

# SUPPLEMENTARY MATERIAL FOR: SIRUS: MAKING RANDOM FORESTS INTERPRETABLE

BY CLÉMENT BÉNARD<sup>\*,†</sup>, GÉRARD BIAU<sup>\*</sup>, SÉBASTIEN DA  
VEIGA<sup>†</sup> AND ERWAN SCORNET<sup>‡</sup>

*Sorbonne Université* <sup>\*</sup>, *Safran Tech* <sup>†</sup> and *Ecole Polytechnique* <sup>‡</sup>

**1. Proof of Theorem 1.** We recall Assumptions (A1)-(A3) and Theorem 1 for the sake of clarity.

- (A1) The subsampling rate  $a_n$  satisfies  $\lim_{n \rightarrow \infty} a_n = \infty$  and  $\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0$ .
- (A2) The number of trees  $M_n$  satisfies  $\lim_{n \rightarrow \infty} M_n = \infty$ .
- (A3)  $\mathbf{X}$  has a strictly positive density  $f$  with respect to the Lebesgue measure. Furthermore, for all  $j \in \{1, \dots, p\}$ , the marginal density  $f^{(j)}$  of  $X^{(j)}$  is continuous, bounded, and strictly positive.

**THEOREM 1.** *If Assumptions (A1)-(A3) are satisfied, then, for all  $\mathcal{P} \in \Pi$ , we have*

$$\lim_{n \rightarrow \infty} \hat{p}_{M_n, n}(\mathcal{P}) = p^*(\mathcal{P}) \quad \text{in probability.}$$

**1.1. Proof of Theorem 1 for a path of one split.** First, we prove Theorem 1 for a path of one split. The proof is extended for a path of two splits in the next subsection and follows the same steps. Thus, we consider  $\mathcal{P}_1 = \{(j_1, r_1, s_1)\}$  a path of one split, where  $j_1 \in \{1, \dots, p\}$ ,  $r_1 \in \{1, \dots, q-1\}$ , and  $s_1 \in \{L, R\}$ . We assume throughout that Assumptions (A1)-(A3) are satisfied.

Before proving Theorem 1, we state five lemmas (Lemma 1 to Lemma 5). Their proof can be found in the Subsection 1.3. Lemma 1 is a preliminary technical result used to state both Lemmas 2 and 4 - case (b).

**LEMMA 1.** *Let  $\mathbf{X}$  be a random variable distributed on  $\mathbb{R}^p$  such that Assumptions (A1) and (A3) are satisfied. Then, for all  $j \in \{1, \dots, p\}$  and all  $r \in \{1, \dots, q-1\}$ , we have*

$$\lim_{n \rightarrow \infty} \sqrt{a_n} \mathbb{P}(q_r^{*(j)} \leq X^{(j)} < \hat{q}_{n,r}^{(j)}) = 0$$

and

$$\lim_{n \rightarrow \infty} \sqrt{a_n} \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X^{(j)} < q_r^{*(j)}) = 0.$$

Lemma 2 is used to prove both consistency (Lemma 3) and convergence rate (Lemma 4) of the CART-splitting criterion when the root node of the tree is cut at an empirical quantile. Lemma 5 is an intermediate result to prove Theorem 1.

LEMMA 2. *If Assumptions (A1) and (A3) are satisfied, then for all  $j \in \{1, \dots, p\}$ , all  $r \in \{1, \dots, q-1\}$ , and all  $H \subseteq \mathbb{R}^p$  such that  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} < q_r^{*(j)}) > 0$  and  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} \geq q_r^{*(j)}) > 0$ , we have*

$$\lim_{n \rightarrow \infty} \sqrt{a_n} (L_{a_n}(H, \hat{q}_{n,r}^{(j)}) - L_{a_n}(H, q_r^{*(j)})) = 0 \quad \text{in probability.}$$

LEMMA 3. *If Assumptions (A1) and (A3) are satisfied, then for all  $j \in \{1, \dots, p\}$ , all  $r \in \{1, \dots, q-1\}$ , and all  $H \subseteq \mathbb{R}^p$  such that  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} < q_r^{*(j)}) > 0$  and  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} \geq q_r^{*(j)}) > 0$ , we have*

$$\lim_{n \rightarrow \infty} L_{a_n}(H, \hat{q}_{n,r}^{(j)}) = L^*(H, q_r^{*(j)}) \quad \text{in probability.}$$

When splitting a node, if the theoretical CART-splitting criterion has multiple maxima, one is randomly selected. This random selection follows a discrete probability law, which is not necessarily uniform and is based on  $\mathbb{P}_{\mathbf{X},Y}$  as specified in Definition 1. In order to derive the limit of the probability that a given split occurs in a  $\Theta$ -random tree in the empirical algorithm, one needs to assess the convergence rate of the empirical CART-splitting criterion when it has multiple maxima.

LEMMA 4. *Consider that Assumptions (A1) and (A3) are satisfied. Let  $\mathcal{C}_1 \subset \{1, \dots, p\} \times \{1, \dots, q-1\}$  be a set of splits of cardinality  $c_1 \geq 2$ , such that, for all  $(j, r) \in \mathcal{C}_1$ ,  $L^*(\mathbb{R}^p, q_r^{*(j)}) \stackrel{\text{def}}{=} L_{\mathcal{C}_1}^*$ , i.e., the theoretical CART-splitting criterion is constant for all splits in  $\mathcal{C}_1$ . Let  $(j_1, r_1) \in \mathcal{C}_1$  and let  $\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)}$  be a random vector where each component is the difference between the empirical CART-splitting criterion for the splits  $(j, r) \in \mathcal{C}_1 \setminus (j_1, r_1)$  and  $(j_1, r_1)$ , that is*

$$\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)} = \left( L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \right)_{(j,r) \in \mathcal{C}_1 \setminus (j_1, r_1)}.$$

(a). *If  $L_{\mathcal{C}_1}^* > 0$ , then we have*

$$\sqrt{a_n} \mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where, for all  $(j, r), (j', r') \in \mathcal{C}_1 \setminus (j_1, r_1)$ , each element of the covariance matrix  $\Sigma$  is defined by  $\Sigma_{(j,r),(j',r')} = \text{Cov}[Z_{j,r}, Z_{j',r'}]$ , with

$$\begin{aligned} Z_{j,r} &= (Y - \mathbb{E}[Y|X^{(j_1)} < q_{r_1}^{*(j_1)}] \mathbb{1}_{X^{(j_1)} < q_{r_1}^{*(j_1)}} \\ &\quad - \mathbb{E}[Y|X^{(j_1)} \geq q_{r_1}^{*(j_1)}] \mathbb{1}_{X^{(j_1)} \geq q_{r_1}^{*(j_1)}})^2 \\ &\quad - (Y - \mathbb{E}[Y|X^{(j)} < q_r^{*(j)}] \mathbb{1}_{X^{(j)} < q_r^{*(j)}} - \mathbb{E}[Y|X^{(j)} \geq q_r^{*(j)}] \mathbb{1}_{X^{(j)} \geq q_r^{*(j)}})^2. \end{aligned}$$

Besides, for all  $(j, r) \in \mathcal{C}_1$ ,  $\mathbb{V}[Z_{j,r}] > 0$ .

(b). If  $L_{\mathcal{C}_1}^* = 0$ , then we have

$$a_n \mathbf{L}_{n, \mathcal{D}_1}^{(\mathcal{C}_1)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} h_{\mathcal{D}_1}(\mathbf{V}),$$

where  $\mathbf{V}$  is a Gaussian vector of covariance matrix  $\text{Cov}[\mathbf{Z}]$ . If  $\mathcal{C}_1$  is explicitly written  $\mathcal{C}_1 = \{(j_k, r_k)\}_{k=1, \dots, c_1}$ ,  $\mathbf{Z}$  is defined, for  $k \in \{1, \dots, c_1\}$ , by

$$\begin{aligned} Z_{2k-1} &= \frac{1}{\sqrt{p_{L,k}}} (Y - \mathbb{E}[Y]) \mathbb{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}} \\ Z_{2k} &= \frac{1}{\sqrt{p_{R,k}}} (Y - \mathbb{E}[Y]) \mathbb{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}}, \end{aligned}$$

where  $p_{L,k} = \mathbb{P}(X^{(j_k)} < q_{r_k}^{*(j_k)})$ ,  $p_{R,k} = \mathbb{P}(X^{(j_k)} \geq q_{r_k}^{*(j_k)})$ , and  $h_{\mathcal{D}_1}$  is a multivariate quadratic form defined as

$$h_{\mathcal{D}_1} : \begin{pmatrix} x_1 \\ \vdots \\ x_{2c_1} \end{pmatrix} \rightarrow \begin{pmatrix} x_3^2 + x_4^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2k-1}^2 + x_{2k}^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2c_1-1}^2 + x_{2c_1}^2 - x_1^2 - x_2^2 \end{pmatrix}.$$

Besides, the variance of each component of  $h_{\mathcal{D}_1}(\mathbf{V})$  is strictly positive.

**DEFINITION 1** (Theoretical splitting procedure). Let  $\theta_1^{(V)}$  be the set of eligible variables to split the root node of a theoretical random tree. The set of best theoretical cuts at the root node is defined as

$$\mathcal{C}_1^*(\theta_1^{(V)}) = \underset{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\}}{\text{argmax}} L^*(\mathbb{R}^p, q_r^{*(j)}).$$

If  $\mathcal{C}_1^*(\theta_1^{(V)})$  has multiple elements, then  $(j_1, r_1)$  is randomly drawn with probability

$$(1.1) \quad \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = \Phi_{\theta_1^{(V)}, (j_1, r_1)}(\mathbf{0}),$$

where  $\Phi_{\theta_1^{(V)}, (j_1, r_1)}$  is the cdf of the limit law defined in Lemma 4 for  $\mathcal{C}_1 = \mathcal{C}_1^*(\theta_1^{(V)})$ . This definition is extended for the second split in Definition 2.

Recall that the randomness in a tree can be decomposed as  $\Theta = (\Theta^{(S)}, \Theta^{(V)})$ , where  $\Theta^{(S)}$  corresponds to the subsampling and  $\Theta^{(V)}$  is related to the variable selection.  $\Theta^{(V)}$  takes values in the finite set  $\Omega^{(V)} = \{1, \dots, p\}^{3 \times \text{mtree}}$ .

LEMMA 5. *If Assumptions (A1)-(A3) are satisfied, then for all  $\theta^{(V)} \in \Omega^{(V)}$ , we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}).$$

We are now equipped to prove Theorem 1 in the case of one single split. Recall that

$$(1.2) \quad \mathbb{E}[\hat{p}_{M_n, n}(\mathcal{P}_1)] = \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)).$$

Since  $\Theta^{(V)}$  takes values in the finite set  $\Omega^{(V)}$ , according to Lemma 5, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)) \\ &= \lim_{n \rightarrow \infty} \sum_{\theta^{(V)} \in \Omega^{(V)}} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \mathbb{P}_{\Theta^{(V)}}(\Theta^{(V)} = \theta^{(V)}) \\ &= \sum_{\theta^{(V)} \in \Omega^{(V)}} \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) \mathbb{P}_{\Theta^{(V)}}(\Theta^{(V)} = \theta^{(V)}) \\ &= \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta)). \end{aligned}$$

Therefore,

$$\lim_{n \rightarrow \infty} \mathbb{E}[\hat{p}_{M_n, n}(\mathcal{P}_1)] = p^*(\mathcal{P}_1).$$

To finish the proof, we just have to show that  $\lim_{n \rightarrow \infty} \mathbb{V}[\hat{p}_{M_n, n}(\mathcal{P}_1)] = 0$ .

The law of total variance gives

$$\begin{aligned} \mathbb{V}[\hat{p}_{M_n, n}(\mathcal{P}_1)] &= \mathbb{E}[\mathbb{V}[\hat{p}_{M_n, n}(\mathcal{P}_1) | \mathcal{D}_n]] + \mathbb{V}[\mathbb{E}[\hat{p}_{M_n, n}(\mathcal{P}_1) | \mathcal{D}_n]] \\ &= \mathbb{E} \left[ \mathbb{V} \left[ \frac{1}{M_n} \sum_{\ell=1}^{M_n} \mathbf{1}_{\mathcal{P}_1 \in T(\Theta_\ell, \mathcal{D}_n)} \middle| \mathcal{D}_n \right] \right] + \mathbb{V}[p_n(\mathcal{P}_1)] \end{aligned}$$

$$\begin{aligned}
 \mathbb{V}[\hat{p}_{M_n, n}(\mathcal{P}_1)] &= \frac{1}{M_n} \mathbb{E}[\mathbb{V}[\mathbf{1}_{\mathcal{P}_1 \in T(\Theta_1, \mathcal{D}_n)} | \mathcal{D}_n]] + \mathbb{V}[p_n(\mathcal{P}_1)] \\
 &= \frac{1}{M_n} \mathbb{E}[p_n(\mathcal{P}_1) - p_n(\mathcal{P}_1)^2] + \mathbb{V}[p_n(\mathcal{P}_1)], \\
 (1.3) \quad &= \frac{1}{M_n} \mathbb{E}[p_n(\mathcal{P}_1)](1 - \mathbb{E}[p_n(\mathcal{P}_1)]) + \left(1 - \frac{1}{M_n}\right) \mathbb{V}[p_n(\mathcal{P}_1)].
 \end{aligned}$$

Following the approach of [Mentch and Hooker \(2016\)](#),  $p_n(\mathcal{P}_1)$  is a complete infinite order U-statistic with the kernel  $\mathbb{E}[\mathbf{1}_{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)} | \Theta^{(S)}, \mathcal{D}_n]$ . From [Hoeffding \(1948\)](#),

$$\mathbb{V}[p_n(\mathcal{P}_1)] \leq \frac{a_n}{n} \xi_{a_n, a_n},$$

where  $\xi_{a_n, a_n} = \mathbb{V}[\mathbb{E}[\mathbf{1}_{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)} | \Theta^{(S)}, \mathcal{D}_n] | \Theta^{(S)}]$ . Since  $\xi_{a_n, a_n}$  is bounded and  $\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0$ ,

$$\lim_{n \rightarrow \infty} \mathbb{V}[p_n(\mathcal{P}_1)] = 0.$$

Using equality (1.3), since  $p_n(\mathcal{P}_1)$  is bounded and  $\lim_{n \rightarrow \infty} M_n = \infty$ ,

$$\lim_{n \rightarrow \infty} \mathbb{V}[\hat{p}_{M_n, n}(\mathcal{P}_1)] = 0.$$

Finally,

$$\begin{aligned}
 &\lim_{n \rightarrow \infty} \mathbb{E}[(\hat{p}_{M_n, n}(\mathcal{P}_1) - p^*(\mathcal{P}_1))^2] \\
 &= \lim_{n \rightarrow \infty} \mathbb{V}[\hat{p}_{M_n, n}(\mathcal{P}_1)] + (\mathbb{E}[\hat{p}_{M_n, n}(\mathcal{P}_1)] - p^*(\mathcal{P}_1))^2 = 0,
 \end{aligned}$$

which concludes the proof.

*1.2. Proof of Theorem 1 for a path of two splits.* The proof of Theorem 1 is extended for a path of two splits. We consider  $\mathcal{P}_1 = \{(j_1, r_1, s_1)\}$  a path of one split and  $\mathcal{P}_2 = \{(j_k, r_k, s_k), k = 1, 2\}$  a path of two splits, where  $j_1, j_2 \in \{1, \dots, p\}$ ,  $r_1, r_2 \in \{1, \dots, q-1\}$  and  $s_1, s_2 \in \{L, R\}$ . We assume assumptions (A1)-(A3) are satisfied.

The path  $\mathcal{P}_2 = \{(j_1, r_1, s_1), (j_2, r_2, s_2)\}$  can occur in trees where the split at the root node is  $(j_1, r_1)$  and the split of one of the child node is  $(j_2, r_2)$ , and in trees where the splits are made in the reversed order,  $(j_2, r_2)$  at the root node and  $(j_1, r_1)$  at one of the child node. Since these two events are disjoint,  $\mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n))$  is the sum of the probability of these two events. Without loss of generality, we will consider in the entire proof that the split at the root node is  $(j_1, r_1)$ . Lemmas 6 - 9 below extend Lemmas 2 - 5 to the case where the tree path contains two splits.

We need to introduce additional notations, first, the theoretical hyperrectangle based on a path  $\mathcal{P}$  by

$$H^*(\mathcal{P}) = \left\{ \mathbf{x} \in \mathbb{R}^p : \begin{cases} x^{(j_k)} < q_{r_k}^{*(j_k)} & \text{if } s_k = L \\ x^{(j_k)} \geq q_{r_k}^{*(j_k)} & \text{if } s_k = R \end{cases}, k \in 1, \dots, d \right\},$$

with  $d \in \{1, 2\}$ , the empirical counterpart of  $\hat{H}_n(\mathcal{P})$  defined in (2.3). Furthermore, since from assumption (A3),  $\mathbf{X}$  has a strictly positive density, then for  $j \in \{1, \dots, p\} \setminus j_1$ , and  $r \in \{1, \dots, q-1\}$ ,  $\mathbb{P}(\mathbf{X} \in H^*(\mathcal{P}_1), X^{(j)} < q_r^{*(j)}) > 0$  and  $\mathbb{P}(\mathbf{X} \in H^*(\mathcal{P}_1), X^{(j)} \geq q_r^{*(j)}) > 0$ . When  $j = j_1$ , the second cut is performed along the same direction as the first one. In that case, depending on the side  $s_1$  of the first cut and the cut positions  $r_1$  and  $r$ , one of the two child node can be empty with probability one. For example, the hyperrectangle associated to the path  $\{(1, 2, L), (1, 3, R)\}$  is empty. In SIRUS, such splits are not considered to find the best cut for a node at the second level of the tree. Thus we define  $\mathcal{C}_{\mathcal{P}_1}$  the set of possible splits for the second cut

$$\begin{aligned} \mathcal{C}_{\mathcal{P}_1} = & \{(j, r), j \in \{1, \dots, p\} \setminus j_1, r \in \{1, \dots, q-1\}\} \\ & \cup \{(j_1, r), \text{ s.t. } r < r_1 \text{ if } s_1 = L, \text{ and } r > r_1 \text{ if } s_1 = R\}, \end{aligned}$$

and  $\mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) = \{(j, r) \in \mathcal{C}_{\mathcal{P}_1} \text{ s.t. } j \in \theta_2^{(V)}\}$  when the split directions are restricted to  $\theta_2^{(V)} \subset \{1, \dots, p\}$ .

LEMMA 6. *If Assumptions (A1) and (A3) are satisfied, then for all  $(j, r) \in \mathcal{C}_{\mathcal{P}_1}$ , we have*

$$\lim_{n \rightarrow \infty} \sqrt{a_n} (L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)})) = 0 \quad \text{in probability.}$$

LEMMA 7. *If Assumptions (A1) and (A3) are satisfied, then for all  $(j, r) \in \mathcal{C}_{\mathcal{P}_1}$ , we have*

$$\lim_{n \rightarrow \infty} L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) = L^*(H^*(\mathcal{P}_1), q_r^{*(j)}) \quad \text{in probability.}$$

LEMMA 8. *Consider that Assumptions (A1) and (A3) are satisfied. Let  $\mathcal{C}_1 \subset \{1, \dots, p\} \times \{1, \dots, q-1\}$  and  $\mathcal{C}_2 \subset \mathcal{C}_{\mathcal{P}_1}$  be two sets of splits of cardinality  $c_1 \geq 1$  and  $c_2 \geq 2$ , such that the theoretical CART-splitting criterion is constant for all splits in  $\mathcal{C}_1$  on one hand, and in  $\mathcal{C}_2$  on the other hand, i.e.,*

$$\forall l \in \{1, 2\}, \quad \forall (j, r) \in \mathcal{C}_l, \quad L^*(H_l, q_r^{*(j)}) \stackrel{\text{def}}{=} L_{\mathcal{C}_l}^*,$$

where  $H_1 = \mathbb{R}^p$  and  $H_2 = H^*(\mathcal{P}_1)$ . Let  $(j_1, r_1) \in \mathcal{C}_1$ ,  $(j_2, r_2) \in \mathcal{C}_2$ , and let  $\mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1, \mathcal{C}_2)}$  a the random vector where each component is the difference between the empirical CART-splitting criterion for the splits  $(j, r) \in \mathcal{C}_1 \setminus (j_1, r_1)$  and  $(j_1, r_1)$  for the first  $c_1 - 1$  components, and for the splits  $(j, r) \in \mathcal{C}_2 \setminus (j_2, r_2)$  and  $(j_2, r_2)$  for the remaining  $c_2 - 1$  components, that is

$$\mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1, \mathcal{C}_2)} = \begin{pmatrix} [L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)})]_{(j,r) \in \mathcal{C}_1 \setminus (j_1, r_1)} \\ [L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)})]_{(j,r) \in \mathcal{C}_2 \setminus (j_2, r_2)} \end{pmatrix}.$$

(a). If  $L_{\mathcal{C}_1}^* > 0$  and  $L_{\mathcal{C}_2}^* > 0$ , then we have

$$\sqrt{a_n} \mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1, \mathcal{C}_2)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

where for  $l, l' \in \{1, 2\}$ , for all  $(j, r) \in \mathcal{C}_l \setminus (j_l, r_l)$ ,  $(j', r') \in \mathcal{C}_{l'} \setminus (j_{l'}, r_{l'})$ , each element of the covariance matrix  $\Sigma$  is defined by  $\Sigma_{(j,r,l),(j',r',l')} = \text{Cov}[Z_{j,r,l}, Z_{j',r',l'}]$ , with

$$Z_{j,r,l} = \frac{1}{\mathbb{P}(\mathbf{X} \in H_l)} (Y - \mu_{L,r_l}^{(j_i)} \mathbf{1}_{X^{(j_i)} < q_{r_l}^{*(j_i)}} - \mu_{R,r_l}^{(j_i)} \mathbf{1}_{X^{(j_i)} \geq q_{r_l}^{*(j_i)}})^2 \mathbf{1}_{\mathbf{X} \in H_l} \\ - \frac{1}{\mathbb{P}(\mathbf{X} \in H_l)} (Y - \mu_{L,r}^{(j)} \mathbf{1}_{X^{(j)} < q_r^{*(j)}} - \mu_{R,r}^{(j)} \mathbf{1}_{X^{(j)} \geq q_r^{*(j)}})^2 \mathbf{1}_{\mathbf{X} \in H_l},$$

$\mu_{L,r}^{(j)} = \mathbb{E}[Y | X^{(j)} < q_r^{*(j)}, \mathbf{X} \in H_l]$ ,  $\mu_{R,r}^{(j)} = \mathbb{E}[Y | X^{(j)} \geq q_r^{*(j)}, \mathbf{X} \in H_l]$ . Besides, for all  $l \in \{1, 2\}$  and for all  $(j, r) \in \mathcal{C}_l$ ,  $\mathbb{V}[Z_{j,r,l}] > 0$ .

(b). If  $L_{\mathcal{C}_1}^* = L_{\mathcal{C}_2}^* = 0$ , then we have

$$a_n \mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1, \mathcal{C}_2)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} h_{\mathcal{P}_2}(\mathbf{V}),$$

where  $\mathbf{V}$  is a gaussian vector of covariance matrix  $\text{Cov}[\mathbf{Z}]$ . If  $\mathcal{C}_1$  and  $\mathcal{C}_2$  are explicitly written  $\mathcal{C}_1 = \{(j_k, r_k)\}_{k \in J_1}$ , and  $\mathcal{C}_2 = \{(j_k, r_k)\}_{k \in J_2}$ , with  $J_1 = \{1, \dots, c_1 + 1\} \setminus 2$  and  $J_2 = \{2\} \cup \{c_1 + 2, \dots, c_1 + c_2\}$ ,  $\mathbf{Z}$  is defined, for  $l \in \{1, 2\}$  and  $k \in J_l$ , by

$$Z_{2k-1} = \frac{1}{\sqrt{p_{L,k} \mathbb{P}(\mathbf{X} \in H_l)}} (Y - \mathbb{E}[Y | \mathbf{X} \in H_l]) \mathbf{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}} \mathbf{1}_{\mathbf{X} \in H_l} \\ Z_{2k} = \frac{1}{\sqrt{p_{R,k} \mathbb{P}(\mathbf{X} \in H_l)}} (Y - \mathbb{E}[Y | \mathbf{X} \in H_l]) \mathbf{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}} \mathbf{1}_{\mathbf{X} \in H_l},$$

where  $p_{L,k} = \mathbb{P}(X^{(j_k)} < q_{r_k}^{*(j_k)}, \mathbf{X} \in H_l)$ ,  $p_{R,k} = \mathbb{P}(X^{(j_k)} \geq q_{r_k}^{*(j_k)}, \mathbf{X} \in H_l)$ , and  $h_{\mathcal{P}_2}$  is a multivariate quadratic form defined as

$$h_{\mathcal{P}_2} : \begin{pmatrix} x_1 \\ \vdots \\ x_{2(c_1+c_2)} \end{pmatrix} \rightarrow \begin{pmatrix} x_5^2 + x_6^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2c_1+1}^2 + x_{2c_1+2}^2 - x_1^2 - x_2^2 \\ x_{2c_1+3}^2 + x_{2c_1+4}^2 - x_3^2 - x_4^2 \\ \vdots \\ x_{2(c_1+c_2)-1}^2 + x_{2(c_1+c_2)}^2 - x_3^2 - x_4^2 \end{pmatrix}.$$

Besides, the variance of each component of  $h_{\mathcal{P}_2}(\mathbf{V})$  is strictly positive.

(c). If  $L_{\mathcal{C}_1}^* > 0$  and  $L_{\mathcal{C}_2}^* = 0$ , then we have

$$a_n \mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1, \mathcal{C}_2)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} h'_{\mathcal{P}_2}(\mathbf{V}),$$

where  $\mathbf{V}$  is a gaussian vector of covariance matrix  $\text{Cov}[\mathbf{Z}]$ , and  $\mathbf{Z}$  is defined as, for  $k \in J_1$ ,

$$\begin{aligned} Z_{2k-1} &= (Y - \mathbb{E}[Y | X^{(j_k)} < q_{r_k}^{*(j_k)}])^2 \mathbb{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}} \\ Z_{2k} &= (Y - \mathbb{E}[Y | X^{(j_k)} \geq q_{r_k}^{*(j_k)}])^2 \mathbb{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}}, \end{aligned}$$

for  $k \in J_2$ ,

$$\begin{aligned} Z_{2k-1} &= \frac{Y - \mathbb{E}[Y | \mathbf{X} \in H^*(\mathcal{P}_1)]}{\sqrt{p_{L,k} \mathbb{P}(\mathbf{X} \in H^*(\mathcal{P}_1))}} \mathbb{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}, \mathbf{X} \in H^*(\mathcal{P}_1)} \\ Z_{2k} &= \frac{Y - \mathbb{E}[Y | \mathbf{X} \in H^*(\mathcal{P}_1)]}{\sqrt{p_{R,k} \mathbb{P}(\mathbf{X} \in H^*(\mathcal{P}_1))}} \mathbb{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}, \mathbf{X} \in H^*(\mathcal{P}_1)}, \end{aligned}$$

and  $h'_{\mathcal{P}_2}$  is a multivariate quadratic form defined as

$$h'_{\mathcal{P}_2} : \begin{pmatrix} x_1 \\ \vdots \\ x_{2(c_1+c_2)} \end{pmatrix} \rightarrow \begin{pmatrix} x_1 + x_2 - x_5 - x_6 \\ \vdots \\ x_1 + x_2 - x_{2c_1+1} - x_{2c_1+2} \\ x_{2c_1+3}^2 + x_{2c_1+4}^2 - x_3^2 - x_4^2 \\ \vdots \\ x_{2(c_1+c_2)-1}^2 + x_{2(c_1+c_2)}^2 - x_3^2 - x_4^2 \end{pmatrix}.$$

Besides, the variance of each component of  $h'_{\mathcal{P}_2}(\mathbf{V})$  is strictly positive.



(d).  $L_{\mathcal{C}_1}^* = 0$  and  $L_{\mathcal{C}_2}^* > 0$ . Symmetric to case (c).

DEFINITION 2 (Theoretical splitting procedure at children nodes). *Let  $\theta^{(V)} = (\theta_1^{(V)}, \theta_2^{(V)}, \cdot) \in \Omega^{(V)}$  be the sets of eligible variables to split the nodes of a theoretical random tree. The set of best theoretical cuts at the left children node along the variables in  $\theta_2^{(V)}$  is defined as*

$$\mathcal{C}_2^*(\theta_2^{(V)}) = \underset{(j,r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)})}{\operatorname{argmax}} L^*(H^*(\mathcal{P}_1), q_r^{*(j)}).$$

If  $\mathcal{C}_2^*(\theta_2^{(V)})$  has multiple elements, then  $(j_2, r_2)$  is randomly drawn with probability

$$(1.4) \quad \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = \frac{\Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\boldsymbol{\theta})}{\mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})},$$

where  $\mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})$  is defined from Definition 1, and  $\Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}$  is the cdf of the limit law defined in Lemma 8 for  $\mathcal{C}_1 = \mathcal{C}_1^*(\theta_1^{(V)})$  and  $\mathcal{C}_2 = \mathcal{C}_2^*(\theta_2^{(V)})$ .

LEMMA 9. *If Assumptions (A1)-(A3) are satisfied, then for all  $\theta^{(V)} \in \Omega^{(V)}$ , we have*

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})$$

Finally, the proof of Theorem 1 in the two-splits scenario is the same as in the single-split scenario.

### 1.3. Proofs of intermediate lemmas.

PROOF OF LEMMA 1. Set  $j \in \{1, \dots, p\}$ , and  $r \in \{1, \dots, q - 1\}$ . We define the marginal cumulative distribution function  $F^{(j)}$  of  $X^{(j)}$ ,  $F^{(j)}(x) = \mathbb{P}(X^{(j)} < x)$ , and  $F_n^{(j)}$  the empirical c.d.f.

$$F_n^{(j)}(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i^{(j)} \leq x}.$$

We adapt an inequality from Serfling (2009) (section 2.3.2 page 75) to bound

the following conditional probability for all  $\varepsilon > 0$

$$\begin{aligned}
& \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)} | X_1^{(j)} = q_r^{*(j)} + \varepsilon) \\
&= \mathbb{P}(q_r^{*(j)} + \varepsilon < \hat{q}_{n,r}^{(j)} | X_1^{(j)} = q_r^{*(j)} + \varepsilon) \\
&\leq \mathbb{P}(F_n^{(j)}(q_r^{*(j)} + \varepsilon) \leq F_n^{(j)}(\hat{q}_{n,r}^{(j)}) | X_1^{(j)} = q_r^{*(j)} + \varepsilon) \\
&\leq \mathbb{P}\left(1 + \sum_{i=2}^n \mathbf{1}_{X_i^{(j)} \leq q_r^{*(j)} + \varepsilon} \leq \left\lceil \frac{n \cdot r}{q} \right\rceil\right) \\
(1.5) \quad &\leq \mathbb{P}\left(\sum_{i=2}^n \mathbf{1}_{X_i^{(j)} \leq q_r^{*(j)} + \varepsilon} - (n-1)F^{(j)}(q_r^{*(j)} + \varepsilon)\right)
\end{aligned}$$

$$(1.6) \quad \leq \left\lceil \frac{n \cdot r}{q} \right\rceil - 1 - (n-1)F^{(j)}(q_r^{*(j)} + \varepsilon)$$

Since  $f$  is continuous and strictly positive, there exists three constants  $c_1, c_2, \eta > 0$  such that for all  $x \in [q_r^{*(j)}, q_r^{*(j)} + \eta]$ ,  $c_1 \leq f^{(j)}(x) \leq c_2$ . Thus, for all  $\varepsilon < \eta$ , we have

$$F^{(j)}(q_r^{*(j)} + \varepsilon) - F^{(j)}(q_r^{*(j)}) = \int_{q_r^{*(j)}}^{q_r^{*(j)} + \varepsilon} f^{(j)}(x) dx,$$

which leads to

$$c_1 \varepsilon \leq F^{(j)}(q_r^{*(j)} + \varepsilon) - F^{(j)}(q_r^{*(j)}) \leq c_2 \varepsilon.$$

Consequently,

$$\begin{aligned}
& \left\lceil \frac{n \cdot r}{q} \right\rceil - 1 - (n-1)F^{(j)}(q_r^{*(j)} + \varepsilon) \\
&\leq \left\lceil \frac{n \cdot r}{q} \right\rceil - 1 - (n-1)(c_1 \varepsilon + F^{(j)}(q_r^{*(j)})) \\
&\leq \left\lceil \frac{n \cdot r}{q} \right\rceil - 1 - (n-1)c_1 \varepsilon - \frac{(n-1) \cdot r}{q} \\
&\leq 1 - (n-1)c_1 \varepsilon.
\end{aligned}$$

For  $n > 1 + \frac{1}{c_1 \varepsilon}$ , we can apply Hoeffding inequality to 1.6,

$$\begin{aligned}
& \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)} | X_1^{(j)} = q_r^{*(j)} + \varepsilon) \\
&\leq \mathbb{P}\left(\sum_{i=2}^n \mathbf{1}_{X_i^{(j)} \leq q_r^{*(j)} + \varepsilon} - (n-1)F^{(j)}(q_r^{*(j)} + \varepsilon) \leq 1 - (n-1)c_1 \varepsilon\right) \\
&\leq e^{-\frac{2}{n}(1-(n-1)c_1 \varepsilon)^2} \\
(1.7) \quad &\leq C e^{-2nc_1^2 \varepsilon^2},
\end{aligned}$$

where  $C = e^{2c_1\eta(1+2c_1\eta)}$ . By definition, we have

$$\begin{aligned} \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)}) &= \int_{]0,\infty[} \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)} | X_1^{(j)} = q_r^{*(j)} + \varepsilon) \\ &\quad \times f^{(j)}(q_r^{*(j)} + \varepsilon) d\varepsilon. \end{aligned}$$

To bound the previous integral, we break it down in three parts. Since  $f^{(j)}$  is bounded by  $c_2$  on  $[q_r^{*(j)}, q_r^{*(j)} + \eta]$ , for  $n > 1 + \frac{1}{c_1\eta}$  we use inequality 1.7 to get

$$\begin{aligned} \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)}) &\leq \int_{]0, \frac{1}{(n-1)c_1}] c_2 d\varepsilon \\ &\quad + \int_{] \frac{1}{(n-1)c_1}, \eta[} c_2 C e^{-2nc_1^2\varepsilon^2} d\varepsilon \\ &\quad + \int_{] \eta, \infty[} C e^{-2nc_1^2\eta^2} f^{(j)}(q_r^{*(j)} + \varepsilon) d\varepsilon. \end{aligned}$$

In the second integral, we introduce the following change of variable  $u = \sqrt{2nc_1}\varepsilon$

$$\begin{aligned} \int_{] \frac{1}{(n-1)c_1}, \eta[} c_2 C e^{-2nc_1^2\varepsilon^2} d\varepsilon &= \frac{c_2 C}{c_1 \sqrt{2n}} \int_{] \frac{\sqrt{2n}}{(n-1)}, \sqrt{2nc_1}\eta[} e^{-u^2} du \\ &\leq \frac{c_2 C}{c_1 \sqrt{2n}} \int_{]0, \infty[} e^{-u^2} du \leq \frac{\sqrt{\pi} c_2 C}{2c_1 \sqrt{2n}}, \end{aligned}$$

and therefore we can write

$$\sqrt{a_n} \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)}) \leq \frac{c_2 \sqrt{a_n}}{(n-1)c_1} + \frac{\sqrt{\pi a_n} c_2 C}{2c_1 \sqrt{2n}} + C \sqrt{a_n} e^{-2nc_1^2\eta^2}$$

From Assumption (A1),  $\lim_{n \rightarrow \infty} \frac{a_n}{n} = 0$ , and then

$$\lim_{n \rightarrow \infty} \sqrt{a_n} \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)}) = 0.$$

The case  $\lim_{n \rightarrow \infty} \sqrt{a_n} \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X_1^{(j)} < q_r^{*(j)}) = 0$  is similar.  $\square$

1.3.1. *Case 1:  $\mathcal{P}_1$ .*

PROOF OF LEMMA 2. Let  $j \in \{1, \dots, p\}$ ,  $r \in \{1, \dots, q-1\}$ , and  $H \subseteq \mathbb{R}^p$  such that  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} < q_r^{*(j)}) > 0$  and  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} \geq q_r^{*(j)}) > 0$ . Let

$$\Delta_{n,r}^{(j)} = \sqrt{a_n} (L_{a_n}(H, \hat{q}_{n,r}^{(j)}) - L_{a_n}(H, q_r^{*(j)}))$$

that is

$$\begin{aligned} \Delta_{n,r}^{(j)} = & -\frac{\sqrt{a_n}}{N_n(H)} \left[ \sum_{i=1}^{a_n} (Y_i - \bar{Y}_{H_L} \mathbf{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} - \bar{Y}_{H_R} \mathbf{1}_{X_i^{(j)} \geq \hat{q}_{n,r}^{(j)}})^2 \mathbf{1}_{\mathbf{x}_i \in H} \right. \\ & \left. - \sum_{i=1}^{a_n} (Y_i - \bar{Y}_{H_L^*} \mathbf{1}_{X_i^{(j)} < q_r^{*(j)}} - \bar{Y}_{H_R^*} \mathbf{1}_{X_i^{(j)} \geq q_r^{*(j)}})^2 \mathbf{1}_{\mathbf{x}_i \in H} \right] \end{aligned}$$

where, for a generic hyperrectangle  $H$ , we define  $N_n(H) = \sum_{i=1}^{a_n} \mathbf{1}_{\mathbf{x}_i \in H}$ , and

$$H_L = \{\mathbf{x} \in H : x^{(j)} < \hat{q}_{n,r}^{(j)}\} \quad \text{and} \quad \bar{Y}_{H_L} = \frac{1}{N_n(H_L)} \sum_{i=1}^{a_n} Y_i \mathbf{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbf{1}_{\mathbf{x}_i \in H},$$

with the convention  $\bar{Y}_{H_L} = 0$  if  $H_L$  is empty. The theoretical quantities  $H_L^*$  and  $\bar{Y}_{H_L^*}$  are defined similarly by replacing the empirical quantile by its population version. We define symmetrically  $H_R, H_R^*, \bar{Y}_{H_R}, \bar{Y}_{H_R^*}$ .

Simple calculations show that

$$\begin{aligned} \Delta_{n,r}^{(j)} = & \frac{\sqrt{a_n}}{N_n(H)} (\bar{Y}_{H_L}^2 N_n(H_L) - \bar{Y}_{H_L^*}^2 N_n(H_L^*)) \\ (1.8) \quad & + \frac{\sqrt{a_n}}{N_n(H)} (\bar{Y}_{H_R}^2 N_n(H_R) - \bar{Y}_{H_R^*}^2 N_n(H_R^*)) \end{aligned}$$

The first term in equation (1.8) can be rewritten as

$$\begin{aligned} & \frac{\sqrt{a_n}}{N_n(H)} (\bar{Y}_{H_L}^2 N_n(H_L) - \bar{Y}_{H_L^*}^2 N_n(H_L^*)) \\ & = \frac{\sqrt{a_n}}{N_n(H) N_n(H_L) N_n(H_L^*)} \sum_{i,k,l=1}^{a_n} Y_i Y_k \mathbf{1}_{\mathbf{x}_i \in H, \mathbf{x}_k \in H} \\ & \quad \times (\mathbf{1}_{X_l^{(j)} < q_r^{*(j)}} \mathbf{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbf{1}_{X_k^{(j)} < \hat{q}_{n,r}^{(j)}} - \mathbf{1}_{X_l^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbf{1}_{X_i^{(j)} < q_r^{*(j)}} \mathbf{1}_{X_k^{(j)} < q_r^{*(j)}}). \end{aligned}$$

Since  $Y_i \in \{0, 1\}$ , we have the following bound

$$\begin{aligned}
 & \frac{\sqrt{a_n}}{N_n(H)} |\bar{Y}_{H_L}^2 N_n(H_L) - \bar{Y}_{H_L^*}^2 N_n(H_L^*)| \\
 & \leq \frac{\sqrt{a_n}}{N_n(H)N_n(H_L)N_n(H_L^*)} \sum_{i,k,l=1}^{a_n} |\mathbb{1}_{X_l^{(j)} < q_r^{*(j)}} \mathbb{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbb{1}_{X_k^{(j)} < \hat{q}_{n,r}^{(j)}} \\
 & \quad - \mathbb{1}_{X_l^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \mathbb{1}_{X_k^{(j)} < q_r^{*(j)}}|,
 \end{aligned}$$

and finally

$$(1.9) \quad \frac{\sqrt{a_n}}{N_n(H)} |\bar{Y}_{H_L}^2 N_n(H_L) - \bar{Y}_{H_L^*}^2 N_n(H_L^*)| \leq \frac{a_n^3}{N_n(H)N_n(H_L)N_n(H_L^*)} W_{n,r}^{(j)},$$

where

$$(1.10) \quad W_{n,r}^{(j)} = \frac{\sqrt{a_n}}{a_n^3} \sum_{i,k,l=1}^{a_n} |\mathbb{1}_{X_l^{(j)} < q_r^{*(j)}} \mathbb{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbb{1}_{X_k^{(j)} < \hat{q}_{n,r}^{(j)}} \\
 \quad - \mathbb{1}_{X_l^{(j)} < \hat{q}_{n,r}^{(j)}} \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \mathbb{1}_{X_k^{(j)} < q_r^{*(j)}}|.$$

A close inspection of the terms inside the sum of (1.10) reveals that

$$\begin{aligned}
 \mathbb{E}[W_{n,r}^{(j)}] & \leq \frac{\sqrt{a_n}}{a_n^3} \sum_{i,k,l=1}^{a_n} \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X_i^{(j)} < q_r^{*(j)}) + \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X_k^{(j)} < q_r^{*(j)}) \\
 & \quad + \mathbb{P}(q_r^{*(j)} \leq X_l^{(j)} < \hat{q}_{n,r}^{(j)}) + \mathbb{P}(q_r^{*(j)} \leq X_i^{(j)} < \hat{q}_{n,r}^{(j)}) \\
 & \quad + \mathbb{P}(q_r^{*(j)} \leq X_k^{(j)} < \hat{q}_{n,r}^{(j)}) + \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X_l^{(j)} < q_r^{*(j)}) \\
 & \leq 3\sqrt{a_n} \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X_1^{(j)} < q_r^{*(j)}) + 3\sqrt{a_n} \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)}),
 \end{aligned}$$

which tends to zero, according to Lemma 1. Thus, in probability,

$$(1.11) \quad \lim_{n \rightarrow \infty} W_{n,r}^{(j)} = 0.$$

Regarding the remaining terms in inequality (1.9), by the law of large numbers, in probability,

$$(1.12) \quad \lim_{n \rightarrow \infty} \frac{N_n(H)}{a_n} = \mathbb{P}(\mathbf{X} \in H), \quad \lim_{n \rightarrow \infty} \frac{N_n(H_L^*)}{a_n} = \mathbb{P}(\mathbf{X} \in H_L^*).$$

Additionally,

$$\begin{aligned} \mathbb{E} \left[ \left| \frac{N_n(H_L)}{a_n} - \frac{N_n(H_L^*)}{a_n} \right| \right] &\leq \mathbb{E} \left[ \frac{1}{a_n} \sum_{i=1}^{a_n} \mathbf{1}_{X_i^{(j)} \in H} \left| \mathbf{1}_{X_i^{(j)} \leq \hat{q}_{n,r}^{(j)}} - \mathbf{1}_{X_i^{(j)} \leq q_r^{*(j)}} \right| \right] \\ &\leq \mathbb{P}(\hat{q}_{n,r}^{(j)} \leq X_1^{(j)} < q_r^{*(j)}) + \mathbb{P}(q_r^{*(j)} \leq X_1^{(j)} < \hat{q}_{n,r}^{(j)}), \end{aligned}$$

which tends to zero, according to Lemma 1. Therefore, in probability,

$$(1.13) \quad \lim_{n \rightarrow \infty} \frac{N_n(H_L)}{a_n} - \frac{N_n(H_L^*)}{a_n} = 0.$$

Since  $\mathbb{P}(\mathbf{X} \in H) > 0$  and  $\mathbb{P}(\mathbf{X} \in H_L^*) > 0$  by assumption, we can combine (1.11)-(1.13) to obtain, in probability,

$$(1.14) \quad \lim_{n \rightarrow \infty} \frac{a_n^3}{N_n(H)N_n(H_L)N_n(H_L^*)} = \frac{1}{\mathbb{P}(\mathbf{X} \in H)\mathbb{P}(\mathbf{X} \in H_L^*)^2}.$$

Using (1.11) and (1.14) and inequality (1.9), we obtain, in probability,

$$\lim_{n \rightarrow \infty} \frac{\sqrt{a_n}}{N_n(H)} \left| \bar{Y}_{H_L}^2 N_n(H_L) - \bar{Y}_{H_L^*}^2 N_n(H_L^*) \right| = 0.$$

Similar results can be derived for the other term in equation (1.8), which allows us to conclude that, in probability,

$$\lim_{n \rightarrow \infty} \sqrt{a_n} (L_{a_n}(H, \hat{q}_{n,r}^{(j)}) - L_{a_n}(H, q_r^{*(j)})) = 0.$$

□

**PROOF OF LEMMA 3.** Let  $j \in \{1, \dots, p\}$ ,  $r \in \{1, \dots, q-1\}$  and  $H \subseteq \mathbb{R}^p$  such that  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} < q_r^{*(j)}) > 0$  and  $\mathbb{P}(\mathbf{X} \in H, X^{(j)} \geq q_r^{*(j)}) > 0$ .

$$L_{a_n}(H, \hat{q}_{n,r}^{(j)}) = L_{a_n}(H, q_r^{*(j)}) + (L_{a_n}(H, \hat{q}_{n,r}^{(j)}) - L_{a_n}(H, q_r^{*(j)}))$$

From the law of large number, in probability,

$$\lim_{n \rightarrow \infty} L_{a_n}(H, q_r^{*(j)}) = L^*(H, q_r^{*(j)}).$$

Thus, according to Lemma 2, in probability,

$$\lim_{n \rightarrow \infty} L_{a_n}(H, \hat{q}_{n,r}^{(j)}) = L^*(H, q_r^{*(j)}).$$

□

**PROOF OF LEMMA 4.** We consider  $\mathcal{C}_1$ , a set of splits of cardinality  $c_1 \geq 2$  satisfying, for all  $(j, r) \in \mathcal{C}_1$ ,  $L^*(\mathbb{R}^p, q_r^{*(j)}) \stackrel{\text{def}}{=} L_{\mathcal{C}_1}^*$ . Fix  $(j_1, r_1) \in \mathcal{C}_1$ , we recall that

$$\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)} = \left( L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \right)_{(j,r) \in \mathcal{C}_1 \setminus (j_1, r_1)}.$$

Case (a):  $L_{\mathcal{C}_1}^* > 0$ . We first consider the following decomposition for  $(j, r) \in \mathcal{C}_1$ ,

$$\begin{aligned} L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) &= L_{a_n}(\mathbb{R}^p, q_r^{*(j)}) + (L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, q_r^{*(j)})) \\ &= \frac{1}{a_n} \sum_{i=1}^n (Y_i - \bar{Y})^2 - \frac{1}{a_n} \sum_{i=1}^n (Y_i - \bar{Y}_L^* \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} - \bar{Y}_R^* \mathbb{1}_{X_i^{(j)} \geq q_r^{*(j)}})^2 \\ &\quad + L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, q_r^{*(j)}), \end{aligned}$$

where

$$N_{n,L}^* = \sum_{i=1}^{a_n} \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \quad \text{and} \quad \bar{Y}_L^* = \frac{1}{N_{n,L}^*} \sum_{i=1}^{a_n} Y_i \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}}$$

( $\bar{Y}_R^*$ ,  $N_{n,R}^*$  are defined symmetrically). Letting  $\mu_{L,r}^{(j)} = \mathbb{E}[Y | X^{(j)} < q_r^{*(j)}]$  (and  $\mu_{R,r}^{(j)}$  symmetrically), the first two terms of the last decomposition are standard variance estimates and we can write

(1.15)

$$\begin{aligned} L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) &= \frac{1}{a_n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \\ &\quad - \frac{1}{a_n} \sum_{i=1}^n (Y_i - \mu_{L,r}^{(j)} \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} - \mu_{R,r}^{(j)} \mathbb{1}_{X_i^{(j)} \geq q_r^{*(j)}})^2 + R_{n,r}^{(j)}, \end{aligned} \tag{1.16}$$

where

$$\begin{aligned} R_{n,L}^{(j)} &= \frac{N_{n,L}^*}{a_n} (\bar{Y}_L^* - \mu_{L,r}^{(j)})^2 + \frac{N_{n,R}^*}{a_n} (\bar{Y}_R^* - \mu_{L,r}^{(j)})^2 \\ &\quad + L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, q_r^{*(j)}). \end{aligned} \tag{1.17}$$

Using the Central limit theorem, in probability,

$$\lim_{n \rightarrow \infty} \sqrt{a_n} \frac{N_{L,r}^*}{a_n} (\bar{Y}_{L,r}^* - \mu_{L,r}^{(j)})^2 = 0. \tag{1.18}$$

The same result holds for the second term of (1.17), and using Lemma 2 for the third term of (1.17), we get that, in probability,

$$\lim_{n \rightarrow \infty} \sqrt{a_n} (L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, q_r^{*(j)})) = 0.$$

Finally,

$$\lim_{n \rightarrow \infty} \sqrt{a_n} R_{n,r}^{(j)} = 0, \quad \text{in probability.}$$

Using Equation (1.16), each component of  $\mathbf{L}_{n,\mathcal{P}_1}^{(\mathcal{C}_1)}$  writes, with  $(j, r) \in \mathcal{C}_1 \setminus (j_1, r_1)$ ,

$$\begin{aligned} & L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \\ &= \frac{1}{a_n} \sum_{i=1}^n \left( Y_i - \mu_{L,r_1}^{(j_1)} \mathbb{1}_{X_i^{(j_1)} < q_{r_1}^{*(j_1)}} - \mu_{R,r_1}^{(j_1)} \mathbb{1}_{X_i^{(j_1)} \geq q_{r_1}^{*(j_1)}} \right)^2 \\ &\quad - \left( Y_i - \mu_{L,r}^{(j)} \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} - \mu_{R,r}^{(j)} \mathbb{1}_{X_i^{(j)} \geq q_r^{*(j)}} \right)^2 \\ &\quad + R_{n,r}^{(j)} - R_{n,r_1}^{(j_1)} \end{aligned}$$

We can apply the multivariate Central limit theorem and Slutsky's theorem to obtain,

$$\sqrt{a_n} \mathbf{L}_{n,\mathcal{P}_1}^{(\mathcal{C}_1)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \Sigma)$$

where for all  $(j, r), (j', r') \in \mathcal{C}_1 \setminus (j_1, r_1)$ , each element of the covariance matrix  $\Sigma$  is defined by  $\Sigma_{(j,r),(j',r')} = \text{Cov}[Z_{j,r}, Z_{j',r'}]$ , with

$$\begin{aligned} Z_{j,r} &= \left( Y - \mu_{L,r_1}^{(j_1)} \mathbb{1}_{X^{(j_1)} < q_{r_1}^{*(j_1)}} - \mu_{R,r_1}^{(j_1)} \mathbb{1}_{X^{(j_1)} \geq q_{r_1}^{*(j_1)}} \right)^2 \\ &\quad - \left( Y - \mu_{L,r}^{(j)} \mathbb{1}_{X^{(j)} < q_r^{*(j)}} - \mu_{R,r}^{(j)} \mathbb{1}_{X^{(j)} \geq q_r^{*(j)}} \right)^2. \end{aligned}$$

Since  $L_{\mathcal{C}_1}^* > 0$ , we have for all  $(j, r) \in \mathcal{C}_1$ ,  $\mu_{L,r}^{(j)} \neq \mu_{R,r}^{(j)}$ . Besides, according to assumption (A3),  $\mathbf{X}$  has a strictly positive density. Consequently, the variance of  $Z_{j,r}$  is strictly positive. This concludes the first case.

*Case (b):*  $L_{\mathcal{C}_1}^* = 0$ . Fix  $(j, r) \in \mathcal{C}_1$ . Since  $L^*(\mathbb{R}^p, q_r^{*(j)}) = 0$ , we have

$$\mathbb{E}[Y] = \mathbb{E}[Y|X^{(j)} < q_r^{*(j)}] = \mathbb{E}[Y|X^{(j)} \geq q_r^{*(j)}] \stackrel{\text{def}}{=} \mu.$$

Then, simple calculations show that  $L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})$  writes

$$L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) = -(\bar{Y} - \mu)^2 + \underbrace{\frac{N_{n,L}}{a_n} (\bar{Y}_L - \mu)^2}_{\delta_L} + \underbrace{\frac{N_{n,R}}{a_n} (\bar{Y}_R - \mu)^2}_{\delta_R},$$

where

$$N_{n,L} = \sum_{i=1}^{a_n} \mathbb{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} \quad \text{and} \quad \bar{Y}_L = \frac{1}{N_{n,L}} \sum_{i=1}^{a_n} Y_i \mathbb{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}}$$



( $N_{n,R}, \bar{Y}_R$  are defined similarly for the other cell). Letting  $p_{L,r}^{(j)} = \mathbb{P}(X^{(j)} < q_r^{*(j)})$  and  $p_{R,r}^{(j)} = \mathbb{P}(X^{(j)} \geq q_r^{*(j)})$  with  $p_{L,r}^{(j)}, p_{R,r}^{(j)} > 0$ , we have

$$\begin{aligned} \delta_L &= \frac{N_{n,L}}{a_n} (\bar{Y}_L - \mu)^2 \\ &= \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \mu)^2 - 2 \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L) (\bar{Y}_L - \mu) + \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L)^2 \\ &= \frac{1}{p_{L,r}^{(j)}} \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \right)^2 + R_{L,r}^{(j)}, \end{aligned}$$

where

$$\begin{aligned} R_{L,r}^{(j)} &= \left( \frac{a_n N_{n,L}}{N_{n,L}^{*2}} - \frac{1}{p_{n,L}^{(j)}} \right) \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \right)^2 \\ &\quad - 2 \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L) (\bar{Y}_L - \mu) + \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L)^2 \end{aligned}$$

By the law of large numbers,  $\lim_{n \rightarrow \infty} \frac{N_{n,L}^*}{a_n} = p_{L,r}^{(j)}$  in probability. Using Equation (1.13) in the proof of Lemma 2, it comes that, in probability,  $\lim_{n \rightarrow \infty} \frac{N_{n,L}}{a_n} = p_{L,r}^{(j)}$ , and consequently  $\lim_{n \rightarrow \infty} \frac{a_n N_{n,L}}{N_{n,L}^{*2}} = \frac{1}{p_{L,r}^{(j)}}$ . Since  $\sqrt{a_n} \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}}$  converges in distribution to a normal distribution by the Central limit theorem,

$$\lim_{n \rightarrow \infty} a_n \left( \frac{a_n N_{n,L}}{N_{n,L}^{*2}} - \frac{1}{p_{L,r}^{(j)}} \right) \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \right)^2 = 0, \quad \text{in probability.}$$

Furthermore, as for Equation (1.10) in the proof of Lemma 2,

$$\begin{aligned} &\sqrt{a_n} |\bar{Y}_L^* - \bar{Y}_L| \\ &\leq \frac{a_n^2}{N_{n,L} N_{n,L}^*} \underbrace{\frac{\sqrt{a_n}}{a_n^2} \sum_{i=1, l=1}^{a_n} Y_i \left| \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \mathbb{1}_{X_l^{(j)} < \hat{q}_r^{(j)}} - \mathbb{1}_{X_i^{(j)} < \hat{q}_r^{(j)}} \mathbb{1}_{X_l^{(j)} < q_r^{*(j)}} \right|}_{\varepsilon_L}, \end{aligned}$$

and

$$\mathbb{E}[\varepsilon_L] \leq 2\sqrt{a_n} \mathbb{P}(\hat{q}_r^{(j)} \leq X^{(j)} < q_r^{*(j)}) + 2\sqrt{a_n} \mathbb{P}(q_r^{*(j)} \leq X^{(j)} < \hat{q}_r^{(j)}).$$

According to Lemma 1, the right hand side term converges to 0. Then, in probability,  $\lim_{n \rightarrow \infty} \varepsilon_L = 0$ . Additionally,  $\lim_{n \rightarrow \infty} \frac{a_n^2}{N_{n,L} N_{n,L}^*} = \frac{1}{p_{L,r}^{(j)2}}$ , and then, in

probability,

$$(1.19) \quad \lim_{n \rightarrow \infty} \sqrt{a_n}(\bar{Y}_L^* - \bar{Y}_L) = 0.$$

The second term of  $a_n R_{L,r}^{(j)}$  writes

$$\begin{aligned} & -a_n \times 2 \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L) (\bar{Y}_L^* - \mu) \\ & = -2 \frac{N_{n,L}}{a_n} \times \sqrt{a_n} (\bar{Y}_L^* - \bar{Y}_L) \times \sqrt{a_n} (\bar{Y}_L^* - \mu), \end{aligned}$$

where in probability,  $\lim_{n \rightarrow \infty} 2 \frac{N_{n,L}}{a_n} = p_{L,r}^{(j)}$ ,  $\lim_{n \rightarrow \infty} \sqrt{a_n} (\bar{Y}_L^* - \bar{Y}_L) = 0$  according to equation 1.19, and  $\sqrt{a_n} (\bar{Y}_L^* - \mu)$  converges to a normal random variable from the central limit theorem. By Slutsky theorem, in probability,  $\lim_{n \rightarrow \infty} -a_n \times 2 \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L) (\bar{Y}_L^* - \mu) = 0$ . Finally for the third term of  $a_n R_{L,r}^{(j)}$  we also use equation 1.19 to conclude that in probability

$$\lim_{n \rightarrow \infty} a_n \times \frac{N_{n,L}}{a_n} (\bar{Y}_L^* - \bar{Y}_L)^2 = \lim_{n \rightarrow \infty} \frac{N_{n,L}}{a_n} [\sqrt{a_n} (\bar{Y}_L^* - \bar{Y}_L)]^2 = 0$$

Consequently,

$$\lim_{n \rightarrow \infty} a_n R_{L,r}^{(j)} = 0.$$

Symmetrically, we also have

$$\delta_R = \frac{1}{p_R} \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} \geq q_r^{*(j)}} \right)^2 + R_{R,r}^{(j)},$$

with  $\lim_{n \rightarrow \infty} a_n R_{R,r}^{(j)} = 0$ , in probability.

Each component of  $\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)}$  writes, with  $(j, r) \in \mathcal{C}_1 \setminus (j_1, r_1)$ ,

$$\begin{aligned} L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) &= \frac{1}{p_{L,r}^{(j)}} \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} < q_r^{*(j)}} \right)^2 \\ &+ \frac{1}{p_{R,r}^{(j)}} \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j)} \geq q_r^{*(j)}} \right)^2 - \frac{1}{p_{L,r_1}^{(j_1)}} \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j_1)} < q_{r_1}^{*(j_1)}} \right)^2 \\ &- \frac{1}{p_{R,r_1}^{(j_1)}} \left( \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j_1)} \geq q_{r_1}^{*(j_1)}} \right)^2 + R_{L,r}^{(j)} + R_{R,r}^{(j)} - R_{L,r_1}^{(j_1)} - R_{R,r_1}^{(j_1)}. \end{aligned}$$

We explicitly write  $\mathcal{C}_1 = \{(j_k, r_k)\}_{k=1, \dots, c_1}$ . Then  $\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)}$  can be decomposed as

$$a_n \mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)} = h_{\mathcal{P}_1}(\mathbf{V}_n) + \mathbf{R}_{n, \mathcal{P}_1},$$

where for  $k \in \{1, \dots, c_1\}$ ,

$$\begin{aligned} V_{n, 2k-1} &= \sqrt{\frac{a_n}{p_{L, r_k}^{(j_k)}}} \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j_k)} < q_{r_k}^{*(j_k)}}, \\ V_{n, 2k} &= \sqrt{\frac{a_n}{p_{R, r_k}^{(j_k)}}} \frac{1}{a_n} \sum_{i=1}^{a_n} (Y_i - \mu) \mathbb{1}_{X_i^{(j_k)} \geq q_{r_k}^{*(j_k)}}. \end{aligned}$$

$h_{\mathcal{P}_1}$  is a multivariate quadratic form defined as

$$h_{\mathcal{P}_1} : \begin{pmatrix} x_1 \\ \vdots \\ x_{2c_1} \end{pmatrix} \rightarrow \begin{pmatrix} x_3^2 + x_4^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2k-1}^2 + x_{2k}^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2c_1-1}^2 + x_{2c_1}^2 - x_1^2 - x_2^2 \end{pmatrix}.$$

and  $R_{n, \mathcal{P}_1, k} = R_{L, r_k}^{(j_k)} + R_{R, r_k}^{(j_k)} - R_{L, r_1}^{(j_1)} - R_{R, r_1}^{(j_1)}$ .

From the multivariate central limit theorem,  $\mathbf{V}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbf{V}$ , where  $\mathbf{V}$  is a gaussian vector of covariance matrix  $\text{Cov}[\mathbf{Z}]$ , and  $\mathbf{Z}$  is defined as, for  $k \in \{1, \dots, c_1\}$ ,

$$Z_{2k-1} = \frac{1}{\sqrt{p_{L, k}}} (Y - \mathbb{E}[Y]) \mathbb{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}}, \quad Z_{2k} = \frac{1}{\sqrt{p_{R, k}}} (Y - \mathbb{E}[Y]) \mathbb{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}},$$

with the simplified notations  $p_{L, k} = p_{L, r_k}^{(j_k)}$  and  $p_{R, k} = p_{R, r_k}^{(j_k)}$ .

Finally, since  $\lim_{n \rightarrow \infty} \mathbf{R}_{n, \mathcal{P}_1} = \mathbf{0}$  in probability, from Slutsky's theorem and the continuous mapping theorem,  $a_n \mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} h_{\mathcal{P}_1}(\mathbf{V})$ . Note that, since  $\mathbf{X}$  has a strictly positive density, each component of  $h_{\mathcal{P}_1}(\mathbf{V})$  has a strictly positive variance.  $\square$

**PROOF OF LEMMA 5.** Consider a path  $\mathcal{P} = (j_1, r_1, \cdot)$ . Set  $\theta^{(V)} = (\theta_1^{(V)}, \cdot, \cdot) \in \Omega^{(V)}$ , a realization of the randomization of the split direction. Recalling that the best split in a random tree is the one maximizing the

CART-splitting criterion, condition on  $\Theta^{(V)} = \theta^{(V)}$ ,

$$(1.20) \quad \{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)\} = \bigcap_{\substack{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \\ \setminus (j_1, r_1)}} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}.$$

We recall that, given  $\theta^{(V)}$ , we define the set of best theoretical cuts along the variables in  $\theta_1^{(V)}$  as

$$\mathcal{C}_1^*(\theta_1^{(V)}) = \operatorname{argmax}_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\}} L^*(\mathbb{R}^p, q_r^{*(j)}).$$

Obviously if  $(j_1, r_1) \notin \theta_1^{(V)} \times \{1, \dots, q-1\}$ , the probability to select  $\mathcal{P}_1$  in the empirical and theoretical tree is null. In the sequel, we assume that  $(j_1, r_1) \in \theta_1^{(V)} \times \{1, \dots, q-1\}$  and distinguish between four cases:  $(j_1, r_1)$  is not among the best theoretical cuts  $\mathcal{C}_1^*(\theta_1^{(V)})$ , is the only element in  $\mathcal{C}_1^*(\theta_1^{(V)})$ , is one element of  $\mathcal{C}_1^*(\theta_1^{(V)})$  with a positive value of the theoretical CART-splitting criterion, or finally, is one element of  $\mathcal{C}_1^*(\theta_1^{(V)})$  that all have a null value of the theoretical CART-splitting criterion.

*Case 1.* We assume that  $(j_1, r_1) \notin \mathcal{C}_1^*(\theta_1^{(V)})$ . By definition of the theoretical random forest,

$$(1.21) \quad \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = 0$$

Let  $(j^*, r^*) \in \mathcal{C}_1^*(\theta_1^{(V)})$ , thus

$$\varepsilon = L^*(\mathbb{R}^p, q_{r^*}^{*(j^*)}) - L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) > 0.$$

Using equation (1.20), we have:

$$\begin{aligned} & \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ & \leq \mathbb{P}(L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r^*}^{(j^*)})) \\ & \leq \mathbb{P}(L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) - L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) - \epsilon > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r^*}^{(j^*)}) - L^*(\mathbb{R}^p, q_{r^*}^{*(j^*)})) \\ & \leq \mathbb{P}(L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) - L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) \\ & \quad - (L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r^*}^{(j^*)}) - L^*(\mathbb{R}^p, q_{r^*}^{*(j^*)})) > \epsilon) \end{aligned}$$

Therefore, according to Lemma 3,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = 0 = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})$$

*Case 2.* We assume that  $\mathcal{C}_1^*(\theta_1^{(V)}) = \{(j_1, r_1)\}$ . By definition of the theoretical random forest,

$$(1.22) \quad \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = 1.$$

Conditional on  $\Theta^{(V)} = \theta^{(V)}$ ,

$$\{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)\}^c = \bigcup_{\substack{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \\ \setminus (j_1, r_1)}} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \leq L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\},$$

which leads to

$$(1.23) \quad \begin{aligned} & 1 - \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ & \leq \sum_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \setminus (j_1, r_1)} \mathbb{P}(L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \leq L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})). \end{aligned}$$

From Lemma 3, for all  $j \in \theta_0^{(V)}$ ,  $r \in \{1, \dots, q-1\}$  such that  $(j, r) \neq (j_1, r_1)$ , in probability,

$$(1.24) \quad \lim_{n \rightarrow \infty} L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) = L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) - L^*(\mathbb{R}^p, q_r^{*(j)}) > 0.$$

Using inequality (1.23) and equation (1.24), we finally obtain,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = 1 = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}).$$

*Case 3.* We assume that  $(j_1, r_1) \in \mathcal{C}_1^*(\theta_1^{(V)})$ ,  $|\mathcal{C}_1^*(\theta_1^{(V)})| > 1$ , and  $L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) > 0$ . On one hand, conditional on  $\Theta^{(V)} = \theta^{(V)}$ ,

$$\{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)\} \subset \bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}.$$

On the other hand, conditional on  $\Theta^{(V)} = \theta^{(V)}$ ,

$$\begin{aligned} \{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)\}^c &= \bigcup_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \leq L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\} \\ &\quad \bigcup_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \setminus \mathcal{C}_1^*(\theta_1^{(V)})} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \leq L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}. \end{aligned}$$

Combining the two previous inclusions,

$$\begin{aligned}
0 &\leq \mathbb{P}\left(\bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}\right) \\
&\quad - \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\
&\leq \sum_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \setminus \mathcal{C}_1^*(\theta_1^{(V)})} \mathbb{P}(L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \leq L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})).
\end{aligned}$$

Using the same reasoning as in **Case 2**, we get

$$\begin{aligned}
\lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}\right) \\
- \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = 0.
\end{aligned}$$

We define the random vector  $\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1^*)}$  where each component is the difference between the empirical CART-splitting criterion for the splits  $(j, r) \in \mathcal{C}_1^* \setminus (j_1, r_1)$  and  $(j_1, r_1)$ ,

$$\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1^*)} = \left( L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \right)_{(j,r) \in \mathcal{C}_1^* \setminus (j_1, r_1)},$$

then

$$\mathbb{P}\left(\bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}\right) = \mathbb{P}(\mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1^*)} < \mathbf{0})$$

From Lemma 4 (case (a)),

$$\sqrt{a_n} \mathbf{L}_{n, \mathcal{P}_1}^{(\mathcal{C}_1^*)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma).$$

where for all  $(j, r), (j', r') \in \mathcal{C}_1^* \setminus (j_1, r_1)$ , each element of the covariance matrix  $\Sigma$  is defined by

$$\Sigma_{(j,r), (j',r')} = \text{Cov}[Z_{j,r}, Z_{j',r'}],$$

with

$$\begin{aligned}
Z_{j,r} &= \left( Y - \mu_{L,r_1}^{(j_1)} \mathbf{1}_{X^{(j_1)} < q_{r_1}^{*(j_1)}} - \mu_{R,r_1}^{(j_1)} \mathbf{1}_{X^{(j_1)} \geq q_{r_1}^{*(j_1)}} \right)^2 \\
&\quad - \left( Y - \mu_{L,r}^{(j)} \mathbf{1}_{X^{(j)} < q_r^{*(j)}} - \mu_{R,r}^{(j)} \mathbf{1}_{X^{(j)} \geq q_r^{*(j)}} \right)^2,
\end{aligned}$$

$\mu_{L,r}^{(j)} = \mathbb{E}[Y|X^{(j)} < q_r^{*(j)}]$ ,  $\mu_{R,r}^{(j)} = \mathbb{E}[Y|X^{(j)} \geq q_r^{*(j)}]$ , and the variance of  $Z_{j,r}$  is strictly positive. If  $\Phi_{\theta_1^{(V)},(j_1,r_1)}$  is the c.d.f. of the multivariate normal distribution of covariance matrix  $\Sigma$ , we can conclude

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) &= \lim_{n \rightarrow \infty} \mathbb{P}(\sqrt{a_n} \mathbf{L}_{n, \mathcal{P}_1}^{(C_1^*)} < \mathbf{0}) \\
 &= \Phi_{\theta_1^{(V)},(j_1,r_1)}(\mathbf{0}),
 \end{aligned}$$

where

$$\sum_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)})} \Phi_{\theta_1^{(V)},(j,r)}(\mathbf{0}) = 1.$$

According to Definition 1, in the theoretical random forest, if  $\mathcal{C}_1^*(\theta_1^{(V)})$  has multiple elements,  $(j_1, r_1)$  is randomly drawn with probability

$$\mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = \Phi_{\theta_1^{(V)},(j_1,r_1)}(\mathbf{0}),$$

that is

$$\begin{aligned}
 \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) &= \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) \\
 &= \Phi_{\theta_1^{(V)},(j_1,r_1)}(\mathbf{0}).
 \end{aligned}$$

We can notice that, in the specific case where  $\mathcal{C}_1^*(\theta_1^{(V)})$  has two elements, they are both selected with equal probability  $\frac{1}{2}$ . For more than two elements, the weights are not necessary equal, it depends on the covariance matrix  $\Sigma$ .

*Case 4.* We assume that all candidate splits have a null value for the theoretical CART-splitting criterion, i.e. for  $(j, r) \in \theta_1^{(V)} \times \{1, \dots, q-1\}$ ,  $L^*(\mathbb{R}^p, q_r^{*(j)}) = 0$ . Consequently  $\mathcal{C}_1^*(\theta_1^{(V)}) = \theta_1^{(V)} \times \{1, \dots, q-1\}$ . By definition

$$\mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathbf{L}_{n, \mathcal{P}_1}^{(C_1^*)} < \mathbf{0}).$$

According to Lemma 4 (case (b)),

$$a_n \mathbf{L}_{n, \mathcal{P}_1}^{(C_1^*)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} h_{\mathcal{P}_1}(\mathbf{V}),$$

where  $\mathbf{V}$  is a gaussian vector of covariance matrix  $\text{Cov}[\mathbf{Z}]$ . If  $\mathcal{C}_1^*(\theta_1^{(V)})$  is explicitly written  $\mathcal{C}_1^*(\theta_1^{(V)}) = \{(j_k, r_k)\}_{k=1, \dots, c_1}$ ,  $\mathbf{Z}$  is defined as, for  $k \in$

$\{1, \dots, c_1\}$ ,

$$Z_{2k-1} = \frac{1}{\sqrt{p_{L,k}}} (Y - \mathbb{E}[Y]) \mathbb{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}}$$

$$Z_{2k} = \frac{1}{\sqrt{p_{R,k}}} (Y - \mathbb{E}[Y]) \mathbb{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}},$$

$p_{L,k} = \mathbb{P}(X^{(j_k)} < q_{r_k}^{*(j_k)})$ ,  $p_{R,k} = \mathbb{P}(X^{(j_k)} \geq q_{r_k}^{*(j_k)})$ , and  $h_{\mathcal{P}_1}$  is a multivariate quadratic form defined as

$$h_{\mathcal{P}_1} : \begin{pmatrix} x_1 \\ \vdots \\ x_{2c_1} \end{pmatrix} \rightarrow \begin{pmatrix} x_3^2 + x_4^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2k-1}^2 + x_{2k}^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2c_1-1}^2 + x_{2c_1}^2 - x_1^2 - x_2^2 \end{pmatrix}.$$

and the variance of each component of  $h_{\mathcal{P}_1}(\mathbf{V})$  is strictly positive. If  $\Phi_{\theta_1^{(V)}, (j_1, r_1)}$  is the cdf of  $h_{\mathcal{P}_1}(\mathbf{V})$ , then as in **Case 3**,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) &= \Phi_{\theta_1^{(V)}, (j_1, r_1)}(\mathbf{0}) \\ &= \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}). \end{aligned}$$

□

### 1.3.2. Case 2: $\mathcal{P}_2$ .

PROOF OF LEMMA 6. Let  $(j, r) \in \mathcal{C}_{\mathcal{P}_1}$ .

$$\begin{aligned} &\sqrt{a_n} (L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)})) \\ &= \sqrt{a_n} [L_{a_n}(H^*(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)})] \\ &\quad + \sqrt{a_n} [L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})]. \end{aligned}$$

Since  $(j, r) \in \mathcal{C}_{\mathcal{P}_1}$ ,  $\mathbb{P}(\mathbf{X} \in H^*(\mathcal{P}_1) | X^{(j)} < q_r^{*(j)}) > 0$  and  $\mathbb{P}(\mathbf{X} \in H^*(\mathcal{P}_1) | X^{(j)} \geq q_r^{*(j)}) > 0$ . Then, we can directly apply Lemma 2 to the first term of this decomposition, which shows that, in probability

$$\lim_{n \rightarrow \infty} \sqrt{a_n} (L_{a_n}(H^*(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)})) = 0.$$

We expand the second term

$$\sqrt{a_n} (L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})) = I_1 - I_2 - I_3 + I_4,$$



where

$$I_1 = \frac{\sqrt{a_n}}{N_n(\hat{H}_n(\mathcal{P}_1))} \sum_{i=1}^{a_n} (Y_i - \bar{Y}_{\hat{H}_n(\mathcal{P}_1)})^2 \mathbf{1}_{\mathbf{X}_i \in \hat{H}_n(\mathcal{P}_1)}$$

$$I_2 = \frac{\sqrt{a_n}}{N_n(H^*(\mathcal{P}_1))} \sum_{i=1}^{a_n} (Y_i - \bar{Y}_{H^*(\mathcal{P}_1)})^2 \mathbf{1}_{\mathbf{X}_i \in H^*(\mathcal{P}_1)}$$

$$I_3 = \frac{\sqrt{a_n}}{N_n(\hat{H}_n(\mathcal{P}_1))} \sum_{i=1}^{a_n} (Y_i - \bar{Y}_{\hat{H}_L} \mathbf{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} - \bar{Y}_{\hat{H}_R} \mathbf{1}_{X_i^{(j)} \geq \hat{q}_{n,r}^{(j)}})^2 \mathbf{1}_{\mathbf{X}_i \in \hat{H}_n(\mathcal{P}_1)}$$

$$I_4 = \frac{\sqrt{a_n}}{N_n(H^*(\mathcal{P}_1))} \sum_{i=1}^{a_n} (Y_i - \bar{Y}_{H_L^*} \mathbf{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}} - \bar{Y}_{H_R^*} \mathbf{1}_{X_i^{(j)} \geq \hat{q}_{n,r}^{(j)}})^2 \mathbf{1}_{\mathbf{X}_i \in H^*(\mathcal{P}_1)}$$

with  $\hat{H}_L = \{\mathbf{x} \in \hat{H}_n(\mathcal{P}_1) : x^{(j)} < \hat{q}_{n,r}^{(j)}\}$ ,  $H_L^* = \{\mathbf{x} \in H^*(\mathcal{P}_1) : x^{(j)} < \hat{q}_{n,r}^{(j)}\}$ , and for all  $H \subseteq \mathbb{R}^p$

$$N_n(H) = \frac{1}{a_n} \sum_{i=1}^{a_n} \mathbf{1}_{\mathbf{X}_i \in H}, \quad \bar{Y}_H = \frac{1}{N_n(H)} \sum_{i=1}^{a_n} Y_i \mathbf{1}_{\mathbf{X}_i \in H}.$$

We define symmetrically  $\hat{H}_R$  and  $H_R^*$ . We obtain

$$\begin{aligned} & \sqrt{a_n} (L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})) \\ &= \underbrace{\sqrt{a_n} (\bar{Y}_{H^*(\mathcal{P}_1)}^2 - \bar{Y}_{\hat{H}_n(\mathcal{P}_1)}^2)}_{\Delta_{n,1}} \\ &+ \underbrace{\sqrt{a_n} \frac{\bar{Y}_{\hat{H}_L}^2 N_n(\hat{H}_L) N_n(H^*(\mathcal{P}_1)) - \bar{Y}_{H_L^*}^2 N_n(H_L^*) N_n(\hat{H}_n(\mathcal{P}_1))}{N_n(\hat{H}_n(\mathcal{P}_1)) N_n(H^*(\mathcal{P}_1))}}_{\Delta_{n,2}} \\ &+ \underbrace{\sqrt{a_n} \frac{\bar{Y}_{\hat{H}_R}^2 N_n(\hat{H}_R) N_n(H^*(\mathcal{P}_1)) - \bar{Y}_{H_R^*}^2 N_n(H_R^*) N_n(\hat{H}_n(\mathcal{P}_1))}{N_n(\hat{H}_n(\mathcal{P}_1)) N_n(H^*(\mathcal{P}_1))}}_{\Delta_{n,3}} \end{aligned}$$

We first consider  $\Delta_{n,1}$ . Simple calculations show that

$$\begin{aligned} \Delta_{n,1} &= \frac{\sqrt{a_n}}{N_n(H^*(\mathcal{P}_1))^2 N_n(\hat{H}_n(\mathcal{P}_1))^2} \\ &\times \sum_{i,k,l,m} Y_i Y_k [\mathbf{1}_{\mathbf{X}_i \in H^*(\mathcal{P}_1), \mathbf{X}_k \in H^*(\mathcal{P}_1), \mathbf{X}_l \in \hat{H}_n(\mathcal{P}_1), \mathbf{X}_m \in \hat{H}_n(\mathcal{P}_1)} \\ &\quad - \mathbf{1}_{\mathbf{X}_i \in \hat{H}_n(\mathcal{P}_1), \mathbf{X}_k \in \hat{H}_n(\mathcal{P}_1), \mathbf{X}_l \in H^*(\mathcal{P}_1), \mathbf{X}_m \in H^*(\mathcal{P}_1)}] \end{aligned}$$

We consider the case  $s_1 = L$ , ( $s_1 = R$  is similar). Since  $Y_i \in \{0, 1\}$ ,

$$\begin{aligned} |\Delta_{n,1}| &\leq \frac{\sqrt{a_n}}{N_n(H^*(\mathcal{P}_1))^2 N_n(\hat{H}_n(\mathcal{P}_1))^2} \\ &\quad \times \sum_{i,k,l,m} \left| \mathbb{1}_{X_i^{(j_1)} < q_{r_1}^{*(j_1)}, X_k^{(j_1)} < q_{r_1}^{*(j_1)}, X_l^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_m^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}} \right. \\ &\quad \left. - \mathbb{1}_{X_i^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_k^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_l^{(j_1)} < q_{r_1}^{*(j_1)}, X_m^{(j_1)} < q_{r_1}^{*(j_1)}} \right| \end{aligned}$$

As in the proof of Lemma 2, according to Lemma 1,  $\lim_{n \rightarrow \infty} \Delta_{n,1} = 0$ , in probability. Since  $\Delta_{n,2}$  and  $\Delta_{n,3}$  are the same quantities computed on each of the two daughter nodes, we study  $\Delta_{n,2}$  only.

$$\begin{aligned} \Delta_{n,2} &= \frac{\sqrt{a_n}(N_n(\hat{H}_L)N_n(H_L^*))^{-1}}{N_n(\hat{H}_n(\mathcal{P}_1))N_n(H^*(\mathcal{P}_1))} \\ &\quad \times \sum_{i,k,l,m} Y_i Y_k \left[ \mathbb{1}_{\mathbf{X}_i \in \hat{H}_L, \mathbf{X}_k \in \hat{H}_L, \mathbf{X}_l \in H_L^*, \mathbf{X}_m \in H^*(\mathcal{P}_1)} \right. \\ &\quad \left. - \mathbb{1}_{\mathbf{X}_i \in H_L^*, \mathbf{X}_k \in H_L^*, \mathbf{X}_l \in \hat{H}_L, \mathbf{X}_m \in \hat{H}_n(\mathcal{P}_1)} \right] \\ &= \frac{\sqrt{a_n}(N_n(\hat{H}_L)N_n(H_L^*))^{-1}}{N_n(\hat{H}_n(\mathcal{P}_1))N_n(H^*(\mathcal{P}_1))} \sum_{i,k,l,m} Y_i Y_k \mathbb{1}_{X_i^{(j)} < \hat{q}_{n,r}^{(j)}, X_k^{(j)} < \hat{q}_{n,r}^{(j)}, X_l^{(j)} < \hat{q}_{n,r}^{(j)}} \\ &\quad \times \left[ \mathbb{1}_{X_i^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_k^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_l^{(j_1)} < q_{r_1}^{*(j_1)}, X_m^{(j_1)} < q_{r_1}^{*(j_1)}} \right. \\ &\quad \left. - \mathbb{1}_{X_i^{(j_1)} < q_{r_1}^{*(j_1)}, X_k^{(j_1)} < q_{r_1}^{*(j_1)}, X_l^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_m^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}} \right]. \end{aligned}$$

Therefore

$$\begin{aligned} |\Delta_{n,2}| &\leq \frac{\sqrt{a_n}(N_n(\hat{H}_L)N_n(H_L^*))^{-1}}{N_n(\hat{H}_n(\mathcal{P}_1))N_n(H^*(\mathcal{P}_1))} \\ &\quad \times \sum_{i,k,l,m} \left| \mathbb{1}_{X_i^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_k^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_l^{(j_1)} < q_{r_1}^{*(j_1)}, X_m^{(j_1)} < q_{r_1}^{*(j_1)}} \right. \\ &\quad \left. - \mathbb{1}_{X_i^{(j_1)} < q_{r_1}^{*(j_1)}, X_k^{(j_1)} < q_{r_1}^{*(j_1)}, X_l^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}, X_m^{(j_1)} < \hat{q}_{n,r_1}^{(j_1)}} \right|. \end{aligned}$$

As in the proof of Lemma 2, according to Lemma 1,  $\lim_{n \rightarrow \infty} \Delta_{n,2} = 0$ , in probability, which concludes the proof, since  $\Delta_{n,3}$  can be studied in the same manner.  $\square$

PROOF OF LEMMA 7. Let  $(j, r) \in \mathcal{C}_{\mathcal{P}_1}$ .

$$(1.25) \quad \begin{aligned} L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) &= L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)}) \\ &\quad + [L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)})] \end{aligned}$$

According to Lemma 6, the second term in equation (1.25) converges to 0 in probability. From the law of large numbers, in probability,

$$\lim_{n \rightarrow \infty} L_{a_n}(H^*(\mathcal{P}_1), q_r^{*(j)}) = L^*(H^*(\mathcal{P}_1), q_r^{*(j)}),$$

which concludes the proof.  $\square$

PROOF OF LEMMA 8. Similar to the case with  $\mathcal{P}_1$  (Lemma 3), where Lemma 6 is used instead of Lemma 2.  $\square$

PROOF OF LEMMA 9. Consider a path  $\mathcal{P}_2 = \{(j_1, r_1, L), (j_2, r_2, \cdot)\}$ . Set  $\theta^{(V)} = (\theta_1^{(V)}, \theta_2^{(V)})$ , a realization of  $\Theta^{(V)} = (\Theta_0^{(V)}, \Theta_L^{(V)})$ , the randomization of the split direction at the root node and its left child node— see Section 2, and note that we omit  $\Theta_R^{(V)}$  to simplify notations. Then,  $\theta_1^{(V)}$  and  $\theta_2^{(V)}$  denotes the set of eligible variables for respectively the first and second split. We also consider  $\mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) \subset \mathcal{C}_{\mathcal{P}_1}$  the set of eligible second splits.

Recalling that the best split in a random tree is the one maximizing the CART-splitting criterion, conditional on  $\Theta^{(V)} = \theta^{(V)}$ ,

$$\begin{aligned} \{\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n)\} &= \bigcap_{\substack{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \\ \setminus (j_1, r_1)}} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\} \\ &\quad \bigcap_{(j,r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) \setminus (j_2, r_2)} \{L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})\} \end{aligned}$$

Recall that  $\mathcal{C}_1^*(\theta_1^{(V)}) = \operatorname{argmax}_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\}} L^*(\mathbb{R}^p, q_r^{*(j)})$ , and similarly

$$\mathcal{C}_2^*(\theta_2^{(V)}) = \operatorname{argmax}_{(j,r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)})} L^*(H^*(\mathcal{P}_1), q_r^{*(j)}).$$

Obviously if  $(j_1, r_1) \notin \theta_1^{(V)} \times \{1, \dots, q-1\}$  or  $(j_2, r_2) \notin \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)})$ , the probability to select  $\mathcal{P}_2$  in the empirical and theoretical tree is null. In the sequel, we assume that  $(j_1, r_1) \in \theta_1^{(V)} \times \{1, \dots, q-1\}$  and  $(j_2, r_2) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)})$  and distinguish between cases, depending on whether  $(j_1, r_1) \in \mathcal{C}_1^*(\theta_1^{(V)})$  or not and  $(j_2, r_2) \in \mathcal{C}_2^*(\theta_2^{(V)})$  or not, as well as the cardinality of  $\mathcal{C}_1^*(\theta_1^{(V)})$  and  $\mathcal{C}_2^*(\theta_2^{(V)})$ , and whether the maximum of the theoretical CART-splitting criterion is null or not.

*Case 1.* We assume that  $(j_1, r_1) \notin \mathcal{C}_1^*(\theta_1^{(V)})$ . Hence, the theoretical decision tree satisfies

$$\mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = 0.$$

According to Lemma 5, we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ & \leq \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ & = 0 \\ & = \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}). \end{aligned}$$

*Case 2.* We assume that  $(j_2, r_2) \notin \mathcal{C}_2^*(\theta_2^{(V)})$ . Again, for the theoretical decision tree,

$$\mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = 0.$$

Letting  $(j^*, r^*) \in \mathcal{C}_2^*(\theta_2^{(V)})$ ,

$$\varepsilon = L^*(H^*(\mathcal{P}_1), q_{r^*}^{*(j^*)}) - L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}).$$

Therefore,

$$\begin{aligned} & \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ & \leq \mathbb{P}(L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) > L_{a_n}(H^*(\mathcal{P}_1), \hat{q}_{n,r^*}^{*(j^*)})) \\ & \leq \mathbb{P}(L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) - L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) - \epsilon \\ & \quad > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r^*}^{*(j^*)}) - L^*(H^*(\mathcal{P}_1), q_{r^*}^{*(j^*)})) \\ & \leq \mathbb{P}(L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) - L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) \\ & \quad - (L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r^*}^{*(j^*)}) - L^*(H^*(\mathcal{P}_1), q_{r^*}^{*(j^*)})) > \epsilon). \end{aligned}$$

Consequently, according to Lemma 7,

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = 0 = \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}).$$

*Case 3.* We assume that  $(j_1, r_1) \in \mathcal{C}_1^*(\theta_1^{(V)})$  and  $\mathcal{C}_2^*(\theta_2^{(V)}) = \{(j_2, r_2)\}$ , i.e.  $(j_2, r_2)$  is the unique maximum of the theoretical CART-splitting criterion for the cell  $H^*(\mathcal{P}_1)$ . By definition of the theoretical decision tree,

$$\mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})$$

Conditional on  $\{\Theta^{(V)} = \theta^{(V)}\}$ ,

$$(1.26) \quad \begin{aligned} \{\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n)\} &= \{\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n)\} \\ &\cap \bigcap_{(j,r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) \setminus (j_2, r_2)} \{L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})\}. \end{aligned}$$

Consequently,

$$(1.27) \quad \begin{aligned} &\mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ &\geq \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ &- \sum_{(j,r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) \setminus (j_2, r_2)} \mathbb{P}(L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) \leq L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})). \end{aligned}$$

For  $(j, r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) \setminus (j_2, r_2)$ ,

$$(1.28) \quad L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) - L^*(H^*(\mathcal{P}_1), q_r^{*(j)}) > 0.$$

Thus, using inequalities (1.27) and (1.28), and according to Lemma 7,

$$\lim_{n \rightarrow \infty} \mathbb{P}(L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) \leq L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})) = 0,$$

and thus, using (1.26) and (1.27),

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) \\ &= \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}), \end{aligned}$$

where the second inequality is a direct consequence of Lemma 5.

*Case 4.* For the first split, we assume  $(j_1, r_1) \in \mathcal{C}_1^*(\theta_1^{(V)})$  with  $L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) > 0$ , and for the second split,  $(j_2, r_2) \in \mathcal{C}_2^*(\theta_2^{(V)})$  with  $|\mathcal{C}_2^*(\theta_2^{(V)})| > 1$  and  $L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) > 0$ .

On one hand, conditional on the event  $\{\Theta^{(V)} = \theta^{(V)}\}$ ,

$$(1.29) \quad \begin{aligned} \{\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n)\} &= \bigcap_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\} \\ &\cap \bigcap_{(j,r) \in \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)}) \setminus (j_2, r_2)} \{L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})\}. \end{aligned}$$

Using equation (1.29) to find a subset and a superset of  $\{\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n)\}$ , we obtain

$$\begin{aligned}
0 &\geq \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\
&\quad - \mathbb{P}\left(\bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}\right. \\
&\quad \quad \left. \bigcap_{(j,r) \in \mathcal{C}_2^*(\theta_2^{(V)}) \setminus (j_2, r_2)} \{L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})\}\right) \\
&\geq \sum_{(j,r) \in \theta_1^{(V)} \times \{1, \dots, q-1\} \setminus \mathcal{C}_1^*(\theta_1^{(V)})} \mathbb{P}(L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) \leq L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})) \\
&\quad + \sum_{(j,r) \in \theta_2^{(V)} \times \{1, \dots, q-1\} \setminus \mathcal{C}_2^*(\theta_2^{(V)})} \mathbb{P}(L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) \leq L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}))
\end{aligned}$$

We proved in **Case 3** that the limit of the last two terms of the previous inequality is zero, in probability. Therefore,

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) \\
&= \lim_{n \rightarrow \infty} \mathbb{P}\left(\bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)})\}\right. \\
(1.30) \quad &\quad \left. \bigcap_{(j,r) \in \mathcal{C}_2^*(\theta_2^{(V)}) \setminus (j_2, r_2)} \{L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)}) > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)})\}\right).
\end{aligned}$$

We define the random vector  $\mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)}$  (we drop  $\theta^{(V)}$  to lighten notations) where each component is the difference between the empirical CART-splitting criterion for the splits  $(j, r) \in \mathcal{C}_1^* \setminus (j_1, r_1)$  and  $(j_1, r_1)$  for the first  $|\mathcal{C}_1^*| - 1$  components, and for the splits  $(j, r) \in \mathcal{C}_2^* \setminus (j_2, r_2)$  and  $(j_2, r_2)$  for the remaining  $|\mathcal{C}_2^*| - 1$  components, i.e.,

$$\mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)} = \left( \begin{array}{c} [L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r}^{(j)}) - L_{a_n}(\mathbb{R}^p, \hat{q}_{n,r_1}^{(j_1)})]_{(j,r) \in \mathcal{C}_1^* \setminus (j_1, r_1)} \\ [L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r}^{(j)}) - L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n,r_2}^{(j_2)})]_{(j,r) \in \mathcal{C}_2^* \setminus (j_2, r_2)} \end{array} \right).$$

Then, we can write

$$\begin{aligned}
 & \mathbb{P} \left( \bigcap_{(j,r) \in \mathcal{C}_1^*(\theta_1^{(V)}) \setminus (j_1, r_1)} \{L_{a_n}(\mathbb{R}^p, \hat{q}_{n, r_1}^{(j_1)}) > L_{a_n}(\mathbb{R}^p, \hat{q}_{n, r}^{(j)})\} \right. \\
 & \quad \left. \bigcap_{(j,r) \in \mathcal{C}_2^*(\theta_2^{(V)}) \setminus (j_2, r_2)} \{L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n, r_2}^{(j_2)}) > L_{a_n}(\hat{H}_n(\mathcal{P}_1), \hat{q}_{n, r}^{(j)})\} \right) \\
 (1.31) \quad & = \mathbb{P}(\mathbf{L}_{n,2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)} < \mathbf{0})
 \end{aligned}$$

According to Lemma 8,

$$\sqrt{a_n} \mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma)$$

where for  $l, l' \in \{1, 2\}$ , for all  $(j, r) \in \mathcal{C}_l^* \setminus (j_l, r_l)$ ,  $(j', r') \in \mathcal{C}_{l'}^* \setminus (j_{l'}, r_{l'})$ , each element of the covariance matrix  $\Sigma$  is defined by  $\Sigma_{(j,r,l),(j',r',l')} = \text{Cov}[Z_{j,r,l}, Z_{j',r',l'}]$ , with

$$\begin{aligned}
 Z_{j,r,l} &= \frac{1}{\mathbb{P}(\mathbf{X} \in H_l)} (Y - \mu_{L,r_l}^{(j_l)} \mathbf{1}_{X^{(j_l)} < q_{r_l}^{*(j_l)}} - \mu_{R,r_l}^{(j_l)} \mathbf{1}_{X^{(j_l)} \geq q_{r_l}^{*(j_l)}})^2 \mathbf{1}_{\mathbf{X} \in H_l} \\
 & \quad - \frac{1}{\mathbb{P}(\mathbf{X} \in H_l)} (Y - \mu_{L,r}^{(j)} \mathbf{1}_{X^{(j)} < q_r^{*(j)}} - \mu_{R,r}^{(j)} \mathbf{1}_{X^{(j)} \geq q_r^{*(j)}})^2 \mathbf{1}_{\mathbf{X} \in H_l},
 \end{aligned}$$

$\mu_{L,r}^{(j)} = \mathbb{E}[Y|X^{(j)} < q_r^{*(j)}, \mathbf{X} \in H_l]$ ,  $\mu_{R,r}^{(j)} = \mathbb{E}[Y|X^{(j)} \geq q_r^{*(j)}, \mathbf{X} \in H_l]$ , and the variance of  $Z_{j,r,l}$  is strictly positive.

Letting  $\Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}$  be the c.d.f. of the multivariate normal distribution with covariance matrix  $\Sigma$ , and using equalities (1.30) and (1.31),

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = \Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\mathbf{0}).$$

We can check that

$$\sum_{(j,r) \in \mathcal{C}_2^*(\theta^{(V)})} \Phi_{\mathcal{P}_1, \theta^{(V)}, (j,r)}(\mathbf{0}) = \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}).$$

In the theoretical random forest, the first cut  $(j_1, r_1)$  is randomly selected with probability  $\mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})$  (see the proof of Lemma 5). For the second cut, according to Definition 2, if  $\mathcal{C}_2^*(\theta_2^{(V)})$  has multiple elements,  $(j_2, r_2)$  is randomly drawn with probability

$$\frac{\Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\mathbf{0})}{\mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})}$$

Since the random selection at the root node of the tree and its children nodes are independent in the theoretical algorithm,  $\mathcal{P}_2$  is selected with probability

$$\begin{aligned} \mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) &\times \frac{\Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\mathbf{0})}{\mathbb{P}(\mathcal{P}_1 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)})} \\ &= \Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\mathbf{0}). \end{aligned}$$

Ultimately,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) &= \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) \\ &= \Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\mathbf{0}). \end{aligned}$$

*Case 5.* We assume that  $(j_1, r_1) \in \mathcal{C}_1^*(\theta_1^{(V)})$  and  $(j_2, r_2) \in \mathcal{C}_2^*(\theta_2^{(V)})$ , and that the theoretical CART-splitting criterion is null for both splits:  $L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) = 0$  and  $L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) = 0$ .

Consequently  $\mathcal{C}_1^*(\theta_1^{(V)}) = \theta_1^{(V)} \times \{1, \dots, q-1\}$ , and  $\mathcal{C}_2^*(\theta_2^{(V)}) = \mathcal{C}_{\mathcal{P}_1}(\theta_2^{(V)})$ . Using the same notations defined in **Case 4**, we have by definition

$$\mathbb{P}(\mathcal{P}_1 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) = \mathbb{P}(\mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)} < \mathbf{0}).$$

According to Lemma 8 (case (b)),

$$a_n \mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} h_{\mathcal{P}_2}(\mathbf{V}),$$

where  $\mathbf{V}$  is a gaussian vector of covariance matrix  $\text{Cov}[\mathbf{Z}]$ . If  $\mathcal{C}_1^*$  and  $\mathcal{C}_2^*$  are explicitly written  $\mathcal{C}_1^* = \{(j_k, r_k)\}_{k \in J_1}$ , and  $\mathcal{C}_2^* = \{(j_k, r_k)\}_{k \in J_2}$ , with  $J_1 = \{1, \dots, c_1 + 1\} \setminus 2$  and  $J_2 = \{2\} \cup \{c_1 + 2, \dots, c_1 + c_2\}$ ,  $\mathbf{Z}$  is defined as, for  $l \in \{1, 2\}$  and  $k \in J_l$ ,

$$\begin{aligned} Z_{2k-1} &= \frac{1}{\sqrt{p_{L,k} \mathbb{P}(\mathbf{X} \in H_l)}} (Y - \mathbb{E}[Y | \mathbf{X} \in H_l]) \mathbb{1}_{X^{(j_k)} < q_{r_k}^{*(j_k)}} \mathbb{1}_{\mathbf{X} \in H_l}, \\ Z_{2k} &= \frac{1}{\sqrt{p_{R,k} \mathbb{P}(\mathbf{X} \in H_l)}} (Y - \mathbb{E}[Y | \mathbf{X} \in H_l]) \mathbb{1}_{X^{(j_k)} \geq q_{r_k}^{*(j_k)}} \mathbb{1}_{\mathbf{X} \in H_l}, \end{aligned}$$

$p_{L,k} = \mathbb{P}(X^{(j_k)} < q_{r_k}^{*(j_k)}, \mathbf{X} \in H_l)$ ,  $p_{R,k} = \mathbb{P}(X^{(j_k)} \geq q_{r_k}^{*(j_k)}, \mathbf{X} \in H_l)$ .  $h_{\mathcal{P}_2}$  is a multivariate quadratic form defined as

$$h_{\mathcal{P}_2} : \begin{pmatrix} x_1 \\ \vdots \\ x_{2(c_1+c_2)} \end{pmatrix} \rightarrow \begin{pmatrix} x_5^2 + x_6^2 - x_1^2 - x_2^2 \\ \vdots \\ x_{2c_1+1}^2 + x_{2c_1+2}^2 - x_1^2 - x_2^2 \\ x_{2c_1+3}^2 + x_{2c_1+4}^2 - x_3^2 - x_4^2 \\ \vdots \\ x_{2(c_1+c_2)-1}^2 + x_{2(c_1+c_2)}^2 - x_3^2 - x_4^2 \end{pmatrix},$$



and the variance of each component of  $h_{\mathcal{P}_2}(\mathbf{V})$  is strictly positive.

$\Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}$  is now defined as the cdf of  $h_{\mathcal{P}_2}(\mathbf{V})$ , and the end of the proof is identical to **Case 4**. We conclude

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{P}_2 \in T(\Theta, \mathcal{D}_n) | \Theta^{(V)} = \theta^{(V)}) &= \mathbb{P}(\mathcal{P}_2 \in T^*(\Theta) | \Theta^{(V)} = \theta^{(V)}) \\ &= \Phi_{\mathcal{P}_1, \theta^{(V)}, (j_2, r_2)}(\mathbf{0}). \end{aligned}$$

*Case 6.* We assume  $(j_1, r_1) \in \mathcal{C}_1^*(\theta_1^{(V)})$ ,  $(j_2, r_2) \in \mathcal{C}_2^*(\theta_2^{(V)})$  and  $|\mathcal{C}_2^*(\theta_2^{(V)})| > 1$  as in **Case 4**, but either  $L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) = 0$  and  $L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) > 0$ , or  $L^*(\mathbb{R}^p, q_{r_1}^{*(j_1)}) > 0$  and  $L^*(H^*(\mathcal{P}_1), q_{r_2}^{*(j_2)}) = 0$ .

The same reasoning than for **Cases 4** and **5** applies where the limit law of  $\mathbf{L}_{n, \mathcal{P}_2}^{(\mathcal{C}_1^*, \mathcal{C}_2^*)}$  has both gaussian and  $\chi$ -square components and is given by case (c) or case (d) of Lemma 8.  $\square$

## 2. Proof of Theorem 2.

We recall Theorem 2 for the sake of clarity.

**THEOREM 2.** *If  $p_0 \in [0, 1] \setminus \mathcal{U}_n$  and  $\mathcal{D}'_n = \mathcal{D}_n$ , then, conditional on  $\mathcal{D}_n$ , we have*

$$(2.1) \quad \lim_{M \rightarrow \infty} \hat{S}_{M, n, p_0} = 1 \quad \text{in probability.}$$

In addition for  $p_0 < \max \mathcal{U}_n$ ,

$$1 - \mathbb{E}[\hat{S}_{M, n, p_0} | \mathcal{D}_n] \underset{M \rightarrow \infty}{\sim} \sum_{\mathcal{P} \in \Pi} \frac{\Phi(Mp_0, M, p_n(\mathcal{P}))(1 - \Phi(Mp_0, M, p_n(\mathcal{P})))}{\frac{1}{2} \sum_{\mathcal{P}' \in \Pi} \mathbb{1}_{p_n(\mathcal{P}') > p_0} + \mathbb{1}_{p_n(\mathcal{P}') > p_0 - \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})} (p_0 - p_n(\mathcal{P}))}},$$

where  $\Phi(Mp_0, M, p_n(\mathcal{P}))$  is the cdf of a binomial distribution with parameter  $p_n(\mathcal{P})$ ,  $M$  trials, evaluated at  $Mp_0$ , and, for all  $\mathcal{P}, \mathcal{P}' \in \Pi$ ,

$$\sigma_n(\mathcal{P}) = \sqrt{p_n(\mathcal{P})(1 - p_n(\mathcal{P}))},$$

and

$$\rho_n(\mathcal{P}, \mathcal{P}') = \frac{\text{Cov}(\mathbb{1}_{\mathcal{P} \in T(\Theta, \mathcal{D}_n)}, \mathbb{1}_{\mathcal{P}' \in T(\Theta, \mathcal{D}_n)} | \mathcal{D}_n)}{\sigma_n(\mathcal{P})\sigma_n(\mathcal{P}')}.$$

Let  $p_0 \in [0, \max \mathcal{U}_n) \setminus \mathcal{U}_n$  and  $\mathcal{D}'_n = \mathcal{D}_n$ . Before proving Theorem 2, we need the following two lemmas.

LEMMA 10. *Let  $F$  be the hypergeometric function. Then, for  $(a, c) \in \mathbb{Z}^2$  and  $\mathcal{P} \in \Pi$  such that  $p_n(\mathcal{P}) > p_0$ , we have*

$$\lim_{M \rightarrow \infty} \frac{F(M + a, 1, M(1 - p_0) + c, 1 - p_n(\mathcal{P}))}{F(M + 1, 1, M(1 - p_0) + 1, 1 - p_n(\mathcal{P}))} = 1.$$

LEMMA 11. *Let  $\mathcal{P}' \in \Pi$ . For all  $\mathcal{P} \in \Pi$  such that  $p_n(\mathcal{P}) > p_0$ , we have*

$$\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) > p_0, \mathcal{D}_n) = \mathbf{1}_{p_n(\mathcal{P}') > p_0}$$

and

$$\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) = \mathbf{1}_{p_n(\mathcal{P}') > p_0 - \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})}} \times (p_0 - p_n(\mathcal{P}))$$

Symmetrically, for all  $\mathcal{P} \in \Pi$  such that  $p_n(\mathcal{P}) \leq p_0$ , we have

$$\begin{aligned} \lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) &= \mathbf{1}_{p_n(\mathcal{P}') > p_0}, \\ \lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) > p_0, \mathcal{D}_n) &= \mathbf{1}_{p_n(\mathcal{P}') > p_0 - \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})}} \times (p_0 - p_n(\mathcal{P})) \end{aligned}$$

We are now in a position to prove Theorem 2.

PROOF OF THEOREM 2. The first statement, identity (2.1), is proved similarly to Corollary 2, using the law of large numbers instead of Theorem 1. For the second statement, we first recall that, by definition,

$$\begin{aligned} \hat{S}_{M,n,p_0} &= \frac{2 \sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0 \cap \hat{p}'_{M,n}(\mathcal{P}) > p_0}}{\sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0} + \mathbf{1}_{\hat{p}'_{M,n}(\mathcal{P}) > p_0}} \\ &= 1 - \frac{\sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0 \cap \hat{p}'_{M,n}(\mathcal{P}) \leq p_0} + \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) \leq p_0 \cap \hat{p}'_{M,n}(\mathcal{P}) > p_0}}{\sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0} + \mathbf{1}_{\hat{p}'_{M,n}(\mathcal{P}) > p_0}}. \end{aligned}$$

Taking the expectation conditional on  $\mathcal{D}_n$  gives

$$\begin{aligned} \mathbb{E}[\hat{S}_{M,n,p_0} | \mathcal{D}_n] &= 1 - 2 \mathbb{E} \left[ \frac{\sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0 \cap \hat{p}'_{M,n}(\mathcal{P}) \leq p_0}}{\sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0} + \mathbf{1}_{\hat{p}'_{M,n}(\mathcal{P}) > p_0}} \middle| \mathcal{D}_n \right] \\ &= 1 - 2 \mathbb{E} \left[ \frac{U_M}{V_M + V'_M} \middle| \mathcal{D}_n \right], \end{aligned}$$

where  $U_M = \sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0 \cap \hat{p}'_{M,n}(\mathcal{P}) \leq p_0}$ ,  $V_M = \sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0}$ , and  $V'_M = \sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}'_{M,n}(\mathcal{P}) > p_0}$ . Note that

$$\begin{aligned} \mathbb{E}[V_M | \mathcal{D}_n] &= \sum_{\mathcal{P} \in \Pi} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) > p_0 | \mathcal{D}_n) \xrightarrow{M \rightarrow \infty} \sum_{\mathcal{P} \in \Pi} \mathbf{1}_{p_n(\mathcal{P}) > p_0}, \\ \mathbb{E}[U_M | \mathcal{D}_n] &= \sum_{\mathcal{P} \in \Pi} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) > p_0 | \mathcal{D}_n) \mathbb{P}(\hat{p}'_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n) \xrightarrow{M \rightarrow \infty} 0. \end{aligned}$$

Also,

$$\begin{aligned} &\mathbb{E} \left[ \frac{U_M}{V_M + V'_M} \middle| \mathcal{D}_n \right] \\ &= \sum_{m, m'} \frac{1}{m + m'} \mathbb{E}[U_M | V_M = m, V'_M = m', \mathcal{D}_n] \\ &\quad \times \mathbb{P}(V_M = m | \mathcal{D}_n) \mathbb{P}(V'_M = m' | \mathcal{D}_n) \\ &= \sum_{m, m'} \frac{1}{m + m'} \mathbb{E} \left[ \sum_{\mathcal{P} \in \Pi} \mathbf{1}_{\hat{p}_{M,n}(\mathcal{P}) > p_0 \cap \hat{p}'_{M,n}(\mathcal{P}) \leq p_0} \middle| V_M = m, V'_M = m', \mathcal{D}_n \right] \\ &\quad \times \mathbb{P}(V_M = m | \mathcal{D}_n) \mathbb{P}(V'_M = m' | \mathcal{D}_n) \\ &= \sum_{m, m'} \frac{1}{m + m'} \sum_{\mathcal{P} \in \Pi} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) > p_0 | V_M = m, \mathcal{D}_n) \\ &\quad \times \mathbb{P}(\hat{p}'_{M,n}(\mathcal{P}) \leq p_0 | V'_M = m', \mathcal{D}_n) \mathbb{P}(V_M = m | \mathcal{D}_n) \mathbb{P}(V'_M = m' | \mathcal{D}_n) \\ &= \sum_{m, m'} \frac{1}{m + m'} \sum_{\mathcal{P} \in \Pi} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) > p_0, V_M = m | \mathcal{D}_n) \\ &\quad \times \mathbb{P}(\hat{p}'_{M,n}(\mathcal{P}) \leq p_0, V'_M = m' | \mathcal{D}_n) \\ &= \sum_{\mathcal{P} \in \Pi} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) > p_0 | \mathcal{D}_n) \mathbb{P}(\hat{p}'_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n) \\ &\quad \times \left[ \sum_{m, m'} \frac{1}{m + m'} \mathbb{P}(V_M = m | \hat{p}_{M,n}(\mathcal{P}) > p_0, \mathcal{D}_n) \right. \\ &\quad \left. \times \mathbb{P}(V'_M = m' | \hat{p}'_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \right]. \end{aligned}$$

For all  $\mathcal{P} \in \Pi$ ,

$$\begin{aligned} &\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) > p_0 | \mathcal{D}_n) \mathbb{P}(\hat{p}'_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n) \\ &= \Phi(Mp_0, M, p_n(\mathcal{P})) (1 - \Phi(Mp_0, M, p_n(\mathcal{P}))), \end{aligned}$$

where  $\Phi$  is the cdf of the binomial distribution. As a direct consequence of Lemma 11,

$$\begin{aligned} & \lim_{M \rightarrow \infty} \sum_{m, m'} \frac{1}{m + m'} \mathbb{P}(V_M = m | \hat{p}_{M,n}(\mathcal{P}) > p_0, \mathcal{D}_n) \\ & \quad \times \mathbb{P}(V_M = m' | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ & = \frac{1}{\sum_{\mathcal{P}' \in \Pi} \mathbb{1}_{p_n(\mathcal{P}') > p_0} + \mathbb{1}_{p_n(\mathcal{P}') + \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})} (p_0 - p_n(\mathcal{P})) > p_0}}, \end{aligned}$$

which yields

$$\begin{aligned} & 1 - \mathbb{E}[\hat{S}_{M,n,p_0} | \mathcal{D}_n] \\ & \stackrel{M \rightarrow \infty}{\sim} \sum_{\mathcal{P} \in \Pi} \frac{2\Phi(Mp_0, M, p_n(\mathcal{P}))(1 - \Phi(Mp_0, M, p_n(\mathcal{P})))}{\sum_{\mathcal{P}' \in \Pi} \mathbb{1}_{p_n(\mathcal{P}') > p_0} + \mathbb{1}_{p_n(\mathcal{P}') + \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})} (p_0 - p_n(\mathcal{P})) > p_0}}. \end{aligned}$$

This is the desired result.  $\square$

### 2.1. Proof of intermediate lemmas.

PROOF OF LEMMA 10. Cvitković, Smith and Pande (2017) provides an asymptotic expansion of the hypergeometric function  $F$  in the case where the first and third parameters goes to infinity with a constant ratio. For  $a, c, z, \varepsilon \in \mathbb{R}$ ,  $b \notin \mathbb{Z} \setminus \mathbb{N}$ , such that  $\varepsilon > 1$ , and  $z\varepsilon < 1$ , Cvitković, Smith and Pande (2017) gives in the section 2.2.2 (end of page 10)

$$(2.2) \quad F(a + \varepsilon\lambda, b, c + \lambda, z) \underset{|\lambda| \rightarrow \infty}{\sim} \frac{1}{(1 - \varepsilon z)^b}.$$

We can then derive the limit of the following ratio

$$(2.3) \quad \lim_{|\lambda| \rightarrow \infty} \frac{F(a + \varepsilon\lambda, b, c + \lambda, z)}{F(1 + \varepsilon\lambda, b, 1 + \lambda, z)} = 1$$

We use 2.3 in the specific case where  $b = 1$ ,  $a, c \in \mathbb{Z}$ ,  $\varepsilon = \frac{1}{1-p_0} > 1$ ,  $z = 1 - p_n(\mathcal{P})$  for  $\mathcal{P} \in \Pi$  such that  $p_n(\mathcal{P}) > p_0$  (and then  $z\varepsilon < 1$ ), and  $\lambda = M(1 - p_0)$ , it follows that

$$(2.4) \quad \lim_{M \rightarrow \infty} \frac{F(M + a, 1, M(1 - p_0) + c, 1 - p_n(\mathcal{P}))}{F(M + 1, 1, M(1 - p_0) + 1, 1 - p_n(\mathcal{P}))} = 1$$

$\square$

PROOF OF LEMMA 11. Fix  $\mathcal{D}_n$ . Let  $\mathcal{P}', \mathcal{P} \in \Pi$ . In what follows, when there is no ambiguity, we will replace  $T(\Theta, \mathcal{D}_n)$  by  $T_n(\Theta)$  to lighten notations.

Case 1:  $p_n(\mathcal{P}) > p_0$ .

$$\begin{aligned}
 & \mathbb{E}[\hat{p}_{M,n}(\mathcal{P}') | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] \\
 = & \mathbb{E}\left[\frac{1}{M} \sum_{l=1}^M \mathbf{1}_{\mathcal{P}' \in T_n(\Theta_l)} \mid \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n\right] \\
 = & \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
 & \times \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
 & + \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
 & \times (1 - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)) \\
 = & \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
 & + \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{D}_n) \\
 (2.5) \quad & \times (1 - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)).
 \end{aligned}$$

since

$$\begin{aligned}
 & \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
 (2.6) \quad & = \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n).
 \end{aligned}$$

because, conditional on  $\mathcal{D}_n$ , the events  $\mathcal{P}' \in T_n(\Theta_1), \dots, \mathcal{P}' \in T_n(\Theta_M)$  are independent. We can rewrite,

$$\begin{aligned}
 & \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{D}_n) \\
 & = \frac{\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_1) | \mathcal{D}_n)}{1 - p_n(\mathcal{P})} \\
 & = \frac{(1 - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \mathcal{P}' \in T_n(\Theta_1), \mathcal{D}_n)) p_n(\mathcal{P}')}{1 - p_n(\mathcal{P})} \\
 (2.7) \quad & = \frac{p_n(\mathcal{P}')}{1 - p_n(\mathcal{P})} - \frac{p_n(\mathcal{P})}{1 - p_n(\mathcal{P})} \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n),
 \end{aligned}$$

yielding, using equation (2.5),

$$\begin{aligned}
 & \mathbb{E}[\hat{p}_{M,n}(\mathcal{P}') | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] \\
 = & \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \left( \frac{\mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)}{1 - p_n(\mathcal{P})} \right. \\
 & \left. - \frac{p_n(\mathcal{P})}{1 - p_n(\mathcal{P})} \right) + \frac{p_n(\mathcal{P}')}{1 - p_n(\mathcal{P})} (1 - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)).
 \end{aligned}$$

Besides, by definition of the correlation

$$\rho_n(\mathcal{P}, \mathcal{P}') = \frac{\text{Cov}(\mathbb{1}_{\mathcal{P} \in T_n(\Theta)}, \mathbb{1}_{\mathcal{P}' \in T_n(\Theta)} | \mathcal{D}_n)}{\sigma_n(\mathcal{P})\sigma_n(\mathcal{P}')},$$

simple calculations show that

$$(2.9) \quad \begin{aligned} & \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \\ &= p_n(\mathcal{P}') + \rho_n(\mathcal{P}, \mathcal{P}') \sqrt{\frac{p_n(\mathcal{P}')}{p_n(\mathcal{P})} (1 - p_n(\mathcal{P})) (1 - p_n(\mathcal{P}'))}, \end{aligned}$$

which, together with equation (2.8) leads to,

$$(2.10) \quad \begin{aligned} & \mathbb{E}[\hat{p}_{M,n}(\mathcal{P}') | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] \\ &= p_n(\mathcal{P}') + \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})} (\mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ & \quad - p_n(\mathcal{P})). \end{aligned}$$

Regarding the probability in the right-hand side of equation (2.10), we have

$$\begin{aligned} & \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= p_n(\mathcal{P}) \frac{\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n)}{\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n)} \\ &= p_n(\mathcal{P}) \frac{\mathbb{P}((M-1)\hat{p}_{M-1,n}(\mathcal{P}) \leq Mp_0 - 1 | \mathcal{D}_n)}{\mathbb{P}(M\hat{p}_{M,n}(\mathcal{P}) \leq Mp_0 | \mathcal{D}_n)} \\ &= p_n(\mathcal{P}) \frac{\Phi(Mp_0 - 1, M-1, p_n(\mathcal{P}))}{\Phi(Mp_0, M, p_n(\mathcal{P}))}. \end{aligned}$$

Using standard formulas,  $\Phi$  can be expressed with the incomplete beta function,

$$\Phi(k, M, p) = I_{1-p}(M-k, k+1) = \frac{B_{1-p}(M-k, k+1)}{B(M-k, k+1)},$$

and the regularized beta function is related to the hypergeometric function  $F$ , for  $a > 0$ ,  $b > 0$ , and  $p \in [0, 1]$  (Olver et al., 2010),

$$B_{1-p}(a, b) = \frac{(1-p)^a p^b}{a} F(a+b, 1, a+1, 1-p).$$

Then, we can express the cdf of the binomial distribution using the hypergeometric function, and it follows

$$(2.11) \quad \begin{aligned} & \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= p_0 \frac{F(M, 1, M(1-p_0) + 1, 1 - \hat{p}_n(\mathcal{P}))}{F(M+1, 1, M(1-p_0) + 1, 1 - \hat{p}_n(\mathcal{P}))} \end{aligned}$$

According to Lemma 10,

$$(2.12) \quad \lim_{M \rightarrow \infty} \frac{F(M, 1, M(1-p_0) + 1, 1 - p_n(\mathcal{P}))}{F(M+1, 1, M(1-p_0) + 1, 1 - p_n(\mathcal{P}))} = 1.$$

Consequently,

$$(2.13) \quad \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \in T(\Theta_1, \mathcal{D}_n) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) = p_0,$$

and using this limiting result with equation (2.10) yields,

$$(2.14) \quad \begin{aligned} \lim_{M \rightarrow \infty} \mathbb{E}[\hat{p}_{M,n}(\mathcal{P}') | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] &= p_n(\mathcal{P}') + \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})} \\ &\quad \times (p_0 - p_n(\mathcal{P})). \end{aligned}$$

Regarding the conditional variance,

$$\begin{aligned} & \mathbb{V}[\hat{p}_{M,n}(\mathcal{P}') | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] \\ &= \mathbb{V}\left[\frac{1}{M} \sum_{l=1}^M \mathbf{1}_{\mathcal{P}' \in T_n(\Theta_l)} | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n\right] \\ &= \frac{1}{M} \mathbb{V}[\mathbf{1}_{\mathcal{P}' \in T(\Theta_1, \mathcal{D}_n)} | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] \\ &\quad + \left(1 - \frac{1}{M}\right) \text{Cov}(\mathbf{1}_{\mathcal{P}' \in T_n(\Theta_1)}, \mathbf{1}_{\mathcal{P}' \in T_n(\Theta_2)} | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\leq \frac{1}{M} + C_M \end{aligned}$$

where

$$\begin{aligned} C_M &= \text{Cov}(\mathbf{1}_{\mathcal{P}' \in T_n(\Theta_1)}, \mathbf{1}_{\mathcal{P}' \in T_n(\Theta_2)} | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P}' \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad - \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad \times \mathbb{P}(\mathcal{P}' \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \end{aligned}$$

Then, we follow the same reasoning that leads to equation (2.13). We can fully expand  $C_M$  using Bayes formula, depending whether  $\mathcal{P} \in T_n(\Theta_1)$  or  $\mathcal{P} \in T_n(\Theta_2)$ . Note that, since all the trees are independent conditional on  $\mathcal{D}_n$ , we can reduce the conditioning event  $\{\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2), \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n\}$  to  $\{\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2), \mathcal{D}_n\}$ , then

$$\begin{aligned}
C_M &= \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P}' \in T_n(\Theta_2) | \mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2), \mathcal{D}_n) \\
&\quad \times \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
&\quad - (\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \\
&\quad \quad \times \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n))^2 \\
&+ 2[\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P}' \in T_n(\Theta_2) | \mathcal{P} \in T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2), \mathcal{D}_n) \\
&\quad \times \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
&\quad - \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \\
&\quad \quad \times \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
&\quad \quad \times \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{D}_n) \\
&\quad \quad \times \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)] \\
&+ \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P}' \in T_n(\Theta_2) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2), \mathcal{D}_n) \\
&\quad \times \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\
&\quad - (\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{D}_n) \\
&\quad \quad \times \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n))^2
\end{aligned}$$

Conditional on  $\mathcal{D}_n$ ,  $T_n(\Theta_1)$  and  $T_n(\Theta_2)$  are independent, then

$$\begin{aligned}
&\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P}' \in T_n(\Theta_2) | \mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2), \mathcal{D}_n) \\
&= \frac{\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P}' \in T_n(\Theta_2), \mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \mathcal{D}_n)}{\mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \mathcal{D}_n)} \\
&= \frac{\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_1) | \mathcal{D}_n) \mathbb{P}(\mathcal{P}' \in T_n(\Theta_2), \mathcal{P} \in T_n(\Theta_2) | \mathcal{D}_n)}{\mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \mathcal{D}_n) \mathbb{P}(\mathcal{P} \in T_n(\Theta_2) | \mathcal{D}_n)} \\
&= \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \mathbb{P}(\mathcal{P}' \in T_n(\Theta_2) | \mathcal{P} \in T_n(\Theta_2), \mathcal{D}_n) \\
&= \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n)^2
\end{aligned}$$



we can rewrite  $C_M$

$$\begin{aligned} C_M &= \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n)^2 \times \Delta_{M,1} \\ &\quad + 2\mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \in T_n(\Theta_1), \mathcal{D}_n) \\ &\quad \quad \times \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{D}_n) \times \Delta_{M,2} \\ &\quad + \mathbb{P}(\mathcal{P}' \in T_n(\Theta_1) | \mathcal{P} \notin T_n(\Theta_1), \mathcal{D}_n)^2 \times \Delta_{M,3}, \end{aligned}$$

where

$$\begin{aligned} \Delta_{M,1} &= \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)^2, \\ \Delta_{M,2} &= \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad \quad (1 - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)), \\ \Delta_{M,3} &= \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad - \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)^2. \end{aligned}$$

We first consider the term

$$\begin{aligned} \Delta_{M,1} &= \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &\quad - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)^2 \end{aligned}$$

Equation (2.13) directly gives,

$$(2.15) \quad \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)^2 = p_0^2.$$

On the other hand

$$\begin{aligned} &\mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= p_n(\mathcal{P})^2 \frac{\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2), \mathcal{D}_n)}{\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n)} \\ &= p_n(\mathcal{P})^2 \frac{\Phi(Mp_0 - 2, M - 2, p_n(\mathcal{P}))}{\Phi(Mp_0, M, p_n(\mathcal{P}))}. \end{aligned}$$

Again, as for equation (2.11), we can express the cdf of the binomial distribution using the hypergeometric function  $F$

$$(2.16) \quad \begin{aligned} &\mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= p_0^2 \left( 1 + \frac{p_0 - 1}{p_0(M - 1)} \right) \frac{F(M - 1, 1, M(1 - p_0) + 1, 1 - p_n(\mathcal{P}))}{F(M + 1, 1, M(1 - p_0) + 1, 1 - p_n(\mathcal{P}))}, \end{aligned}$$

and from Lemma 10,

$$\lim_{M \rightarrow \infty} \frac{F(M-1, 1, M(1-p_0)+1, 1-p_n(\mathcal{P}))}{F(M+1, 1, M(1-p_0)+1, 1-p_n(\mathcal{P}))} = 1,$$

that is

$$(2.17) \quad \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) = p_0^2.$$

Using equations (2.15) and (2.17), we conclude

$$\lim_{M \rightarrow \infty} \Delta_{M,1} = 0.$$

We follow the same reasoning for  $\Delta_{M,3}$ , equation (2.13) gives

$$(2.18) \quad \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)^2 = (1-p_0)^2.$$

On the other hand,

$$\begin{aligned} & \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= (1-p_0)^2 \left(1 - \frac{p_0}{M-1}\right) \frac{F(M-1, 1, M(1-p_0)-11, 1-p_n(\mathcal{P}))}{F(M+1, 1, M(1-p_0)+1, 1-p_n(\mathcal{P}))} \end{aligned}$$

From Lemma 10,

$$(2.19) \quad \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) = (1-p_0)^2$$

And finally  $\lim_{M \rightarrow \infty} \Delta_{M,3} = 0$ . The term  $\Delta_{M,2}$  can be treated in a similar way, since equation (2.13) gives

$$\begin{aligned} & \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \in T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= p_0(1-p_0). \end{aligned}$$

Simple identity shows

$$\begin{aligned} & \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= \frac{1}{2} \left(1 - \mathbb{P}(\mathcal{P} \notin T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \right. \\ & \quad \left. - \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \in T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \right). \end{aligned}$$

Taking the limit of the previous equation and using equations (2.17) and (2.19), we get

$$(2.20) \quad \lim_{M \rightarrow \infty} \mathbb{P}(\mathcal{P} \in T_n(\Theta_1), \mathcal{P} \notin T_n(\Theta_2) | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) = p_0(1 - p_0).$$

Using (2.13) and (2.20),  $\lim_{M \rightarrow \infty} \Delta_{M,2} = 0$ . Since  $\Delta_{M,1}, \Delta_{M,2}, \Delta_{M,3} \rightarrow 0$ , we obtain  $\lim_{M \rightarrow \infty} C_M = 0$ , that is,

$$(2.21) \quad \lim_{M \rightarrow \infty} \mathbb{V}[\hat{p}_{M,n}(\mathcal{P}') | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n] = 0.$$

Finally combining equations (2.14) and (2.21),

$$\begin{aligned} & \lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= \mathbb{1}_{p_n(\mathcal{P}') + \rho_n(\mathcal{P}, \mathcal{P}') \frac{\sigma_n(\mathcal{P}')}{\sigma_n(\mathcal{P})} (p_0 - p_n(\mathcal{P})) > p_0} \end{aligned}$$

*Case 2:*  $p_n(\mathcal{P}) \leq p_0$ . By the law of large numbers,  $\lim_{M \rightarrow \infty} \hat{p}_{M,n}(\mathcal{P}) = p_n(\mathcal{P})$  in probability, and consequently  $\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) \leq p_0) = 1$ . Additionally, we can simply write

$$\begin{aligned} & \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) \\ &= \frac{\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0, \hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n)}{\mathbb{P}(\hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n)} \end{aligned}$$

Again, by the law of large numbers,  $\lim_{M \rightarrow \infty} \hat{p}_{M,n}(\mathcal{P}') = p_n(\mathcal{P}')$  in probability. Then, if  $p_n(\mathcal{P}') > p_0$ ,  $\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0) = 1$ , and it follows that  $\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0, \hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n) = 1$ . If  $p_n(\mathcal{P}') \leq p_0$ ,  $\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0) = 0$ , and consequently  $\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0, \hat{p}_{M,n}(\mathcal{P}) \leq p_0 | \mathcal{D}_n) = 0$ . This can be compacted under the form

$$\lim_{M \rightarrow \infty} \mathbb{P}(\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) \leq p_0, \mathcal{D}_n) = \mathbb{1}_{p_n(\mathcal{P}') > p_0}.$$

The proof for the case  $\mathbb{P}[\hat{p}_{M,n}(\mathcal{P}') > p_0 | \hat{p}_{M,n}(\mathcal{P}) > p_0, \mathcal{D}_n]$  is similar.  $\square$

## REFERENCES

- CVITKOVIĆ, M., SMITH, A. S. and PANDE, J. (2017). Asymptotic expansions of the hypergeometric function with two large parameters application to the partition function of a lattice gas in a field of traps. *Journal of Physics A: Mathematical and Theoretical* **50** 265206.
- HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics* **19** 293–325.
- MENTCH, L. and HOOKER, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *Journal of Machine Learning Research* **17** 841–881.
- OLVER, F. W. J., LOZIER, D. W., BOISVERT, R. F. and CLARK, C. W. (2010). *NIST Handbook of Mathematical Functions Hardback and CD-ROM*. Cambridge University Press.
- SERFLING, R. J. (2009). *Approximation Theorems of Mathematical Statistics* **162**. John Wiley & Sons.

C. BÉNARD  
 S. DA VEIGA  
 SAFRAN TECH, MODELLING & SIMULATION  
 RUE DES JEUNES BOIS  
 CHTEAUFORT, 78114 MAGNY-LES-HAMEAUX, FRANCE  
 E-MAIL: [clement.benard@safrangroup.com](mailto:clement.benard@safrangroup.com)  
[sebastien.da-veiga@safrangroup.com](mailto:sebastien.da-veiga@safrangroup.com)

G. BIAU  
 LPSM, SORBONNE UNIVERSITÉ  
 4 PLACE JUSSIEU  
 75005 PARIS, FRANCE  
 E-MAIL: [gerard.biau@upmc.fr](mailto:gerard.biau@upmc.fr)

E. SCORNET  
 CMAP, ÉCOLE POLYTECHNIQUE  
 ROUTE DE SACLAY  
 91128 PALAISEAU CEDEX, FRANCE  
 E-MAIL: [erwan.scornet@polytechnique.edu](mailto:erwan.scornet@polytechnique.edu)