The Redundancy of Two–Part Codes for Finite-Length Parametric Sources

Ahmad Beirami and Faramarz Fekri
School of Electrical and Computer Engineering
Georgia Institute of Technology, Atlanta GA 30332, USA
Email: {beirami, fekri}@ece.gatech.edu

In this paper, we investigate the redundancy in the universal compression of finite-length smooth parametric sources. Rissanen demonstrated that for a smooth parametric source with \(d\) unknown parameters, the expected redundancy for regular codes is asymptotically given by \(\frac{d}{2} \log n + o(\log n)\) for almost all sources [1]. Clarke and Barron derived the “minimax expected redundancy” for memoryless sources, which is the maximum redundancy of the best code over the space of source parameters [2], [3]. However, the minimax redundancy is for a particular parameter value, which does not provide much insight about different source parameters. In [4], we derived a lower bound on the compression of finite-length memoryless sequences using a probabilistic treatment. In this paper, we extend our analysis to smooth parametric sequences. We focus on two–part codes with an asymptotic \(O(1)\) extra redundancy [5]. We also require that the length function be regular, which is not restrictive since all codes that we know are regular [1].

We derive a lower bound on the probability that the source is compressed with redundancy greater than any redundancy level \(R_0\), i.e., we find a lower bound on \(P[R_n(l_{2p}, \theta) > R_0]\), where \(R_n(l_{2p}, \theta)\) is the redundancy in the compression of a parametric sequence of length \(n\) using a two–part length function \(l_{2p}\) for the source parameter \(\theta\). In other words, we derive a lower bound on the probability measure of the sources that are not compressible with a redundancy smaller than a certain fraction of \(\frac{d}{2} \log n\):

**Theorem 1** Let \(\epsilon\) be a real number such that \(0 < \epsilon < 1\). Then, the probability that \(R_n(l_{2p}, \theta)\) is greater than \((1 - \epsilon)\frac{d}{2} \log n\) is lower bounded as

\[
P\left[ \frac{R_n(l_{2p}, \theta)}{\frac{d}{2} \log n} \geq 1 - \epsilon \right] \geq 1 - \frac{C_d}{\int |I(\theta)|^{\frac{1}{2}} d\theta} \left( \frac{d}{e n^\epsilon} \right)^{\frac{d}{2}},
\]

where \(C_d\) is the volume of \(d\)-dimensional unit ball, \(I(\theta)\) is the fisher information matrix.

Further, we precisely characterize the minimax redundancy of universal coding for parametric sources when a two–part length function is considered. Let \(g(d)\) denote the extra redundancy incurred by the two–part assumption. Then, \(g(d) = \log \Gamma \left( \frac{d}{2} + 1 \right) - \frac{d}{2} \log \left( \frac{d}{2e} \right) + o(1)\), where \(\Gamma\) is the gamma function. This extra redundancy is negligible compared to the main term \(\left( \frac{d}{2} \log n \right)\) when the number of source parameters is large.

**REFERENCES**