A LEMMA ON THE COMPRESSIBILITY OF MODIFIABLE BINARY SEQUENCES

YING-ZONG HUANG (2005)

We are interested in compressing a binary sequence not all of whose bits are important (some can be replaced).

Proposition. *Let* $\mathbf{s}$ *be a data sequence drawn i.i.d. Bernoulli with parameter* $p_s \leq \frac{1}{2}$. *Let* $\mathbf{q}$ *be a side-information sequence drawn i.i.d. Bernoulli with parameter* p_q, *independent of* $\mathbf{s}$, *interpreted as "care/don't care" markers on* $\mathbf{s}$. *(*$\mathbf{q_i} = 1$ *means* $\mathbf{s_i}$ *must be preserved;* $\mathbf{q_i} = 0$ *means* $\mathbf{s_i}$ *can be replaced.) Let* $\tilde{\mathbf{s}}$ *be the modified random sequence obtained after applying some fill-in scheme* f *to* $\mathbf{s}$. *On average, any fill-in scheme that only looks at the past (or only the future) of the data sequence cannot do better than "0-filling" in entropy rate.*

Remark. The idea is, if the fill-in scheme only depends on the past (i.e. a causal / streaming scheme), then any bit that the side-information says cannot be changed will still be independent of all previous bits even if some of those previous bits were modified. So the entropy rate cannot be driven below the innovation arising from the occurrence of unchangeable bits.

Proof. The following inequality holds:

$$\begin{aligned}
& \Pr(\tilde{\mathbf{s}}_\mathbf{n} = j | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = \tilde{s}_1^{n-1}) \\
= \; & p_q \Pr(\tilde{\mathbf{s}}_\mathbf{n} = j | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = \tilde{s}_1^{n-1}, \mathbf{q_n} = 1) + (1 - p_q) \Pr(\tilde{\mathbf{s}}_\mathbf{n} = j | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = \tilde{s}_1^{n-1}, \mathbf{q_n} = 0) \\
\geq \; & p_q \Pr(\tilde{\mathbf{s}}_\mathbf{n} = j | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = \tilde{s}_1^{n-1}, \mathbf{q_n} = 1) \\
= \; & p_q \Pr(\mathbf{s_n} = j | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = \tilde{s}_1^{n-1}) \\
= \; & p_q \Pr(\mathbf{s_n} = j | f_{n-1}(\mathbf{s}_\mathbf{1}^{\mathbf{n-1}}, \mathbf{q}) = \tilde{s}_1^{n-1}) \\
= \; & p_q \Pr(\mathbf{s_n} = j)
\end{aligned}$$

where $\tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = f_{n-1}(\mathbf{s}_\mathbf{1}^{\mathbf{n-1}}, \mathbf{q})$ expresses the causality constraint on the data sequence.

The entropy rate of $\tilde{\mathbf{s}}$ is defined as $\lim_{n\to\infty} H(\tilde{\mathbf{s}}_\mathbf{n} | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}})$.

$$\begin{aligned}
H(\tilde{\mathbf{s}}_\mathbf{n} | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}}) &= E_{\tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}}}[H(\tilde{\mathbf{s}}_\mathbf{n} | \tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}} = \tilde{s}_1^{n-1})] \\
&\geq E_{\tilde{\mathbf{s}}_\mathbf{1}^{\mathbf{n-1}}}[H_b(p_q \min_j \Pr(\mathbf{s_n} = j))] \\
&= H_b(p_q \min_j \Pr(\mathbf{s_n} = j)) \\
&= H_b(p_q p_s)
\end{aligned}$$

The last expression is therefore the lower bound on the entropy rate. It corresponds to a fill-in scheme that throws all the weight onto the more probable data symbol if a bit can be changed, in other words, "0-filling." □

Remark. "0-filling" is a causal fill-in scheme, therefore, it is the best causal fill-in scheme. An optimal fill-in scheme not subject to constraints should reach an entropy rate of $p_q H_b(p_s) \leq H_b(p_q p_s)$ [1], equal to that of encoding only the unchangeable bits, but it would require some non-causality.

References

[1] Martinian, Wornell, Zamir, "Source Coding with Distortion Side Information at the Encoder." *Data Compression Conference*, 2004.