A new parametric information-gain criterion for tree-based machine learning algorithms

Diogo Costa[1], **Vasco Costa**[2], **and Eugénio M. Rocha**[1,2]

[1]**Center for Research & Development in Mathematics and Applications, University of Aveiro, Aveiro, Portugal**
[2]**Department of Mathematics, University of Aveiro, Aveiro, Portugal**

Corresponding author:
Diogo Costa[1]

Email address: d.costa@ua.pt

PROOFS OF KNOWN ENTROPIES

Considering a set of probabilities $P = (P_1, \ldots, P_n) \in [0,1]^n$ with $\sum_{i=1}^{n} P_i = 1$, where n is the number of unique classes in the set S, the generalized entropy $\hat{E}$ is defined by

$$\hat{E}_{\zeta_1,\zeta_2}^{\alpha,\beta,\gamma}(P) = \left(\hat{\varphi}_{\zeta_1,\zeta_2}^{\alpha,\beta,\gamma}(P) - k_0\right)\left(k_1 - k_0\right)^{-1}, \tag{1}$$

where $(\alpha, \beta, \gamma, \zeta_1, \zeta_2) \in (\mathbb{R}_0^+)^5$ are adequate parameters,

$$\hat{\varphi}_{\zeta_1,\zeta_2}^{\alpha,\beta,\gamma}(P) = \zeta_2\text{-ln}\left(\sum_{i=1}^{n} P_i^{\alpha}\left[-\zeta_1\text{-ln}(P_i^{\beta})\right]^{\gamma}\right), \tag{2}$$

and the q-logarithm, for any $q \in \mathbb{R}_0^+ \backslash \{1\}$ and $x > 0$, is given by

$$q\text{-ln}(x) = \frac{x^{1-q} - 1}{1-q}. \tag{3}$$

For $q = 1$, the q-ln(x) coincides with the natural logarithm ln(x), by computing the limit as $q \to +1$, using the L'Hôpital's rule, and the fact that $dx^{1-q}/dq = -\ln(x)x^{1-q}$. Constants k_0 and k_1 represent the minimum and maximum theoretical values of $\hat{\varphi}$, respectively. Accordingly, $k_0 = min(\hat{\varphi})$, obtained whenever $\exists_j : P_j = 1$. Alternatively, $k_1 = max(\hat{\varphi})$, occurring whenever each event is equally probable, i.e., $P_i = \frac{1}{n}$.

Taking into account the parametrization of (1), to ensure convergence, a computation direction is selected as

$$\hat{\varphi}_{\bar{\zeta}_1,\bar{\zeta}_2}^{\bar{\alpha},\bar{\beta},\bar{\gamma}}(P) \doteq \lim_{\alpha\to\bar{\alpha}}\left(\lim_{\beta\to\bar{\beta}}\left(\lim_{\gamma\to\bar{\gamma}}\left(\lim_{\zeta_1\to\bar{\zeta}_1}\left(\lim_{\zeta_2\to\bar{\zeta}_2}\varphi_{\zeta_1,\zeta_2}^{\alpha,\beta,\gamma}(P)\right)\right)\right)\right). \tag{4}$$

In what follows and for simplicity of proofs, we will omit the corresponding limits since they can be inferred from the context.

Proof of Gini Index Retrieval

Considering that the transformation of the Gini index (Jost, 2006; Kotsiantis, 2013) given as

$$Gini(S) = 1 - \sum_{i=1}^{n} P_i^2, \tag{5}$$

is obtained by using $\alpha = 0$, $\beta = 2$, $\gamma = 1$, $\zeta_1 = 0$, and $\zeta_2 = 0$ in (1), we obtain

$$\zeta_1\text{-}\ln(P_i^\beta) = \frac{P_i^{\beta(1-\zeta_1)} - 1}{1-\zeta_1} = \frac{P_i^{\beta(1-0)} - 1}{1-0} = P_i^\beta - 1 = P_i^2 - 1, \tag{6}$$

and,

$$\begin{aligned}
\hat{\varphi}_{0,0}^{0,2,1}(P) \equiv \zeta_2\text{-}\ln\left(\sum_{i=1}^{n} P_i^\alpha \Big[-\zeta_1\text{-}\ln(P_i^\beta)\Big]^\gamma\right) &= \frac{\left(\sum_{i=1}^{n} P_i^\alpha \Big[-\zeta_1\text{-}\ln(P_i^\beta)\Big]^\gamma\right)^{(1-\zeta_2)} - 1}{1-\zeta_2} \\
&= \frac{\left(\sum_{i=1}^{n} P_i^0 \Big[-(P_i^2-1)\Big]^1\right)^{(1-0)} - 1}{1-0} \\
&= \left(\sum_{i=1}^{n} -(P_i^2-1)\right) - 1 \\
&= -1 - \sum_{i=1}^{n} (P_i^2 - 1) \\
&= n - 1 - \sum_{i=1}^{n} P_i^2.
\end{aligned} \tag{7}$$

Thus, if we now compute k_0 and k_1 as

$$k_0 = min(\hat{\varphi}) = n - 1 - \sum_{i=1}^{1} 1^2 = n - 2, \tag{8}$$

$$k_1 = max(\hat{\varphi}) = n - 1 - \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 = n - 1 - n\frac{1}{n^2} = n - 1 - \frac{1}{n}, \tag{9}$$

we obtain the normalized entropy value,

$$\hat{E}_{0,0}^{0,2,1}(P) = \frac{\left(n - 1 - \sum_{i=1}^{n} P_i^2\right) - (n-2)}{(n - 1 - \frac{1}{n}) - (n-2)} = \frac{n}{n-1}\left(1 - \sum_{i=1}^{n} P_i^2\right). \tag{10}$$

Matching the generic definition of the Gini index after a similar normalization, with

$$k_{0_{Gini}} = min(Gini) = 1 - \sum_{i=1}^{1} (1)^2 = 1 - 1 = 0, \tag{11}$$

$$k_{1_{Gini}} = max(Gini) = 1 - \sum_{i=1}^{n} \left(\frac{1}{n}\right)^2 = 1 - n\left(\frac{1}{n^2}\right) = 1 - \frac{1}{n}. \tag{12}$$

and, thus,

$$Gini_{normed} = (Gini - k_{0_{Gini}})(k_{1_{Gini}} - k_{0_{Gini}})^{-1} = \frac{1 - \sum_{i=1}^{n} P_i^2 - 0}{1 - \frac{1}{n} - 0} = \frac{n}{n-1}\left(1 - \sum_{i=1}^{n} P_i^2\right). \tag{13}$$

Proof of Shannon Entropy Retrieval

Considering that the transformation of the Shannon entropy (Shannon, 1948) given as

$$H(S) = -\sum_{i=1}^{n} P_i \log_2 P_i, \tag{14}$$

is obtained by using $\alpha = 1$, $\beta = 1$, $\gamma = 1$, $\zeta_1 = 1$, and $\zeta_2 = 0$ in (1), we obtain

$$\zeta_1\text{-}\ln(P_i^{\beta}) = \ln(P_i^{\beta}), \tag{15}$$

as for $\zeta_1 = 1$, ζ_1-$\ln(x)$ coincides with the natural logarithm $\ln(x)$, and,

$$\begin{aligned}
\hat{\varphi}_{1,0}^{1,1,1}(P) \equiv \zeta_2\text{-}\ln\left(\sum_{i=1}^{n} P_i^{\alpha}\left[-\zeta_1\text{-}\ln(P_i^{\beta})\right]^{\gamma}\right) &= \frac{\left(\sum_{i=1}^{n} P_i^{\alpha}\left[-\zeta_1\text{-}\ln(P_i^{\beta})\right]^{\gamma}\right)^{(1-\zeta_2)} - 1}{1-\zeta_2} \\
&= \frac{\left(\sum_{i=1}^{n} P_i^{1}\left[-\ln(P_i)\right]^{1}\right)^{(1-0)} - 1}{1-0} \\
&= -1 - \sum_{i=1}^{n} P_i \ln(P_i).
\end{aligned} \tag{16}$$

Thus, if we now compute k_0 and k_1 as

$$k_0 = min(\hat{\varphi}) = -1 - \sum_{i=1}^{1} \ln(1) = -1, \tag{17}$$

$$k_1 = max(\hat{\varphi}) = -1 - \sum_{i=1}^{n} \frac{1}{n} \ln\frac{1}{n} = -1 + \ln(n), \tag{18}$$

we obtain the normalized entropy value,

$$\hat{E}_{1,0}^{1,1,1}(P) = \frac{\left(-1 - \sum_{i=1}^{n} P_i \ln(P_i)\right) - (-1)}{(-1+\ln(n)) - (-1)} = -\frac{\sum P_i \ln(P_i)}{\ln(n)}. \tag{19}$$

Matching the generic definition of Shannon entropy after a similar normalization, with

$$k_{0_H} = min(H) = -\sum_{i=1}^{1} \log 1 = 0, \tag{20}$$

$$k_{1_H} = max(H) = -\sum_{i=1}^{n} \frac{1}{n} \log\frac{1}{n} = \ln(n). \tag{21}$$

and, thus,

$$H_{normed} = (H - k_{0_H})(k_{1_H} - k_{0_H})^{-1} = \frac{-\sum_{i=1}^{n} P_i \log P_i - 0}{\ln(n) - 0} = -\frac{\sum P_i \ln(P_i)}{\ln(n)}. \tag{22}$$

Proof of Tsallis Entropy Retrieval

Considering that the transformation of the Tsallis entropy (Tsallis, 1988) given as

$$S_q(S) = k\frac{1}{q-1}\left(1-\sum_{i=1}^{n} P_i^q\right), \tag{23}$$

is obtained by using $\alpha = w$, $\beta = 0$, $\gamma = 0$, $\zeta_1 = 0$, and $\zeta_2 = 0$ in (1), we obtain

$$\zeta_1\text{-ln}(P_i^\beta) = \frac{P_i^{\beta^{(1-\zeta_1)}}-1}{1-\zeta_1} = \frac{P_i^{\beta^{(1-0)}}-1}{1-0} = P_i^\beta - 1 \tag{24}$$

and,

$$\begin{aligned}\hat{\varphi}_{0,0}^{w,0,0}(P) \equiv \zeta_2\text{-ln}\left(\sum_{i=1}^{n} P_i^\alpha\left[-\zeta_1\text{-ln}(P_i^\beta)\right]^\gamma\right) &= \frac{\left(\sum_{i=1}^{n} P_i^\alpha\left[-P_i^\beta+1\right]^0\right)-1}{1-0} \\ &= -1+\sum_{i=1}^{n} P_i^w\end{aligned} \tag{25}$$

Thus, if we now compute k_0 and k_1 as

$$k_0 = min(\hat{\varphi}) = -1+\sum_{i=1}^{1} 1^w = -1+1 = 0 \tag{26}$$

$$k_1 = max(\hat{\varphi}) = -1+\sum_{i=1}^{n}\left(\frac{1}{n}\right)^w = -1+\frac{n}{n^w} \tag{27}$$

we obtain the normalized entropy value,

$$\hat{E}_{0,0}^{w,0,0}(P) = \frac{-1+\sum_{i=1}^{n} P_i^w - 0}{-1+\frac{n}{n^w}-0} = \frac{-1+\sum_{i=1}^{n} P_i^w}{-1+\frac{n}{n^w}} = \frac{1-\sum_{i=1}^{n} P_i^w}{1-\frac{n}{n^w}} \tag{28}$$

Matching the generic definition of Tsallis entropy (considering that parameter q in Tallis' entropy takes on value w, S_w, with $k = 1$) after a similar normalization, with

$$k_{0_{S_w}} = min(S_w) = \frac{1}{w-1}\left[1-\sum_{i=1}^{1} 1^w\right] = 0, \tag{29}$$

$$k_{1_{S_w}} = max(S_w) = \frac{1}{w-1}\left[1-\sum_{i=1}^{n}\left(\frac{1}{n}\right)^w\right] = \frac{1}{w-1}\left(1-\frac{n}{n^w}\right). \tag{30}$$

and, thus,

$$S_{w_{normed}} = (S_w - k_{0_{S_w}})(k_{1_{S_w}} - k_{0_{S_w}})^{-1} = \frac{\frac{1}{w-1}\left(1-\sum_{i=1}^{n} P_i^w\right)-0}{\frac{1}{w-1}\left(1-\frac{n}{n^w}\right)-0} = \frac{1-\sum_{i=1}^{n} P_i^w}{1-\frac{n}{n^w}}. \tag{31}$$

Proof of Rényi Entropy Retrieval

Considering that the transformation of the Rényi entropy (Rényi, 1961), of order α of a set S, and with $0 < \alpha < \infty$ and $\alpha \neq 1$ as

$$H_\alpha(S) = \frac{1}{1-\alpha} \log\left(\sum_{i=1}^{n} P_i^\alpha\right), \tag{32}$$

is obtained by using $\alpha = w$, $\beta = 0$, $\gamma = 0$, $\zeta_1 = 0$, and $\zeta_2 = 1$ in (1), we obtain

$$\zeta_1\text{-}\ln(P_i^\beta) = \frac{P_i^{\beta(1-\zeta_1)} - 1}{1-\zeta_1} = \frac{P_i^{\beta(1-0)} - 1}{1-0} = P_i^\beta - 1, \tag{33}$$

and,

$$\begin{aligned}\hat{\varphi}_{0,1}^{w,0,0}(P) \equiv \zeta_2\text{-}\ln\left(\sum_{i=1}^{n} P_i^\alpha \Big[-\zeta_1\text{-}\ln(P_i^\beta)\Big]^\gamma\right) &= \ln\Big(\sum_{i=1}^{n} P_i^\alpha [-P_i^\beta + 1]^0\Big) \\ &= \ln\Big(\sum_{i=1}^{n} P_i^w\Big)\end{aligned} \tag{34}$$

as for $\zeta_2 = 1$, ζ_2-$\ln(x)$ coincides with the natural logarithm $\ln(x)$. Thus, if we now compute k_0 and k_1 as

$$k_0 = min(\hat{\varphi}) = \ln\Big(\sum_{i=1}^{1} 1^w\Big) = 0, \tag{35}$$

$$k_1 = max(\hat{\varphi}) = \ln\Big(\sum_{i=1}^{n} \Big(\frac{1}{n}\Big)^w\Big) = \ln(n^{1-w}), \tag{36}$$

we obtain the normalized entropy value,

$$\hat{E}_{0,1}^{w,0,0}(P) = \frac{\ln\Big(\sum_{i=1}^{n} P_i^w\Big) - 0}{\ln(n^{1-w}) - 0} = \frac{\ln\Big(\sum_{i=1}^{n} P_i^w\Big)}{\ln(n^{1-w})}. \tag{37}$$

Matching the generic definition of Rényi entropy (considering that parameter α in Rényi's entropy takes on value w, H_w) after a similar normalization, with

$$k_{0_{H_w}} = min(H_w) = \frac{1}{1-w} \ln\left(\sum_{i=1}^{1} 1^w\right) = 0, \tag{38}$$

$$k_{1_{H_w}} = max(H_w) = \frac{1}{1-w} \ln\left(\sum_{i=1}^{n} \Big(\frac{1}{n}\Big)^w\right) = \frac{1}{1-w} \ln(n^{1-w}), \tag{39}$$

and, thus,

$$H_{w_{normed}} = (H_w - k_{0_{H_w}})(k_{1_{H_w}} - k_{0_{H_w}})^{-1} = \frac{\frac{1}{1-w} \ln\left(\sum_{i=1}^{n} P_i^w\right) - 0}{\frac{1}{1-w} \ln(n^{1-w}) - 0} = \frac{\ln\Big(\sum_{i=1}^{n} P_i^w\Big)}{\ln(n^{1-w})}. \tag{40}$$

Proof of Sharma-Mittal Entropy Retrieval

Considering that the transformation of the Sharma-Mittal entropy (Akturk et al., 2007) is given as

$$S_{SM}(S) = \frac{1}{1-r}\left[\left(\sum_{i=1}^{n} P_i^q\right)^{\frac{1-r}{1-q}} - 1\right], \tag{41}$$

is obtained by using $\alpha = q$, $\beta = 0$, $\gamma = 0$, $\zeta_1 = 0$, and $\zeta_2 = \frac{r-q}{1-q}$ in (1), we obtain

$$\zeta_1\text{-ln}(P_i^\beta) = \frac{P_i^{\beta(1-\zeta_1)} - 1}{1-\zeta_1} = \frac{P_i^{\beta(1-0)} - 1}{1-0} = P_i^\beta - 1, \tag{42}$$

and,

$$\begin{aligned}\hat{\varphi}^{q,0,0}_{0,\frac{r-q}{1-q}}(P) \equiv \zeta_2\text{-ln}\left(\sum_{i=1}^{n} P_i^\alpha\left[-\zeta_1\text{-ln}(P_i^\beta)\right]^\gamma\right) &= \zeta_2\text{-ln}\left(\sum_{i=1}^{n} P_i^\alpha[-P_i^\beta + 1]^0\right)\\ &= \frac{\left(\sum_{i=1}^{n} P_i^q\right)^{1-\frac{r-q}{1-q}} - 1}{1-\frac{r-q}{1-q}}\\ &= \frac{1-q}{1-r}\left[\left(\sum_{i=1}^{n} P_i^q\right)^{\frac{1-r}{1-q}} - 1\right].\end{aligned} \tag{43}$$

Thus, if we now compute k_0 and k_1 as

$$k_0 = min(\hat{\varphi}) = \frac{1-q}{1-r}\left[\left(\sum_{i=1}^{1} 1^q\right)^{\frac{1-r}{1-q}} - 1\right] = 0, \tag{44}$$

$$k_1 = max(\hat{\varphi}) = \frac{1-q}{1-r}\left[\left(\sum_{i=1}^{n}\left(\frac{1}{n}\right)^q\right)^{\frac{1-r}{1-q}} - 1\right] = \frac{1-q}{1-r}\left[\left(\frac{n}{n^q}\right)^{\frac{1-r}{1-q}} - 1\right], \tag{45}$$

we obtain the normalized entropy value,

$$\hat{E}^{q,0,0}_{0,\frac{r-q}{1-q}}(P) = \frac{\frac{1-q}{1-r}\left[\left(\sum_{i=1}^{n} P_i^q\right)^{\frac{1-r}{1-q}} - 1\right] - 0}{\frac{1-q}{1-r}\left[\left(\frac{n}{n^q}\right)^{\frac{1-r}{1-q}} - 1\right] - 0} = \frac{\left(\sum_{i=1}^{n} P_i^q\right)^{\frac{1-r}{1-q}} - 1}{\left(\frac{n}{n^q}\right)^{\frac{1-r}{1-q}} - 1}. \tag{46}$$

Matching the generic definition of Sharma-Mittal entropy after a similar normalization, with

$$k_{0_{S_{SM}}} = min(S_{SM}) = \frac{1}{1-r}\left[\left(\sum_{i=1}^{1} 1^q\right)^{\frac{1-r}{1-q}} - 1\right] = 0, \tag{47}$$

$$k_{1_{S_{SM}}} = max(S_{SM}) = \frac{1}{1-r}\left[\left(\sum_{i=1}^{n}\left(\frac{1}{n}\right)^q\right)^{\frac{1-r}{1-q}} - 1\right] = \frac{1}{1-r}\left[\left(\frac{n}{n^q}\right)^{\frac{1-r}{1-q}} - 1\right], \tag{48}$$

and, thus,

$$S_{SM_{normed}} = (S_{SM} - k_{0_{S_{SM}}})(k_{1_{S_{SM}}} - k_{0_{S_{SM}}})^{-1} = \frac{\frac{1}{1-r}\left[\left(\sum_{i=1}^{n} P_i^q\right)^{\frac{1-r}{1-q}} - 1\right] - 0}{\frac{1}{1-r}\left[\left(\frac{n}{n^q}\right)^{\frac{1-r}{1-q}} - 1\right] - 0} = \frac{\left(\sum_{i=1}^{n} P_i^q\right)^{\frac{1-r}{1-q}} - 1}{\left(\frac{n}{n^q}\right)^{\frac{1-r}{1-q}} - 1}. \quad (49)$$

REFERENCES

Akturk, E., Bagci, G. B., and Sever, R. (2007). Is Sharma-Mittal entropy really a step beyond Tsallis and Renyi entropies?

Jost, L. (2006). Entropy and diversity. *Oikos*, 113(2):363–375.

Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4):261–283.

Rényi, A. (1961). On measures of entropy and information. pages 547–561.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423.

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1):479–487.