# Proof $\Lambda > 0$

Christian Hilscher

July 5, 2021

A single decision tree is consistent if $MDI(X_j, \mathbf{t}) \to \infty \; \forall j \in S$. We can lower bound this by $MDI(X_j; \mathbf{t}) \geq \Lambda_j K_j(\mathbf{t})$.

Now we prove that $\Lambda > 0$ since it can be argued that $K_j(\mathbf{t}) \to \infty$ for deeply grown trees.

## Notation

One can write

$$\Delta(s; \mathbf{t}) = P(\mathbf{t}_L)P(\mathbf{t}_R)\left[\mathbb{E}[Y|\mathbf{X} \in \mathbf{t}, X \leq s] - \mathbb{E}[Y|\mathbf{X} \in \mathbf{t}, X > s]\right]^2$$

Define

$$\Xi(s; \mathbf{t}) = P(\mathbf{t}_L)P(\mathbf{t}_R)\left[\mathbb{E}[Y|\mathbf{X} \in \mathbf{t}, X \leq s] - \mathbb{E}[Y|\mathbf{X} \in \mathbf{t}, X > s]\right] \tag{1}$$

such that

$$\Delta(s; \mathbf{t}) = \frac{[\Xi(s; \mathbf{t})]^2}{P(\mathbf{t}_L)P(\mathbf{t}_R)}$$

## Weighted approach

Here we deviate from Kuslowski (2020) and instead of analyzing $\Delta(s; \mathbf{t})$ we consider a weighted variant:

$$\begin{aligned} \Delta_\alpha(s; \mathbf{t}) &= [4P(\mathbf{t}_L)P(\mathbf{t}_R)]^\alpha \Delta(s; \mathbf{t}) \qquad (2) \\ &= \omega \; \Delta(s; \mathbf{t}) \end{aligned}$$

From now on we consider an interval $[a, b]$ where $0 \leq a < b \leq 1$.
Instead of writing $P(\mathbf{t}_L)$ one can also write $P(s|\mathbf{t})$ to emphasize the dependence of the

split-point.

Taking (2) adapted with the weights then yields

$$\begin{aligned}\Delta_\alpha(s;\mathbf{t}) &= [4P(s|\mathbf{t})(1-P(s|\mathbf{t}))]^\alpha \frac{[\Xi(s;\mathbf{t})]^2}{P(s|\mathbf{t})(1-P(s|\mathbf{t}))} \\ &= \omega\ \frac{(\int_a^s p(s|\mathbf{t})\bar{G}(\tilde{s},\mathbf{t})d\tilde{s})^2}{P(s|\mathbf{t})(1-P(s|\mathbf{t}))} \qquad (3)\end{aligned}$$

where $p(s|\mathbf{t})$ is the density function of $X|\mathbf{X}\in\mathbf{t}$. Since $\Delta_\alpha(s^*,\mathbf{t})$ is the maximum of (3), any average over possible split points is smaller. Taking any prior $\Pi$ on $[0,1]$ with the corresponding density $\pi$, one can thus write

$$\begin{aligned}\Delta_\alpha(s^*;\mathbf{t}) &\geq \int_a^b \omega\ \frac{(\int_a^s p(s|\mathbf{t})\bar{G}(\tilde{s};\mathbf{t})d\tilde{s})^2}{P(s|\mathbf{t})(1-P(s|\mathbf{t}))}\ \pi\frac{\left(\frac{s-a}{b-a}\right)}{b-a}ds \\ &= \int_0^1 \omega\ \frac{(\int_0^s (b-a)p(a+\tilde{s}(b-a)|\mathbf{t})\bar{G}(a+\tilde{s}(b-a);\mathbf{t})d\tilde{s})^2}{P(a+s(b-a)|\mathbf{t})(1-P(a+s(b-a)|\mathbf{t}))}\Pi ds \qquad (4)\end{aligned}$$

With a uniform prior, $\pi(s)=\mathbb{1}_{s\in[0,1]}$, and by assumption positive and continuous $p_X(\cdot)$ we have that

$$\begin{aligned}\lim_{(a,b)\to(c,c)}(b-a)p(a+s(b-a)|\mathbf{t}) &= \frac{p_X(c)}{p_X(c)} = 1 \\ \lim_{(a,b)\to(c,c)} P(a+s(b-a)|\mathbf{t}) &= s\frac{p_X(c)}{p_X(c)} = s\end{aligned}$$

## Digression: Auxiliary re-write

For simplifying things later on, let

$$D(s) = \frac{\bar{F}(a+s(b-a);\mathbf{t})-\bar{F}(c;\mathbf{t})}{(a+s(b-a)-c)^R} \qquad \text{and} \qquad \delta = \frac{c-a}{b-a}$$

and note that this way

$$\frac{\bar{G}(a+\tilde{s}(b-a);\mathbf{t})}{(b-a)^R} = D(\tilde{s})(\tilde{s}-\delta)^R - \int_0^1 D(\tilde{\tilde{s}})(\tilde{\tilde{s}}-\delta)^R d\tilde{\tilde{s}} \qquad (5)$$

Approximating $\bar{F}(a+s(b-a);\mathbf{t})$ the point $s=c$ by a Taylor expansion and yields

$$
\begin{aligned}
\bar{F}(a+s(b-a);\mathbf{t})-\bar{F}(c;\mathbf{t}) &= \bar{F}^{(1)}(c;\mathbf{t})(a+s(b-a)-c)+...+\frac{\bar{F}^{(R)}(c;\mathbf{t})}{R!}(a+s(b-a)-c)^R \\
\frac{\bar{F}(a+s(b-a);\mathbf{t})-\bar{F}(c;\mathbf{t})}{(a+s(b-a)-c)^R} &= \bar{F}^{(1)}(c;\mathbf{t})\frac{(a+s(b-a)-c)}{(a+s(b-a)-c)^R}+...+\frac{\bar{F}^{(R)}(c;\mathbf{t})}{R!} \qquad (6)
\end{aligned}
$$

and (6) is equal to $D(s)$. Taking the limit then gives

$$
\lim_{(a,b)\to(c,c)} D(s) = \frac{\bar{F}^{(R)}(c;\mathbf{t})}{R!} \qquad (7)
$$

We can assume without loss of generality that $\bar{F}^{(R)}(c;\mathbf{t}) > 0$. If then there exists $\varepsilon > 0$ such that $\sqrt{(b-c)^2+(a-c)^2} < \varepsilon$, it holds because of uniform continuity that

$$
\left| D(s) - \frac{\bar{F}^{(R)}(c;\mathbf{t})}{R!} \right| < min\left\{ \frac{\bar{F}^{(R)}(c;\mathbf{t})}{2R!}, \frac{1}{\delta^2} \right\} \qquad (8)
$$

## Continuation main part

We are now in the position to reformulate

$$
\begin{aligned}
\Delta_\alpha(s^*;\mathbf{t}) &\geq \int_0^1 \omega\ \frac{(\int_0^s (b-a)p(a+\tilde{s}(b-a)|\mathbf{t})\bar{G}(a+\tilde{s}(b-a);\mathbf{t})d\tilde{s})^2}{P(a+s(b-a)|\mathbf{t})(1-P(a+s(b-a)|\mathbf{t}))}\Pi ds \\
&= \int_0^1 \omega \frac{(\int_0^s D(\tilde{s})(\tilde{s}-\delta)^R d\tilde{s} - s\int_0^1 D(\tilde{\tilde{s}})(\tilde{\tilde{s}}-\delta)^R)d\tilde{\tilde{s}})^2}{s(1-s)} \qquad (9) \\
& \qquad (10)
\end{aligned}
$$

Since $s(1-s) \leq \frac{1}{4}$ and using Jensen's inequality it holds that (9) is at least

$$
\begin{aligned}
& 4\left(\int_0^1 \omega(\int_0^s D(\tilde{s})(\tilde{s}-\delta)^R d\tilde{s} - s\int_0^1 D(\tilde{\tilde{s}})(\tilde{\tilde{s}}-\delta)^R)d\tilde{\tilde{s}})^2 ds\right) \\
&= 4\left(\int_0^1 [s(1-s)]^\alpha(\int_0^s D(\tilde{s})(\tilde{s}-\delta)^R d\tilde{s} - s\int_0^1 D(\tilde{\tilde{s}})(\tilde{\tilde{s}}-\delta)^R)d\tilde{\tilde{s}})^2 ds\right) \\
&\leq 4\int_0^1 [s(1-s)]^\alpha ds \ \left(\int_0^1 (\int_0^s D(\tilde{s})(\tilde{s}-\delta)^R d\tilde{s} - s\int_0^1 D(\tilde{\tilde{s}})(\tilde{\tilde{s}}-\delta)^R)d\tilde{\tilde{s}})^2 ds\right)
\end{aligned}
$$

which, after applying Fubini's Theorem and using $\tilde{s}, \tilde{\tilde{s}} = x$, can be written as

$$
\begin{aligned}
&4\int_0^1 [s(1-s)]^\alpha ds \ \left(\int_0^1 (\int_0^s D(x)(x-\delta)^R ds - s\int_0^1 D(x)(x-\delta)^R)ds)^2 dx\right) \\
&= 4\int_0^1 [s(1-s)]^\alpha ds \ \left(\int_0^1 D(x)(x-\delta)^R [\int_0^s ds - s\int_0^1 ds]^2 dx\right) \\
&= 4\ \frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)} \int_0^1 D(x)(x-\delta)^R (x-\frac{1}{2})dx
\end{aligned}
$$

where $\Gamma(\cdot)$ is the Gamma function.
From here on proof of Kuslowski applies just with the pre-multiplication of $\frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)}$. Following his steps, it holds that

$$
\liminf_{(a(\mathbf{t}),b(\mathbf{t}))\to(c,c)} \left\{ \frac{\Delta_\alpha(s^*;\mathbf{t})}{\left(\frac{(b(\mathbf{t})-a(\mathbf{t}))^R|\bar{F}^{(R)}(c;\mathbf{t})|}{R!}\right)^2} \right\} \geq \Delta_R \tag{11}
$$

with

$$
\Delta_R = \frac{\Gamma(\alpha+1)^2}{\Gamma(2\alpha+2)} \ \int_0^1 \Delta(s;[0,1])ds > 0
$$