On Theory for BART (Appendix)

Veronika Ročková and Enakshi Saha
University of Chicago

1 Proof of Theorem 7.1

The proof follows from Lemma 6.1, Lemma 5.1 and a modification proof of Theorem 5.1 of RP17. Below, we outline the backbone of the proof and highlight those places where the proof of RP17 had to be modified. Our approach consists of establishing conditions (7), (8) and (9) for $\varepsilon_n = n^{-\alpha/(2\alpha+p)}\log^{1/2} n$. The first step requires constructing the sieve $\mathcal{F}_n \subset \mathcal{F}$. For a given $n \in \mathbb{N}$, $T \in \mathbb{N}$ and a suitably large integer k_n (chosen later), we define the sieve as follows:

$$\mathcal{F}_n = \bigcup_{\boldsymbol{K}: K^t \leq k_n} \bigcup_{\mathcal{E} \in \mathcal{VE}^{\boldsymbol{K}}} \mathcal{F}(\mathcal{E}), \tag{1}$$

where $\mathcal{F}(\mathcal{E})$ consists of all functions $f_{\mathcal{E},\boldsymbol{B}}$ of the form (3) that are supported on a δ-*valid* ensemble $\mathcal{E}$. All δ-valid ensembles consisting of T trees of sizes $\boldsymbol{K} = (K^1, \dots, K^T)'$ are denoted with $\mathcal{VE}^{\boldsymbol{K}}$. The sieve (1) is different from the one in the proof of Theorem 5.1 of RP17. Their sieve consisted of all ensembles whose *total* number of leaves was smaller than k_n. Here, we allow for each tree individually to have up to k_n leaves.

Regarding Condition (7), RP17 in Section 9.1 obtain an upper bound on the covering number for $\mathcal{F}(\mathcal{E})$ as well as the cardinality of $\mathcal{VE}^{\boldsymbol{K}}$ which together yield (for some $D > 0$)

$$\log N\left(\tfrac{\varepsilon}{36}, \left\{f_{\mathcal{E},\boldsymbol{B}} \in \mathcal{F}_n : \|f_{\mathcal{E},\boldsymbol{B}} - f_0\|_n < \varepsilon\right\}, \|.\|_n\right) \tag{2}$$

$$< (k_n+1)T\log(n\,p\,k_n) + D\,T\,k_n \log\left(108\sqrt{T\,k_n}\,n^{1+\delta/2}\right).$$

With the choice $k_n = \lfloor \widetilde{C} n\varepsilon_n^2/\log n \rfloor \asymp n^{p/(2\alpha+p)}$ (for a large enough constant $\widetilde{C} > 0$), fixed $T \in \mathbb{N}$ and assuming $p \lesssim \log^{1/2} n$, the Condition 7 will be met.

Next, we wish to show that the prior assigns enough mass around the truth in the sense that

$$\Pi(f_{\mathcal{E},\boldsymbol{B}} \in \mathcal{F} : \|f_{\mathcal{E},\boldsymbol{B}} - f_0\|_n \leq \varepsilon_n) \geq \mathrm{e}^{-d\,n\varepsilon_n^2} \tag{3}$$

for some large enough $d > 2$. We establish this condition by finding a lower bound on the prior probability in (3), using only step functions supported on a single ensemble. According to Lemma 10.1 of RP17 there exists a 1-valid tree ensemble $f_{\widehat{\mathcal{E}},\widehat{\boldsymbol{B}}}$ that approximates f_0 well in the sense that

$$\|f_0 - f_{\widehat{\mathcal{E}},\widehat{\boldsymbol{B}}}\|_n \leq f_{0\mathcal{H}^\alpha} C\,p/\widehat{K}^{\alpha/p}$$

for some $C > 0$, where $\|f_0\|_{\mathcal{H}^\alpha}$ is the Hölder norm and where $\widehat{K} = 2^{s\,p}$ for some $s \in \mathbb{N}$. Next, we find the smallest $\widehat{K}$ such that $f_{0\mathcal{H}^\alpha} C\,p/\widehat{K}^{\alpha/p} < \varepsilon_n/2$. This value will be denoted by a_n and it satisfies

$$\left(\frac{2C_0 p}{\varepsilon_n}\right)^{\frac{p}{\alpha}} \leq a_n \leq \left(\frac{2C_0 p}{\varepsilon_n}\right)^{\frac{p}{\alpha}} + 1. \tag{4}$$

Under the assumption $p \lesssim \log^{1/2} n$ we have $a_n \asymp n^{p/(2\alpha+p)}$. Denote by $\widehat{\mathcal{E}}$ the approximating ensemble described in Section 6. Next, we denote with $\widehat{\boldsymbol{K}} = (\widehat{K}^1, \dots, \widehat{K}^T)'$ the vector of tree sizes, where $\log_2 a_n + 1 \leq \widehat{K}^t \leq a_n$. Then we can lower-bound the left-hand side of (3) with

$$\pi(\widehat{\mathcal{E}})\Pi\left(f_{\widehat{\mathcal{E}},\boldsymbol{B}} \in \mathcal{F}(\widehat{\mathcal{E}}) : \|f_{\widehat{\mathcal{E}},\boldsymbol{B}} - f_0\|_n \leq \varepsilon_n\right), \tag{5}$$

where $\mathcal{F}(\widehat{\mathcal{E}})$ consists of all additive tree functions supported on $\widehat{\mathcal{E}}$. In Section 6 we show that $\pi(\widehat{\mathcal{E}}) > \mathrm{e}^{-c_2\,n\varepsilon_n^2}$. Moreover, RP17 in Section 10.2 show that, for some $C > 0$,

$$\Pi\left(f_{\widehat{\mathcal{E}},\boldsymbol{B}} \in \mathcal{F}(\widehat{\mathcal{E}}) : \|f_{\widehat{\mathcal{E}},\boldsymbol{B}} - f_0\|_n \leq \varepsilon_n\right) \tag{6}$$

$$> \Pi\left(\boldsymbol{B} \in \mathbb{R}^{\widetilde{a}_n} : \|\boldsymbol{B} - \widehat{\boldsymbol{B}}\|_2 < \frac{\varepsilon_n}{2}\frac{1}{C\sqrt{\widetilde{a}_n}}\right), \tag{7}$$

where $\widetilde{a}_n = \sum_{t=1}^T \widehat{K}^t \leq T\,a_n$ and where $\widehat{\boldsymbol{B}} \in \mathbb{R}^{\widetilde{a}_n}$ are the steps of the approximating additive trees from Lemma 10.1 of RP17. This can be further lower-bounded with

$$\mathrm{e}^{-\frac{\varepsilon_n^2}{8C^2\widetilde{a}_n} - a_n(C_2\|f_0\|_\infty^2 + \log 2)}\left(\frac{\varepsilon_n^2}{4C^2\widetilde{a}_n}\right)^{\frac{\widetilde{a}_n}{2}}\left(\frac{2}{\widetilde{a}_n}\right)^{\widetilde{a}_n/2+1}. \tag{8}$$

Under the assumption $\|f_0\|_\infty \lesssim \log^{1/2} n$, this term is larger than $\mathrm{e}^{-D\,\widetilde{a}_n \log n}$ for some $D > 0$. Since $\widetilde{a}_n \lesssim n\varepsilon_n^2$, there exists $d > 0$ such that $\Pi(f_{\mathcal{E},\boldsymbol{B}} \in \mathcal{F} : \|f_{\mathcal{E},\boldsymbol{B}} - f_0\|_n \leq \varepsilon_n) > \mathrm{e}^{-d\,n\varepsilon_n^2}$.

Lastly, Condition (9) entails showing that $\Pi(\mathcal{F}\backslash\mathcal{F}_n) = o(\mathrm{e}^{-(d+2)\,n\varepsilon_n^2})$ for d deployed in the previous paragraph. It suffices to show that

$$\Pi\left(\bigcup_{t=1}^{T}\{K^t > k_n\}\right)\,\mathrm{e}^{(d+2)\,n\varepsilon_n^2} \to 0.$$

Under the independent Galton-Watson prior on each tree partition, Corollary 5.2 implies that the probability above can be upper-bounded with $\sum_{t=1}^{T}\Pi(K^t > k_n) \lesssim T\,\mathrm{e}^{-C_K\,k_n\log k_n}$. With $k_n \asymp n\varepsilon_n^2/\log n$ and a fixed $T\in\mathbb{N}$, we have $T\,\mathrm{e}^{-C_K\,k_n\log k_n+(d+2)\,n\varepsilon_n^2} \to 0$ for C_K large enough.