k}\sum_{\ell=1}^{k}\xi_\ell(\widetilde{X})-\overline{\xi}(\widetilde{X})\Big|\ge\frac{\varepsilon}{18}\,\Big|\,\widetilde{X}=\widetilde{x}\Big)d\mathbb{P}_{\widetilde{X}}(\widetilde{x})\\
&\le\int_{\mathcal{X}}2\exp\Big(-\frac{k\varepsilon^2[S_{\mathrm{Y}}(t|x)]^2}{162}\Big)d\mathbb{P}_{\widetilde{X}}(\widetilde{x})\\
&=2\exp\Big(-\frac{k\varepsilon^2[S_{\mathrm{Y}}(t|x)]^2}{162}\Big)\\
&\le 2\exp\Big(-\frac{k\varepsilon^2\theta^2}{162}\Big).
\end{aligned}$$

□

C.6. Proof of Lemma C.6

Recall from equation (11) in Lemma C.5's proof that

$$\mathbb{E}[U_1(t|x)|\widetilde{X}]=\mathbb{E}_{Y,\delta}\Big[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\mathrm{Y}}(Y|x)}\,\Big|\,X\in\mathcal{B}^o_{x,\rho(x,\widetilde{X})}\Big],$$

where $\widetilde{X}$ is the $(k+1)$-st nearest neighbor of x. With abbreviation $\mathcal{B}^o:=\mathcal{B}^o_{x,\rho(x,\widetilde{X})}$,

$$\begin{aligned}
&|\mathbb{E}[U_1(t|x)|\widetilde{X}]-\log S(t|x)|\\
&=|\mathbb{E}[U_1(t|x)-\log S(t|x)\mid\widetilde{X}]|\\
&=\Big|\frac{\int_{\mathcal{B}^o}\{\mathbb{E}_{Y,\delta}[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\mathrm{Y}}(Y|x)}\,|X=x']-\log S(t|x)\}d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o)}\Big|\\
&\le\frac{\int_{\mathcal{B}^o}|\mathbb{E}_{Y,\delta}[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\mathrm{Y}}(Y|x)}\,|X=x']-\log S(t|x)|d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o)}
\end{aligned}\tag{12}$$

As we show next, for any $x'\in\mathcal{B}_{x,h^*}$,

$$\Big|\mathbb{E}_{Y,\delta}\Big[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\mathrm{Y}}(Y|x)}\,\Big|\,X=x'\Big]-\log S(t|x)\Big|\le\frac{\varepsilon}{18},\tag{13}$$

which, combined with inequality (12) and noting that $\rho(x,\widetilde{X})\le(h^*)^\alpha$, implies that

$$|\mathbb{E}[U_1(t|x)|\widetilde{X}]-\log S(t|x)|\le\frac{\varepsilon}{18}.$$

This means that conditioning on event $\mathcal{E}^{k\text{-NN}}_{\text{far neighbors}}(x)$ not happening, we deterministically have $|\mathbb{E}[U_1(t|x)|\widetilde{X}]-\log S(t|x)|\le\varepsilon/18$.

We now just need to show that inequality (13) holds. First, note that $\log S(t|x)$ is equal to the following expectation:

$$\mathbb{E}_{Y,\delta}\Big[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\mathrm{Y}}(Y|x)}\,\Big|\,X=x\Big]$$

$$
\begin{aligned}
&= -\int_0^t \Big[\int_s^\infty \frac{1}{S_{\text{Y}}(s|x)} d\mathbb{P}_{C|X=x}(c)\Big] d\mathbb{P}_{T|X=x}(s) \\
&= -\int_0^t \frac{1}{S_{\text{Y}}(s|x)} \Big[\int_s^\infty d\mathbb{P}_{C|X=x}(c)\Big] d\mathbb{P}_{T|X=x}(s) \\
&= -\int_0^t \frac{1}{S_{\text{Y}}(s|x)} S_{\text{C}}(s|x) f_{\text{T}}(s|x) ds \\
&= -\int_0^t \frac{1}{S(s|x)S_{\text{C}}(s|x)} S_{\text{C}}(s|x) f_{\text{T}}(s|x) ds \\
&= -\int_0^t \frac{1}{S(s|x)} f_{\text{T}}(s|x) ds \\
&= \log S(t|x) - \log \underbrace{S(0|x)}_{1} \\
&= \log S(t|x), \qquad (14)
\end{aligned}
$$

where we have used the fact that $\frac{d}{dx}\log S(t|x) = -\frac{f_{\text{T}}(t|x)}{S(t|x)}$ since S is 1 minus the CDF and f is the PDF of distribution $\mathbb{P}_{T|X=x}$.

For any x' within distance h^* of x, using an integral calculation similar to the one above,

$$
\begin{aligned}
&\Big|\mathbb{E}_{Y,\delta}\Big[-\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\text{Y}}(Y|x)} \,\Big|\, X = x'\Big] - \log S(t|x)\Big| \\
&= \Big| -\int_0^t \frac{1}{S_{\text{Y}}(s|x)} S_{\text{C}}(s|x') f_{\text{T}}(s|x') ds \\
&\qquad + \int_0^t \frac{1}{S_{\text{Y}}(s|x)} S_{\text{C}}(s|x) f_{\text{T}}(s|x) ds \Big| \\
&= \Big| \int_0^t \frac{1}{S_{\text{Y}}(s|x)} (S_{\text{C}}(s|x) f_{\text{T}}(s|x) - S_{\text{C}}(s|x') f_{\text{T}}(s|x')) ds \Big| \\
&\le \int_0^t \frac{1}{S_{\text{Y}}(s|x)} \big| S_{\text{C}}(s|x) f_{\text{T}}(s|x) - S_{\text{C}}(s|x') f_{\text{T}}(s|x') \big| ds.
\end{aligned}
$$

Using the fact that $S_{\text{Y}}(\cdot|x)$ monotonically decreases, Assumptions A3 and A4 (in particular, recall that $S_{\text{C}}(s|\cdot)f_{\text{T}}(s|\cdot)$ is Hölder continuous with parameters $(\lambda_{\text{T}} + f_{\text{T}}^*\lambda_{\text{C}} s)$ and α), and the choice of critical distance $h^* \le [\frac{\varepsilon\theta}{18(\lambda_{\text{T}}\tau + (f_{\text{T}}^*\lambda_{\text{C}}\tau^2)/2)}]^{1/\alpha}$,

$$
\begin{aligned}
&\int_0^t \frac{1}{S_{\text{Y}}(s|x)} \big| S_{\text{C}}(s|x) f_{\text{T}}(s|x) - S_{\text{C}}(s|x') f_{\text{T}}(s|x') \big| ds \\
&\le \frac{1}{S_{\text{Y}}(t|x)} \int_0^t \big| S_{\text{C}}(s|x) f_{\text{T}}(s|x) - S_{\text{C}}(s|x') f_{\text{T}}(s|x') \big| ds \\
&\le \frac{1}{S_{\text{Y}}(t|x)} \int_0^t (\lambda_{\text{T}} + f_{\text{T}}^*\lambda_{\text{C}} s)\rho(x,x')^\alpha ds \\
&\le \frac{1}{S_{\text{Y}}(t|x)} \int_0^t (\lambda_{\text{T}} + f_{\text{T}}^*\lambda_{\text{C}} s)(h^*)^\alpha ds \\
&= \frac{(h^*)^\alpha}{S_{\text{Y}}(t|x)} \Big(\lambda_{\text{T}} t + \frac{f_{\text{T}}^*\lambda_{\text{C}} t^2}{2}\Big) \\
&\le \frac{(h^*)^\alpha}{\theta} \Big(\lambda_{\text{T}} \tau + \frac{f_{\text{T}}^*\lambda_{\text{C}} \tau^2}{2}\Big) \\
&\le \frac{\big([\frac{\varepsilon\theta}{18(\lambda_{\text{T}}\tau + (f_{\text{T}}^*\lambda_{\text{C}}\tau^2)/2)}]^{1/\alpha}\big)^\alpha}{\theta} \Big(\lambda_{\text{T}} \tau + \frac{f_{\text{T}}^*\lambda_{\text{C}} \tau^2}{2}\Big) \\
&= \frac{\varepsilon}{18}.
\end{aligned}
$$

which establishes inequality (13). □

C.7. Proof of Lemma C.7

Recall the description of randomness in the proof of Lemma C.5. Let $\widetilde{X}$ denote the $(k+1)$-st nearest neighbor. Since bad event $\mathcal{E}^{k\text{-NN}}_{\text{far neighbors}}(x)$ does not happen, we know that $\rho(x, \widetilde{X}) \le (h^*)^\alpha$. This means that, using the fact that $S_{\text{Y}}(s|\cdot)$ is Hölder continuous with parameters $(\lambda_{\text{T}} + \lambda_{\text{C}})s$ and α,

$$
\begin{aligned}
&|S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\text{Y}}(s|x)|\widetilde{X}]| \\
&= |S_{\text{Y}}(s|x) - \mathbb{P}(Y > s \mid X \in \mathcal{B}^o_{x,\rho(x,\widetilde{X})})| \\
&= \Big| S_{\text{Y}}(s|x) - \frac{\int_{\mathcal{B}^o_{x,\rho(x,\widetilde{X})}} S_{\text{Y}}(s|x') d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o_{x,\rho(x,\widetilde{X})})} \Big| \\
&= \Big| \frac{\int_{\mathcal{B}^o_{x,\rho(x,\widetilde{X})}} [S_{\text{Y}}(s|x) - S_{\text{Y}}(s|x')] d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o_{x,\rho(x,\widetilde{X})})} \Big| \\
&\le \frac{\int_{\mathcal{B}^o_{x,\rho(x,\widetilde{X})}} |S_{\text{Y}}(s|x) - S_{\text{Y}}(s|x')| d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o_{x,\rho(x,\widetilde{X})})} \\
&\le \frac{\int_{\mathcal{B}^o_{x,\rho(x,\widetilde{X})}} (\lambda_{\text{T}} + \lambda_{\text{C}}) s \rho(x,x')^\alpha d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o_{x,\rho(x,\widetilde{X})})} \\
&\le \frac{(\lambda_{\text{T}} + \lambda_{\text{C}}) s (h^*)^\alpha \int_{\mathcal{B}^o_{x,\rho(x,\widetilde{X})}} d\mathbb{P}_X(x')}{\mathbb{P}_X(\mathcal{B}^o_{x,\rho(x,\widetilde{X})})} \\
&= (\lambda_{\text{T}} + \lambda_{\text{C}}) s (h^*)^\alpha.
\end{aligned}
$$

Taking the supremum of both sides over $s \in [0, \tau]$, multiplying through by $\frac{2}{\theta^2}$, and noting that $h^* \le [\frac{\varepsilon\theta^2}{36(\lambda_{\text{T}} + \lambda_{\text{C}})\tau}]^{1/\alpha}$, we obtain

$$
\begin{aligned}
&\frac{2}{\theta^2} \sup_{s\in[0,\tau]} |S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}^{k\text{-NN}}_{\text{Y}}(s|x)|\widetilde{X}]| \\
&\le \frac{2}{\theta^2} (\lambda_{\text{T}} + \lambda_{\text{C}}) \tau (h^*)^\alpha \\
&\le \frac{2}{\theta^2} (\lambda_{\text{T}} + \lambda_{\text{C}}) \tau \Big(\Big[\frac{\varepsilon\theta^2}{36(\lambda_{\text{T}} + \lambda_{\text{C}})\tau}\Big]^{1/\alpha}\Big)^\alpha = \frac{\varepsilon}{18}. \quad \square
\end{aligned}
$$

C.8. Proof of Lemma C.8

We abbreviate the set of k nearest training subjects $\mathcal{N}_{k\text{-NN}}(x)$ as the set $\mathcal{I}$. Since bad event $\mathcal{E}^{k\text{-NN}}_{\text{bad }\tau}(x)$ does not happen, we have $d^+_{\mathcal{I}}(\tau) > k\theta/2$. Note that $|U_3(t|x)| = \sum_{i\in\mathcal{I}} \Xi_i$, where

$$
\Xi_i := \delta_i \mathbb{1}\{Y_i \le t\} \sum_{\ell=2}^{\infty} \frac{1}{\ell (d^+_{\mathcal{I}}(Y_i) + 1)^\ell}.
$$

Using the fact that $d^+_{\mathcal{I}}$ monotonically decreases, and that $\sum_{\ell=2}^{\infty} \frac{1}{\ell(z+1)^\ell} = \log(1+\frac{1}{z}) - \frac{1}{z+1} \le \frac{1}{(z+1)^2}$ for all $z \ge 0.46241$,

$$\begin{aligned}\Xi_i &\le \sum_{\ell=2}^{\infty} \frac{1}{\ell(d^+_{\mathcal{I}}(t)+1)^\ell} \le \frac{1}{(d^+_{\mathcal{I}}(t)+1)^2} \\ &\le \frac{1}{(d^+_{\mathcal{I}}(\tau)+1)^2} \le \frac{1}{(d^+_{\mathcal{I}}(\tau))^2} \le \frac{4}{k^2\theta^2}.\end{aligned}$$

Lastly, using the assumption that $k \ge \frac{72}{\varepsilon\theta^2}$,

$$|U_3(t|x)| = \sum_{i\in\mathcal{I}} \Xi_i \le \frac{4|\mathcal{I}|}{k^2\theta^2} = \frac{4k}{k^2\theta^2} = \frac{4}{k\theta^2} \le \frac{\varepsilon}{18}. \quad \square$$

C.9. Technical Changes to the Analysis by Földes & Rejtö (1981)

Our event $\mathcal{E}^{k\text{-NN}}_{\text{bad }\tau}(x)$ not happening ensures that $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau) > k\theta/2$. Földes and Rejtö instead condition on two separate bad events, the first being $\{\max_{i\in\mathcal{N}_{k\text{-NN}}(x)} Y_i \le \tau\}$. When this bad event does not happen, then the number of survivors beyond time τ satisfies $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau) \ge 1$. This is a bit too weak of a requirement on $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau)$. As a result, Földes and Rejtö condition on a second bad event not happening to guarantee that (slightly rephrased to be in our setup's context) $d^+_{\mathcal{N}_{k\text{-NN}}(x)}(\tau) > k[S_{\text{Y}}(\tau|x)]^2$ (they ensure that this holds with high probability using Bernstein's inequality). Effectively this means that they have an extra bad event that they condition on not happening.

Next, in the partitioning of $[0,\tau]$ into $L(\varepsilon)$ pieces, Földes and Rejtö actually have all bad events except $\{\max_{i\in\mathcal{N}_{k\text{-NN}}(x)} Y_i \le \tau\}$ being repeated for $t = \eta_1,\dots,\eta_{L(\varepsilon)}$. Put another way, their final bound is looser since they multiply many more terms by $L(\varepsilon)$.

Lastly, Földes and Rejtö use versions of the DKW and Bernstein's inequalities with vintage constants that have since been improved. Notably, nowadays the DKW inequality generally refers to the refinement by Massart (1990).

D. Proof of Corollary 3.1

The basic idea of the proof is to solve for ε and n that satisfy both: i) sufficient conditions (3) with error probability set to be equal to $\gamma = 1/n^2$, and ii) $\varepsilon \le \frac{18\Lambda(r^*)^\alpha}{\theta}$. The choice of γ is not special and is chosen so that summing it from $n=1$ to $n=\infty$ results in a finite number, upon which the Borel-Cantelli lemma finishes the proof. (This proof would still work but with different constants if $\gamma = 1/n^\nu$ for any $\nu > 1$ due to convergence of hyperharmonic series.) There is a small technical hiccup of making sure that there is a valid integer to set k to be. The rest is a fair amount of algebra involving the Lambert W function. We provide the details for just the k-NN case below.

Let $\gamma = 1/n^2$. Recall that $h^* = (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$. Then one can easily check that each of the terms in bound (2) is at most $\gamma/4$ when k and n satisfy

$$\frac{648}{\varepsilon^2\theta^4}\log\frac{32n^2}{\varepsilon} \le k \le \frac{np_{\min}}{2}\Big(\frac{\varepsilon\theta}{18\Lambda}\Big)^{d/\alpha}.$$

We shall show how to set $\varepsilon \in (0, \frac{18\Lambda(r^*)^\alpha}{\theta}]$ as a function of n (along with additional conditions on n) such that

$$\frac{649}{\varepsilon^2\theta^4}\log\frac{32n^2}{\varepsilon} \le \frac{np_{\min}}{2}\Big(\frac{\varepsilon\theta}{18\Lambda}\Big)^{d/\alpha}. \tag{15}$$

Having the constant 649 is intentional. When inequality (15) holds, then

$$\frac{648}{\varepsilon^2\theta^4}\log\frac{32n^2}{\varepsilon} + 1 < \frac{649}{\varepsilon^2\theta^4}\log\frac{32n^2}{\varepsilon} \le \frac{np_{\min}}{2}\Big(\frac{\varepsilon\theta}{18\Lambda}\Big)^{d/\alpha},$$

which guarantees there to be at least one integer between $\frac{648}{\varepsilon^2\theta^4}\log\frac{32}{\varepsilon\gamma}$ and $\frac{np_{\min}}{2}(\frac{\varepsilon\theta}{18\Lambda})^{d/\alpha}$. Hence, a valid choice for k is

$$k = \Big\lfloor \frac{np_{\min}}{2}\Big(\frac{\varepsilon\theta}{18\Lambda}\Big)^{d/\alpha}\Big\rfloor.$$

The following pair of lemmas help us obtain a choice for ε as well as conditions on how large n should be; these lemmas are fundamentally about the Lambert W function.

Lemma D.1 (Lemma 3.6.11 of Chen & Shah 2018, combined with Theorem 2.1 of Hoorfar & Hassani 2008)**.** *Let W_0 be the principal branch of the Lambert W function. For any $a>0, b>0, c>0$, and $z\in(0,b)$, we have*

$$z^c \ge a\log\frac{b}{z}$$

if either of the following is true:

(a) We have

$$z \ge b\exp\Big(-\frac{1}{c}W_0\Big(\frac{cb^c}{a}\Big)\Big). \tag{16}$$

(b) We have

$$\frac{cb^c}{a} \ge e \quad \text{and} \quad z \ge \Big[\frac{a}{c}\log\Big(\frac{cb^c}{a}\Big)\Big]^{1/c}.$$

Proof. This lemma with only part (a) is precisely Lemma 3.6.11 of Chen & Shah (2018). Under the assumption that $\frac{cb^c}{a} \ge e$, then applying Theorem 2.1 of Hoorfar & Hassani (2008),

$$W_0\Big(\frac{cb^c}{a}\Big) \ge \log\Big(\frac{cb^c}{a}\Big) - \log\log\Big(\frac{cb^c}{a}\Big).$$

Thus, a sufficient condition to guarantee that inequality (16) holds is to ask that

$$z \geq b \exp\Big(-\frac{1}{c}\Big[\log\Big(\frac{cb^c}{a}\Big) - \log\log\Big(\frac{cb^c}{a}\Big)\Big]\Big) = \Big[\frac{a}{c}\log\Big(\frac{cb^c}{a}\Big)\Big]^{1/c}. \qquad \square$$

Lemma D.2. *Let W_{-1} be the lower branch of the Lambert W function. For any $a > 0$, $b > 0$, and $z > 0$,*

$$z \geq a \log z + b$$

if any of the following is true:

(a) We have $\frac{b}{a} + \log a \leq 1$.

(b) We have $\frac{b}{a} + \log a > 1$ and

$$z \geq -aW_{-1}\Big(-\frac{1}{ae^{b/a}}\Big).$$

(c) We have $\frac{b}{a} + \log a > 1$ and

$$z \geq a\big(1 + \sqrt{2\log(ae^{b/a-1})} + \log(ae^{b/a-1})\big).$$

Proof. To prove (a), using the assumption that $\frac{b}{a} + \log a \leq 1$, and recalling that $\log z \leq z - 1$ for all $z > 0$,

$$\begin{aligned} a\log z + b &= a\Big(\log\frac{z}{a} + \frac{b}{a} + \log a\Big) \\ &\leq a\Big(\log\frac{z}{a} + 1\Big) \\ &\leq a\Big(\frac{z}{a} - 1 + 1\Big) \\ &= z. \end{aligned}$$

To prove (b), first off, note that under the assumption that $\frac{b}{a} + \log a > 1$, then $-\frac{1}{e} < -\frac{1}{ae^{b/a}} < 0$, so $W_{-1}(-\frac{1}{ae^{b/a}})$ is well-defined. Next, assumption $z \geq -aW_{-1}\big(-\frac{1}{ae^{b/a}}\big)$ can be rewritten as

$$-\frac{z}{a} \leq W_{-1}\Big(-\frac{1}{ae^{b/a}}\Big). \qquad (17)$$

At this point, noting that the inverse of W_{-1} (namely $W_{-1}^{-1}(s) = se^s$, where $s \in (-\infty, -1]$) is a monotonically decreasing function, applying the inverse of W_{-1} to both sides of the above inequality yields

$$-\frac{z}{a}e^{-z/a} \geq -\frac{1}{ae^{b/a}}.$$

Rearranging terms yields $z \geq a\log z + b$, as desired.

Lastly, the proof for (c) just builds on (b). Using Theorem 1 of Chatzigeorgiou (2013),

$$W_{-1}\Big(-\frac{1}{ae^{b/a}}\Big) > -1-\sqrt{2\log(ae^{b/a-1})}-\log(ae^{b/a-1}).$$

A sufficient condition that guarantees inequality (17) to hold is that

$$-\frac{z}{a} \leq -1 - \sqrt{2\log(ae^{b/a-1})} - \log(ae^{b/a-1}),$$

i.e.,

$$z \geq a\big(1 + \sqrt{2\log(ae^{b/a-1})} + \log(ae^{b/a-1})\big). \qquad \square$$

Using Lemma D.1 (with $a = \frac{2\cdot 649}{n\theta^4 p_{\min}(\frac{\theta}{18\Lambda})^{d/\alpha}}$, $b = 32n^2$, $c = \frac{d}{\alpha} + 2$, and $z = \varepsilon$) and a bit of algebra, inequality (15) holds if

$$\begin{aligned} n &\geq \Big(\frac{e}{\chi}\Big)^{\frac{1}{\frac{2d}{\alpha}+5}}, \\ \varepsilon &\geq \Big[\frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)\theta^4 p_{\min}(\frac{\theta}{18\Lambda})^{\frac{d}{\alpha}}} \cdot \frac{1}{n}\cdot\log(\chi^{\frac{1}{\frac{2d}{\alpha}+5}}n)\Big]^{\frac{1}{\frac{d}{\alpha}+2}}, \end{aligned}$$

where

$$\chi := \frac{(\frac{d}{\alpha}+2)(32)^{\frac{d}{\alpha}+2}\theta^4 p_{\min}(\frac{\theta}{18\Lambda})^{\frac{d}{\alpha}}}{2\cdot 649}. \qquad (18)$$

In particular, we shall choose

$$\varepsilon = \Big[\frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)\theta^4 p_{\min}(\frac{\theta}{18\Lambda})^{\frac{d}{\alpha}}} \cdot \frac{1}{n}\cdot\log(\chi^{\frac{1}{\frac{2d}{\alpha}+5}}n)\Big]^{\frac{1}{\frac{d}{\alpha}+2}}.$$

To make sure that $\varepsilon \leq \frac{18\Lambda(r^*)^\alpha}{\theta}$, we require that

$$n \geq \frac{2\cdot 649}{(\frac{d}{\alpha}+2)p_{\min}(18\theta\Lambda)^2(r^*)^{2\alpha+d}}\big[(\tfrac{2d}{\alpha}+5)\log n + \log\chi\big]. \qquad (19)$$

Using Lemma D.2 (with $a = \frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)p_{\min}(18\theta\Lambda)^2(r^*)^{2\alpha+d}}$, $b = \frac{2\cdot 649\cdot\log\chi}{(\frac{d}{\alpha}+2)p_{\min}(18\theta\Lambda)^2(r^*)^{2\alpha+d}}$, and $z = n$), and defining

$$u := \log\Big(\Big[\frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)p_{\min}(18\theta\Lambda)^2(r^*)^{2\alpha+d}}\Big]\frac{\chi^{\frac{1}{\frac{2d}{\alpha}+5}}}{e}\Big),$$

then condition (19) holds if $u \leq 0$ or, in the event that $u > 0$, if we further constrain n to satisfy

$$n \geq \frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)p_{\min}(18\theta\Lambda)^2(r^*)^{2\alpha+d}}(1+\sqrt{2u}+u).$$

In summary, define

$$\begin{aligned} c_1 &:= \frac{1}{2}p_{\min}^{\frac{2\alpha}{2\alpha+d}}\Big(\frac{649(5\alpha+2d)}{162(2\alpha+d)}\Big)^{\frac{d}{2\alpha+d}}\Big(\frac{1}{\Lambda\theta}\Big)^{\frac{2d}{2\alpha+d}}, \\ c_2 &:= \chi^{\frac{1}{2d+5}} = \Big[\frac{512(2\alpha+d)p_{\min}\theta^4}{649\alpha}\Big(\frac{16\theta}{9\Lambda}\Big)^{\frac{d}{\alpha}}\Big]^{\frac{\alpha}{5\alpha+2d}}, \end{aligned}$$

$$
\begin{aligned}
c_3 &:= \Big[\frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)\theta^4 p_{\min}(\frac{\theta}{18\Lambda})^{\frac{d}{\alpha}}}\Big]^{\frac{1}{\frac{d}{\alpha}+2}} \\
&= \Big[\frac{1298(5\alpha+2d)}{(2\alpha+d)p_{\min}\theta^4}\Big(\frac{18\Lambda}{\theta}\Big)^{\frac{d}{\alpha}}\Big]^{\frac{\alpha}{2\alpha+d}} \\
c_4 &:= \frac{2\cdot 649\cdot(\frac{2d}{\alpha}+5)}{(\frac{d}{\alpha}+2)p_{\min}(18\theta\Lambda)^2(r^*)^{2\alpha+d}} \\
&= \frac{649(5\alpha+2d)}{162(2\alpha+d)p_{\min}(r^*)^{2\alpha+d}\theta^2\Lambda^2}.
\end{aligned}
$$

Note that $u = \log(c_2 c_4/e)$. Set

$$
n_0 := \begin{cases} \lceil \frac{e^{\alpha/(5\alpha+2d)}}{c_2} \rceil & \text{if } \frac{c_2c_4}{e} \le 1, \\ \lceil \max\{ \frac{e^{\alpha/(5\alpha+2d)}}{c_2}, \\ \quad c_4(1+\sqrt{2\log\frac{c_2c_4}{e}}+\log\frac{c_2c_4}{e})\}\rceil & \text{if } \frac{c_2c_4}{e} > 1. \end{cases}
$$

Then for any $n \ge n_0$, the conditions that we discussed for n are met, so we can choose

$$
\begin{aligned}
k_n &:= \Big\lfloor c_1 n^{\frac{2\alpha}{2\alpha+d}}\big(\log(c_2 n)\big)^{\frac{d}{2\alpha+d}}\Big\rfloor, \\
\varepsilon_n &:= c_3\Big(\frac{\log(c_2n)}{n}\Big)^{\frac{\alpha}{2\alpha+d}} \qquad (20)
\end{aligned}
$$

to achieve

$$
\mathbb{P}\Big(\sup_{t\in[0,\tau]}|\widehat{S}^{k_n\text{-NN}}(t|x) - S(t|x)| \ge \varepsilon_n\Big) \le \frac{1}{n^2}.
$$

As a result, we have

$$
\begin{aligned}
&\sum_{n=1}^{\infty}\mathbb{P}\Big(\sup_{t\in[0,\tau]}|\widehat{S}^{k_n\text{-NN}}(t|x) - S(t|x)| \ge \varepsilon_n\Big) \\
&\le n_0 + \sum_{n=n_0}^{\infty}\frac{1}{n^2} \le n_0 + \sum_{n=1}^{\infty}\frac{1}{n^2} = n_0 + \frac{\pi^2}{6} < \infty,
\end{aligned}
$$

so by the Borel-Cantelli lemma,

$$
\mathbb{P}\Big(\limsup_{n\to\infty}\Big\{\sup_{t\in[0,\tau]}|\widehat{S}^{k_n\text{-NN}}(t|x) - S(t|x)| \ge \varepsilon_n\Big\}\Big) = 0.
$$

E. Proof of Theorem A.1

The proof of the fixed-radius NN estimator is similar to that of the k-NN estimator and actually does not require the more nuanced analysis of Chaudhuri & Dasgupta (2014). In particular, in proving the k-NN estimator guarantee, we took the expectation $\mathbb{E}[\cdot|\widetilde{X}]$, where $\widetilde{X}$ was the feature vector of the $(k+1)$-st nearest neighbor of x. This conditioning made the k nearest neighbors appear i.i.d. The analysis for the fixed-radius NN estimator is simpler in that with the threshold distance $h > 0$ fixed, the training data that land within distance h are i.i.d. as is. However, the bad events do slightly change since now there could be no neighbors found within distance h of x. Whereas previously the number of neighbors was fixed, now the number of neighbors being random. Thus, instead of conditioning on the $(k+1)$-st nearest neighbor, we now condition on the number of neighbors.

We focus on the proof of the main fixed-radius NN estimator nonasymptotic bound (7). The proof of the strong consistency result is the same as that of the k-NN estimator with the only change being that we do not need to worry about k (in proving the k-NN strong consistency result, Corollary 3.1, we had a short extra step that makes sure that k can be chosen to be a valid integer; for establishing the fixed-radius NN strong consistency result, we do not need this extra step although even if we use it, the choices for c_1, c_2, c_3, and n_0 still work). We then pick the threshold distance to be $h = h^* = (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$, where ε is chosen as in equation (20).

We proceed to proving the nonasymptotic bound (7). Let $x \in \text{supp}(\mathbb{P}_X)$ and $N_{x,h} = |\mathcal{N}_{\text{NN}(h)}(x)|$ denote the number of neighbors found within distance h of x. Using the same reasoning as for the k-NN estimator,

$$
\begin{aligned}
&\log \widehat{S}^{\text{NN}(h)}(t|x) \\
&= \log \prod_{i\in\mathcal{N}_{\text{NN}(h)}(x)}\Big(\frac{d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(Y_i)}{d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(Y_i)+1}\Big)^{\delta_i\mathbb{1}\{Y_i\le t\}} \\
&= V_1(t|x) + V_2(t|x) + V_3(t|x),
\end{aligned}
$$

where

$$
\begin{aligned}
V_1(t|x) &= -\frac{1}{N_{x,h}}\sum_{\substack{i\in\mathcal{N}_{\text{NN}(h)}(x)\\ \text{s.t. } Y_i\le t}}\delta_i\frac{1}{S_{\text{Y}}(Y_i|x)}, \\
V_2(t|x) &= -\frac{1}{N_{x,h}}\sum_{\substack{i\in\mathcal{N}_{\text{NN}(h)}(x)\\ \text{s.t. } Y_i\le t}}\delta_i\Big[\frac{N_{x,h}}{d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(Y_i)+1} \\
&\qquad\qquad - \frac{1}{S_{\text{Y}}(Y_i|x)}\Big], \\
V_3(t|x) &= -\sum_{\substack{i\in\mathcal{N}_{\text{NN}(h)}(x)\\ \text{s.t. } Y_i\le t}}\delta_i\sum_{\ell=2}^{\infty}\frac{1}{\ell(d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(Y_i)+1)^{\ell}}.
\end{aligned}
$$

Defining $\widehat{S}_{\text{Y}}^{\text{NN}(h)}(s|x) := \frac{d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(s)}{N_{x,h}}$, then the bad events are:

- $\mathcal{E}^{\text{NN}(h)}_{\text{few neighbors}}(x) := \{N_{x,h} \le \frac{n\mathbb{P}_X(\mathcal{B}_{x,h})}{2}\}$
- $\mathcal{E}^{\text{NN}(h)}_{\text{bad }\tau}(x) := \{d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(\tau) \le \frac{N_{x,h}\theta}{2}\}$
- $\mathcal{E}^{\text{NN}(h)}_{\text{bad EDF}}(x) :=$ $\{\sup_{s\ge 0}|\widehat{S}_{\text{Y}}^{\text{NN}(h)}(s|x) - \mathbb{E}[\widehat{S}_{\text{Y}}^{\text{NN}(h)}(s|x)|N_{x,h}]| > \frac{\varepsilon\theta^2}{36}\}$
- $\mathcal{E}^{\text{NN}(h)}_{\text{bad }V_1}(t,x) := \{|V_1(t|x) - \mathbb{E}[V_1(t|x)|N_{x,h}]| \ge \frac{\varepsilon}{18}\}$

Once all of these bad events do not happen, then applying a very similar proof to the k-NN estimator yields Theorem 3.1. Note that as before, we actually want $\mathcal{E}_{\text{bad } V_1}^{\text{NN}(h)}(t,x)$ to hold for a finite collection of times $t=\eta_1,\dots,\eta_{L(\varepsilon)}$ within interval $[0,\tau]$.

We remark that the union bounding over the bad events is done slightly differently for the fixed-radius NN estimator. In particular, at least one of the bad events happening can actually be written as the union over the following events:

- $\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)$
- $\mathcal{E}_{\text{bad } \tau}^{\text{NN}(h)}(x) \cap [\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c$
- $\mathcal{E}_{\text{bad EDF}}^{\text{NN}(h)}(x) \cap [\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c$
- $\mathcal{E}_{\text{bad } V_1}^{\text{NN}(h)}(t,x) \cap [\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c$ for $t=\eta_1,\dots,\eta_{L(\varepsilon)}$

We use the fact that for any two events $\mathcal{E}_1$ and $\mathcal{E}_2$, $\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2)=\mathbb{P}(\mathcal{E}_1)\mathbb{P}(\mathcal{E}_2|\mathcal{E}_1)\le\mathbb{P}(\mathcal{E}_2|\mathcal{E}_1)$. Then

$$\begin{aligned}
&\mathbb{P}(\text{at least one bad event happens})\\
&\quad\le \mathbb{P}\big(\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)\big)\\
&\qquad+\mathbb{P}\big(\mathcal{E}_{\text{bad } \tau}^{\text{NN}(h)}(x)\cap[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\
&\qquad+\mathbb{P}\big(\mathcal{E}_{\text{bad EDF}}^{\text{NN}(h)}(x)\cap[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\
&\qquad+\sum_{\ell=1}^{L(\varepsilon)}\mathbb{P}\big(\mathcal{E}_{\text{bad } V_1}^{\text{NN}(h)}(\eta_\ell,x)\cap[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\
&\quad\le \mathbb{P}\big(\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)\big)\\
&\qquad+\mathbb{P}\big(\mathcal{E}_{\text{bad } \tau}^{\text{NN}(h)}(x)\,\big|\,[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\
&\qquad+\mathbb{P}\big(\mathcal{E}_{\text{bad EDF}}^{\text{NN}(h)}(x)\,\big|\,[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\
&\qquad+\sum_{\ell=1}^{L(\varepsilon)}\mathbb{P}\big(\mathcal{E}_{\text{bad } V_1}^{\text{NN}(h)}(\eta_\ell,x)\,\big|\,[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big).
\end{aligned}$$

The rest of this section is on giving upper bounds for the four different probability terms that appear on the RHS, and also on why when all of these bad events do not happen, we indeed have $|\widehat{S}^{\text{NN}(h)}(t|x)-S(t|x)|\le\varepsilon/3$ for any $t\in[0,\tau]$, which using the argument from proving Theorem 3.1 with carefully chosen points $\eta_1,\dots,\eta_{L(\varepsilon)}$ is sufficient to guarantee that $\sup_{t\in[0,\tau]}|\widehat{S}^{\text{NN}(h)}(t|x)-S(t|x)|\le\varepsilon$.

Lemma E.1. *Under Assumption A1, let $x\in supp(\mathbb{P}_X)$. Let $N_{x,h}$ be the number of nearest neighbors found within distance h of x. Then*

$$\mathbb{P}\big(\mathcal{E}_{\textit{few neighbors}}^{\text{NN}(h)}(x)\big)\le\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})}{8}\Big).$$

Proof. Since $N_{x,h}\sim\text{Binomial}(n,\mathbb{P}_X(\mathcal{B}_{x,h}))$, the claim follows from applying Chernoff's inequality for the binomial distribution. □

Lemma E.2. *Under Assumptions A1–A3, let $x\in supp(\mathbb{P}_X)$. We have*

$$\begin{aligned}&\mathbb{P}\big(\mathcal{E}_{\textit{bad } \tau}^{\text{NN}(h)}(x)\,\big|\,[\mathcal{E}_{\textit{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\&\quad\le\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\theta}{16}\Big).\end{aligned}$$

Proof. By conditioning on $N_{x,h}=k$ for any $k\in\{1,\dots,n\}$, then a proof similar to that of Lemma C.1 yields

$$\mathbb{P}\Big(d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(\tau)\le\frac{k\theta}{2}\,\Big|\,N_{x,h}=k\Big)\le\exp\Big(-\frac{k\theta}{8}\Big).$$

We now use a worst-case argument that appears many times in later proofs. Let k_0 be the smallest integer larger than $\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h})$. Then

$$\begin{aligned}
&\mathbb{P}\big(\mathcal{E}_{\text{bad } \tau}^{\text{NN}(h)}(x)\,\big|\,[\mathcal{E}_{\text{few neighbors}}^{\text{NN}(h)}(x)]^c\big)\\
&\quad=\mathbb{P}\Big(d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(\tau)\le\frac{N_{x,h}\theta}{2}\,\Big|\,N_{x,h}\ge k_0\Big)\\
&\quad=\frac{\left[\begin{array}{l}\sum_{k=k_0}^n\mathbb{P}(N_{x,h}=k)\\ \times\mathbb{P}\big(d^+_{\mathcal{N}_{\text{NN}(h)}(x)}(\tau)\le\frac{k\theta}{2}\,\big|\,N_{x,h}=k\big)\end{array}\right]}{\mathbb{P}(N_{x,h}\ge k_0)}\\
&\quad\le\frac{\sum_{k=k_0}^n\mathbb{P}(N_{x,h}=k)\exp\big(-\frac{k_0\theta}{8}\big)}{\mathbb{P}(N_{x,h}\ge k_0)}\\
&\quad=\exp\big(-\frac{k_0\theta}{8}\big)\\
&\quad\le\exp\big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\theta}{16}\big).
\end{aligned}$$
□

Lemma E.3. *Under Assumptions A1–A3, let $x\in supp(\mathbb{P}_X)$ and $t\in[0,\tau]$. When bad events $\mathcal{E}_{\textit{few neighbors}}^{\text{NN}(h)}(x)$ and $\mathcal{E}_{\textit{bad } \tau}^{\text{NN}(h)}(x)$ do not happen,*

$$\begin{aligned}
&|V_2(t|x)|\\
&\le\frac{2}{N_{x,h}\theta^2}\\
&\quad+\frac{2}{\theta^2}\sup_{s\in[0,\tau]}|S_{\text{Y}}(s|x)-\mathbb{E}[\widehat{S}_{\text{Y}}^{\text{NN}(h)}(s|x)\,|\,N_{x,h}]|\\
&\quad+\frac{2}{\theta^2}\sup_{s\ge0}|\widehat{S}_{\text{Y}}^{\text{NN}(h)}(s|x)-\mathbb{E}[\widehat{S}_{\text{Y}}^{\text{NN}(h)}(s|x)\,|\,N_{x,h}]|,
\end{aligned}$$

where the RHS is a function of random variable $N_{x,h}$ (which is greater than 0 since bad event $\mathcal{E}_{\textit{few neighbors}}^{\text{NN}(h)}(x)$ does not happen).

Proof. See the proof of Lemma C.3 as given in Appendix C.3, where we replace $\mathcal{I}=\mathcal{N}_{k\text{-NN}}(x)$ with $\mathcal{I}=\mathcal{N}_{\text{NN}(h)}(x)$, k with $N_{x,h}$, and bad event $\mathcal{E}_{\text{bad } \tau}^{k\text{-NN}}(x)$ with $\mathcal{E}_{\text{bad } \tau}^{\text{NN}(h)}(x)$. Also instead of using expectation $\mathbb{E}[\,\cdot\,|\widetilde{X}]$ (i.e., conditioning on the $(k+1)$-st nearest neighbor), we use $\mathbb{E}[\,\cdot\,|N_{x,h}]$. □

Lemma E.4. *Under Assumptions A1–A3, let $x \in$ supp$(\mathbb{P}_X)$. We have*

$$\mathbb{P}\big(\mathcal{E}^{\text{NN}(h)}_{\textit{bad EDF}}(x) \,\big|\, [\mathcal{E}^{\text{NN}(h)}_{\textit{few neighbors}}(x)]^c\big) \le 2\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^4}{1296}\Big).$$

Proof. By conditioning on $N_{x,h} = k$ for any $k \in \{1,\dots,n\}$, then $1 - \widehat{S}^{\text{NN}(h)}_{\text{Y}}(s|x)$ is an empirical distribution with samples drawn i.i.d. from CDF $1 - \mathbb{P}(Y > s \,|\, X \in \mathcal{B}_{x,h})$. By the DKW inequality,

$$\mathbb{P}\big(\mathcal{E}^{\text{NN}(h)}_{\text{bad EDF}}(x) \,\big|\, N_{x,h} = k\big) \le 2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big).$$

A worst-case argument similar to the one in the ending of Lemma E.2's proof says that $\mathbb{P}\big(\mathcal{E}^{\text{NN}(h)}_{\text{bad EDF}}(x) \,\big|\, N_{x,h} > \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h})\big)$ satisfies the above inequality with k replaced by $\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h})$. □

Lemma E.5. *Under Assumptions A1–A3, let $x \in$ supp$(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. We have*

$$\mathbb{P}\big(\mathcal{E}^{\text{NN}(h)}_{\textit{bad } V_1}(t,x) \,\big|\, [\mathcal{E}^{\text{NN}(h)}_{\textit{few neighbors}}(x)]^c\big) \le 2\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^2}{324}\Big).$$

Proof. Note that $V_1(t|x)$ and $\mathbb{E}[V_1(t|x) \,|\, N_{x,h}]$ can both be written as functions of random variable $N_{x,h}$, provided that $N_{x,h}$ is positive. Specifically,

$$V_1(t|x) = \frac{1}{N_{x,h}}\sum_{\ell=1}^{N_{x,h}} \xi_\ell,$$

where random variables $\xi_1,\dots,\xi_{N_{x,h}}$ are sampled i.i.d. from the same distribution as random variable $-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\text{Y}}(Y|x)}$ (where feature vector X is sampled from $\mathbb{P}_X$ restricted to ball $\mathcal{B}_{x,h}$, and observed time Y and censoring indicator δ as sampled as usual conditioned on X). Each ξ_ℓ is bounded in $[-\frac{1}{S_{\text{Y}}(t|x)}, 0]$ and has expectation

$$\begin{aligned}\overline{\xi}(N_{x,h}) &:= \mathbb{E}_{Y,\delta}\Big[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\text{Y}}(Y|x)} \,\Big|\, X \in \mathcal{B}_{x,h}\Big] \\ &= \mathbb{E}[V_1(t|x) \,|\, N_{x,h}].\end{aligned}$$

Thus, using Hoeffding's inequality, for any $k \in \{1,\dots,n\}$,

$$\begin{aligned}&\mathbb{P}\Big(|V_1(t|x) - \mathbb{E}[V_1(t|x) \,|\, N_{x,h}]| \ge \frac{\varepsilon}{18} \,\Big|\, N_{x,h} = k\Big) \\ &= \mathbb{P}\Big(\Big|\frac{1}{N_{x,h}}\sum_{\ell=1}^{N_{x,h}} \xi_\ell - \overline{\xi}(N_{x,h})\Big| \ge \frac{\varepsilon}{18} \,\Big|\, N_{x,h} = k\Big) \\ &\le 2\exp\Big(-\frac{k\varepsilon^2[S_{\text{Y}}(t|x)]^2}{162}\Big) \le 2\exp\Big(-\frac{k\varepsilon^2\theta^2}{162}\Big).\end{aligned}$$

A worst-case argument similar to the one in the ending of Lemma E.2's proof yields

$$\mathbb{P}\big(\mathcal{E}^{\text{NN}(h)}_{\text{bad } V_1}(t,x) \,\big|\, [\mathcal{E}^{\text{NN}(h)}_{\text{few neighbors}}(x)]^c\big) \le \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^2}{324}\Big).$$ □

Then when none of the bad events happen,

$$\begin{aligned}&|\log \widehat{S}^{\text{NN}(h)}(t|x) - \log S(t|x)| \\ &\le |V_1(t|x) - \mathbb{E}[V_1(t|x) \,|\, N_{x,h}]| \\ &\quad + |\mathbb{E}[V_1(t|x) \,|\, N_{x,h}] - \log S(t|x)| + \frac{2}{N_{x,h}\theta^2} \\ &\quad + \frac{2}{\theta^2}\sup_{s\in[0,\tau]} |S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}^{\text{NN}(h)}_{\text{Y}}(s|x) \,|\, N_{x,h}]| \\ &\quad + \frac{2}{\theta^2}\sup_{s\ge 0} |\widehat{S}^{\text{NN}(h)}_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}^{\text{NN}(h)}_{\text{Y}}(s|x) \,|\, N_{x,h}]|, \\ &\quad + V_3(t|x).\end{aligned}$$

The 1st and 5th terms on the RHS are at most $\frac{\varepsilon}{18}$ since bad events $\mathcal{E}^{\text{NN}(h)}_{\text{bad } V_1}(t,x)$ and $\mathcal{E}^{\text{NN}(h)}_{\text{bad EDF}}(x)$ do not happen (these bad events also rely on $\mathcal{E}^{\text{NN}(h)}_{\text{few neighbors}}(x)$ not happening so that $N_{x,h} > 0$). The theorem assumes that $n \ge \frac{144}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,h})}$, so the 3rd term is at most $\frac{2}{N_{x,h}\theta^2} < \frac{4}{n\mathbb{P}_X(\mathcal{B}_{x,h})\theta^2} \le \frac{\varepsilon}{36} < \frac{\varepsilon}{18}$. The 2nd, 4th, and 6th terms can be bounded in a similar manner as we did for the k-NN estimator.

Lemma E.6. *Under Assumptions A1–A4 (this lemma uses Hölder continuity of $S_{\text{C}}(t|\cdot)f_{\text{T}}(t|\cdot)$), let $x \in$ supp$(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. If bad event $\mathcal{E}^{\text{NN}(h)}_{\textit{few neighbors}}(x)$ does not happen, and the threshold distance satisfies $h \le [\frac{\varepsilon\theta}{18(\lambda_{\text{T}}\tau + (f^*_{\text{T}}\lambda_{\text{C}}\tau^2)/2)}]^{1/\alpha}$, then*

$$|\mathbb{E}[V_1(t|x) \,|\, N_{x,h}] - \log S(t|x)| \le \frac{\varepsilon}{18}.$$

Proof. See the proof for Lemma C.6 as given in Appendix C.6. The main change is that we don't have to condition on the $(k+1)$-st nearest neighbor of x. Instead, conditioning on $N_{x,h} = k$ for integer k in $(\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h}), n]$, then $V_1(t|x)$ is the average of k i.i.d. bounded random variables each with expectation $\mathbb{E}_{Y,\delta}[-\frac{\delta\mathbb{1}\{Y\le t\}}{S_{\text{Y}}(Y|x)} \,|\, X \in \mathcal{B}_{x,h}]$. □

Lemma E.7. *Under Assumptions A1–A4 (this lemma uses Hölder continuity of $S_{\text{Y}}(t|\cdot)$), let $x \in$ supp$(\mathbb{P}_X)$ and $\varepsilon \in (0,1)$. If bad event $\mathcal{E}^{\text{NN}(h)}_{\textit{few neighbors}}(x)$ does not happen, and the threshold distance satisfies $h \le [\frac{\varepsilon\theta^2}{36(\lambda_{\text{T}}+\lambda_{\text{C}})\tau)}]^{1/\alpha}$, then*

$$\frac{2}{\theta^2}\sup_{s\in[0,\tau]} |S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}^{\text{NN}(h)}_{\text{Y}}(s|x) \,|\, N_{x,h}]| \le \frac{\varepsilon}{18}.$$

Proof. See the proof for Lemma C.7 as given in Appendix C.7. Once again, the main change is that we don't have to condition on the $(k+1)$-st nearest neighbor of x. Instead, conditioning on $N_{x,h} = k$ for integer k in $(\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,h}), n]$, then $1 - \widehat{S}_{\mathrm{Y}}^{\mathrm{NN}(h)}(s|x)$ is an empirical distribution constructed based on i.i.d. samples from CDF $1 - \mathbb{E}[\widehat{S}_{\mathrm{Y}}^{\mathrm{NN}(h)}(s|x) \mid N_{x,h} = k] = 1 - \mathbb{P}(Y > s \mid X \in \mathcal{B}_{x,h})$. □

Lemma E.8. *Under Assumptions A1–A3, let $x \in supp(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. If bad events $\mathcal{E}^{\mathrm{NN}(h)}_{\textit{few neighbors}}(x)$ and $\mathcal{E}^{\mathrm{NN}(h)}_{\textit{bad } \tau}(x)$ do not happen, and $n \geq \frac{144}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,h})}$, then $|V_3(t|x)| \leq \varepsilon/18$.*

Proof. See the proof of Lemma C.8 as given in Appendix C.8, where we replace $\mathcal{I} = \mathcal{N}_{k\text{-NN}}(x)$ with $\mathcal{I} = \mathcal{N}_{\mathrm{NN}(h)}(x)$, k with $N_{x,h}$, and bad event $\mathcal{E}^{k\text{-NN}}_{\text{bad } \tau}(x)$ with $\mathcal{E}^{\mathrm{NN}(h)}_{\text{bad } \tau}(x)$. □

F. Proof of Theorem 3.2

First off, we state a longer version of the kernel pointwise theorem that includes a strong consistency result. This is the version of the theorem we prove in this section.

Theorem F.1 (Kernel pointwise guarantees)**.** *Under Assumptions A1–A5, let $\varepsilon \in (0,1)$ be a user-specified error tolerance. Suppose that the threshold distance satisfies $h \in (0, \frac{1}{\phi}(\frac{\varepsilon\theta}{18\Lambda_K})^{1/\alpha}]$, and the number of training data satisfies $n \geq \frac{144}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,\phi h})\kappa}$. For any $x \in supp(\mathbb{P}_X)$,*

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,\tau]} |\widehat{S}^K(t|x;h) - S(t|x)| > \varepsilon\Big) \\
&\quad\leq \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\theta}{16}\Big) + \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})}{8}\Big) \\
&\qquad + \frac{216}{\varepsilon\theta^2\kappa}\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\varepsilon^2\theta^4\kappa^4}{11664}\Big) \\
&\qquad + \frac{8}{\varepsilon}\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\varepsilon^2\theta^2\kappa^2}{324}\Big).
\end{aligned}$$

Moreover, if there exist constants $p_{\min} > 0$, $d > 0$, and $r^* > 0$ such that $\mathbb{P}_X(\mathcal{B}_{x,r}) \geq p_{\min}r^d$ for all $r \in (0, r^*]$, then we get the same strong consistency behavior as in Theorem A.1 with the numbers c_1', c_2 and c_3 replaced by $c_1'' = \Theta\big(\frac{1}{\phi(\theta\Lambda_K\kappa^2)^{2/(2\alpha+d)}}\big)$, $c_2'' = \Theta\big(\frac{1}{(\theta\Lambda_K)^{d/(5\alpha+2d)}\kappa^{(d-2\alpha)/(5\alpha+2d)}}\big)$, and $c_3'' = \Theta\big(\frac{(\Lambda_K)^{d/(2\alpha+d)}}{\theta^{(4\alpha+d)/(2\alpha+d)}\kappa^{4\alpha/(2\alpha+d)}}\big)$.

For the kernel estimator, there is a fair amount more notation to keep track of. To keep the equations from becoming unwieldy, we adopt the following abbreviations. First off, the training subjects with nonzero kernel weight are precisely the ones with feature vectors landing in the ball $\mathcal{B}_{x,\phi h}$. We denote the number of these subjects as $N := N_{x,\phi h} \sim \text{Binomial}(n, \mathbb{P}_X(\mathcal{B}_{x,\phi h}))$. We denote their data points as $(X_{(1)}, Y_{(1)}, \delta_{(1)})$, …, $(X_{(N)}, Y_{(N)}, \delta_{(N)})$; we treat the ordering of these points as uniform at random (the points could be thought of as being generated i.i.d. first by sampling a feature vector X from $\mathbb{P}_X$ restricted to $\mathcal{B}_{x,\phi h}$, and then sampling observed time Y and censoring indicator δ as usual). We use the abbreviations $K_{(i)} := K(\frac{\rho(x,X_{(i)})}{h})$, $d_K^+(t) := \sum_{j=1}^N K_{(j)}\mathbb{1}\{Y_{(j)} > t\}$, and

$$\widehat{S}_{\mathrm{Y}}^K(t) := \frac{d_K^+(t)}{\sum_{\ell=1}^N K_{(\ell)}} = \sum_{j=1}^N \frac{K_{(j)}}{\sum_{\ell=1}^N K_{(\ell)}}\mathbb{1}\{Y_{(j)} > t\}.$$

Let $\mathbb{E}_{\{Y\}}$ denote the expectation only over the nearest neighbors' observed times $Y_{(1)}, \dots, Y_{(N)}$ (so we are conditioning on $N, X_{(1)}, \dots X_{(N)}$). Similarly, we let $\mathbb{E}_{\{Y,\delta\}}$ denote the expectation only over only the nearest neighbors' observed times and censoring indicators $(Y_{(1)}, \delta_{(1)}), \dots, (Y_{(N)}, \delta_{(N)})$.

Using the same reasoning as for the k-NN estimator,

$$\begin{aligned}
\log \widehat{S}^K(t|x;h) &= \log \prod_{i=1}^N \Big(\frac{d_K^+(Y_{(i)})}{d_K^+(Y_{(i)}) + K_{(i)}}\Big)^{\delta_{(i)}\mathbb{1}\{Y_{(i)}\leq t\}} \\
&= W_1(t|x) + W_2(t|x) + W_3(t|x),
\end{aligned}$$

where

$$W_1(t|x) = -\sum_{i=1}^N \frac{K_{(i)}\delta_{(i)}\mathbb{1}\{Y_{(i)} \leq t\}\frac{1}{S_{\mathrm{Y}}(Y_{(i)}|x)}}{\sum_{j=1}^N K_{(j)}},$$

$$W_2(t|x) = -\sum_{i=1}^N \frac{K_{(i)}\delta_{(i)}\mathbb{1}\{Y_{(i)} \leq t\}\big[\frac{\sum_{\ell=1}^N K_{(\ell)}}{d_K^+(Y_{(i)})+K_{(i)}} - \frac{1}{S_{\mathrm{Y}}(Y_{(i)}|x)}\big]}{\sum_{j=1}^N K_{(j)}},$$

$$W_3(t|x) = -\sum_{i=1}^N \delta_{(i)}\mathbb{1}\{Y_{(i)} \leq t\}\sum_{\ell=2}^\infty \frac{1}{\ell(\frac{d_K^+(Y_{(i)})}{K_{(i)}} + 1)^\ell}.$$

The bad events are as follows:

- $\mathcal{E}^{\mathrm{NN}(\phi h)}_{\text{few neighbors}}(x)$ is the same bad event as for the fixed-radius NN estimator except using threshold distance ϕh instead of h
- $\mathcal{E}^{\mathrm{NN}(\phi h)}_{\text{bad } \tau}(x)$ is another bad event borrowed from the fixed-radius NN estimator
- $\mathcal{E}^{\text{kernel}}_{\text{bad weighted EDF}}(x) := \big\{\sup_{s\geq 0}\big|\widehat{S}_{\mathrm{Y}}^K(t) - \mathbb{E}_{\{Y\}}\big[\widehat{S}_{\mathrm{Y}}^K(t)\big]\big| > \frac{\varepsilon\theta^2K(\phi)}{36K(0)}\big\}$ is analogous to event $\mathcal{E}^{\mathrm{NN}(h)}_{\text{bad EDF}}(x)$
- $\mathcal{E}^{\text{kernel}}_{\text{bad } W_1}(t,x) := \{|W_1(t|x) - \mathbb{E}_{\{Y,\delta\}}[W_1(t|x)]| \geq \frac{\varepsilon}{18}\}$ is analogous to event $\mathcal{E}^{\mathrm{NN}(h)}_{\text{bad } V_1}(t,x)$, and as before we ask that this holds at specific points $t = \eta_1, \dots, \eta_{L(\varepsilon)}$ (using the same construction as in the proof of Theorem 3.1)

We show how to prevent bad events $\mathcal{E}^{\text{kernel}}_{\text{bad weighted EDF}}(x)$ and $\mathcal{E}^{\text{kernel}}_{\text{bad } W_1}(t,x)$in the next two lemmas.

Lemma F.1. *Under Assumptions A1–A3 and A5, let $x \in$ supp$(\mathbb{P}_X)$ and $\varepsilon \in (0,1)$. Then*

$$\begin{aligned}&\mathbb{P}\big(\mathcal{E}^{\textit{kernel}}_{\textit{bad weighted EDF}}(x) \,\big|\, [\mathcal{E}^{\text{NN}(\phi h)}_{\textit{few neighbors}}(x)]^c\big)\\ &\quad\le \frac{216K(0)}{\varepsilon\theta^2 K(\phi)} \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\varepsilon^2\theta^4K^4(\phi)}{11664K^4(0)}\Big).\end{aligned}$$

Proof. Conditioned on $N, X_{(1)}, \dots, X_{(N)}$ with N positive (recall that $K_{(i)}$ depends on $X_{(i)}$), then $\widehat{S}^K_{\text{Y}}(t)$ appears to be constructed from independent weighted samples, where the weights are deterministic and, moreover, $1 - \widehat{S}^K_{\text{Y}}(t)$ is precisely a weighted empirical distribution with samples drawn i.i.d. from CDF $1 - \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)]$. Thus, by conditioning on the event

$$\mathcal{A} := \{N = k, X_{(1)} = x_{(1)}, \dots, X_{(k)} = x_{(k)}\}$$

for any integer $k \in \{1, \dots, n\}$, and any choices for $x_{(1)}, \dots, x_{(k)} \in \mathcal{B}_{x,\phi h}$, we can then apply Proposition 3.1 (with $\ell = k$ and noting that $\sum_{i=1}^k w_i \ge kK(\phi)$ and $\sum_{i=1}^k w_i^2 \le kK^2(0)$) to get

$$\begin{aligned}&\mathbb{P}\Big(\sup_{t\ge0} |\widehat{S}^K_{\text{Y}}(t) - \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)]| > \frac{\varepsilon\theta^2K(\phi)}{36K(0)} \,\Big|\, \mathcal{A}\Big)\\ &\quad\le \frac{216K(0)}{\varepsilon\theta^2K(\phi)} \exp\Big(-\frac{k\varepsilon^2\theta^4K^4(\phi)}{5832K^4(0)}\Big).\end{aligned}$$

This inequality holds for all $x_{(1)}, \dots, x_{(k)} \in \mathcal{B}_{x,\phi h}$, so we can marginalize over $X_{(1)}, \dots, X_{(k)}$ to get:

$$\begin{aligned}&\mathbb{P}\Big(\sup_{t\ge0} |\widehat{S}^K_{\text{Y}}(t) - \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)]| > \frac{\varepsilon\theta^2K(\phi)}{36K(0)} \,\Big|\, N = k\Big)\\ &\quad\le \frac{216K(0)}{\varepsilon\theta^2K(\phi)} \exp\Big(-\frac{k\varepsilon^2\theta^4K^4(\phi)}{5832K^4(0)}\Big).\end{aligned}$$

Finally, conditioned on $[\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)]^c = \{N > \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\}$, a worst-case argument similar to the one used at the end of Lemma E.2's proof yields the claim. □

Lemma F.2. *Under Assumptions A1–A3 and A5, let $x \in$ supp$(\mathbb{P}_X)$, $t \in [0, \tau]$, and $\varepsilon \in (0,1)$. We have*

$$\begin{aligned}&\mathbb{P}\big(\mathcal{E}^{\textit{kernel}}_{\textit{bad } W_1}(t,x) \,\big|\, [\mathcal{E}^{\text{NN}(\phi h)}_{\textit{few neighbors}}(x)]^c\big)\\ &\quad\le 2\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\varepsilon^2\theta^2K^2(\phi)}{324K^2(0)}\Big).\end{aligned}$$

Proof. The proof is similar to that of Lemma E.5. Note that

$$W_1(t|x) = \sum_{i=1}^N \underbrace{-\frac{K_{(i)}}{\sum_{j=1}^k K_{(j)}} \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}}{S_{\text{Y}}(Y_{(i)}|x)}}_{\text{bounded in } \big[-\big(\frac{K_{(i)}}{\sum_{j=1}^N K_{(j)}}\big)\frac{1}{S_{\text{Y}}(t|x)},\, 0\big]}.$$

Conditioned on $N, X_{(1)}, \dots, X_{(N)}$ with N positive, then $W_1(t|x)$ becomes a sum over independent random variables. Meanwhile, $\mathbb{E}_{\{Y,\delta\}}[W_1(t|x)]$ is precisely the expectation of $W_1(t|x)$ conditioned on $N, X_{(1)}, \dots, X_{(N)}$. Hence, by conditioning on the event

$$\mathcal{A} := \{N = k, X_{(1)} = x_{(1)}, \dots, X_{(k)} = x_{(k)}\}$$

for any $k \in \{1, \dots, n\}$, and any choices of $x_{(1)}, \dots, x_{(k)} \in \mathcal{B}_{x,\phi h}$, and denoting $w_{(i)} := K(\frac{\rho(x,x_{(i)})}{h})$, Hoeffding's inequality gives

$$\begin{aligned}&\mathbb{P}\Big(|W_1(t|x) - \mathbb{E}_{\{Y,\delta\}}[W_1(t|x)]| \ge \frac{\varepsilon}{18} \,\Big|\, \mathcal{A}\Big)\\ &\quad\le 2\exp\Big(-\frac{\varepsilon^2(\sum_{j=1}^k w_{(j)})^2[S_{\text{Y}}(t|x)]^2}{162\sum_{i=1}^k w_{(i)}^2}\Big)\\ &\quad\le 2\exp\Big(-\frac{k\varepsilon^2K^2(\phi)[S_{\text{Y}}(t|x)]^2}{162K^2(0)}\Big)\\ &\quad\le 2\exp\Big(-\frac{k\varepsilon^2K^2(\phi)\theta^2}{162K^2(0)}\Big).\end{aligned}$$

We complete the proof the same way as in Lemma F.1's proof, marginalizing over $X_{(1)}, \dots, X_{(k)}$ and using a worst-case analysis argument to replace k with $\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,\phi h})$. □

Now that we have the bad events sorted out, the argument for why them not happening guarantees that $\sup_{t\in[0,\tau]} |\widehat{S}^K(t|x;h) - S(t|x)| \le \varepsilon$ proceeds in the same manner as for the k-NN and fixed-radius NN analyses. We first upper-bound $|W_2(t|x)|$.

Lemma F.3. *Under Assumptions A1–A3 and A5, let $x \in$ supp$(\mathbb{P}_X)$ and $t \in [0, \tau]$. When bad events $\mathcal{E}^{\text{NN}(\phi h)}_{\textit{few neighbors}}(x)$ and $\mathcal{E}^{\text{NN}(\phi h)}_{\textit{bad } \tau}(x)$ do not happen,*

$$\begin{aligned}|W_2(t|x)| &\le \frac{2K(0)}{NK(\phi)\theta^2}\\ &\quad + \frac{2K(0)}{K(\phi)\theta^2} \sup_{t\in[0,\tau]} |\mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)] - S_{\text{Y}}(t|x)|\\ &\quad + \frac{2K(0)}{K(\phi)\theta^2} \sup_{t\ge0} |\widehat{S}^K_{\text{Y}}(t) - \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)]|.\end{aligned}$$

Thus, when bad events $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ and $\mathcal{E}^{\text{NN}(\phi h)}_{\text{bad } \tau}(x)$ do not happen,

$$\begin{aligned}
&|\widehat{S}^K(t|x;h) - \log S(t|x)| \\
&\quad \le |W_1(t|x) - \mathbb{E}_{\{Y,\delta\}}[W_1(t|x)]| \\
&\qquad + |\mathbb{E}_{\{Y,\delta\}}[W_1(t|x)] - \log S(t|x)| + \frac{2K(0)}{NK(\phi)\theta^2} \\
&\qquad + \frac{2K(0)}{K(\phi)\theta^2} \sup_{t\in[0,\tau]} |\mathbb{E}_{\{Y\}}[\widehat{S}^K_{\mathrm{Y}}(t)] - S_{\mathrm{Y}}(t|x)| \\
&\qquad + \frac{2K(0)}{K(\phi)\theta^2} \sup_{t\ge 0} |\widehat{S}^K_{\mathrm{Y}}(t) - \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\mathrm{Y}}(t)]| \\
&\qquad + |W_3(t|x)|. \qquad (21)
\end{aligned}$$

If we can upper-bound each of the RHS terms by $\varepsilon/18$, then we would be done since the rest of the proof is identical to the ending of the k-NN proof.

On the RHS of inequality (21), the 1st and 5th terms are at most $\frac{\varepsilon}{18}$ when bad events $\mathcal{E}^{\text{kernel}}_{\text{bad } W_1}(t,x)$ and $\mathcal{E}^{\text{kernel}}_{\text{bad weighted EDF}}(x)$ do not happen. The 5th term is less than $\frac{\varepsilon}{18}$ when $n \ge \frac{144K(0)}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,\phi h})K(\phi)}$ and $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ does not happen (so $N > \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,\phi h})$).

The rest of the section is on proving Lemma F.3 and then bounding the 2nd, 4th, and 6th RHS terms (Lemmas F.4, F.5, and F.6).

Proof of Lemma F.3. When bad events $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ and $\mathcal{E}^{\text{NN}(\phi h)}_{\text{bad } \tau}(x)$ do not happen, we are guaranteed that N is an integer within $(\frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,\phi h}), n]$, and $d^+_K(\tau) \ge K(\phi)d^+_{\mathcal{N}_{\text{NN}(\phi h)}}(\tau) > K(\phi)\frac{N\theta}{2}$. Using Hölder's inequality and a bit of algebra,

$$\begin{aligned}
&|W_2(t|x)| \\
&= \Bigg| \sum_{i=1}^N \Big(\tfrac{K_{(i)}}{\sum_{j=1}^N K_{(j)}}\Big)\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\} \\
&\qquad \times \Bigg[\frac{\sum_{\ell=1}^N K_{(\ell)}}{d^+_K(Y_{(i)}) + K_{(i)}} - \frac{1}{S_{\mathrm{Y}}(Y_{(i)}|x)}\Bigg]\Bigg| \\
&\le \max_{i=1,\dots,N} \Bigg|\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}\Big[\tfrac{\sum_{\ell=1}^N K_{(\ell)}}{d^+_K(Y_{(i)})+K_{(i)}} - \tfrac{1}{S_{\mathrm{Y}}(Y_{(i)}|x)}\Big]\Bigg| \\
&= \max_{i=1,\dots,N} \Bigg|\Upsilon_{(i)}\Bigg[\Phi(Y_{(i)}) + \Psi(Y_{(i)}) + \frac{K_{(i)}}{\sum_{\ell=1}^N K_{(\ell)}}\Bigg]\Bigg|,
\end{aligned}$$

where

$$\begin{aligned}
\Upsilon_{(i)} &:= \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}\sum_{\ell=1}^N K_{(\ell)}}{(d^+_K(Y_{(i)}) + K_{(i)})S_{\mathrm{Y}}(Y_{(i)}|x)}, \\
\Phi(t) &:= \widehat{S}^K_{\mathrm{Y}}(t) - \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\mathrm{Y}}(t)], \\
\Psi(t) &:= \mathbb{E}_{\{Y\}}[\widehat{S}^K_{\mathrm{Y}}(t)] - S_{\mathrm{Y}}(t|x).
\end{aligned}$$

We can keep upper-bounding to get:

$$\begin{aligned}
|W_2(t|x)| \le &\Big[\max_{i=1,\dots,N}\Upsilon_{(i)}\Big]\sup_{s\ge 0}|\Phi(s)| \\
&+ \Big[\max_{i=1,\dots,N}\Upsilon_{(i)}\Big]\sup_{s\in[0,\tau]}|\Psi(s)| \\
&+ \max_{i=1,\dots,N}\frac{\Upsilon_{(i)}K_{(i)}}{\sum_{\ell=1}^N K_{(\ell)}}. \qquad (22)
\end{aligned}$$

We upper-bound $\max_{i=1,\dots,N}\Upsilon_{(i)}$ by upper-bounding $\Upsilon_{(i)}$ for every i:

$$\begin{aligned}
\Upsilon_{(i)} &= \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}\sum_{\ell=1}^N K_{(\ell)}}{(d^+_K(Y_{(i)}) + K_{(i)})S_{\mathrm{Y}}(Y_{(i)}|x)} \\
&\le \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}\sum_{\ell=1}^N K_{(\ell)}}{(d^+_K(t) + K_{(i)})S_{\mathrm{Y}}(t|x)} \\
&\le \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}\sum_{\ell=1}^N K_{(\ell)}}{(d^+_K(\tau) + K_{(i)})S_{\mathrm{Y}}(\tau|x)} \\
&\le \frac{\mathbb{1}\{Y_{(i)} \le t\}\sum_{\ell=1}^N K_{(\ell)}}{(d^+_K(\tau) + K_{(i)})\theta} \\
&\le \frac{\mathbb{1}\{Y_{(i)} \le t\}\sum_{\ell=1}^N K_{(\ell)}}{d^+_K(\tau)\theta} \\
&< \frac{\mathbb{1}\{Y_{(i)} \le t\}NK(0)}{K(\phi)\frac{N\theta}{2}\theta} \\
&= \frac{2K(0)}{K(\phi)\theta^2}. \qquad (23)
\end{aligned}$$

Next, we bound $\frac{\Upsilon_{(i)}K_{(i)}}{\sum_{\ell=1}^N K_{(\ell)}}$:

$$\begin{aligned}
\frac{\Upsilon_{(i)}K_{(i)}}{\sum_{\ell=1}^N K_{(\ell)}} &= \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}K_{(i)}}{(d^+_K(Y_{(i)}) + K_{(i)})S_{\mathrm{Y}}(Y_{(i)}|x)} \\
&\le \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}K_{(i)}}{(d^+_K(t) + K_{(i)})S_{\mathrm{Y}}(t|x)} \\
&\le \frac{K_{(i)}}{(d^+_K(t) + K_{(i)})S_{\mathrm{Y}}(t|x)} \\
&\le \frac{K_{(i)}}{d^+_K(t)S_{\mathrm{Y}}(t|x)} \\
&\le \frac{K_{(i)}}{d^+_K(\tau)S_{\mathrm{Y}}(\tau|x)} \\
&\le \frac{K_{(i)}}{d^+_K(\tau)\theta} \\
&< \frac{K(0)}{K(\phi)\frac{N\theta}{2}\theta} \\
&= \frac{2K(0)}{K(\phi)\theta^2 N}. \qquad (24)
\end{aligned}$$

Combining inequalities (22), (23), and (24) finishes the proof. □

Lemma F.4. *Under Assumptions A1–A5, let $x \in supp(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. When bad event $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ does not hold, and the threshold distance satisfies $h \le \frac{1}{\phi}[\frac{\varepsilon\theta}{18(\lambda_{\text{T}}\tau+(f^*_{\text{T}}\lambda_{\text{C}}\tau^2)/2)}]^{1/\alpha}$,*

$$|\mathbb{E}_{\{Y,\delta\}}[W_1(t|x)] - \log S(t|x)| \le \frac{\varepsilon}{18}.$$

Proof. Note that $\mathbb{E}_{Y,\delta}[W_1(t|x)]$ is a function of random variables $N, X_{(1)}, \dots, X_{(N)}$. Since bad event $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ does not happen, we know $N > \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,\phi h})$. We have

$$\begin{aligned}
&\mathbb{E}_{\{Y,\delta\}}[W_1(t|x)] \\
&= \mathbb{E}_{\{Y,\delta\}}\Big[-\sum_{i=1}^N \frac{K_{(i)}}{\sum_{j=1}^N K_{(j)}} \frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}}{S_{\text{Y}}(Y_{(i)}|x)}\Big] \\
&= \sum_{i=1}^N \frac{K_{(i)}}{\sum_{j=1}^N K_{(j)}} \mathbb{E}_{Y_{(i)},\delta_{(i)}}\Big[-\frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}}{S_{\text{Y}}(Y_{(i)}|x)}\Big],
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{E}_{Y_{(i)},\delta_{(i)}}\Big[-\frac{\delta_{(i)}\mathbb{1}\{Y_{(i)} \le t\}}{S_{\text{Y}}(Y_{(i)}|x)}\Big] \\
&= -\int_0^t \Big[\int_s^\infty \frac{1}{S_{\text{Y}}(s|x)} d\mathbb{P}_{C|X=x_{(i)}}(c)\Big] d\mathbb{P}_{T|X=X_{(i)}}(s) \\
&= -\int_0^t \frac{1}{S_{\text{Y}}(s|x)} S_{\text{C}}(s|X_{(i)}) f_{\text{T}}(s|X_{(i)}) ds.
\end{aligned}$$

Recall from equation (14) that

$$\log S(t|x) = -\int_0^t \frac{1}{S_{\text{Y}}(s|x)} S_{\text{C}}(s|x) f_{\text{T}}(s|x) ds.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}_{\{Y,\delta\}}[W_1(t|x)] - \log S(t|x) \\
&= \sum_{i=1}^N \frac{K_{(i)}}{\sum_{j=1}^N K_{(j)}} \\
&\quad \times \int_0^t \frac{1}{S_{\text{Y}}(s|x)}[S_{\text{C}}(s|x)f_{\text{T}}(s|x) - S_{\text{C}}(s|X_{(i)})f_{\text{T}}(s|X_{(i)})]ds.
\end{aligned}$$

Thus, using Hölder's inequality and since $S_{\text{C}}(s|\cdot)f_{\text{T}}(s|\cdot)$ is Hölder continuous with parameters $(\lambda_{\text{T}} + f^*_{\text{T}}\lambda_{\text{C}}s)$ and α,

$$\begin{aligned}
&\big|\mathbb{E}_{\{Y,\delta\}}[W_1(t|x)] - \log S(t|x)\big| \\
&\le \max_{i=1,\dots,N} \Big|\int_0^t \frac{1}{S_{\text{Y}}(s|x)}[S_{\text{C}}(s|x)f_{\text{T}}(s|x) \\
&\qquad\qquad - S_{\text{C}}(s|X_{(i)})f_{\text{T}}(s|X_{(i)})]ds\Big| \\
&\le \max_{i=1,\dots,N} \int_0^t \frac{1}{S_{\text{Y}}(s|x)}|S_{\text{C}}(s|x)f_{\text{T}}(s|x) \\
&\qquad\qquad - S_{\text{C}}(s|X_{(i)})f_{\text{T}}(s|X_{(i)})|ds \\
&\le \frac{1}{S_{\text{Y}}(t|x)} \max_{i=1,\dots,N} \int_0^t (\lambda_{\text{T}} + f^*_{\text{T}}\lambda_{\text{C}}s)\rho(x, X_{(i)})^\alpha ds \\
&\le \frac{1}{S_{\text{Y}}(t|x)} \max_{i=1,\dots,N} \int_0^t (\lambda_{\text{T}} + f^*_{\text{T}}\lambda_{\text{C}}s)(\phi h)^\alpha ds \\
&= \frac{(\phi h)^\alpha}{S_{\text{Y}}(t|x)}\Big(\lambda_{\text{T}}t + \frac{f^*_{\text{T}}\lambda_{\text{C}}t^2}{2}\Big) \\
&\le \frac{(\phi h)^\alpha}{\theta}\Big(\lambda_{\text{T}}\tau + \frac{f^*_{\text{T}}\lambda_{\text{C}}\tau^2}{2}\Big) \\
&\le \frac{\varepsilon}{18},
\end{aligned}$$

where the last inequality uses the fact that $h \le \frac{1}{\phi}[\frac{\varepsilon\theta}{18(\lambda_{\text{T}}\tau+(f^*_{\text{T}}\lambda_{\text{C}}\tau^2)/2)}]^{1/\alpha}$. □

Lemma F.5. *Under Assumptions A1–A5, let $x \in supp(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. When bad event $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ does not happen, and the threshold distance satisfies $h \le \frac{1}{\phi}[\frac{\varepsilon\theta^2 K(\phi)}{36(\lambda_{\text{T}}+\lambda_{\text{C}})\tau K(0)}]^{1/\alpha}$,*

$$\frac{2K(0)}{K(\phi)\theta^2} \sup_{t\in[0,\tau]} |\mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)] - S_{\text{Y}}(t|x)| \le \frac{\varepsilon}{18}.$$

Proof. Note that $\mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)]$ is a function of random variables $N, X_{(1)}, \dots, X_{(N)}$. Since bad event $\mathcal{E}^{\text{NN}(\phi h)}_{\text{few neighbors}}(x)$ does not happen, we know $N > \frac{1}{2}n\mathbb{P}_X(\mathcal{B}_{x,\phi h})$. Then

$$\begin{aligned}
\mathbb{E}_{\{Y\}}[\widehat{S}^K_{\text{Y}}(t)] &= \sum_{j=1}^N \frac{K_{(j)}}{\sum_{\ell=1}^N K_{(\ell)}} \mathbb{E}_{Y_{(j)}}[\mathbb{1}\{Y_{(j)} > t\}] \\
&= \sum_{j=1}^N \frac{K_{(j)}}{\sum_{\ell=1}^N K_{(\ell)}} S_{\text{Y}}(t|X_{(j)}).
\end{aligned}$$

Using Hölder's inequality and since $S_{\text{Y}}(t|\cdot)$ is Hölder continuous with parameters $(\lambda_{\text{T}} + \lambda_{\text{C}})t$ and α,

$$\begin{aligned}
&|\mathbb{E}_{\{Y\}}[\widehat{S}_{\text{Y}}^{K}(t)] - S_{\text{Y}}(t|x)| \\
&\quad = \Big| \sum_{j=1}^{N} \frac{K_{(j)}}{\sum_{\ell=1}^{N} K_{(\ell)}} (S_{\text{Y}}(t|X_{(j)}) - S_{\text{Y}}(t|x)) \Big| \\
&\quad \le \max_{j=1,\dots,N} |S_{\text{Y}}(t|X_{(j)}) - S_{\text{Y}}(t|x)| \\
&\quad \le \max_{j=1,\dots,k} (\lambda_{\text{T}} + \lambda_{\text{C}}) t \rho(x, X_{(j)})^{\alpha} \\
&\quad \le (\lambda_{\text{T}} + \lambda_{\text{C}}) t (\phi h)^{\alpha} \\
&\quad \le (\lambda_{\text{T}} + \lambda_{\text{C}}) \tau (\phi h)^{\alpha} \\
&\quad \le \frac{K(\phi)\theta^2}{2K(0)} \cdot \frac{\varepsilon}{18},
\end{aligned}$$

where the last inequality uses the assumption that $h \le \frac{1}{\phi}[\frac{\varepsilon\theta^2 K(\phi)}{36(\lambda_{\text{T}}+\lambda_{\text{C}})\tau K(0)}]^{1/\alpha}$. □

Lemma F.6. *Under Assumptions A1–A3 and A5, let $x \in \text{supp}(\mathbb{P}_X)$, $t \in [0,\tau]$, and $\varepsilon \in (0,1)$. If bad events $\mathcal{E}_{\textit{few neighbors}}^{\text{NN}(\phi h)}(x)$ and $\mathcal{E}_{\textit{bad } \tau}^{\text{NN}(\phi h)}(x)$ do not happen, and the number of training subjects satisfies*

$$n \ge \frac{144K^2(0)}{\varepsilon\theta^2 \mathbb{P}_X(\mathcal{B}_{x,\phi h}) K^2(\phi)},$$

then $|W_3(t|x)| \le \varepsilon/18$.

Proof. We have $|W_3(t|x)| = \sum_{i=1}^{N} \Xi_{(i)}$, where

$$\begin{aligned}
\Xi_{(i)} &:= \delta_{(i)} \mathbb{1}\{Y_{(i)} \le t\} \sum_{\ell=2}^{\infty} \frac{1}{\ell(\frac{d_K^+(Y_{(i)})}{K_{(i)}} + 1)^{\ell}} \\
&\le \delta_{(i)} \mathbb{1}\{Y_{(i)} \le t\} \sum_{\ell=2}^{\infty} \frac{1}{\ell(\frac{d_K^+(t)}{K_{(i)}} + 1)^{\ell}} \\
&\le \sum_{\ell=2}^{\infty} \frac{1}{\ell(\frac{d_K^+(\tau)}{K_{(i)}} + 1)^{\ell}} \\
&\le \sum_{\ell=2}^{\infty} \frac{1}{\ell(\frac{d_K^+(\tau)}{K(0)} + 1)^{\ell}},
\end{aligned}$$

using the facts that d_K^+ monotonically decreases and $K(\frac{\rho(x,X_i)}{h}) \le K(0)$. Since bad events $\mathcal{E}_{\text{few neighbors}}^{\text{NN}(\phi h)}(x)$ and $\mathcal{E}_{\text{bad } \tau}^{\text{NN}(\phi h)}(x)$ do not happen, we have

$$\begin{aligned}
d_K^+(\tau) &\ge K(\phi) d_{\mathcal{N}_{\text{NN}(\phi h)}(x)}^+(\tau) \\
&> K(\phi) \frac{N\theta}{2} \\
&> K(\phi) \frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\theta}{4}.
\end{aligned}$$

Since we assume that $n \ge \frac{144K^2(0)}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,\phi h})K^2(\phi)} \ge \frac{1.84964K(0)}{\theta\mathbb{P}_X(\mathcal{B}_{x,\phi h})K(\phi)}$, then using the above inequality, we have $\frac{d_K^+(\tau)}{K(0)} \ge 0.46241$, which is needed to apply the reasoning from the proof of Lemma C.8 to get

$$\Xi_i \le \sum_{\ell=2}^{\infty} \frac{1}{\ell(\frac{d_K^+(\tau)}{K(0)} + 1)^{\ell}} \le \frac{1}{(\frac{d_K^+(\tau)}{K(0)})^2} \le \frac{4K^2(0)}{K^2(\phi)N^2\theta^2}.$$

Hence,

$$\begin{aligned}
|W_3(t|x)| = \sum_{i=1}^{N} \Xi_{(i)} &\le \frac{4K^2(0)N}{K^2(\phi)N^2\theta^2} = \frac{4K^2(0)}{K^2(\phi)N\theta^2} \\
&< \frac{8K^2(0)}{K^2(\phi)n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\theta^2} \le \frac{\varepsilon}{18},
\end{aligned}$$

where the last inequality uses the assumption that $n \ge \frac{144K^2(0)}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,\phi h})K^2(\phi)}$. □

Lastly, for the strong consistency result, the calculation is nearly the same as for the k-NN case in Appendix D. To have each of the four terms in bound (6) be at most $\frac{1}{4n^2}$, it suffices to have

$$n \ge \frac{11664}{p_{\min}\varepsilon^2\theta^4\kappa^4} \Big(\frac{18\Lambda_K}{\varepsilon\theta}\Big)^{d/\alpha} \log \frac{864n^2}{\varepsilon\theta^2\kappa},$$

where

$$\varepsilon \le \frac{18\Lambda_K(\phi r^*)^{\alpha}}{\theta}.$$

Then with a fair bit of algebra, one can show that the constants that show up in the theorem statement are

$$\begin{aligned}
c_1'' &:= \frac{1}{\phi}\Big[\frac{36(5\alpha+2d)}{(2\alpha+d)p_{\min}(\theta\Lambda_K)^2\kappa^4}\Big]^{1/(2\alpha+d)}, \\
c_2'' &:= \Big[\frac{64(2\alpha+d)p_{\min}}{\alpha\kappa^{(d-2\alpha)/\alpha}}\Big(\frac{48}{\theta\Lambda_K}\Big)^{d/\alpha}\Big]^{\alpha/(5\alpha+2d)}, \\
c_3'' &:= \Big[\frac{11664(5\alpha+2d)(18\Lambda_K)^{d/\alpha}}{(2\alpha+d)p_{\min}\theta^{(4\alpha+d)/\alpha}\kappa^4}\Big]^{\alpha/(2\alpha+d)}
\end{aligned}$$

In particular, define $u'' := \log(\frac{c_2''c_4''}{e})$, where

$$c_4'' := \frac{36(5\alpha+2d)}{(2\alpha+d)p_{\min}(\theta\Lambda_K)^2\kappa^4(\phi r^*)^{2\alpha+d}}.$$

Then for

$$n \ge n_0'' := \begin{cases} \lceil \frac{e^{1/(2d+5)}}{c_2''} \rceil & \text{if } c_2''c_4'' \le e, \\ \max\{\lceil \frac{e^{1/(2d+5)}}{c_2''} \rceil, \\ \quad c_4''(1+\sqrt{2u}+u)\} & \text{if } c_2''c_4'' > e, \end{cases}$$

if we choose

$$h_n := c_1''\Big(\frac{\log(c_2''n)}{n}\Big)^{1/(2\alpha+d)},$$

then

$$\mathbb{P}\Big(\sup_{t\in[0,\tau]}|\widehat{S}^K(t|x;h_n)-S(t|x)|\ge c_3''\Big(\frac{\log(c_2''n)}{n}\Big)^{\frac{\alpha}{2\alpha+d}}\Big)\le\frac{1}{n^2}.$$

As with the end of the proof of Corollary 3.1 as provided in Appendix D, applying the Borel-Cantelli lemma completes the proof.

G. Proof of Proposition 3.1

The proof strategy is similar to that of proving the k-NN estimator guarantee in terms of how the supremum is handled. Let $a := \sup\{t \in \mathbb{R} : F(t) = \varepsilon/3\}$ and $b := \inf\{t \in \mathbb{R} : F(t) = 1 - \varepsilon/3\}$; these exist due continuity of F. We partition interval $[a, b]$ at points $a = \eta_1 < \eta_2 < \cdots < \eta_{L(\varepsilon)} = b$, where:

- $F(\eta_j) - F(\eta_{j-1}) \le \varepsilon/3$ for $j = 2, \dots, L(\varepsilon)$,
- $L(\varepsilon) \le 3/\varepsilon$.

We can always produce $\eta_1, \dots, \eta_{L(\varepsilon)}$ satisfying the above conditions since if we take them to be at points in which F increases by exactly $\varepsilon/3$ in value starting from a (except for the last point $\eta_{L(\varepsilon)}$, where the increase from $\eta_{L(\varepsilon)-1}$ could be less than $\varepsilon/3$), then the most number $L(\varepsilon)$ of interval pieces needed is $\lceil \frac{(1-\varepsilon/3)-\varepsilon/3}{\varepsilon/3} \rceil + 1 = \lceil 3/\varepsilon \rceil - 1 \le 3/\varepsilon$. Then since $\widehat{F}$ is piecewise constant, if we can guarantee that $|\widehat{F}(\eta_j) - F(\eta_j)| \le \varepsilon/3$ for $j = 1, \dots, L(\varepsilon)$, then at any point $t \in \mathbb{R}$, we indeed will have $|\widehat{F}(t) - F(t)| \le \varepsilon$.

Thus, the main task is in showing, for any given $t \in \mathbb{R}$, how to guarantee $|\widehat{F}(t) - F(t)| \le \varepsilon/3$ with high probability, i.e., we want to upper-bound $\mathbb{P}(|\widehat{F}(t) - F(t)| > \varepsilon/3)$. Once we have an upper bound for this probability, then by a union bound,

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in\mathbb{R}}|\widehat{F}(t) - F(t)| > \varepsilon\Big) &\le \mathbb{P}\Big(\bigcup_{j=1}^{L(\varepsilon)}\{|\widehat{F}(\eta_j) - F(\eta_j)| > \varepsilon/3\}\Big) \\ &\le \sum_{j=1}^{L(\varepsilon)} \mathbb{P}(|\widehat{F}(\eta_j) - F(\eta_j)| > \varepsilon/3). \qquad (25)\end{aligned}$$

We now upper-bound $\mathbb{P}(|\widehat{F}(t) - F(t)| > \varepsilon/3)$. Fix $t \in \mathbb{R}$. Note that $\widehat{F}(t)$ is the sum of ℓ independent variables, where the i-th variable is bounded in $[0, \frac{w_i}{\sum_{j=1}^{\ell} w_j}]$ Moreover, by linearity of expectation, $\mathbb{E}[\widehat{F}(t)] = F(t)$. Then applying Hoeffding's inequality,

$$\mathbb{P}(|\widehat{F}(t) - F(t)| > \varepsilon/3) \le 2\exp\Big(-\frac{2\varepsilon^2(\sum_{j=1}^{\ell} w_j)^2}{9\sum_{i=1}^{\ell} w_i^2}\Big) \qquad (26)$$

Putting together inequalities (25) and (26), and noting that $L(\varepsilon) \le 3/\varepsilon$,

$$\begin{aligned}&\mathbb{P}\Big(\sup_{t\in\mathbb{R}}|\widehat{F}(t) - F(t)| > \varepsilon\Big) \\ &\le \frac{6}{\varepsilon}\exp\Big(-\frac{2\varepsilon^2(\sum_{j=1}^{\ell} w_j)^2}{9\sum_{i=1}^{\ell} w_i^2}\Big). \qquad \square\end{aligned}$$

H. Choosing *k* Using a Validation Set

We now present a guarantee that chooses k based on a validation set of size n, sampled in the same manner as the training set. A similar approach can be used to select bandwidth h for the fixed-radius NN and kernel estimators. We denote the validation set as $(X_1', Y_1', \delta_1'), \dots, (X_n', Y_n', \delta_n')$. For the validation data, we minimize a variant of the IPEC score (Gerds & Schumacher, 2006; Lowsky et al., 2013), which requires conditional survival and censoring time tail estimates $\widehat{S}$ and $\widehat{S}_{\text{C}}$ for S and S_{C}. The IPEC score estimates the following mean squared error of $\widehat{S}$, which cannot be directly computed from training and validation data:

$$\text{MSE}(\widehat{S}) := \int_0^\tau \mathbb{E}[(\mathbb{1}\{T > t\} - \widehat{S}(t|X))^2]dt.$$

Provided that estimators $\widehat{S}$ and $\widehat{S}_{\text{C}}$ are consistent, then the IPEC score is a consistent estimator of MSE(S) (Gerds & Schumacher, 2006).

For any two estimators $\widehat{S}$ and $\widehat{S}_{\text{C}}$ of S and S_{C}, and user-specified time horizon $\tau > 0$ and lower bound $\theta_{\text{LB}} > 0$ for θ in Assumption A3, our IPEC score variant is

$$\begin{aligned}&\text{IPEC}(\widehat{S}, \widehat{S}_{\text{C}}; \tau, \theta_{\text{LB}}) \\ &:= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \widehat{W}_i(t)(\mathbb{1}(Y_i' > t) - \widehat{S}(t|X_i'))^2 dt, \quad (27)\end{aligned}$$

where

$$\widehat{W}_i(t) := \begin{cases} \frac{\delta_i'\mathbb{1}\{Y_i' \le t\}}{\widehat{S}_{\text{C}}(Y_i' - |X_i')} + \frac{\mathbb{1}\{Y_i' > t\}}{\widehat{S}_{\text{C}}(t|X_i')} & \text{if } \widehat{S}_{\text{C}}(t|X_i') \ge \theta_{\text{LB}}, \\ \frac{1}{\theta_{\text{LB}}} & \text{otherwise,} \end{cases}$$

and $\widehat{S}_{\text{C}}(t - |x) = \lim_{s\to t^-} \widehat{S}_{\text{C}}(s|x)$ (for our estimators, $\widehat{S}_{\text{C}}$ is piecewise constant so $\widehat{S}_{\text{C}}(t-|x)$ is straightforward to compute). The only difference between this score and the original IPEC score is that in the original IPEC score, there is no parameter θ_{LB} (put another way, $\theta_{\text{LB}} = 0$). We introduce θ_{LB} to prevent division by 0 and so that in our analysis, the worst-case IPEC score is finite (note that $\widehat{W}_i(t) \le 1/\theta_{\text{LB}}$, so the worst-case IPEC score is τ/θ_{LB}, assuming that estimate $\widehat{S}_{\text{C}}$ monotonically decreases and $\widehat{S}$ takes on values between 0 and 1). In practice, θ_{LB} could simply be set to an arbitrarily small but positive constant.

Due to the inherent symmetry in the problem setup, we can readily use the same k-NN estimator devised for estimating S to instead estimate S_{C}. The only difference is that we replace the censoring indicator δ by $1-\delta$. In terms of the theory, the survival and censoring times swap roles. Thus, we can readily obtain estimates $\widehat{S}^{k\text{-NN}}$ and $\widehat{S}_{\text{C}}^{k\text{-NN}}$ of S and S_{C}.

Note that in practice, often the number of censored data can be quite small compared to n, which can make estimating the conditional censoring tail function S_{C} difficult. There may be reason to believe that the censoring mechanism is actually independent of the feature vector, i.e., $S_{\text{C}}(t|x)=\mathbb{P}(C>t|X=x)=\mathbb{P}(C>t)$. In this case, we can estimate S_{C} using, for instance, the standard Kaplan-Meier estimator (with δ replaced by $1-\delta$). Our validation guarantee will not be making this simplifying assumption; however, it can easily be modified to handle the case when the censoring time is independent of the feature vector.

The validation strategy we analyze is as follows: for a user-specified collection $\mathcal{K}$ of number of nearest neighbors to try (e.g., $\mathcal{K}=\{2^j : j=0,1,\dots,\lceil\log n\rceil\}$, or $\mathcal{K}=[n]$), choose $k\in\mathcal{K}$ that minimizes $\text{IPEC}(\widehat{S}^{k\text{-NN}},\widehat{S}_{\text{C}}^{k\text{-NN}};\tau,\theta_{\text{LB}})$. Denote the resulting choice of k as $\widehat{k}$. We have the following guarantee.

Proposition H.1. *Under Assumptions A1–A4, suppose that there exists $p_{\min}>0$, $d>0$, and $r^*>0$ such that $\mathbb{P}_X(\mathcal{B}_{x,r})\geq p_{\min}r^d$ for all $x\in supp(\mathbb{P}_X)$ and $r\in[0,r^*]$. Let $\varepsilon\in(0,1)$ be a desired error tolerance and $\gamma\in(0,1)$ be a error probability tolerance in estimating $\widehat{S}$. Define $\Lambda_{val}:=\max\{\frac{2\tau}{\theta}(\lambda_{\text{T}}+\lambda_{\text{C}}),\lambda_{\text{T}}\tau+\frac{f_{\text{T}}^*\lambda_{\text{C}}\tau^2}{2},\lambda_{\text{C}}\tau+\frac{f_{\text{C}}^*\lambda_{\text{T}}\tau^2}{2}\}$, and*

$$\mathcal{K}^*:=\bigg\{k\in[n]:\frac{648}{\varepsilon^2\theta^4}\log\Big[\frac{4}{\gamma}\Big(\frac{8}{\varepsilon}+2\Big(\frac{3}{\varepsilon}\log\frac{1}{\theta}+1\Big)\Big)\Big]\leq k\leq\frac{1}{2}np_{\min}\Big(\frac{\varepsilon\theta}{18\Lambda_{val}}\Big)^{d/\alpha}\bigg\}.$$

Using the above procedure for selecting $\widehat{k}$, we have

$$\begin{aligned}&\mathbb{E}[\textit{IPEC}(\widehat{S}^{k\text{-NN}},\widehat{S}_{\text{C}}^{k\text{-NN}};\tau,\theta_{LB})]\\&\leq 2e^{\varepsilon}\textit{MSE}(S)+2e^{\varepsilon}\varepsilon^2\tau\\&\quad+\frac{\tau}{\theta_{LB}}\bigg[\gamma+\sqrt{\frac{\log(2|\mathcal{K}|\sqrt{n})}{2n}}+\frac{1}{\sqrt{n}}+\mathbb{1}\{\theta_{LB}>\theta\}\\&\qquad+\mathbb{1}\{\mathcal{K}\cap\mathcal{K}^*=\emptyset\}+\mathbb{1}\Big\{\varepsilon>\frac{18\Lambda_{val}(r^*)^\alpha}{\theta}\Big\}\bigg].\end{aligned}$$

As with our rate of strong consistency results, the desired error tolerance ε and error probability γ should be set to decrease to 0 as a function of n. Also, unsurprisingly the terms in the bound involve parameters in the underlying model that the user does not know in practice. Note that by choosing $\mathcal{K}=\{2^j : j=0,1,\dots,\lceil\log n\rceil\}$, ε and γ to decrease with n toward 0, and assuming that $\theta_{\text{LB}}>\theta$, then as $n\to\infty$, the bound above converges to $\text{MSE}(S)$.

In the bound, the first two terms correspond to approximation error in the IPEC score estimating $\text{MSE}(S)$. Next, τ/θ_{LB} is the worst-case IPEC score. The terms that it is multiplied by are as follows:

- γ is the error probability in estimating S and $\log S_{\text{C}}$
- The two $\widetilde{\mathcal{O}}(n^{-1/2})$ terms both have to do with the $|\mathcal{K}|$ empirical IPEC scores not being close to their means (over randomness in validation data)
- $\theta_{\text{LB}}>\theta$ happens when the user-specified θ_{LB} is not a lower bound for the true θ
- $\mathcal{K}\cap\mathcal{K}^*=\emptyset$ means one of two things: either the number of training data n is too small, or $\mathcal{K}$ is chosen poorly so that it does not contain any members of $\mathcal{K}^*$, which consists of good choices for the number of nearest neighbors k (e.g., if $\mathcal{K}=\{2^j : j=0,1,\dots,\lceil\log n\rceil\}$, then by having the number of training data n be large enough that $\mathcal{K}^*$ contains a power of 2, we can ensure $\mathcal{K}\cap\mathcal{K}^*$ to be nonempty)
- $\varepsilon>\frac{18\Lambda_{\text{val}}(r^*)^\alpha}{\theta}$ happens when the error tolerance chosen is too large

Note that our analysis requires that we simultaneously have an additive error guarantee for $\widehat{S}^{k\text{-NN}}$ and a multiplicative error guarantee for $\widehat{S}_{\text{C}}^{k\text{-NN}}$. We use the following lemma.

Lemma H.1. *Under Assumptions A1–A4, let $\varepsilon\in(0,1)$ be a user-specified error tolerance and define critical distance $h^*=(\frac{\varepsilon\theta}{18\Lambda_{val}})^{1/\alpha}$. For any feature vector $x\in supp(\mathbb{P}_X)$ and any choice of number of nearest neighbors $k\in[\frac{72}{\varepsilon\theta^2},\frac{n\mathbb{P}_X(\mathcal{B}_{x,h^*})}{2}]$, we have, over randomness in the training data,*

$$\begin{aligned}&\mathbb{P}\bigg(\Big\{\sup_{t\in[0,\tau]}|\widehat{S}^{k\text{-NN}}(t|x)-S(t|x)|>\varepsilon\Big\}\\&\quad\cup\Big\{\sup_{t\in[0,\tau]}|\log\widehat{S}_{\text{C}}^{k\text{-NN}}(t|x)-\log S_{\text{C}}(t|x)|>\varepsilon\Big\}\bigg)\\&\quad\leq\exp\Big(-\frac{k\theta}{8}\Big)+\exp\Big(-\frac{np_{\min}}{8}\Big(\frac{\varepsilon\theta}{18\Lambda_{val}}\Big)^{d/\alpha}\Big)\\&\qquad+2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big)\\&\qquad+\Big[\frac{8}{\varepsilon}+2\Big(\frac{3}{\varepsilon}\log\frac{1}{\theta}+1\Big)\Big]\exp\Big(-\frac{k\varepsilon^2\theta^2}{162}\Big).\end{aligned}\tag{28}$$

Proof. This lemma follows readily from the proof of Theorem 3.1 and the remark at the end of Appendix J for how to modify the proof of Theorem 3.1 to handle log. By carefully examining the proof for Theorem 3.1, we see that bad events $\mathcal{E}_{\text{bad }\tau}^{k\text{-NN}}(x)$, $\mathcal{E}_{\text{far neighbors}}^{k\text{-NN}}(x)$, and $\mathcal{E}_{\text{bad EDF}}^{k\text{-NN}}(x)$ for the k-NN estimate $\widehat{S}^{k\text{-NN}}$ of S can actually be shared with the bad events for the k-NN estimate $\log \widehat{S}_{\text{C}}^{k\text{-NN}}$ of $\log S_{\text{C}}$, with the small change that we now replace Λ with Λ_{val} within the choice of h^* (note that Λ_{val} is now symmetric in the survival and censoring time terms, which naturally happens because we estimate tail functions for both).

With the above explanation, note that the first three RHS terms in bound (28) are the same as those of Theorem 3.1. However, bad event $\mathcal{E}_{\text{bad } U_1}^{k\text{-NN}}(t, X)$ (which is controlled at no larger than $8/\varepsilon$ time points) has to be changed for estimating $\log S_{\text{C}}$ instead (as discussed in Appendix J, the number of time points for controlling the log is at most $2(\frac{3}{\varepsilon}\log\frac{1}{\theta}+1)$ instead of $8/\varepsilon$). Thus, the fourth RHS term in bound (28) union bounds over the final k-NN regression pieces of estimators $\widehat{S}^{k\text{-NN}}$ and $\log \widehat{S}_{\text{C}}^{k\text{-NN}}$. $\square$

Proof of Proposition H.1

Bound (28) is at most γ (by making each of the four RHS terms at most $\gamma/4$) when k, n, and ε satisfy

$$\frac{648}{\varepsilon^2\theta^4}\log\left[\frac{4}{\gamma}\Big(\frac{8}{\varepsilon}+2\Big(\frac{3}{\varepsilon}\log\frac{1}{\theta}+1\Big)\Big)\right] \le k \le \frac{1}{2}np_{\min}\Big(\frac{\varepsilon\theta}{18\Lambda_{\text{val}}}\Big)^{d/\alpha}, \tag{29}$$

and

$$\varepsilon \le \frac{18\Lambda_{\text{val}}(r^*)^\alpha}{\theta}. \tag{30}$$

We refer to the bad event of Lemma H.1 as $\mathcal{E}_{\text{bad est}}^{k\text{-NN}}(x)$. The set $\mathcal{K}^*$ precisely corresponds to choices for the number of nearest neighbors that satisfy sufficient condition (29). If $\mathcal{K}\cap\mathcal{K}^*$ is nonempty, then the validation procedure could potentially select some $k \in \mathcal{K}\cap\mathcal{K}^*$. If, furthermore, $\theta_{\text{LB}} \le \theta$, and ε satisfies condition (30), then our performance guarantee comes into effect. For the rest of the proof, we assume that these nice conditions happen; otherwise, we assume a worst-case IPEC score of τ/θ_{LB}.

Throughout the proof, we use the abbreviation $\text{IPEC}(k) := \text{IPEC}(\widehat{S}^{k\text{-NN}}, \widehat{S}_{\text{C}}^{k\text{-NN}}; \tau, \theta_{\text{LB}})$. We denote $\mathbb{E}_n$ to be the expectation over the n training data, and $\mathbb{E}_{n'}$ to be the expectation over the n validation data.

We introduce a bad event for when at least one of the IPEC scores we compute during validation is not sufficiently close to its expectation over randomness in the validation data:

$$\mathcal{E}_{\text{bad IPEC}} := \bigcup_{k\in\mathcal{K}}\Big\{\text{IPEC}(k) \ge \mathbb{E}_{n'}[\text{IPEC}(k)] + \frac{\tau}{\theta_{\text{LB}}}\sqrt{\frac{\log(|\mathcal{K}|\sqrt{n})}{2n}}\Big\}.$$

Note that, over randomness in the validation data, $\text{IPEC}(k)$ is the average of n independent terms each bounded in $[0, \tau/\theta_{\text{LB}}]$. Thus, by Hoeffding's inequality and a union bound over $k \in \mathcal{K}$, we have $\mathbb{P}(\mathcal{E}_{\text{bad IPEC}}) \le 1/\sqrt{n}$.

Let $\widetilde{k} \in \mathcal{K}\cap\mathcal{K}^*$. We will show shortly that $\mathbb{E}[\text{IPEC}(\widetilde{k})]$ is close to $\text{MSE}(S)$. When bad event $\mathcal{E}_{\text{bad IPEC}}$ does not happen, then

$$\text{IPEC}(\widetilde{k}) \le \mathbb{E}_{n'}[\text{IPEC}(\widetilde{k})] + \frac{\tau}{\theta_{\text{LB}}}\sqrt{\frac{\log(|\mathcal{K}|\sqrt{n})}{2n}}.$$

Moreover, by how $\widehat{k}$ is chosen, $\text{IPEC}(\widehat{k}) \le \text{IPEC}(k)$ for all $k \in \mathcal{K}$. In particular, $\text{IPEC}(\widehat{k}) \le \text{IPEC}(\widetilde{k})$. Therefore,

$$\text{IPEC}(\widehat{k}) \le \mathbb{E}_{n'}[\text{IPEC}(\widetilde{k})] + \frac{\tau}{\theta_{\text{LB}}}\sqrt{\frac{\log(2|\mathcal{K}|\sqrt{n})}{2n}}.$$

Taking the expectation $\mathbb{E}_n$ of both sides above over randomness in the training data,

$$\mathbb{E}_n[\text{IPEC}(\widehat{k})] \le \mathbb{E}[\text{IPEC}(\widetilde{k})] + \frac{\tau}{\theta_{\text{LB}}}\sqrt{\frac{\log(2|\mathcal{K}|\sqrt{n})}{2n}}. \tag{31}$$

Much of the rest of the proof is in upper-bounding $\mathbb{E}[\text{IPEC}(\widetilde{k})]$ in terms of the mean squared error achieved by S:

$$\text{MSE}(S) = \int_0^\tau \mathbb{E}_X\big[\mathbb{E}_T[(\mathbb{1}\{T>t\} - S(t|X))^2]\big]dt.$$

As it will be helpful to know what this is equal to, we compute it now. The inner-most expectation inside the integral is

$$\begin{aligned}&\mathbb{E}_T[(\mathbb{1}\{T>t\} - S(t|X))^2]\\ &\quad= \mathbb{E}_T[\mathbb{1}\{T>t\} - 2\mathbb{1}\{T>t\}S(t|X) + (S(t|X))^2]\\ &\quad= S(t|X) - 2(S(t|X))^2 + (S(t|X))^2\\ &\quad= S(t|X)(1-S(t|X)).\end{aligned}$$

Hence,

$$\text{MSE}(S) = \int_0^\tau \mathbb{E}[S(t|X)(1-S(t|X))]dt. \tag{32}$$

We proceed to upper-bounding $\mathbb{E}[\text{IPEC}(\widetilde{k})]$ in terms of $\text{MSE}(S)$. Note that

$$\begin{aligned}&\mathbb{E}_{n'}[\text{IPEC}(\widetilde{k})]\\ &= \frac{1}{n}\sum_{i=1}^n\int_0^\tau \mathbb{E}_{X_i',Y_i',\delta_i'}[\widehat{W}_i(t)(\mathbb{1}(Y_i'>t) - \widehat{S}^{\widetilde{k}\text{-NN}}(t|X_i'))^2]dt.\end{aligned}$$

Since the validation data are i.i.d., let X denote a feature vector sampled from $\mathbb{P}_X$ and denote its observed time and censoring indicator as Y and δ. Then

$$\begin{aligned}&\mathbb{E}_{n'}[\text{IPEC}(\widetilde{k})]\\ &\quad= \int_0^\tau \mathbb{E}_{X,Y,\delta}[\widehat{W}(t)(\mathbb{1}(Y>t) - \widehat{S}^{\widetilde{k}\text{-NN}}(t|X))^2]dt,\end{aligned}$$

where

$$\widehat{W}(t) := \begin{cases} \frac{\delta \mathbb{1}\{Y \le t\}}{\widehat{S}_{\mathrm{C}}(Y-|X)} + \frac{\mathbb{1}\{Y > t\}}{\widehat{S}_{\mathrm{C}}(t|X)} & \text{if } \widehat{S}_{\mathrm{C}}(t|X) \ge \theta_{\mathrm{LB}}, \\ \frac{1}{\theta_{\mathrm{LB}}} & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} &\mathbb{E}[\mathrm{IPEC}(\widetilde{k})] \\ &= \mathbb{E}_n\big[\mathbb{E}_{n'}[\mathrm{IPEC}(\widetilde{k})]\big] \\ &= \int_0^\tau \mathbb{E}_X\Big[\mathbb{E}_n\big[\mathbb{E}_{Y,\delta}[\widehat{W}(t)(\mathbb{1}(Y>t) - \widehat{S}^{\widetilde{k}\text{-NN}}(t|X))^2]\big]\Big]dt \\ &= \int_0^\tau \mathbb{E}_X\big[\mathbb{E}_n[\Xi]\big]dt, \end{aligned} \tag{33}$$

where

$$\Xi := \mathbb{E}_{Y,\delta}[\widehat{W}(t)(\mathbb{1}(Y>t) - \widehat{S}^{k\text{-NN}}(t|X))^2].$$

Note that Ξ is a function of test point X and the training data, and Ξ is upper-bounded by $1/\theta_{\mathrm{LB}}$. The expectation $\mathbb{E}_n[\Xi]$ is a function of X, which we are conditioning on (so we treat it as fixed). Then, denoting $\mathbb{P}_n$ to be probability over the training data, and noting that bad event $\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)$ is also a function of training data,

$$\begin{aligned} \mathbb{E}_n[\Xi] &= \underbrace{\mathbb{E}_n[\Xi \mid \mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]}_{\le 1/\theta_{\mathrm{LB}}} \underbrace{\mathbb{P}_n(\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X))}_{\le \gamma} \\ &\quad + \mathbb{E}_n\big[\Xi \,\big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\big] \underbrace{\mathbb{P}_n([\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c)}_{\le 1} \\ &\le \frac{\gamma}{\theta_{\mathrm{LB}}} + \mathbb{E}_n\big[\Xi \,\big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\big]. \end{aligned} \tag{34}$$

When bad event $\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)$ does not happen, we simultaneously have

$$\sup_{t\in[0,\tau]} |\widehat{S}^{\widetilde{k}\text{-NN}}(t|X) - S(t|X)| \le \varepsilon,$$

and

$$\sup_{t\in[0,\tau]} |\log \widehat{S}_{\mathrm{C}}^{\widetilde{k}\text{-NN}}(t|X) - \log S_{\mathrm{C}}(t|X)| \le \varepsilon.$$

Hence,

$$\begin{aligned} &(\mathbb{1}(Y>t) - \widehat{S}^{k\text{-NN}}(t|X))^2 \\ &\quad\le (|\mathbb{1}(Y>t) - S(t|X)| + |S(t|X) - \widehat{S}^{k\text{-NN}}(t|X)|)^2 \\ &\quad\le (|\mathbb{1}(Y>t) - S(t|X)| + \varepsilon)^2 \\ &\quad\le 2((\mathbb{1}(Y>t) - S(t|X))^2 + \varepsilon^2), \end{aligned}$$

and

$$\begin{aligned} \widehat{W}(t) &= \frac{\delta \mathbb{1}\{Y \le t\}}{\widehat{S}_{\mathrm{C}}^{k\text{-NN}}(Y-|X)} + \frac{\mathbb{1}\{Y > t\}}{\widehat{S}_{\mathrm{C}}^{k\text{-NN}}(t|X)} \\ &\le \frac{\delta \mathbb{1}\{Y \le t\}}{\widehat{S}_{\mathrm{C}}^{k\text{-NN}}(Y|X)} + \frac{\mathbb{1}\{Y > t\}}{\widehat{S}_{\mathrm{C}}^{k\text{-NN}}(t|X)} \\ &\le e^{\varepsilon}\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)} + e^{\varepsilon}\frac{\mathbb{1}\{Y > t\}}{S_{\mathrm{C}}(t|X)}. \end{aligned}$$

Then

$$\begin{aligned} &\widehat{W}(t)(\mathbb{1}(Y>t) - \widehat{S}^{k\text{-NN}}(t|X))^2 \\ &\le e^{\varepsilon}\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)} 2((\mathbb{1}(Y>t) - S(t|X))^2 + \varepsilon^2) \\ &\quad + e^{\varepsilon}\frac{\mathbb{1}\{Y > t\}}{S_{\mathrm{C}}(t|X)} 2((\mathbb{1}(Y>t) - S(t|X))^2 + \varepsilon^2) \\ &= 2e^{\varepsilon}\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)}((S(t|X))^2 + \varepsilon^2) \\ &\quad + 2e^{\varepsilon}\frac{\mathbb{1}\{Y > t\}}{S_{\mathrm{C}}(t|X)}((1 - S(t|X))^2 + \varepsilon^2), \end{aligned}$$

so

$$\begin{aligned} &\mathbb{E}_n\big[\Xi \,\big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\big] \\ &= 2e^{\varepsilon}(S(t|X))^2 + \varepsilon^2) \\ &\qquad \times \mathbb{E}_n\Big[\mathbb{E}_{Y,\delta}\Big[\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)}\Big] \,\Big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\Big] \\ &\quad + \frac{2e^{\varepsilon}((1 - S(t|X))^2 + \varepsilon^2)}{S_{\mathrm{C}}(t|X)} \\ &\qquad \times \mathbb{E}_n\Big[\mathbb{E}_{Y,\delta}\big[\mathbb{1}\{Y > t\}\big] \,\Big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\Big]. \end{aligned} \tag{35}$$

Next, note that we are currently conditioning on X and $[\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c$. With this conditioning, $\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)}$ (which does not depend on training data) is independent of $[\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c$. Thus

$$\begin{aligned} &\mathbb{E}_n\Big[\mathbb{E}_{Y,\delta}\Big[\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)}\Big] \,\Big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\Big] \\ &= \mathbb{E}_{Y,\delta}\Big[\frac{\delta \mathbb{1}\{Y \le t\}}{S_{\mathrm{C}}(Y|X)}\Big] \\ &= \int_0^t \int_s^\infty \frac{1}{S_{\mathrm{C}}(s|X)} d\mathbb{P}_{C|X}(c) d\mathbb{P}_{T|X}(s) \\ &= \int_0^t \frac{S_{\mathrm{C}}(s|X)}{S_{\mathrm{C}}(s|X)} d\mathbb{P}_{T|X}(s) \\ &= \int_0^t d\mathbb{P}_{T|X}(s) \\ &= 1 - S(t|X). \end{aligned} \tag{36}$$

Similarly,

$$\begin{aligned} &\mathbb{E}_n\Big[\mathbb{E}_{Y,\delta}\big[\mathbb{1}\{Y > t\}\big] \,\Big|\, [\mathcal{E}_{\text{bad est}}^{\widetilde{k}\text{-NN}}(X)]^c\Big] \\ &\quad = \mathbb{E}_Y[\mathbb{1}\{Y > t\}] = S_{\mathrm{Y}}(t|X) = S(t|X)S_{\mathrm{C}}(t|X). \end{aligned} \tag{37}$$

Putting together inequality (35) with equations (36) and (37),

$$
\begin{aligned}
&\mathbb{E}_n\big[\Xi \,\big|\, [\mathcal{E}^{\widetilde{k}\text{-NN}}_{\text{bad est}}(X)]^c\big] \\
&\quad\le 2e^{\varepsilon}(S(t|X))^2+\varepsilon^2)(1-S(t|X)) \\
&\qquad+2e^{\varepsilon}((1-S(t|X))^2+\varepsilon^2)S(t|X). \\
&\quad= 2e^{\varepsilon}S(t|X)(1-S(t|X))+2e^{\varepsilon}\varepsilon^2. \qquad (38)
\end{aligned}
$$

Finally, putting together equation (33) with inequalities (34) and (38) and also using equation (32),

$$
\begin{aligned}
&\mathbb{E}[\text{IPEC}(\widetilde{k})] \\
&= \int_0^{\tau}\mathbb{E}_X\big[\mathbb{E}_n[\Xi]\big]dt \\
&\le \int_0^{\tau}\mathbb{E}_X\Big[\frac{\gamma}{\theta_{\text{LB}}}+\mathbb{E}_n\big[\Xi \,\big|\, [\mathcal{E}^{\widetilde{k}\text{-NN}}_{\text{bad est}}(X)]^c\big]\Big]dt \\
&\le \int_0^{\tau}\mathbb{E}_X\Big[\frac{\gamma}{\theta_{\text{LB}}}+2e^{\varepsilon}S(t|X)(1-S(t|X))+2e^{\varepsilon}\varepsilon^2\Big]dt \\
&= 2e^{\varepsilon}\text{MSE}(S)+2e^{\varepsilon}\varepsilon^2\tau+\frac{\gamma\tau}{\theta_{\text{LB}}}.
\end{aligned}
$$

Combining this with inequality (31), we get

$$
\begin{aligned}
&\mathbb{E}_n[\text{IPEC}(\widehat{k})] \\
&\le 2e^{\varepsilon}\text{MSE}(S)+2e^{\varepsilon}\varepsilon^2\tau+\frac{\tau}{\theta_{\text{LB}}}\Big[\gamma+\sqrt{\frac{\log(2|\mathcal{K}|\sqrt{n})}{2n}}\Big].
\end{aligned}
$$

This holds with probability at least $1-1/\sqrt{n}$ over randomness in the validation data and provided that $\theta_{\text{LB}}\le\theta$, $\mathcal{K}\cap\mathcal{K}^*\neq\emptyset$, and $\varepsilon\le\frac{18\Lambda_{\text{val}}(r^*)^{\alpha}}{\theta}$.

I. Additional Example Distribution Satisfying Assumptions A1–A4

Example I.1 (Weibull regression)**.** *We generalize the exponential regression model of Example 3.1. As before, $\mathcal{X}=\mathbb{R}^d$, and $\mathbb{P}_X$ is a Borel probability measure with compact, convex support. We now take the hazard function to be $h_{\text{T}}(t|x)=q(h_{\text{T},0})^q t^{q-1}\exp(x^\top\beta_{\text{T}})$ for parameters $q>0$, $h_{\text{T},0}>0$, and $\beta_{\text{T}}\in\mathbb{R}^d$ (choosing $q=1$ yields Example 3.1). Following a similar integral calculation as in Example 3.1, we have $S(t|x)=\exp(-(h_{\text{T},0}e^{x^\top\beta_{\text{T}}}t)^q)$, so the conditional survival time distribution $\mathbb{P}_{T|X=x}$ corresponds to a Weibull distribution with shape parameter q and scale parameter $[h_{\text{T},0}e^{x^\top\beta_{\text{T}}}]^{-1}$. We similarly define the conditional censoring time distribution using hazard function $h_{\text{C}}(t|x)=q(h_{\text{C},0})^q t^{q-1}\exp(x^\top\beta_{\text{C}})$ using the same $q>0$ as for the survival time but different parameters $h_{\text{C},0}>0$ and $\beta_{\text{C}}\in\mathbb{R}^d$. In this case, the observed time $Y=\min\{T,C\}$ conditioned on $X=x$ has a Weibull distribution with shape parameter q and scale parameter $1/\omega'(x)$, where*

$$
\omega'(x):=\big[\big(h_{\text{T},0}e^{x^\top\beta_{\text{T}}}\big)^q+\big(h_{\text{C},0}e^{x^\top\beta_{\text{C}}}\big)^q\big]^{1/q}.
$$

The median of this distribution is $[(\log 2)^{1/q}]/\omega'(x)$. Thus, Assumption A3 is satisfied with $\theta=1/2$ and $\tau=\min_{x\in supp(\mathbb{P}_X)}\{[(\log 2)^{1/q}]/\omega'(x)\}$. Lastly, for Assumption A4, we can again take the Lipschitz constant for $f_{\text{T}}(t|\cdot)$ to be $\lambda_{\text{T}}=\sup_{x\in supp(\mathbb{P}_X),t\in[0,\tau]}\|\frac{\partial f_{\text{T}}(t|x)}{\partial x}\|_2$. We can similarly choose the Lipschitz constant for $f_{\text{C}}(t|\cdot)$.

J. Nearest Neighbor and Kernel Variants of the Nelson-Aalen Estimator

The Nelson-Aalen estimator estimates the marginal cumulative hazard function $H_{\text{marg}}(t)=-\log S_{\text{marg}}(t)=-\log\mathbb{P}(T>t)$ (Nelson, 1969; Aalen, 1978). We first give the general form of the Nelson-Aalen estimator, restricted to training subjects $\mathcal{I}\in[n]$. Recall that among training subjects $\mathcal{I}$, the set of unique death times is $\mathcal{Y}_{\mathcal{I}}$. At time $t\ge 0$, the number of deaths is $d_{\mathcal{I}}(t)$ and the number of subjects at risk is $n_{\mathcal{I}}(t)$. Then the Nelson-Aalen estimator restricted to subjects $\mathcal{I}$ is given by

$$
\widehat{H}^{\text{NA}}(t|\mathcal{I}):=\sum_{t'\in\mathcal{Y}_{\mathcal{I}}}\frac{d_{\mathcal{I}}(t')\mathbb{1}\{t'\le t\}}{n_{\mathcal{I}}(t')}.
$$

Thus, the Nelson-Aalen-based k-NN and fixed-radius NN estimates for the (conditional) cumulative hazard function $H(t|x)=-\log S(t|x)$ are $\widehat{H}^{k\text{-NN}}(t|x):=\widehat{H}^{\text{NA}}(t|\mathcal{N}_{k\text{-NN}}(x))$ and $\widehat{H}^{\text{NN}(h)}(t|x):=\widehat{H}^{\text{NA}}(t|\mathcal{N}_{\text{NN}(h)}(x))$.

Recalling that for kernel K and bandwidth $h>0$, the kernel versions of the unique death times, number of deaths, and number of subjects at risk are denoted $\mathcal{Y}_K(x;h)$, $d_K(t|x;h)$, and $n_K(t|x;h)$, then the Nelson-Aalen-based kernel estimate for $H(t|x)$ is

$$
\widehat{H}^{K}(t|x;h):=\sum_{t'\in\mathcal{Y}_K(x;h)}\frac{d_K(t'|x;h)\mathbb{1}\{t'\le t\}}{n_K(t'|x;h)}.
$$

As already discussed in our analysis outline (Section B), the main change to our proofs to obtain nonasymptotic guarantees for these Nelson-Aalen-based estimators is quite simple: for any of the Kaplan-Meier-based estimators $\widehat{S}$ we consider, taking the first-order Taylor expansion of $\log\widehat{S}$ is exactly the negated version of the corresponding Nelson-Aalen-based estimator. This is the only high-level change. A few technical changes have to be made to arrive at a guarantee for each Nelson-Aalen-based estimator. We explain these changes only for the k-NN case.

We reuse notation from our analysis outline (Section B). When there are no ties in survival and censoring times, we have

$$
-\widehat{H}^{k\text{-NN}}(t|x)=U_1(t|x)+U_2(t|x).
$$

Importantly, note that we no longer have to worry about the higher-order Taylor series terms $U_3(t|x)$. Thus, rather than using inequality (10), we now have

$$\begin{aligned}
&|\widehat{H}^{k\text{-NN}}(t|x) - H(t|x)| \\
&\quad = |U_1(t|x) - \log S(t|x) + U_2(t|x)| \\
&\quad \le |U_1(t|x) - \mathbb{E}[U_1(t|x)|\widetilde{X}]| \\
&\qquad + |\mathbb{E}[U_1(t|x)|\widetilde{X}] - \log S(t|x)| \\
&\qquad + \frac{2}{k\theta^2} + \frac{2}{\theta^2}\sup_{s\in[0,\tau]}|S_{\text{Y}}(s|x) - \mathbb{E}[\widehat{S}_{\text{Y}}^{k\text{-NN}}(s|x)|\widetilde{X}]| \\
&\qquad + \frac{2}{\theta^2}\sup_{s\ge 0}|\widehat{S}_{\text{Y}}^{k\text{-NN}}(s|x) - \mathbb{E}[\widehat{S}_{\text{Y}}^{k\text{-NN}}(s|x)|\widetilde{X}]|.
\end{aligned}$$

Thus, we have five RHS terms. As before, we want the RHS to be at most $\varepsilon/3$. For simplicity, we use our earlier bounds, which controls each of the RHS terms to be at most $\varepsilon/18$ so that the RHS above is at most $5\varepsilon/18 < \varepsilon/3$.

At this point, another change is needed. Previously we showed that $|\log \widehat{S}^{k\text{-NN}}(t|x) - \log S(t|x)| \le \varepsilon/3$ implies $|\widehat{S}^{k\text{-NN}}(t|x) - S(t|x)| \le \varepsilon/3$. We then used the fact that $S(\cdot|x)$ changes by at most a value of 1 over the interval $[0,\tau]$. Now we do not remove the logs and instead observe that $H(\cdot|x)$ changes by at most a value of $-\log S(\tau|x) \le -\log\theta = \log\frac{1}{\theta}$ over the interval $[0,\tau]$. Thus, when we partition the interval $[0,\tau]$ into $L(\varepsilon)$ pieces such that $0 = \eta_0 < \eta_1 < \cdots < \eta_{L(\varepsilon)} = \tau$, as before, we ask that $|\widehat{H}^{k\text{-NN}}(t|x) - H(t|x)| \le \varepsilon/3$ for $j = 1,\dots,L(\varepsilon)$. However, the bound on $L(\varepsilon)$ changes. By placing the points η_j's at times when $H(t|x)$ changes by exactly $\varepsilon/3$ (except possibly across $[\eta_{L(\varepsilon)-1}, \eta_{L(\varepsilon)}]$, where $H(t|x)$ can change by less), then $L(\varepsilon) = \lceil \frac{\log\frac{1}{\theta}}{\varepsilon/3} \rceil = \lceil \frac{3}{\varepsilon}\log\frac{1}{\theta} \rceil \le \frac{3}{\varepsilon}\log\frac{1}{\theta} + 1$. The rest of the proof is the same.

We now state the resulting pointwise guarantees for the Nelson-Aalen-based k-NN, fixed-radius NN, and kernel estimators.

Theorem J.1 (Nelson-Aalen-based k-NN pointwise bound)**.** *Under Assumptions A1–A4, let $\varepsilon \in (0,1)$ be a user-specified error tolerance and define critical distance $h^* := (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$. For any feature vector $x \in$ supp$(\mathbb{P}_X)$ and any choice of number of nearest neighbors $k \in [\frac{72}{\varepsilon\theta^2}, \frac{n\mathbb{P}_X(\mathcal{B}_{x,h^*})}{2}]$, we have, over randomness in training data,*

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,\tau]}|\widehat{H}^{k\text{-NN}}(t|x) - H(t|x)| > \varepsilon\Big) \\
&\le \exp\Big(-\frac{k\theta}{8}\Big) + \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h^*})}{8}\Big) \\
&\quad + 2\exp\Big(-\frac{k\varepsilon^2\theta^4}{648}\Big) + 2\Big(\frac{3}{\varepsilon}\log\frac{1}{\theta} + 1\Big)\exp\Big(-\frac{k\varepsilon^2\theta^2}{162}\Big).
\end{aligned}$$

Theorem J.2 (Nelson-Aalen-based fixed-radius NN pointwise bound)**.** *Under Assumptions A1–A4, let $\varepsilon \in (0,1)$ be a user-specified error tolerance. Suppose that the threshold distance satisfies $h \in (0, h^*]$ with $h^* := (\frac{\varepsilon\theta}{18\Lambda})^{1/\alpha}$, and the number of training data satisfies $n \ge \frac{144}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,h})}$. For any $x \in$ supp$(\mathbb{P}_X)$,*

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,\tau]}|\widehat{H}^{\text{NN}(h)}(t|x) - H(t|x)| > \varepsilon\Big) \\
&\quad \le \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\theta}{16}\Big) + \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})}{8}\Big) \\
&\qquad + 2\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^4}{1296}\Big) \\
&\qquad + 2\Big(\frac{3}{\varepsilon}\log\frac{1}{\theta} + 1\Big)\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,h})\varepsilon^2\theta^2}{324}\Big).
\end{aligned}$$

Theorem J.3 (Nelson-Aalen-based kernel pointwise bound)**.** *Under Assumptions A1–A5, let $\varepsilon \in (0,1)$ be a user-specified error tolerance. Suppose that the threshold distance satisfies $h \in (0, \frac{1}{\phi}(\frac{\varepsilon\theta}{18\Lambda_K})^{1/\alpha}]$, and the number of training data satisfies $n \ge \frac{144}{\varepsilon\theta^2\mathbb{P}_X(\mathcal{B}_{x,\phi h})\kappa}$. For any $x \in$ supp$(\mathbb{P}_X)$,*

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{t\in[0,\tau]}|\widehat{H}^{K}(t|x;h) - H(t|x)| > \varepsilon\Big) \\
&\quad \le \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\theta}{16}\Big) + \exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})}{8}\Big) \\
&\qquad + \frac{216}{\varepsilon\theta^2\kappa}\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\varepsilon^2\theta^4\kappa^4}{11664}\Big) \\
&\qquad + 2\Big(\frac{3}{\varepsilon}\log\frac{1}{\theta} + 1\Big)\exp\Big(-\frac{n\mathbb{P}_X(\mathcal{B}_{x,\phi h})\varepsilon^2\theta^2\kappa^2}{324}\Big).
\end{aligned}$$

We remark that the slight change in the proof (regarding partitioning $[0,\tau]$ as to handle log space) can actually be applied to any of the nearest neighbor and kernel Kaplan-Meier-based estimators $\widehat{S}$ to guarantee that $\sup_{t\in[0,\tau]}|\log\widehat{S}(t|x) - \log S(t|x)| \le \varepsilon$.

K. Details on Experimental Results

Concordance index calculation. Harrell's concordance index (c-index) (Harrell Jr et al., 1982) is a pairwise-ranking-based accuracy metric for survival analysis. Roughly, it measures the fraction of pairs of subjects that are correctly ordered among pairs that can actually be ordered (not every pair can be ordered due to censoring). As such, the highest c-index is 1, and 0.5 corresponds to a random ordering. Because c-index is ranking based, it requires that a survival estimator provide some way to rank pairs of subjects in terms of who is at greater risk (ties are allowed).

C-index is computed as follows. Suppose that there are n' test subjects with data $(X'_1, Y'_1, \delta'_1), \dots, (X'_{n'}, Y'_{n'}, \delta'_{n'}) \in \mathcal{X} \times \mathbb{R}_+ \times \{0,1\}$. Then:

1. Construct the set of all pairs of test subjects:

$$\mathcal{P} := \{(i,j) : i,j \in [n'] \text{ such that } i < j\}.$$

2. Remove any pair (i, j) from $\mathcal{P}$ for which the earlier observed time among test subjects i and j is censored.

3. Remove any pair (i, j) from $\mathcal{P}$ for which the observed times are tied unless at least one of test subjects i and j has an event indicator value of 1.

4. For each pair (i, j) that remains in $\mathcal{P}$, we compute a score $C_{(i,j)}$ for (i, j) as follows:
   - If $Y_i' \neq Y_j'$: set $C_{(i,j)} := 1$ if the subject with the shorter observed time (which is guaranteed to be a survival time due to step 2) is predicted to be at higher risk among subjects i and j; set $C_{(i,j)} := 1/2$ if the predicted risks are tied between subjects i and j; otherwise, set $C_{(i,j)} := 0$.
   - If $Y_i' = Y_j'$ and $\delta_i' = \delta_j' = 1$: set $C_{(i,j)} := 1$ if the predicted risks are tied between i and j; otherwise, set $C_{(i,j)} := 1/2$.
   - If $Y_i' = Y_j'$ and exactly one of δ_i' or δ_j' is 1: set $C_{(i,j)} = 1$ if the predicted risk is higher for the subject with event indicator set to 1; otherwise set $C_{(i,j)} = 1/2$.

5. Finally, the c-index is given by:

$$\frac{1}{|\mathcal{P}|} \sum_{(i,j)\in\mathcal{P}} C_{(i,j)}.$$

As for how we rank any pair of test subjects in our experimental results, we use the same approach as Ishwaran et al. (2008). Let $Y_1^*, \dots, Y_m^*$ denote the unique observed times among the test subjects. Then test subject i is considered to be at higher risk than test subject j if

$$\sum_{j=1}^{m} \widehat{H}(Y_j^*|X_i') > \sum_{j=1}^{m} \widehat{H}(Y_j^*|X_j'),$$

where $\widehat{H}$ is an estimate of the conditional cumulative hazard function $H(t|x) = -\log S(t|x)$ (we can, for instance, use nearest neighbor and kernel variants of the Nelson-Aalen estimator). (As a remark, other ways of ranking test subjects are possible. For instance, for the i-th test subject, we could estimate the subject's median survival time by finding time $t \geq 0$ such that $\widehat{S}(t|X_i') \approx 1/2$ for some estimate $\widehat{S}$ of conditional survival function S, and then rank the test subjects by predicted median survival times, i.e., shorter predicted median survival time means higher risk.)

Parameter selection grids. For the k-NN estimator, we search for k over integer powers of 2, starting at 4 and up to the size of the training dataset. For the kernel estimator, we first compute the largest pairwise distance $h_{\max}$ seen in the training data. Then we search for kernel bandwidth h from $0.01h_{\max}$ to $h_{\max}$ on an evenly spaced logarithmic scale with 20 grid points. For random survival forests and the adaptive kernel variant, we search over the number of trees (50, 100, 150, 200) and over the max depth (3, 4, 5, 6, 7, 8, and lastly no restriction on max depth).

Extended results. We now present extended experimental results that also include Epanechnikov and truncated Gaussian kernels for the k-NN, CDF-REG, and kernel estimators. The truncated Gaussian kernel is of the form $K(s) = \exp(-\frac{s^2}{2\sigma^2})\mathbb{1}\{s \leq 1\}$ for standard deviation/scale parameter $\sigma > 0$. We have results for $\sigma \in \{1, 2, 3\}$. The concordance indices are reported for the PBC, GBSG2, RECID, and KIDNEY datasets in Figures 2, 3, 4, and 5.

We also report our IPEC score variant given in equation (27) (with $\theta_{\text{LB}} = 10^{-6}$) in Figures 6, 7, 8, and 9. Note that this IPEC score requires a user-specified time horizon τ. For a given dataset, we set the time horizon to be the 75th percentile of the observed times in the training data (when using other percentiles that are at least the 50th percentile, although the IPEC scores can be different, the relative performance between the methods remains about the same). For our IPEC score variant, the algorithms with best performance changes slightly from what we get using the concordance index. Consistently, random survival forests has lower IPEC score than the adaptive kernel method and tends to have the lowest IPEC score for the GBSG2, RECID, and KIDNEY datasets. For these three datasets, the adaptive kernel method tends to have performance that is on par with the second best method. Similar to the case of concordance indices, for the smallest dataset PBC, weighted versions of k-NN using ℓ_2 distance have the best performance.

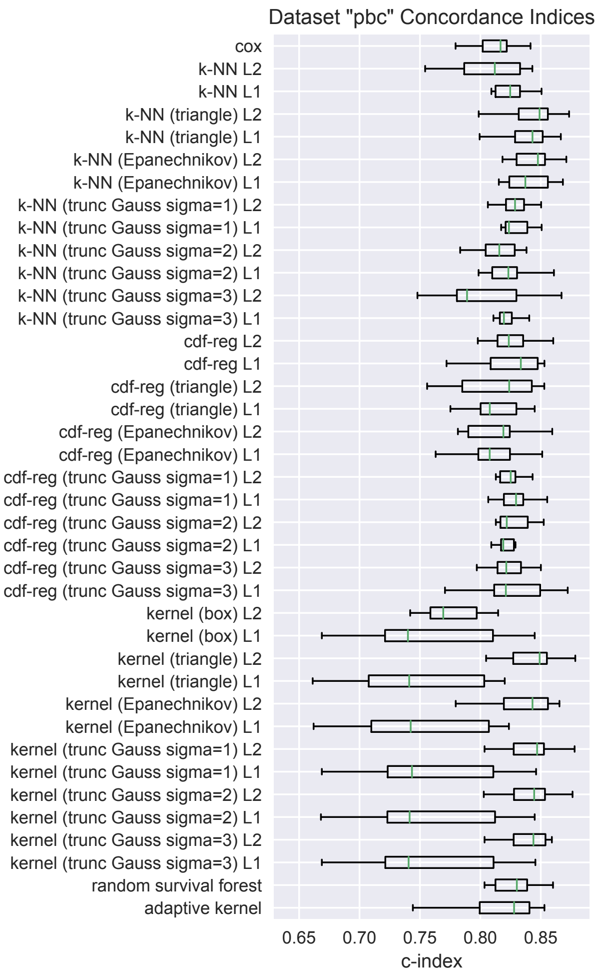

Figure 2. Extended concordance index results for the PBC dataset (higher is better).

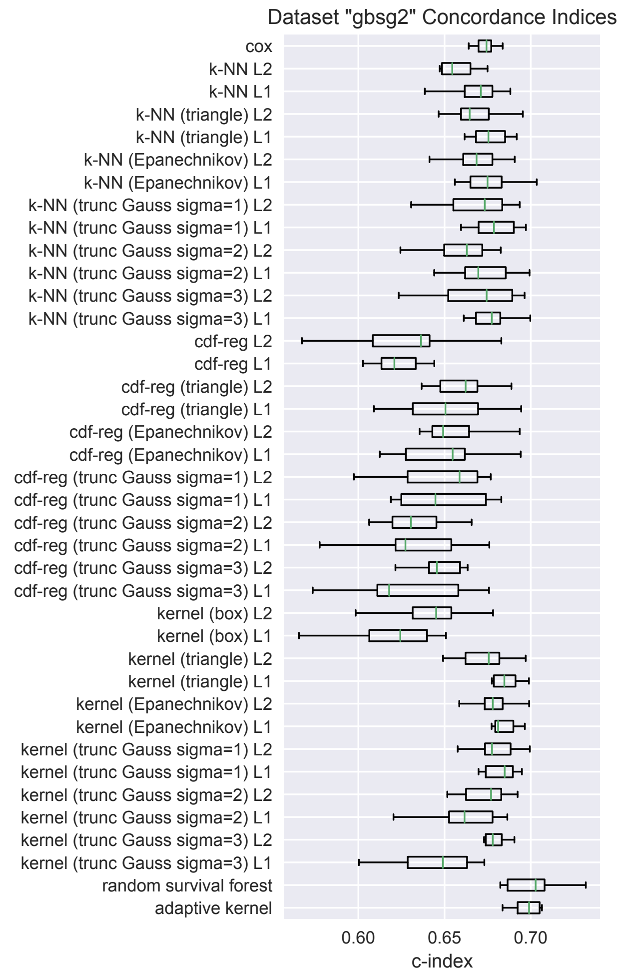

Figure 3. Extended concordance index results for the GBSG2 dataset (higher is better).

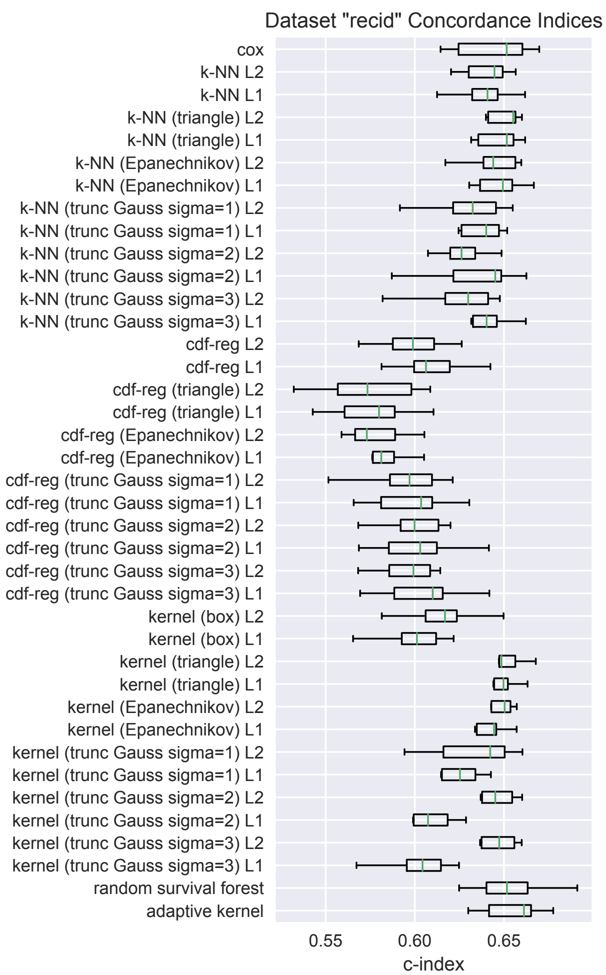

Figure 4. Extended concordance index results for the RECID dataset (higher is better).

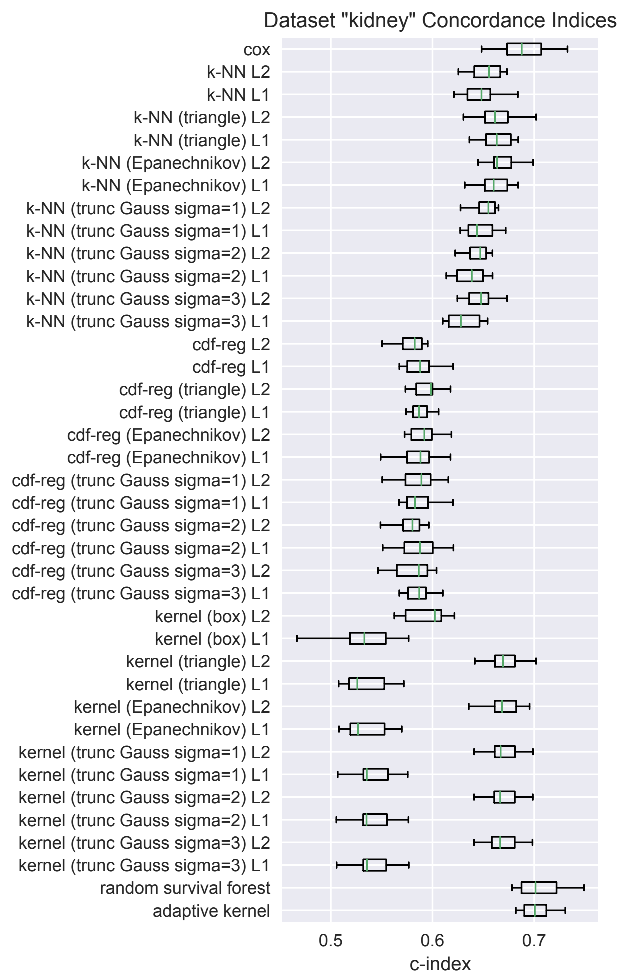

Figure 5. Extended concordance index results for the KIDNEY dataset (higher is better).

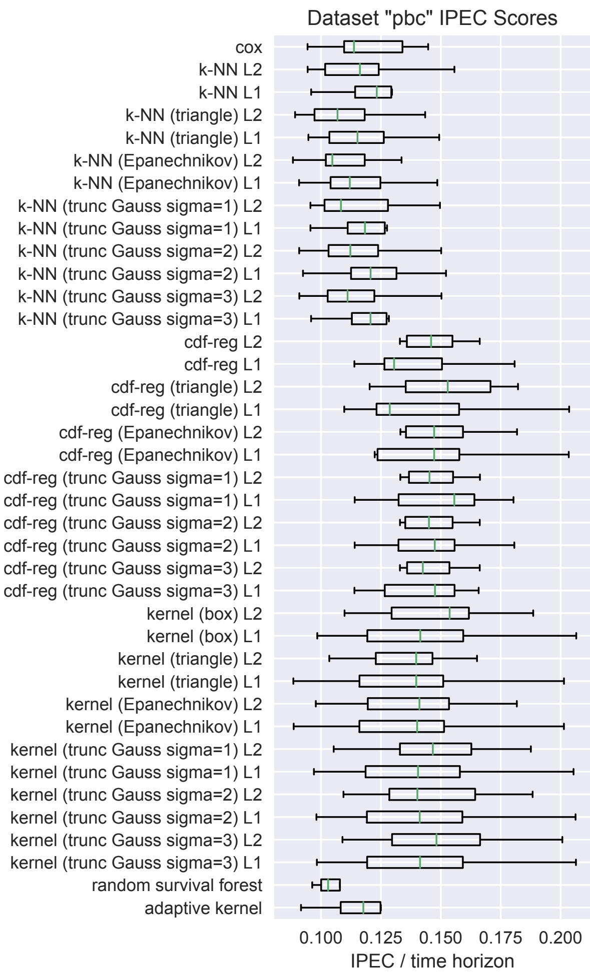

Figure 6. IPEC scores (divided by the time horizon) for the PBC dataset (lower is better).

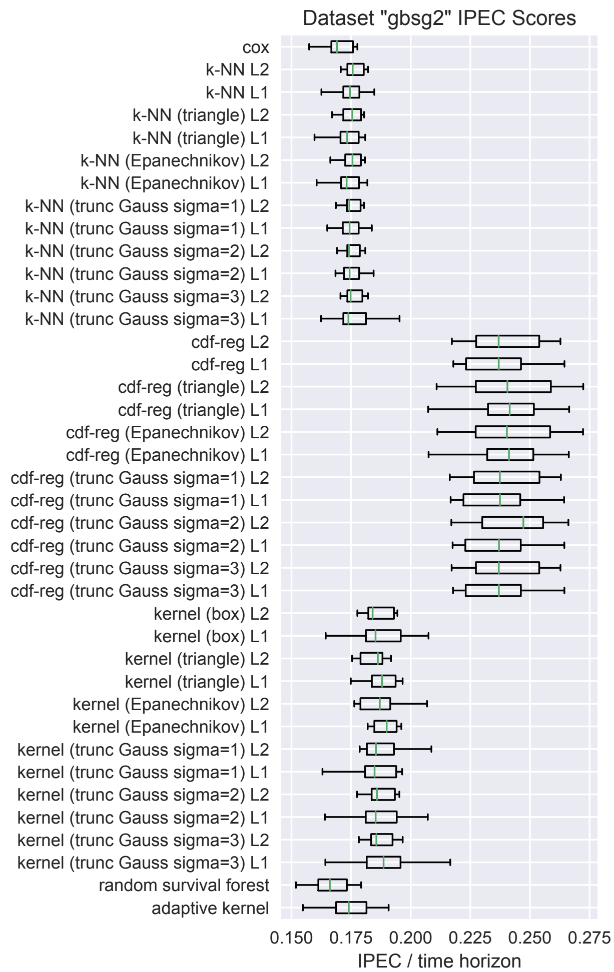

Figure 7. IPEC scores (divided by the time horizon) for the GBSG2 dataset (lower is better).

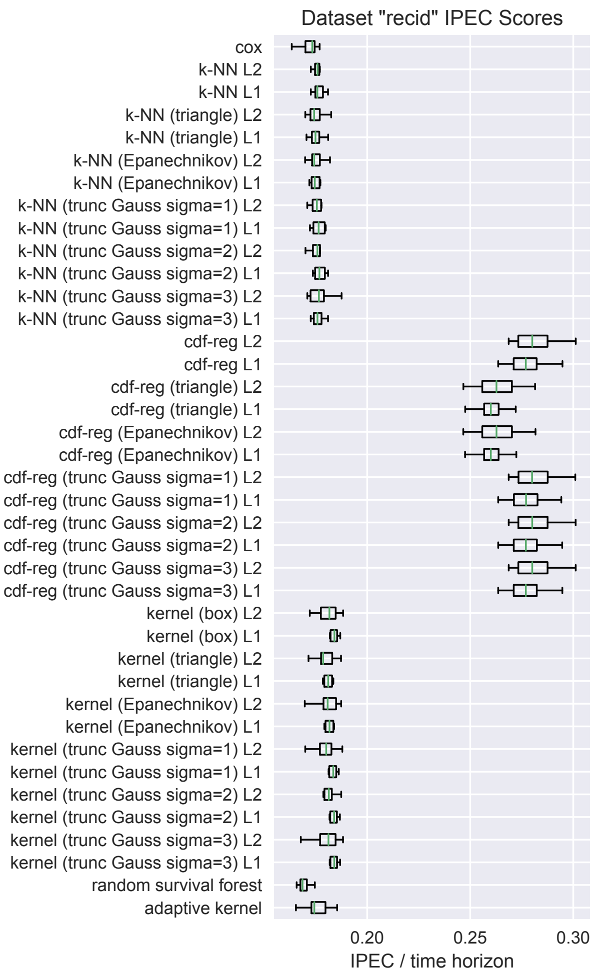

Figure 8. IPEC scores (divided by the time horizon) for the RECID dataset (lower is better).

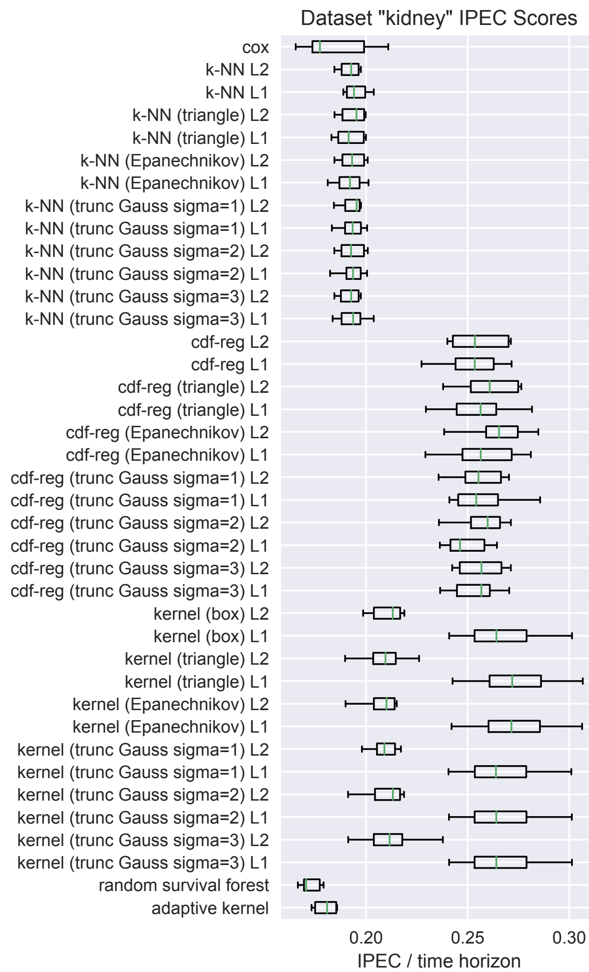

Figure 9. IPEC scores (divided by the time horizon) for the KIDNEY dataset (lower is better).