A. Technical results

Lemma 6. *If $U_1, \ldots U_m$ are iid* $\mathrm{Uniform}[0,1]$ *random variables then*

$$\mathbb{E}\left[\max(\max_{i=1}^{m} U_i, 1 - \min_{i=1}^{m} U_i)\right] = \frac{2m+1}{2m+2}$$

Proof. Let $M_i = \max(U_i, 1 - U_i)$, so M_i are iid $\mathrm{Uniform}[1/2, 1]$ with CDF given by

$$F_{M_i}(x) = 2x - 1$$

for $1/2 \le x \le 1$. Moreover, if $M = \max_{i=1}^{m} M_i$ then $F_M(x) = (2x-1)^m$ since the M_i are iid. The density of M is then

$$f_M(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_M(x) = 2m(2x-1)^{m-1}$$

and its expected value is

$$\mathbb{E}\left[M\right] = \int_{1/2}^{1} x f_M(dx) = \frac{2m+1}{2m+2}$$

which proves the claim. $\qquad\square$

Proposition 7. *For sufficiently large n, every cell of the tree will be cut infinitely often in probability. That is, if K is the distance from the root of the tree to a leaf then $P(K < t) \to 0$ for all t as $n \to \infty$.*

Proof. The splitting mechanism functions by choosing m structure points uniformly at random from the node to be split and searching between their min and max. We will refer to the points selected by the splitting mechanism as active. Without loss of generality we can assume the active points are uniformly distributed on $[0, 1]$ and lower bound the number of estimation points in the smallest child.

Denote the active points $U_1, \ldots, U_m$ and let $U = \max_{i=1}^{m}(\max(U_i, 1 - U_i))$. We know from the calculations in Lemma 6 that

$$\mathbb{P}\left(U \le t\right) = (2t - 1)^m$$

which means that the length of the smallest child is at least $\delta^{1/K} < 1$ with probability $(2(1 - \delta^{1/K}) - 1)^m$, i.e.

$$\mathbb{P}\left(U \le 1 - \delta^{1/K}\right) = (2(1 - \delta^{1/K}) - 1)^m$$

Repeating this argument K times we have that after K splits all sides of all children have length at least δ with probability at least $(2(1 - \delta^{1/K}) - 1)^{Km}$. This bound is derived by assuming that the same dimension is cut at each level of the tree. If different dimensions are cut at different levels the probability that all sides have length at least δ is greater, so the bound holds in those cases also.

This argument shows that every cell at depth K contains a hypercube with sides of length δ with probability at least $(2(1-\delta^{1/K})-1)^{Km}$. Thus for any K and $\epsilon_1 > 0$ we can pick δ such that

$$0 < \delta^{1/K} \le 1 - \frac{1}{2}((1 - \epsilon_1)^{1/Km} + 1)$$

and know that every cell of depth K contains a hypercube with sides of length δ with probability at least $1 - \epsilon_1$. Since the distribution of X has a non-zero density, each of these hypercubes has positive measure with respect to μ_X. Define

$$p = \min_{L \text{ a leaf at depth } K} \mu_X(L) \ .$$

We know $p > 0$ since the minimum is over finitely many leafs and each leaf contains a set of positive measure.

It remains to show that we can choose n large enough so that any set $A \subset [0, 1]^D$ with $\mu_X(A) \ge p$ contains at least k_n estimation points. To this end, fix an arbitrary $A \subset [0, 1]^D$ with $\mu_X(A) = p$. In a data set of size n the number of points which fall in A is $\mathrm{Binomial}(n, p)$. Each point is an estimation point with probability $1/2$, meaning that the number of estimation points, E_n, in A is $\mathrm{Binomial}(n, p/2)$.

Using Hoeffding's inequality we can bound E_n as follows

$$\mathbb{P}\left(E_n < k_n\right) \le \exp\left(-\frac{2}{n}\left(\frac{np}{2} - k_n\right)^2\right) \le \exp\left((k_n - \frac{np}{2})p\right) \ .$$

For this probability to be upper bounded by an arbitrary $\epsilon_2 > 0$ it is sufficient to have

$$\frac{k_n}{n} \leq \frac{p}{2} - \frac{1}{np} \log(\frac{1}{\epsilon_2}) \ .$$

The second term goes to zero as $n \to \infty$ so for sufficiently large n the RHS is positive and since $k_n/n \to 0$ it is always possible to choose n to satisfy this inequality.

In summary, we have shown that if a branch of the tree is grown to depth K then the leaf at the end of this branch contains a set of positive measure with respect to μ_X with arbitrarily high probability. Moreover, we have shown that if n is sufficiently large this leaf will contain at least k_n estimation points.

The only condition which causes our algorithm to terminate leaf expansion is if it is not possible to create child leafs with at least k_n points. Since we can make the probability that any leaf at depth K contains at least this many points arbitrarily high, we conclude that by making n large we can make the probability that all branches are actually grown to depth at least K by our algorithm arbitrarily high as well. Since this argument holds for any K the claim is shown. □