AIT Blog

A Theorem in Algorithmic Rate Distortion Theory

Samuel Epstein*

October 25, 2022

Classical Rate Distortion Theory

We provide a well known classical rate-distortion theory result and then prove one Algorithmic Information Theoretic version of the theorem. The source produces a sequence $X_1, X_2, \dots, X_n$, i.i.d. $p(x)$, over the input alphabet $x \in \mathcal{X}$. The encoder is of the form $f_n(X^n) \in \{1, 2, \dots, 2^{nR}\}$ and the decoder produces an estimate $\hat{X}^n \in \hat{\mathcal{X}}^n$. This is a $(2^{nr}, n)$ rate distortion code. A distortion function is $d : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$. The expected distortion is $D = \sum_{x^n} p(x^n)d(x^n, g_n(f_n(x^n)))$. A rate distortion pair (R, D) is said to be achievable if there exists a sequence of $(2^{nR}, n)$ distortion codes (f_n, g_n) with $\lim_{n\to\infty} \mathbf{E}d(X^n, g_n(f_n(X^n))) \le D$. The rate distortion region for a source is the closure of the set of achievable rate distortion pairs (R, D).

Definition 1 (Rate Distortion Function) *The rate distortion function $R(D)$ is the infinimum of rates R such that (R, D) is in the rate distortion region of the source for a given distortion D.*

Definition 2 (Information Rate Distortion Function) *The information rate distortion function $R^{(I)}(D)$ for a source X with distortion function $d(x, \hat{x})$ is*

$$R^{(I)}(D) = \min_{p(\hat{x}|x) : \sum_{(x,\hat{x})} p(x)p(\hat{x}|x)d(x,\hat{x}) \le D} \mathbf{I}(X; \hat{X}).$$

Theorem 1 *The rate distortion function for an i.i.d. source X with distribution $p(x)$ and bounded distortion function $d(x, \hat{x})$ is equal to the associated information function.*

$$R(D) = R^{(I)}(D).$$

Distortion of Individual Codewords

This section contains a theorem reworking Theorem 2 in [VV10]. The difference is that we explicitly use a distortion function that is partial computable, whereas in [VV10] it is generalized into distortion families. Furthermore the $O(\log n)$ error term in [VV10] is transformed into $\mathbf{I}(\cdot; \mathcal{H})$, where $\mathbf{I}(x; y) = \mathbf{K}(x) - \mathbf{K}(x|y)$, and $\mathcal{H}$ is the halting sequence. $\mathbf{K}$ is the prefix-free Kolmogorov complexity.

*JP Theory Group. samepst@jptheorygroup.org

One algorithmic version of rate distortion theory is as follows. Alice wants to communicate a single message $\mathbf{y}$ to Bob, and they both share the same reference universal Turing machine U. Alice sends a program p to Bob, who decompresses it to a codeword $\mathbf{x} = U(p)$ and this message has distortion $d(\mathbf{x}, \mathbf{y})$. A distortion function $d : \{0,1\}^* \times \{0,1\}^* \to \mathbb{R}^{\infty}_{\geq 0}$ is a non-negative upper semi-computable function. Let $\mathbf{x}$ be a message sent to Bob. The following theorem shows that if $\mathbf{y}$ is non-exotic there exists a message $\mathbf{x}'$ such that $\mathbf{K}(\mathbf{x}') <^{\log} \mathbf{I}(\mathbf{x};\mathbf{y})$, with $d(\mathbf{x}', \mathbf{y}) \leq d(\mathbf{x}, \mathbf{y})$.

Theorem 2 *Relativized to upper semi-computable distortion function d and $R \in \mathbb{R}_{>0}$,*

$$\min_{\mathbf{x}:d(\mathbf{x},\mathbf{y})<R} \mathbf{K}(\mathbf{x}) <^{\log} \min_{\mathbf{x}:d(\mathbf{x},\mathbf{y})<R} \mathbf{I}(\mathbf{x};\mathbf{y}) + \mathbf{I}(\mathbf{y};\mathcal{H}).$$

Proof. We assume the universal Turing machine U is left-total. For more details on left-total machines, the reader is referred to my October 11th blog post or [Eps22a]. We recall that relativization to elementary objects means that the universal Turing machine has access to their encodings on auxilliary tapes and the complexity terms implicity have the encoded objects in the conditional terms. Let $D_\infty = \{\mathbf{x} : d(\mathbf{x}, \mathbf{y}) < R\}$ be the finite or infinite set of codewords that have distortion measure less than R with $\mathbf{y}$. The set D_∞ is enumerable given y, and for total string $b \in \{0,1\}^*$, let D_b be the finite subset of D_∞ enumerated in $\mathbf{bb}(b)$ steps. We recall the following busy beaver function on total b is

$$\mathbf{bb}(b) = \max\{\|x\| : U(p) = x, p \lhd b \text{ or } p \sqsupseteq b\}.$$

Let $i = 1 + \lceil -\log \mathbf{m}(D_\infty)\rceil$ and b be the shortest total string where $i \geq -\log \mathbf{m}(D_b)$. Arguments similar to those used in my October 11th blog post show $\mathbf{K}(b|\mathbf{y}, \|b\|) <^{+} \mathbf{K}(i)$. Theorem 1 of my October 9th blog post, relativized to b results in $\mathbf{x}' \in \{0,1\}^*$, with

$$\mathbf{K}(\mathbf{x}'|b) <^{\log} i + \mathbf{Ks}(D_b|b).$$

Thes stochasticity function is

$$\mathbf{Ks}(a|b) = \min\{\mathbf{K}(P|b) + 3\log \mathbf{d}(a|P, b) : \; P \text{ is an elementary probability measure}\}.$$

An elementary probability measure has finite support and a range in $\mathbb{Q}_{\geq 0}$. The deficiency of randomness function is $\mathbf{d}(a|P, b) = \lfloor -\log P(a)\rfloor - \mathbf{K}(a|P, b)$. By Lemma 2 of my October 9th blog post, where $D = D_b$, relativized to b,

$$\mathbf{K}(\mathbf{x}'|b) <^{\log} i + \mathbf{I}(D_b;\mathcal{H}|b).$$

Using Lemma 2 from [Eps22b], which states $\mathbf{I}(f(a);\mathcal{H}) <^{+} \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$,

$$\mathbf{K}(\mathbf{x}'|b) <^{\log} i + \mathbf{I}(\mathbf{y};\mathcal{H}|b).$$

This is because given y and b, one can produce D_b. We can apply Lemma 1 to this equation, which results in

$$\begin{aligned}
\mathbf{K}(\mathbf{x}') &<^{\log} i + \mathbf{K}(b) + \mathbf{I}(\mathbf{y};\mathcal{H}|b) \\
&<^{\log} i + \mathbf{I}(\mathbf{y};\mathcal{H}) + \mathbf{K}(b|x, \|b\|) \\
&<^{\log} i + \mathbf{I}(\mathbf{y};\mathcal{H}) + \mathbf{K}(i) \\
&<^{\log} i + \mathbf{I}(\mathbf{y};\mathcal{H}). \qquad (1)
\end{aligned}$$

Let $\tau(x) = 2^{i-2}\mathbf{m}(x)[x \in D_\infty]$, where $[A] = 1$ if A is true, otherwise $[A] = 0$. This semi-measure is lower computable, and if $\mathbf{x} \in D_\infty$, then

$$\begin{aligned}\mathbf{K}(\mathbf{x}|\mathbf{y}) &<^+ -\log \tau(\mathbf{x}) + \mathbf{K}(\tau|\mathbf{y})\\ \mathbf{K}(\mathbf{x}|\mathbf{y}) &<^+ \mathbf{K}(\mathbf{x}) - i + \mathbf{K}(i)\\ i &<^{\log} \mathbf{I}(\mathbf{x};\mathbf{y}).\end{aligned} \tag{2}$$

Combining Equations 1 and 2, results in the theorem statemebt, that is there exists a $\mathbf{x}' \in D_\infty$ where for all $\mathbf{x} \in D_\infty$,

$$\mathbf{K}(\mathbf{x}') <^{\log} \mathbf{I}(\mathbf{x};\mathbf{y}) + \mathbf{I}(\mathbf{y};\mathcal{H}).$$

□

Lemma 1 ([Eps21]) *If b is total and b^- is not, then* $\mathbf{I}(x;\mathcal{H}|b) + \mathbf{K}(b) <^{\log} \mathbf{I}(x;\mathcal{H}) + \mathbf{K}(b|x,\|b\|)$.

References

[Eps21] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021.

[Eps22a] S. Epstein. The kolmogorov birthday paradox. *CoRR*, abs/2208.11237, 2022.

[Eps22b] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.

[VV10] N.. Vereshchagin and P. Vitányi. Rate distortion and denoising of individual data using kolmogorov complexity. *IEEE Transactions on Information Theory*, 56(7):3438–3454, 2010.