Supplementary Material for "Sharp Analysis of a Simple Model for Random Forests"

Jason M. Klusowski

In this supplement, we give proofs of Theorem 1, Theorem 4, and Theorem 5 and an auxiliary lemma used in the proof of Theorem 2.

1 PROOFS

1.1 Proof of Theorem 1

Proof. We first decompose the approximation error as follows:

$$\begin{aligned}
&\mathbb{E}[(\overline{Y}(\mathbf{X}) - f(\mathbf{X}))^2] && \text{(S.1)}\\
&= \mathbb{E}\Big[\Big(\sum_{i=1}^{n} \mathbb{E}_{\Theta}[W_i(f(\mathbf{X}_i) - f(\mathbf{X}))] - \mathbf{1}(\mathcal{E}^c)f(\mathbf{X})\Big)^2\Big] \\
&= \mathbb{E}\Big[\Big(\sum_{i=1}^{n} \mathbb{E}_{\Theta}[W_i(f(\mathbf{X}_i) - f(\mathbf{X}))]\Big)^2\Big] + \mathbb{E}[\mathbf{1}(\mathcal{E}^c)|f(\mathbf{X})|^2] && \text{(S.2)}\\
&\leq \mathbb{E}\Big[\Big(\sum_{i=1}^{n} \mathbb{E}_{\Theta}[W_i(f(\mathbf{X}_i) - f(\mathbf{X}))]\Big)^2\Big] + B^2\mathbb{P}(\mathcal{E}^c). && \text{(S.3)}
\end{aligned}$$

Next, by Assumption 2 in the main text, we have that $|f(\mathbf{X}_i) - f(\mathbf{X})| \leq \sum_{j=1}^{d} \|\partial_j f\|_\infty |\mathbf{X}_i^{(j)} - \mathbf{X}^{(j)}|$, and thus, $W_i|f(\mathbf{X}_i) - f(\mathbf{X})| \leq W_i \sum_{j=1}^{d} \|\partial_j f\|_\infty (b_j(\mathbf{X}) - a_j(\mathbf{X}))$. This shows that

$$\begin{aligned}
\sum_{i=1}^{n} W_i|f(\mathbf{X}_i) - f(\mathbf{X})| &\leq \sum_{i=1}^{n} W_i \sum_{j=1}^{d} \|\partial_j f\|_\infty (b_j(\mathbf{X}) - a_j(\mathbf{X})) \\
&\leq \sum_{j=1}^{d} \|\partial_j f\|_\infty (b_j(\mathbf{X}) - a_j(\mathbf{X})).
\end{aligned}$$

Taking expectations with respect to Θ of both sides of this inequality, we may bound the first term in (S.3) by

$$\mathbb{E}\Big[\Big(\sum_{j=1}^{d} \|\partial_j f\|_\infty \mathbb{E}_{\Theta}[b_j(\mathbf{X}) - a_j(\mathbf{X})]\Big)^2\Big].$$

Jensen's inequality for the square function then yields a further upper bound of $d\sum_{j=1}^{d} \|\partial_j f\|_\infty^2 \mathbb{E}[(\mathbb{E}_{\Theta}[b_j(\mathbf{X}) - a_j(\mathbf{X})])^2]$. ∎

1.2 Proof of Theorem 4

Proof. Using (S.2) from the proof of Theorem 1, Jensen's inequality for the square function, and exchangeability of the data, we obtain the following lower bound on the approximation error:

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{i=1}^{n}\mathbb{E}_{\Theta}[W_i(f(\mathbf{X}_i)-f(\mathbf{X}))]\Big)^2\Big] \\
&\geq \mathbb{E}_{\mathbf{X}}\Big[\Big(\sum_{i=1}^{n}\mathbb{E}_{\mathbf{X}_1,\dots,\mathbf{X}_n,\Theta}[W_i(f(\mathbf{X}_i)-f(\mathbf{X}))]\Big)^2\Big] \\
&= n^2\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\dots,\mathbf{X}_n,\Theta}[W_1(f(\mathbf{X}_1)-f(\mathbf{X}))])^2].
\end{aligned}$$

Recall the form of the weights

$$W_1 = \frac{\mathbf{1}(\mathbf{X}_1\in\mathbf{t})}{\sum_{i=1}^{n}\mathbf{1}(\mathbf{X}_i\in\mathbf{t})}\mathbf{1}(\mathcal{E}) = \frac{\mathbf{1}(\mathbf{X}_1\in\mathbf{t})}{1+\sum_{i\geq 2}\mathbf{1}(\mathbf{X}_i\in\mathbf{t})}.$$

Define $T=\sum_{i\geq 2}\mathbf{1}(\mathbf{X}_i\in\mathbf{t})$ and $\Delta_1 = f(\mathbf{X}_1)-f(\mathbf{X})$. By a conditioning argument, we write

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\dots,\mathbf{X}_n,\Theta}[W_1(f(\mathbf{X}_1)-f(\mathbf{X}))])^2] \\
&= \mathbb{E}_{\mathbf{X}}\Big[\Big(\mathbb{E}_{\mathbf{X}_1,\dots,\mathbf{X}_n,\Theta}\Big[\frac{\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\Delta_1}{1+T}\Big]\Big)^2\Big] \\
&= \mathbb{E}_{\mathbf{X}}\Big[\Big(\mathbb{E}_{\Theta}\Big[\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}\Big[\frac{1}{1+T}\Big]\mathbb{E}_{\mathbf{X}_1}\Big[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\Delta_1\Big]\Big]\Big)^2\Big] \\
&= \mathbb{E}_{\mathbf{X}}\Big[\Big(\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}\Big[\frac{1}{1+T}\Big]\Big)^2\Big(\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\Delta_1]\Big)^2\Big],
\end{aligned}$$

where the last line follows from the fact that $\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}[\frac{1}{1+T}]$ is independent of Θ, a consequence of T being conditionally distributed $\mathrm{Bin}(n-1, 2^{-\lceil \log_2 k_n\rceil})$ given $\mathbf{X}$ and Θ. Next, we can use Jensen's inequality on the convex function $x\mapsto 1/(1+x)$ to lower bound

$$\begin{aligned}
\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}\Big[\frac{1}{1+T}\Big] &\geq \frac{1}{1+\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}[T]} \\
&= \frac{1}{1+(n-1)2^{-\lceil \log_2 k_n\rceil}}.
\end{aligned}$$

Hence, we obtain that $n^2\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\dots,\mathbf{X}_n,\Theta}[W_1(f(\mathbf{X}_1)-f(\mathbf{X}))])^2]$ is at least

$$\Big(\frac{n}{1+(n-1)2^{-\lceil \log_2 k_n\rceil}}\Big)^2\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\Delta_1])^2]. \tag{S.4}$$

Next, in giving a lower bound on $\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\Delta_1])^2]$, we will show that

$$\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\langle\boldsymbol{\beta},\mathbf{X}_1-\mathbf{X}\rangle] \tag{S.5}$$

can be written as a weighted sum of d independent Uniform$(0,1)$ variables minus their mean, $1/2$. Consequently, the squared expectation of (S.5) with respect to $\mathbf{X}$ is the sum of the respective variances. Using this, we will show that

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\langle\boldsymbol{\beta},\mathbf{X}_1-\mathbf{X}\rangle])^2] \\
&= \frac{2^{-2\lceil \log_2 k_n\rceil}\sum_{j=1}^{d}|\boldsymbol{\beta}^{(j)}|(1-p_j/2)^{2\lceil \log_2 k_n\rceil}}{12}.
\end{aligned} \tag{S.6}$$

To prove (S.6), observe that

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}_1}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})\langle\boldsymbol{\beta},\mathbf{X}_1-\mathbf{X}\rangle] \\
&= \sum_{j=1}^{d}\mathbb{E}_{\mathbf{X}_1}[\mathbf{1}(\mathbf{X}_1\in\mathbf{t})(\boldsymbol{\beta}^{(j)}(\mathbf{X}_1^{(j)}-\mathbf{X}^{(j)}))] \\
&= \sum_{j=1}^{d}\boldsymbol{\beta}^{(j)}\prod_{j'\neq j}\lambda([a_{j'},b_{j'}])\mathbb{E}_{\mathbf{X}_1^{(j)}}[\mathbf{1}(\mathbf{X}_1^{(j)}\in[a_j(\mathbf{X}),b_j(\mathbf{X})])(\mathbf{X}_1^{(j)}-\mathbf{X}^{(j)})].
\end{aligned} \tag{S.7}$$

Next, note that because $\mathbf{X}^{(j)} \sim \text{Uniform}(0,1)$, we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{X}_1^{(j)}}[\mathbf{1}(\mathbf{X}_1^{(j)} \in [a_j(\mathbf{X}), b_j(\mathbf{X})])(\mathbf{X}_1^{(j)} - \mathbf{X}^{(j)})]\\&= (b_j(\mathbf{X}) - a_j(\mathbf{X}))\Big(\frac{a_j(\mathbf{X}) + b_j(\mathbf{X})}{2} - \mathbf{X}^{(j)}\Big).\end{aligned}$$

Since $b_j(\mathbf{X}) - a_j(\mathbf{X}) = 2^{-K_j}$, we have

$$\begin{aligned}&\mathbb{E}_{\mathbf{X}_1^{(j)}}[\mathbf{1}(\mathbf{X}_1^{(j)} \in [a_j(\mathbf{X}), b_j(\mathbf{X})])(\mathbf{X}_1^{(j)} - \mathbf{X}^{(j)})]\\&= 2^{-K_j}\Big(\frac{a_j(\mathbf{X}) + b_j(\mathbf{X})}{2} - \mathbf{X}^{(j)}\Big).\end{aligned}$$

Combining this with (S.7) and $\prod_{j=1}^d 2^{-K_j} = 2^{-\lceil \log_2 k_n \rceil}$ yields

$$\begin{aligned}&\mathbb{E}_{\mathbf{X}_1}[\mathbf{1}(\mathbf{X}_1 \in \mathbf{t})\langle \boldsymbol{\beta}, \mathbf{X}_1 - \mathbf{X}\rangle]\\&= 2^{-\lceil \log_2 k_n \rceil}\sum_{j=1}^d \boldsymbol{\beta}^{(j)}\Big(\frac{a_j(\mathbf{X}) + b_j(\mathbf{X})}{2} - \mathbf{X}^{(j)}\Big).\end{aligned}$$

Now, by (4) from the main text, which expresses the endpoints of the interval along the j^{th} feature as randomly stopped binary expansions of $\mathbf{X}^{(j)}$, we have

$$\begin{aligned}\frac{a_j(\mathbf{X}) + b_j(\mathbf{X})}{2} - \mathbf{X}^{(j)} &\overset{d}{=} 2^{-K_j-1} - \sum_{k \geq K_j+1} B_k 2^{-k}\\&\overset{d}{=} 2^{-K_j}(1/2 - \sum_{k\geq 1} B_{k+K_j}2^{-k})\\&\overset{d}{=} 2^{-K_j}(1/2 - \widetilde{\mathbf{X}}^{(j)}),\end{aligned}$$

where $\widetilde{\mathbf{X}}$ is uniformly distributed on $[0,1]^d$. Taking expectations with respect to Θ, we have that

$$\begin{aligned}&\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1 \in \mathbf{t})\langle \boldsymbol{\beta}, \mathbf{X}_1 - \mathbf{X}\rangle]\\&\overset{d}{=} 2^{-\lceil \log_2 k_n \rceil}\sum_{j=1}^d \boldsymbol{\beta}^{(j)}(1 - p_j/2)^{\lceil \log_2 k_n \rceil}(1/2 - \widetilde{\mathbf{X}}^{(j)}).\end{aligned} \tag{S.8}$$

Observe that (S.8) is a sum of mean zero independent random variables, and hence, its squared expectation is equal to the sum of the individual variances, viz.,

$$\begin{aligned}&\mathbb{E}_{\mathbf{X}}[(\mathbb{E}_{\mathbf{X}_1,\Theta}[\mathbf{1}(\mathbf{X}_1 \in \mathbf{t})\langle \boldsymbol{\beta}, \mathbf{X}_1 - \mathbf{X}\rangle])^2]\\&= 2^{-2\lceil \log_2 k_n \rceil}\sum_{j=1}^d |\boldsymbol{\beta}^{(j)}|^2(1 - p_j/2)^{2\lceil \log_2 k_n \rceil}\mathrm{Var}(\widetilde{\mathbf{X}}^{(j)})\\&= \frac{2^{-2\lceil \log_2 k_n \rceil}\sum_{j=1}^d |\boldsymbol{\beta}^{(j)}|^2(1 - p_j/2)^{2\lceil \log_2 k_n \rceil}}{12}.\end{aligned} \tag{S.9}$$

Thus, combining (S.4) and (S.9), we have shown that

$$\begin{aligned}&\mathbb{E}[(\overline{Y}(\mathbf{X}) - f(\mathbf{X}))^2]\\&\geq \Big(\frac{n2^{-\lceil \log_2 k_n \rceil}}{1 + (n-1)2^{-\lceil \log_2 k_n \rceil}}\Big)^2 \frac{\sum_{j=1}^d |\boldsymbol{\beta}^{(j)}|^2(1 - p_j/2)^{2\lceil \log_2 k_n \rceil}}{12}\\&\geq \frac{\sum_{j=1}^d |\boldsymbol{\beta}^{(j)}|^2 k_n^{2\log_2(1-p_j/2)}}{96}.\end{aligned}$$

□

1.3 Proof of Theorem 5

Proof. First, note that by (Biau, 2012, Section 5.2, p. 1083-1084),

$$\begin{aligned}
&\mathbb{E}[(\widehat{Y}(\mathbf{X}) - \overline{Y}(\mathbf{X}))^2] \\
&= n\sigma^2 \mathbb{E}[(\mathbb{E}_\Theta[W_1])^2] \\
&= n\sigma^2 \mathbb{E}[\mathbb{E}_\Theta[W_1]\mathbb{E}_{\Theta'}[W_1]] \\
&= \mathbb{E}\Big[\frac{n\sigma^2 \mathbf{1}(\mathbf{X}_1 \in \mathbf{t} \cap \mathbf{t}')}{(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}))(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}'))}\Big] \\
&= \mathbb{E}\Big[\frac{n\sigma^2 \lambda(\mathbf{t} \cap \mathbf{t}')}{(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}))(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}'))}\Big],
\end{aligned}$$

where Θ' is an independent copy of Θ. We first lower bound

$$\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}\Big[\frac{1}{(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}))(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}'))}\Big].$$

via Jensen's inequality, which yields

$$\frac{1}{\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}\big[\big(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t})\big)\big(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}')\big)\big]}.$$

Next, we use linearity of expectation to write

$$\begin{aligned}
&\mathbb{E}_{\mathbf{X}_2,\dots,\mathbf{X}_n}\Big[\Big(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t})\Big)\Big(1+\sum_{i=2}^n \mathbf{1}(\mathbf{X}_i \in \mathbf{t}')\Big)\Big] \\
&= 1 + 2(n-1)2^{-\lceil \log_2 k_n \rceil} + (n-1)(n-2)2^{-2\lceil \log_2 k_n \rceil} + (n-1)\lambda(\mathbf{t} \cap \mathbf{t}') \\
&\le 5n^2/k_n^2,
\end{aligned}$$

where the last inequality follows from $n \ge 2^{\lceil \log_2 k_n \rceil}$ and $\lambda(\mathbf{t} \cap \mathbf{t}') \le 2^{-\lceil \log_2 k_n \rceil}$. Hence, the estimation error $\mathbb{E}[(\widehat{Y}(\mathbf{X}) - \overline{Y}(\mathbf{X}))^2]$ can be lower bounded by

$$\frac{\sigma^2 k_n^2}{5n}\mathbb{E}_{\Theta,\Theta'}[\lambda(\mathbf{t} \cap \mathbf{t}')], \tag{S.10}$$

where Θ' is an independent copy of Θ. Thus by (S.10) and (12) from the main text, we are done if we can show that $\mathbb{E}_{\Theta,\Theta'}[2^{-\frac{1}{2}\sum_{j=1}^d |K_j - K_j'|}]$ has a lower bound similar in form to the upper bound in (13) from the main text. But this follows directly from Lemma S.1, since

$$\begin{aligned}
\mathbb{E}_{\Theta,\Theta'}[2^{-\frac{1}{2}\sum_{j=1}^d |K_j - K_j'|}] &= \mathbb{E}_{\Theta,\Theta'}[2^{-\frac{1}{2}\sum_{j\in\mathcal{P}} |K_j - K_j'|}] \\
&\ge \frac{(47)^{-d_0}}{\prod_{j\in\mathcal{P}} p_j \times (\lceil \log_2 k_n \rceil)^{d_0 - 1}},
\end{aligned}$$

provided $\lceil \log_2 k_n \rceil p_j \ge 1$. □

1.4 Auxiliary lemma

Lemma S.1. *Let $\mathbf{M} = (M_1, \dots, M_k)$ be distributed according to a multinomial distribution with m trials and class probabilities $(p_1, \dots, p_k)$, each of which is nonzero. Let $\mathbf{M}' = (M_1', \dots, M_k')$ be an independent copy. Then,*

$$\mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^k |M_j - M_j'|}] \le \frac{8^k}{\sqrt{m^{k-1}p_1 \cdots p_k}}. \tag{S.11}$$

Furthermore, if $mp_j \ge 1$ for all j, then

$$\mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^k |M_j - M_j'|}] \ge \frac{(47)^{-k}}{m^{k-1}p_1 \cdots p_k}.$$

Proof. The proof requires only elementary facts about the multinomial distribution and Stirling's approximation. First, note that

$$\begin{aligned}
&\mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^{k}|M_j-M_j'|}] \\
&= \sum_{w_1,\dots,w_k} \mathbb{P}\Big(\bigcap_{j=1}^{k}\{|M_j - M_j'| = w_j\}\Big)2^{-\frac{1}{2}\sum_{j=1}^{k} w_j} \\
&\le \sum_{w_1,\dots,w_{k-1}} \sum_{\boldsymbol{\tau}\in\{-1,+1\}^{k-1}} \mathbb{P}\Big(\bigcap_{j=1}^{k-1}\{M_j - M_j' = \tau_j w_j\}\Big)2^{-\frac{1}{2}\sum_{j=1}^{k-1} w_j}. \qquad \text{(S.12)}
\end{aligned}$$

Next, let $p(\mathbf{m}) = \binom{m}{m_1,\dots,m_k}p_1^{m_1}\cdots p_k^{m_k}$ denote the multinomial mass function and let $\mathbf{m}^*$ be ones of its modes. Then, we can bound each probability in (S.12) by

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{j=1}^{k-1}\{M_j - M_j' = \tau_j w_j\}\Big) &= \sum_{\mathbf{m}} p(\mathbf{m})p(\mathbf{m}+\boldsymbol{\tau}\mathbf{w}) \\
&\le p(\mathbf{m}^*).
\end{aligned}$$

Combining these two inequalities, we have

$$\begin{aligned}
&\mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^{k}|M_j-M_j'|}] \\
&\le \sum_{w_1,\dots,w_{k-1}} \sum_{\boldsymbol{\tau}\in\{-1,+1\}^{k-1}} p(\mathbf{m}^*)2^{-\frac{1}{2}\sum_{j=1}^{k-1} w_j} \\
&\le (4+2\sqrt{2})^{k-1}p(\mathbf{m}^*). \qquad \text{(S.13)}
\end{aligned}$$

Next, using a refinement of Stirling's approximation (see for example (Robbins, 1955)), we have $m! \le e\sqrt{2\pi m}(m/e)^m$ and $m_j! \ge \sqrt{2\pi m_j}(m_j/e)^{m_j} \ge e^{-1}\sqrt{2\pi(m_j+1)}((m_j+1)/e)^{m_j}$. Using these inequalities, we upper bound the multinomial coefficient $\binom{m}{m_1,\dots,m_k}$, which in turn yields an upper bound on $p(\mathbf{m}^*)$, namely,

$$p(\mathbf{m}^*) \le \frac{e^{k+1}}{(\sqrt{2\pi})^{k-1}}\sqrt{\frac{m}{(m_1^*+1)\cdots(m_k^*+1)}}(mp_1/(m_1^*+1))^{m_1^*}\cdots(mp_k/(m_k^*+1))^{m_k^*}. \qquad \text{(S.14)}$$

Finally, (Feller, 1968, page 171, Exercise 28, Equation 10.1) states that any mode $\mathbf{m}^*$ of the multinomial distribution satisfies $m_j^* > mp_j - 1$ and hence from (S.14),

$$p(\mathbf{m}^*) \le \frac{e^{k+1}}{(\sqrt{2\pi})^{k-1}}\frac{1}{\sqrt{m^{k-1}p_1\cdots p_k}}. \qquad \text{(S.15)}$$

Putting everything together from (S.13) and (S.15), we have

$$\begin{aligned}
\mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^{k}|M_j-M_j'|}] &\le (4+2\sqrt{2})^{k-1}\frac{e^{k+1}}{(\sqrt{2\pi})^{k-1}}\frac{1}{\sqrt{m^{k-1}p_1\cdots p_k}} \\
&< \frac{8^k}{\sqrt{m^{k-1}p_1\cdots p_k}}.
\end{aligned}$$

For the other direction, we first remark that

$$\begin{aligned}
\mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^{k}|M_j-M_j'|}] &\ge \mathbb{P}(\mathbf{M}=\mathbf{M}') \\
&= \sum_{\mathbf{m}}(p(\mathbf{m}))^2 \\
&\ge (p(\mathbf{m}'))^2, \qquad \text{(S.16)}
\end{aligned}$$

where $m'_j = \lfloor mp_j \rfloor$. Following the same strategy as before for the binomial coefficient $\binom{m}{m_1,\ldots,m_k}$, i.e., $m_j! \leq e\sqrt{2\pi m_j}(m_j/e)^{m_j}$ and $m! \geq \sqrt{2\pi m}(m/e)^m$, we have

$$\begin{aligned} p(\mathbf{m}') &\geq \frac{e^{-k}}{(\sqrt{2\pi})^{k-1}}\sqrt{\frac{m}{m'_1\cdots m'_k}}(mp_1/m'_1)^{m'_1}\cdots(mp_k/m'_k)^{m'_k} \\ &\geq \frac{e^{-k}}{(\sqrt{2\pi})^{k-1}}\frac{1}{\sqrt{m^{k-1}p_1\cdots p_k}}, \end{aligned}$$

provided $mp_j \geq 1$. Applying this inequality to (S.16) yields

$$\begin{aligned} \mathbb{E}[2^{-\frac{1}{2}\sum_{j=1}^k |M_j - M'_j|}] &\geq \Big(\frac{(e\sqrt{2\pi})^{-k}}{\sqrt{m^{k-1}p_1\cdots p_k}}\Big)^2 \\ &\geq \frac{(47)^{-k}}{m^{k-1}p_1\cdots p_k}. \end{aligned}$$

□

References

Biau, G. (2012). Analysis of a random forests model. *Journal of Machine Learning Research*, 13(Apr):1063–1095.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications: Volume I*, volume 1. John Wiley & Sons.

Robbins, H. (1955). A remark on Stirling's formula. *The American Mathematical Monthly*, 62(1):26–29.